Accepted Manuscript, to appear in:
Proceeding of the ACM Conference on International Computing Education Research Vol.1 (ICER 2026 Vol. 1), 2026

Students' Difficulties in Learning about AI from Teachers' Perspectives: Insights from an Action Research Approach in Compulsory CS Education

FRANZ JETZINGER, Technical University of Munich, Germany
TILMAN MICHAELI, Technical University of Munich, Germany

The omnipresence of AI in everyday life highlights the importance of integrating learning about AI into K-12 computer science (CS) curricula to prepare students for the responsible use of this technology. However, empirical research on AI-specific teaching and learning processes in compulsory CS education is still emerging, with limited findings on students' learning difficulties. This paper presents findings from a participatory action research study on students' learning difficulties in compulsory CS lessons about AI in Year 11. The first iteration intentionally foregrounded teachers' perspectives through data from lesson reflection protocols and semi-structured interviews. To complement these perspectives, a post-instruction learning assessment was carried out. Using a grounded theory-informed analysis, we derived 4 core categories of AI-specific difficulties: (1) evaluating whether real-world systems use AI approaches, (2) distinguishing between knowledge-based and data-based approaches, (3) evaluating AI models, and (4) applying AI concepts to real-world examples. Our analysis further revealed possible reasons for these difficulties, including students' preconceptions, the lack of adequate tools, and the complexity of real-world AI systems. In a final workshop, the findings were collaboratively discussed, and implications for addressing the difficulties in future lessons were formulated. The study's findings support the advancement of AI education and offer a foundation for future research on AI-specific teaching and learning processes.

CCS Concepts: • **Social and professional topics → K-12 education**.

Additional Key Words and Phrases: CS Education, K-12, Artificial Intelligence, Action Research

ACM Reference Format:
Franz Jetzinger and Tilman Michaeli. 2026. Students' Difficulties in Learning about AI from Teachers' Perspectives: Insights from an Action Research Approach in Compulsory CS Education. In *Proceedings of the ACM Conference on International Computing Education Research Vol.1 (ICER 2026 Vol. 1), August 11–14, 2026, Uppsala, Sweden.* ACM, New York, NY, USA, 22 pages. https://doi.org/10.1145/3765964.3811641

1 Introduction

Artificial intelligence (AI) is omnipresent in our everyday lives, and we interact with AI systems almost daily. In response to this development, learning *about* AI is increasingly being integrated into K-12 computer science (CS) curricula to enable students to interact with this technology in a reflective and responsible way [53]. However, for K-12 CS education, AI is relatively new compared to other topics. Consequently, it is neither possible to draw on decades of teaching experience nor on an established body of research on how to teach AI in the classroom. Nevertheless, in recent years, numerous teaching materials have been developed, and the body of literature on AI education in K-12 has grown considerably. However, empirical investigations are frequently conducted as case studies with small sample sizes [29], indicating that the field is still emerging and that the empirical evidence base remains limited [46]. Moreover, research

Authors' Contact Information: Franz Jetzinger, franz.jetzinger@tum.de, Technical University of Munich, Computing Education Research Group, Munich, Germany; Tilman Michaeli, tilman.michaeli@tum.de, Technical University of Munich, Computing Education Research Group, Munich, Germany.

Manuscript submitted to ACM

on teaching about AI is primarily situated in elective and extracurricular settings. Compulsory lessons, however, differ from such contexts, for example, with regard to student motivation or assessment practices.

In light of AI's ongoing integration into K-12 CS curricula, research must provide evidence that can inform classroom practice soon. However, conventional research methods often face the challenge that results rarely or slowly reach school practice, for example, due to insufficient external validity [27]. To this end, the present study adopts a participatory action research (PAR) approach that aims both to immediately advance practice [30] and to develop theory on teaching and learning [9]. PAR typically starts with a classroom problem or difficulty, which is then systematically investigated through iterative cycles. To provide early insights, this study takes place in the first year of implementing a compulsory AI curriculum. As a result, potential classroom difficulties have not yet been systematically documented. Furthermore, previous research provides only limited insights into learning difficulties in compulsory AI education. Consequently, this study starts by identifying students' difficulties in learning about AI. To offer a pragmatic entry point for investigating this entirely new topic and to involve teachers as equals, the first PAR cycle focuses on students' difficulties from teachers' perspectives. This approach not only leverages teachers' ability to observe learning processes across multiple lessons and different student groups, it also increases teachers' acceptance of the findings.

By investigating students' difficulties in compulsory AI education primarily from teachers' perspectives, this study provides early empirical insights into classroom experiences of teaching AI in compulsory CS education. Findings can be immediately applied to advance teaching practice in this context and to inform the development of teaching resources and strategies for AI education in other educational contexts. Beyond their practical relevance, the identified difficulties offer an empirical starting point for theory-building on teaching and learning about AI.

2 Related Work

Although research on teaching and learning about AI in K-12 is still emerging, a growing body of work addresses key questions concerning *what* and *how* to teach about AI in schools, as well as *students' perceptions* of AI.

What to Teach about AI. Regarding what to teach in this new and dynamic field, several approaches have been proposed to structure the content from an educational perspective at different levels of abstraction. For example, Long and Magerko define 17 competencies based on existing literature, grouped into five overarching themes ranging from questions such as *What is AI?* and *What can AI do?* to *How do people perceive AI?* [31]. At a more detailed level, Michaeli et al. define concrete learning objectives developed in collaboration with experts from research and practice. These objectives are organised around three perspectives: technological, socio-cultural, and user-oriented [38]. Similar perspectives can be found in UNESCO's AI competency framework, which additionally structures learning objectives into the three progression levels understand, apply, and create [37]. One of the most comprehensive approaches is Touretzky et al.'s guidelines for AI education for K-12, which define learning objectives for different educational stages based on 5 *Big Ideas*: Perception, Representation & Reasoning, Learning, Natural Interaction, and Social Impact [52].

How to Teach AI. Based on these frameworks, various curricula have been developed for different educational levels, including primary (e.g., [20]), middle (e.g., [60]), and high schools (e.g., [61]). The development often takes place through participatory processes involving teachers (e.g., [61], [7]). Within and beyond curricula, numerous teaching resources have been developed to explore how to teach AI in K-12 education. They cover a wide range of AI-related topics using different pedagogical approaches. Some include multiple AI concepts, often with a focus on aspects like ethics [63], while others only teach a single AI concept, such as decision tree learning [10] or large language models [56]. Pedagogical approaches range from unplugged activities [28] to project-based learning [14], active learning [60], and game-based

learning [25]. Given the critical role of tools in CS education, it is not surprising that many approaches for teaching AI use existing tools, like Google's Teachable Machine [47] or Orange Data Mining [39], while others have developed new tools (e.g., [43]). When designing such tools, design principles from established CS education should be considered; for example, Gresse von Wangenheim et al. emphasise the need for low floors and high ceilings tools [12].

Empirical Evidence on Teaching about AI. Many publications evaluate the mentioned curricula and teaching resources at different educational stages using qualitative and quantitative methods [26]. These studies frequently measure motivational aspects and students' attitudes towards AI [62]. For example, Mariescu-Istodor and Jormanainen showed that students perceived the lessons on AI as very interesting, regardless of gender [32]. Similar, Lee et al. reported that students actively participated in an AI course and preferred ethics-related content [24]. Studies that assess cognitive aspects tend to focus on reporting the effectiveness of an implementation or evaluating self-assessments rather than investigating students' understanding. For instance, Kong et al. found that a 14-hour AI course based on project-based learning improved students' problem-solving skills and metacognitive strategies in the context of AI [21]. Likewise, Zhang et al. showed that their ethics- and career-focused approach supported students in understanding AI concepts [63]. As AI education in K-12 is still emerging, it is not surprising that, with few exceptions (e.g., [11]), many studies are conducted outside compulsory education, for example, in summer camps or elective courses [34]. Accordingly, this research often provides detailed descriptions of extracurricular programs or teaching resources and, as [33] notes, is often presented as case studies. While such exploratory work is characteristic of early-stage research (cf. [15]), the field needs to move towards a more systematic investigation of students' understanding and learning difficulties in AI.

Students' Perceptions of AI. Research on students' learning difficulties often builds on the investigation of misconceptions (e.g., [44]) and the need for conceptual change [55]. In the context of AI, a variety of perceptions have been documented [4], but rarely embedded in specific curricula or classroom implementations. These perceptions range widely: students consider AI systems to act completely autonomously [36], attribute human characteristics to AI [54], perceive AI as omnipresent [59], or see AI as a kind of panacea [19]. Although these conceptions vary considerably, Mertala et al. point out that many of them are based on everyday concepts and media representations of AI [36]. This is notable because most studies on students' perceptions examine *pre*conceptions without connection to lessons about AI. Only very few studies investigate how teaching affects these preconceptions. For example, Mühling and Große-Bölting examine shifts in students' mental models [41], and Kreinsen et al. use conceptual change texts to modify conceptions [22]. Closely related to misconceptions is the notion of threshold concepts, which describe core ideas that are transformative but often difficult for learners to grasp [8]. Early research points to aspects like recurrent neural networks and multilayer perceptrons as potential threshold concepts when learning about AI in higher education [2].

In summary, a wide range of curricula and teaching resources have been developed, many of which are implemented in extracurricular or elective settings. Existing empirical studies primarily focus on motivational aspects and overall effectiveness, whereas research on students' perceptions of AI is often conducted independently of teaching and learning processes. Consequently, the question of what might hinder learning processes remains largely unexplored. Although prior work has identified "barriers" such as mathematics and programming when teaching about AI at university [49], comparable research for compulsory K-12 CS education remains limited. This study addresses this gap by investigating students' difficulties in learning about AI in compulsory CS lessons.

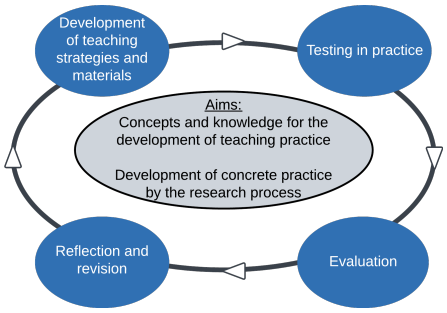
3 Methodology

The aim of this study is to investigate students' difficulties in learning about AI in compulsory CS lessons. We argue that focusing on students' learning difficulties provides a suitable entry point at this early stage of AI education research. By addressing this underexplored area, we contribute to the systematic investigation of teaching and learning processes in formal CS education about AI. In this study, learning difficulties are understood as obstacles that hinder students' understanding, application, or retention of (AI-related) content during lessons (adapted from [58]). Based on this, we formulated the following research questions:

RQ1 Which learning difficulties can be identified when students learn about AI in compulsory CS lessons?
RQ2 How do teachers intend to adapt their AI lessons to address these difficulties in future teaching cycles?

3.1 Study Design

To investigate the research questions within ongoing teaching practice, we adopted the cyclical approach of participatory action research as described by Eilks and Ralle [9]. Following this more research-led approach, teachers and researchers collaborate as partners with complementary roles. While researchers organise and evaluate the process, teachers contribute practical experience by implementing and reflecting on their teaching. In this way, the approach not only supports the iterative refinement of individual teaching practices – typically the aim of action research – but also enables the systematic examination of an educational topic. Eilks and Ralle define four phases: (1) development of teaching materials, (2) testing in practice, (3) evaluation, (4) reflection and revision (Fig. 1a). The present paper reports on the first PAR cycle, in which students' learning difficulties are primarily investigated from teachers' perspectives, complemented by exploratory insights from students. The study was conducted with the participation of 10 teachers (see 3.4).



(a) Cycle of participatory action research according to [9]

Were there any deviations from the planning? ☐ No ☐ Yes **If yes, describe** them.
Did the students have any **difficulties** in **understanding**? ☐ No ☐ Yes **If yes, what** were they?
How much do you agree with the following statements?
Students achieved the **learning objectives**. I… ☐ totally agree, ☐ rather agree. ☐ rather disagree. ☐ totally disagree. **If** you rather or totally **disagree**, **why**? (1-2 aspects)
The **students actively participated** in class. I… ☐ totally agree. ☐ rather agree. ☐ rather disagree. ☐ totally disagree.
Did you notice anything particularly **positive** or **negative**?
Is there anything you would do **differently next time**? What would that be?

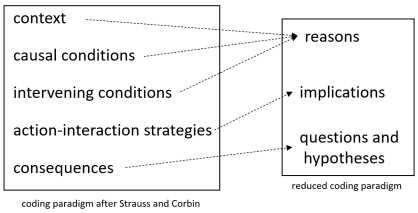
(b) Excerpt from the reflection protocol

Fig. 1. Research design and excerpt from the reflection protocol

In the *first phase*, teachers independently designed their lessons, while the research team developed the data collection instruments (see 3.3). In the *second phase*, the teachers taught their lessons about AI in a total of 20 Year 11 classes at different times during the school year. After each lesson, teachers documented their experiences using a reflection protocol. The completed protocols served as the basis for interviews with the teachers after they completed the entire AI sequence. In addition, students completed a learning assessment, which provided complementary insights into

Source	Sample size
reflecting protocol	115 protocols from 9 teachers
teacher interviews after the sequence	9 teachers (1 of them without protocols)
teaching materials	4 teachers
learning assessment	192 students from 8 classes

(a) Data sources and samples for RQ1



(b) Reducing Strauss and Corbin's coding paradigm

Fig. 2. Data sources of the study (RQ1) and the reduced coding paradigm

their conceptual understanding. Due to organisational constraints, not all teachers were able to conduct the learning assessment in class, but responses from 192 students across 8 classes are available. For data privacy reasons, no additional demographic data was collected on the students. In the *third phase* of the PAR cycle, the research team analysed all collected data sources – including reflection protocols, interviews, learning assessment, and teaching materials provided by 4 teachers. During the initial three phases, the level of participation between teachers and researchers was intentionally limited in order to minimise additional workload for the participating teachers, who were already facing the challenge of teaching the topic for the first time. In the *fourth phase* of the PAR cycle, participation was strengthened in a one-day workshop, where teachers were actively involved in reflecting on the results and deriving implications for future lessons.

3.2 Study Context

This study was conducted in the first year of teaching a new compulsory AI curriculum in Year 11 in the German state of Bavaria. While teaching about AI was newly introduced, the state has a long-standing tradition of compulsory CS education in high schools. Since 2003, CS has been a compulsory subject for all students in Years 6 and 7. CS is also taught two hours a week in Years 9 and 10 to students who have chosen the STEM branch in high school (about 2/3 of all high school students). In 2024/25, a new compulsory CS curriculum was introduced in Year 11, with two hours of CS lessons per week for all students (from all branches). For the first time, this curriculum included compulsory lessons on teaching about AI. In our local context, the term *curriculum* refers to a set of mandatory learning objectives which is defined by the Ministry of Education, along with a recommended period of time to teach them. While achieving the determined objectives is compulsory, concrete implementation is left entirely to the teachers.

Within the Year 11 curriculum, 12–16 lessons are recommended for the AI-related part which covers learning objectives related to the conceptual foundations of AI, the explanation and application of one selected ML algorithm, the analysis of model reliability, very basic aspects of artificial neural networks, and the societal implications of AI (see Tab. 1). In Year 13, AI is also integrated into the curriculum with a particular focus on neural networks and unsupervised learning. Although teachers are free in how they implement the learning objectives in their classrooms, lessons typically follow similar patterns, as we were able to ascertain from interviews and teaching materials (see Tab. 1). Teachers usually begin by discussing what AI is and where AI approaches are applied. They frequently use unplugged activities to introduce basic AI concepts, including machine learning (ML) and knowledge-based approaches. Some teachers use text- or block-based programming languages for very small, well-defined examples. ML algorithms are commonly introduced using small datasets, first by manually executing simple examples and later by implementing them with tools such as Orange Data Mining. Regarding a single perceptron, teachers let students manually calculate

Learning Objectives	Typical Implementation
Students discuss approaches to defining AI, describe various basic ideas of AI methods and the areas in which they can be applied.	Teachers provide numerous examples of AI systems and discuss different definitions of AI. Unplugged materials are often used to introduce the basic ideas of knowledge-based AI systems and ML approaches. Some teachers additionally let students implement small, well-defined programming examples in text- or block-based languages.
Students explain the functionality of a selected ML algorithm (k-nearest neighbours or decision tree learning) in general and for concrete examples.	Mostly, decision tree is chosen as the selected ML algorithm. Teachers usually start with hands-on material and small, prepared datasets; students execute the algorithm manually by calculating the information gain. Teachers also introduce various quality criteria (e.g., precision, recall).
Students analyse the influence of training data and parameters on the reliability of the results of a ML algorithm, using an appropriate tool.	Decision tree learning is typically implemented using tools such as Orange Data Mining, based on small datasets and partially on real-world data.
Students explain how an artificial neuron (perceptron) works and describe the basic structure of a neural network.	Teachers explain the delta-learning rule of a single neuron, partially starting with an unplugged activity. They let students calculate outputs for simple examples and show how multiple neurons interact to classify very simple problems (with a focus on the mathematics of linear equations).
Students implement/simulate a single artificial neuron.	Teachers use different programming languages (e.g., Java, Python, or occasionally Microsoft Excel) to implement a single neuron, or simulation tools such as TensorFlow Playground.
Students comment on selected current applications of AI and assess the opportunities and risks for individuals and society.	Ethical aspects of current AI applications are typically addressed at the beginning and end of the sequence, but sometimes also throughout the whole sequence.

Table 1. Mandatory Learning objectives of the Year 11 AI curriculum and their typical classroom implementation

simple outputs and implement it using different programming languages. Throughout the teaching sequence, especially at the beginning, teachers create space for discussions about ethical aspects of AI.

3.3 Instruments and Data Sources

In order to investigate students' learning difficulties (RQ1), several instruments were developed (see Fig. 2a).

Reflecting Protocol. To support teachers' reflection on each lesson, a protocol was developed (see an excerpt in Fig. 1b). It was designed to be completed in 5-10 minutes to minimise additional workload for teachers. The protocol collected meta-information about the lesson, such as its duration (usually 45 or 90 minutes), its topic, and its function within the whole AI sequence (e.g., introducing new content, deepening). Teachers were asked to document whether they observed any student difficulties and, if so, to describe them. In addition, the teachers should rate certain statements on a 4-point Likert scale based on their observations (e.g., whether students achieved the learning objectives formulated by the teachers or whether students actively participated in class). The protocol ends with two open-text questions about positive or negative aspects the teachers have noticed and whether they would do anything differently next time.

Teacher Interviews. To obtain further answers to RQ1, a semi-structured interview was conducted with each teacher after they had taught the entire AI sequence. In the interview, the teachers were first invited to describe their overall experience teaching the AI-related part of the curriculum. Additionally, teachers were asked if they had prioritised certain curriculum aspects and if they had experienced any challenges. The central part of the interview focused on students' difficulties in learning about AI, as observed by the teachers. To this end, teachers should elaborate on the difficulties documented in the protocol, clarify any ambiguous protocol phrases, and consider any additional difficulties.

Students' Learning Assessment. To supplement teachers' perspectives, a learning assessment was designed to be completed by the students after the lesson sequence. It is intended to provide exploratory, curriculum-aligned insights into students' conceptual understanding rather than to function as a validated instrument. The assessment involves 6 content areas (see Tab. 2), each represented by 2-4 questions, covering aspects like characteristics of AI systems,

principles of ML, and causes of error and bias. In addition, items on concrete ML approaches, such as decision tree learning and neural networks, are included. The items are in closed-response formats to ensure teachers can also use them for in-class assessment without spending too much time on corrections.

The assessment development was closely aligned with the learning objectives specified in the curriculum and informed by established AI literacy frameworks (e.g., [38], [31]). For example, item 3 (see Fig. 3a) was informed by a learning objective defined by [38] and [31]. In addition, existing AI literacy instruments were reviewed to inform item development (e.g., [5], [57], [23], [16]). Where appropriate, items from performance-based instruments that measure cognitive aspects were adapted to the local context and revised in wording; for example, item 2 was conceptually adapted from Hornberger et al. (see Fig. 3b). The resulting item pool was reviewed by external experts from both research and educational practice who were not involved in the study. Based on their feedback, items were revised, clarified, or removed to improve clarity and content coverage.

For which of the following tasks can AI approaches be used beneficially?	*Responses* (%)
object recognition on images	84%
therapy suggestions for certain diseases	58%
sorting a very large amount of data	89%
product recommendations in online stores	82%
creating deep fakes	73%
speech to text conversion	84%
tracking on the Internet with cookies	51%
positioning with GPS	48%

(a) Item 3: Applicability of AI approaches to different tasks

Which of the following statements can be true for AI systems? AI systems...	*Responses* (%)
...follow a goal set by humans.	88%
...can deal with uncertainty.	7%
...can only be developed on supercomputers.	9%
...work on the same principle as the human brain.	46%
...do not make mistakes after learning.	10%
...can improve with experience.	93%
...solve certain tasks better than humans.	87%
...can be self-aware.	13%

(b) Item 2: Behaviour of AI systems

Fig. 3. Two multiple-choice questions of the learning assessment with mean scores for each response option (correct answers are bold)

Teaching Materials. Teachers were invited to share their teaching materials on a voluntary basis to reduce potential bias from selective sharing or unusually polished materials. Four teachers provided materials, including lesson plans and worksheets. In addition, all teachers documented the textbooks they used, with some reporting the employed exercises.

Workshop with Teachers. A structured one-day workshop was developed to support collaborative reflection on the data analysis results and to explore how teachers intend to adapt their lessons to address the identified difficulties (RQ2). The first half of the workshop focused on presenting the analysis results to the teachers and validating them collaboratively. Aspects on which teachers did not fully agree were discussed based on the teachers' experiences and the data. The second half of the workshop focused on how teachers intended to adapt their future lessons to address the identified difficulties. The workshop results were documented in real time using a shared document. Plenary discussions were documented directly, whereas group work results were first presented and subsequently documented in this document.

3.4 Participating Teachers

Through targeted personal contact, 10 CS teachers (1 female, 9 male) from 9 different schools were recruited for the study. They were between 30 and 49 years old and had 2 to 21 years of teaching experience. All but one had mathematics as a second subject. Seven of the participating teachers held a regular teaching degree in CS, while the remaining three held regular degrees in other subjects and completed an intensive 2-year in-service qualification program to become CS teachers. In the German state of Bavaria, teacher education typically requires four to five years of university study,

covering content knowledge (CK) in two subjects (about 100 ECTS each) as well as pedagogical content knowledge (PCK) and pedagogical knowledge (PK). In addition, teachers must complete two years of preparatory service in schools after university. However, the topic of AI was neither part of teachers' university education, nor included in the school curriculum before 2024/25. Therefore, the Ministry of Education coordinated several professional development (PD) opportunities to equip teachers with the CK and PCK needed to teach the AI competencies in the new curriculum [17]. Consequently, all participating teachers attended PD programs in preparation for teaching the new AI-related part of the curriculum. However, this was apart from this study, and they were not explicitly trained in the study's context.

3.5 Data Analysis

All available qualitative and quantitative data (see Fig. 2a) were used for a grounded theory-informed analysis to identify difficulties in learning about AI. We started the analysis by iteratively coding the data from the reflection protocols and interviews using the three coding phases of Strauss and Corbin's grounded theory (GT): open, axial, and selective coding [48]. During the first phase of **open coding**, 218 text passages related to difficulties in learning about AI were marked with in-vivo codes. These codes could be subsumed into 16 *categories*. To focus the analysis on AI-specific learning difficulties, 45 codes referring to non-AI-specific challenges (e.g., insufficient programming skills in general) were grouped into the category *non-AI-specific difficulties*. This category was retained for contextual interpretation but was not included in the subsequent category-building process for AI-specific learning difficulties. In the next step of **axial coding**, a reduced form of Strauss and Corbin's coding paradigm was used to analyse the interrelationships between the categories. In its original form, the coding paradigm includes five aspects: a phenomenon is analysed according to its *context*, *causal conditions*, *intervening conditions*, *action-interaction strategies*, and *consequences*. These five aspects were reduced to three in order to fit the educational context and to ensure that practitioners without a scientific background could easily understand the final results (see Fig. 2b). The reduced coding paradigm allowed us to analyse AI-specific difficulties, which are seen as phenomena in this study, according to three aspects: (1) their *reasons*, (2) *questions and hypotheses* that derived from the data related to the difficulties and (3) their *implications* for practice. In GT, the last phase of **selective coding** is typically used to derive a single core category from all categories. In our context, no single core category emerged during selective coding. Instead, four core categories were identified, each representing a difficulty, possible reasons, questions and hypotheses as well as first implications for the practice.

The entire analysis process followed the core principles of GT, including constant comparison and memo writing. Across all coding phases, all types of data (text passages, in-vivo codes, and categories) were compared with and among each other. Teaching materials and learning assessment results were also included in the comparisons.

Analysis of the Learning Assessment. The learning assessment was analysed descriptively by calculating rate-corrected mean item scores across all students (see Tab. 2), ranging from 0 to 1. For multiple-choice items, each correctly selected option contributed a fraction of 1 proportional to the number of correct options. Incorrectly selected options were subtracted proportionally to the number of distractors. Negative item scores were set to zero. Single-choice items were scored dichotomously (1 = correct, 0 = incorrect). For items with comparatively low mean item scores, mean scores (in percentage) for each response option were additionally calculated to explore common response patterns. These patterns were examined in relation to the qualitative data. As a plausibility check rather than a psychometric validation, we calculated teacher-level means (TLMs) for each item by averaging the item scores of all students taught by one teacher. We report the range of these TLMs across teachers, defined as the difference between the highest and lowest TLM for each item (see Tab. 2).

Quality Assurance Aspects of the Data Analysis. The coding was conducted by the first author. Consistent with GT, no inter-rater reliability was applied [35]. To increase the trustworthiness of the qualitative analysis, several strategies were employed, including peer debriefing and member checking [6]. Interim results of the open and axial coding were discussed in two meetings of the authors' research group, each involving between four and six researchers with experience in qualitative educational research. During these discussions, categories were critically examined, restructured, or merged when conceptual overlap was identified. To obtain an external practitioner perspective, the interim results were additionally discussed with an experienced CS teacher who was not involved in the study.

The final core categories were validated with 6 of the participating teachers during the final workshop (member checking). Each category, including all associated aspects, was presented and discussed in detail. Only aspects that achieved consensus among all participants were retained, while those lacking consensus were revised or discarded.

4 Results

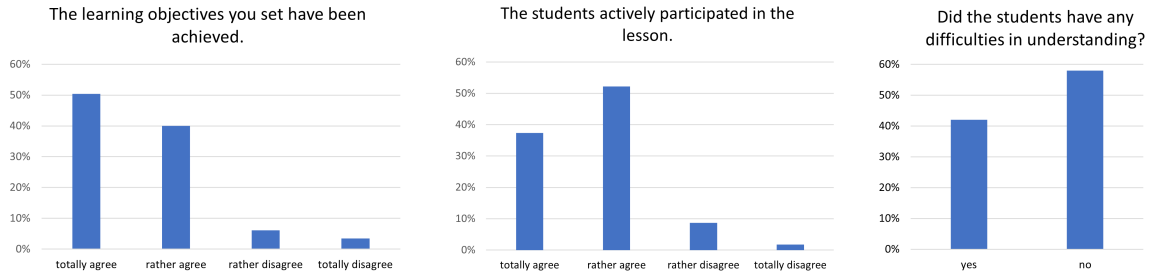
Before presenting the results on students' learning difficulties, we first outline general aspects of the curriculum's implementation to situate the subsequent findings. These aspects draw on data from reflection protocols and interviews. The analysis of reflection protocols indicates that the AI sequence was taught for an average of 16 lesson hours, closely matching the recommended curricular scope. Responses to the closed-ended protocol questions paint a largely positive picture: teachers reported that, in their view, students achieved the intended learning objectives in most cases (see Fig. 4a), and they rated students' active participation as consistently high (see Fig. 4b). Qualitative protocol comments supplement this picture, describing the topic of AI as "very enjoyable" and "satisfying" to teach. In the interviews, teachers also reported that students had "great interest in the topic" and that "active discussions evolved" in the classroom, particularly when societal and ethical aspects were addressed. Regarding specific teaching strategies, teachers highlighted interactive, unplugged activities and hands-on materials as engaging and useful elements for introducing concepts such as reinforcement and supervised learning.

Descriptive results from the learning assessment help to complement the overall picture. Mean item scores ranged from 0.13 to 0.81, with several items achieving a moderately high to high score (see Tab. 2). The ranges of TLMs are mostly below 0.30, suggesting that the items remained interpretable across different classroom settings, rather than being specific to a single teacher. However, this should be interpreted cautiously, as the assessment was designed for supplementary insights into students' difficulties rather than as a validated effectiveness measure.

Despite the overall positive picture, teachers reported observing learning difficulties in approximately 40% of the 115 lessons documented in the reflection protocols (see Fig. 4c).

4.1 RQ1: Students' Difficulties

Although codes related to **non-AI-specific difficulties** were grouped into a separate category early in the analysis, they provide important contextual information and are briefly summarised here. First, teachers reported *problems in mathematics* for two reasons: students had difficulty performing calculations, or sometimes became demotivated when confronted with mathematical formulas. Second, teachers documented *insufficient programming skills* across lessons and in various programming languages (text- and block-based). A third non-AI-specific difficulty reported by teachers is the *high level of heterogeneity* among students in terms of prior knowledge and performance levels, which is a common observation in CS classes. While non-AI-specific difficulties may interact with AI-specific difficulties, analysing these interactions was beyond the scope of the present study, which focuses on identifying AI-specific learning difficulties.



(a) Learning objectives (b) Students' participation (c) Difficulties

Fig. 4. Teachers' ratings reported in the reflection protocols, shown as percentages of all protocols (N = 115)

Table 2. Descriptive item performance for the learning assessment based on student data

	AI Characteristics			**ML Principles**					**Data & Bias**						
Item	1	2	3	4a	4b	4c	5	6	7	8	9a	9b	9c	9d	9e
type	SC	MC	MC	SC	SC	SC	OR	MC	SC	MC	SC	SC	SC	SC	SC
mean item score	.78	.50	.21	.36	.35	.63	.67	.24	.27	.58	.72	.67	.80	.30	.67
range (TLMs)	.68	.13	.07	.19	.17	.12	.21	.10	.22	.17	.46	.30	.22	.15	.18

	Decision Tree				**Perceptron**							**Neural Networks**	
Item	10a	10b	11a	11b	12a	12b	13a	13b	13c	13d	14	15	16
type	SC	SC	MC	MC	SC	SC	SC	SC	SC	SC	OR	MC	MC
mean item score	.78	.13	.54	.27	.81	.72	.54	.51	.37	.22	.26	.16	.40
range (TLMs)	.42	.24	.14	.21	.14	.14	.13	.26	.33	.21	.12	.14	.18

SC = single-choice, **MC** = multiple-choice, **OR** = order; **range (TLMs)**: teacher-level means were calculated by averaging the item scores of all students taught by one teacher. The range represents the difference between the highest and lowest TLM across teacher for each item.

In the following, we provide a detailed description of the four **AI-specific difficulties** that emerged from the analysis of all data. The description follows the core categories' structure derived from the reduced coding paradigm used in axial coding: for each difficulty, possible reasons, upcoming questions, and derived hypotheses are presented. The implications are then presented in section 4.2.

Core Category 1. The first core category that emerged from the data analysis concerns students' difficulty in **evaluating whether real-world systems use AI approaches** (see Fig. 5). Evidence from teachers' reports indicates that students have difficulties deciding whether a system uses an AI approach to solve a specific task and can thus be classified as an AI system. For example, one teacher wrote in the reflection protocol:

> *The classification of examples as 'AI or not' caused (understandable) difficulties for some.*[1] [T4, protocol 1]

The relatively low mean item score of the question for which tasks AI approaches can be used (see Tab. 2, item 3) supports this interpretation. While students correctly associate many examples with AI (e.g., object recognition, product recommendations), they also state that AI approaches are used for sorting a large amount of data (89%) or tracking on the

[1]All quotes have been translated into English by the authors with minimal adjustments to improve comprehensibility.

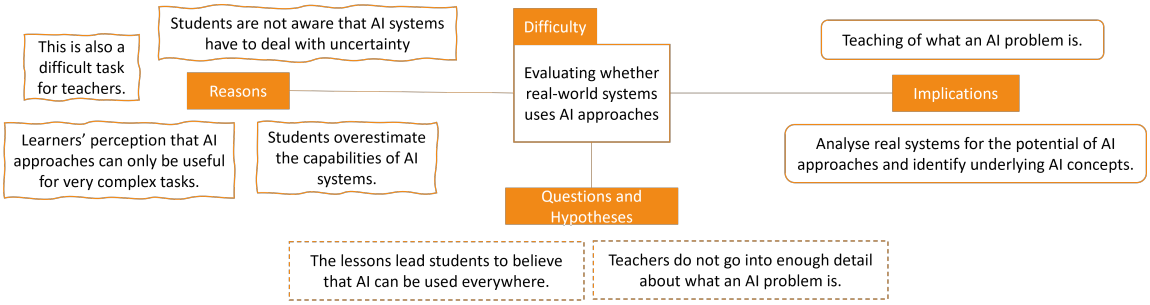


Fig. 5. Core category 1: Difficulty in evaluating whether real-world systems use AI approaches

Internet with cookies (51%) (see Fig. 3a). Although it may sometimes be ambiguous whether systems use AI approaches, these results nevertheless indicate the difficulty of evaluating whether real-world systems use AI approaches.

Reasons. We identified several possible reasons for this difficulty in the data. On the one hand, the data indicate that students are not aware that dealing with uncertainty is an essential feature of AI systems. This is shown, for example, by the results of the learning assessment on the question of which of the given statements apply to AI systems (see Fig. 3b): only 7% of the students stated that AI systems deal with uncertainty. On the other hand, students' perceptions were identified as reasons for this difficulty. For example, teachers reported that students initially overestimate AI systems:

> *So that really stuck with me, that the students [...] overestimated the capabilities of the [AI] tools [...]. I would say that [is] [...] a big misconception I encountered at the beginning.* [T3, interview at 30:50]

Teachers also reported the perception that AI approaches are only used for complex tasks:

> *[...] the less complex ones were not on their radar. So AI is already the ultimate for them, all these ChatGPT things: speech understanding, autonomous driving and so on. That's AI to them, and everything below that [...] somehow has nothing to do with AI.* [T9, interview at 12:37]

The comparison of the reported perceptions with the results of the learning assessment shows that at the beginning of the lesson, students tend to think that AI systems are used only rarely and for very complex tasks. After the lesson, however, they associate many tasks with AI, even those that do not use AI approaches, as the results of the learning assessment show. Another reason we identified for the difficulty in the first core category is grounded in the complexity of real-world systems. In the interviews and the final workshop, teachers described this difficulty as "understandable" (first quote above), noting that even for them, evaluating whether a technology uses AI approaches can be challenging.

Hypotheses. Two hypotheses emerged from the difficulty and its reasons. First, although the lessons change students' preconceptions about where AI approaches are used, they seem to mislead them into thinking that AI approaches can be used profitably everywhere. Second, it seems like teachers did not have discussed in detail what constitutes an AI problem. This hypothesis is supported by the teaching materials and confirmed by the teachers in the final workshop.

Core Category 2. The second core category concerns the difficulty of **distinguishing between knowledge-based and data-based AI approaches** (see Fig. 6). This does not mean that the students have difficulties separating the two terms, rather, they face problems at a conceptual level. It is unclear to them whether and why an AI system or a particular AI approach is data- or knowledge-based, even when they apply an approach themselves. Since there were no questions related to this difficulty in the learning assessment, this difficulty is exclusively derived from the protocols

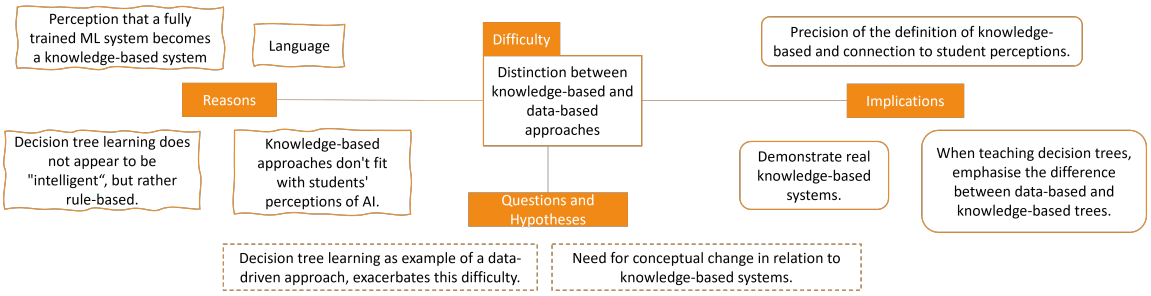


Fig. 6. Core category 2: Difficulty in distinguishing between knowledge-based and data-based approaches

and interviews. For example, in the interview, one teacher reported that "the terms knowledge- and data-based are difficult for students" and that this "was one of the main difficulties that [he] saw in retrospect" [T5, interview at 07:01].

Reasons. Again, several reasons for this difficulty emerged from the data. First, language was identified as a factor that caused difficulties in distinguishing between the two approaches. For example, one teacher reported that the students tend to assume that anything with data must be data-driven and vice versa, that a knowledge-based system does not deal with any data:

> *I think that's a bit due to the terminology, [...] that knowledge-based AI has nothing to do with data, because otherwise it would be data-based. That's where we got stuck a few times.* [T2, interview at 09:28]

Second, teachers reported another reason for this difficulty based on students' general perceptions of AI. Some students do not consider knowledge-based systems to be AI systems because, in their view, AI must always learn something – as one teacher pointed out:

> *The fact that AI does not necessarily mean that the system learns did not fit into students' previous concept of AI.* [T4, protocol 3]

A third reason is the perception that a fully trained data-based AI system becomes a knowledge-based system:

> *Some groups have concluded that a trained AI becomes a knowledge-based AI.* [T1, protocol 5]

This perception has to be seen in the context of decision tree (DT) learning, one of the ML algorithms included in the curriculum. DTs can be created data- and knowledge-based, and many teachers have done both in the classroom. However, when comparing the resulting trees, there seems to be no difference at first glance. A related issue reported by teachers is that executing DT learning manually on a very small dataset does not seem to be different for students from defining rules in a knowledge-based context:

> *[..] difficult to distinguish from [...] a "classical AI," as rules are very obvious in the introductory example and the students, therefore, define rules as in classical AI.* [T6, protocol 5]

Hypotheses. From this difficulty and the reasons, two hypotheses were derived in the context of the data analysis. First, DT learning, as an example of an ML algorithm, could increase this difficulty. The teachers in the workshop are not convinced by this hypothesis and instead emphasise the potential of DTs to differentiate between the two approaches. Second, there is a need for conceptual change regarding knowledge-based systems so that students can adapt their perceptions of AI to integrate them.

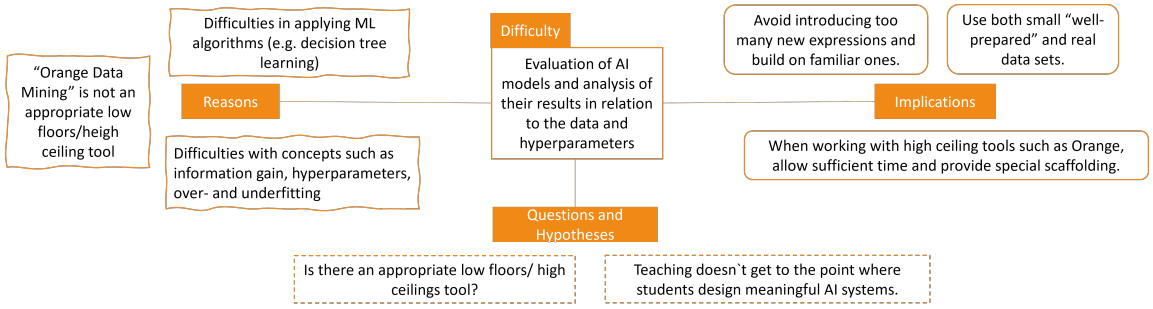


Fig. 7. Core category 3: Difficulty in evaluating AI models and their results in relation to data and hyperparameters

Core Category 3. The third core category concerns difficulties in the context of **evaluating AI models and analysing their results in relation to the data and hyperparameters** (see Fig. 7). This difficulty becomes particularly evident when students implement ML algorithms using tools such as Orange Data Mining. When evaluating the resulting model, students had difficulties interpreting quality metrics and the prediction results (for example, shown in a confusion matrix). As a consequence, they are unable to analyse specific errors of the AI model in relation to the training data or adjust specific hyperparameters to improve the model results. One teacher expressed this in the following statement:

> *They didn't get to the point [...] where they [...] dig in a little bit and say, I'm going to see if I can do something about this and that adjustment screw [hyperparameter; author's note]. [...] they said, OK, now it shows me exactly the same values on the confusion matrix as on the teacher's screen, then I must have done it right [...].* [T6, interview at 23:02]

A second statement from another teacher also shows this difficulty:

> *I wanted them to work with real data and then evaluate this model using the confusion matrix, and then launch an intervention if necessary. But simply putting this workflow together in Orange took too much time, and students did not get to the actual core of what I had planned.* [T3, interview at 13:02]

Reasons. The first reason we identified in the data for this difficulty was that students struggle to execute ML algorithms, such as DT learning. Teachers reported that students faced different problems regardless of the teaching approach, including unplugged activities and manually executing the algorithm on very small datasets:

> *So with the decision trees [...] some things were difficult. Somehow, understanding how to separate the data with this concept.* [T7, interview at 07:29]

The difficulties in applying DT learning can also be seen in the learning assessment. One task there was to apply the algorithm to a very simple dataset by rating the first two splitting rules out of six suggestions (see Fig. 8). The corresponding mean item scores is 0.54 for the first rule (Item 11a) and 0.27 for the second rule (Item 11b); the mean score for correctly identifying both rules is even 0.19, indicating that the task posed challenges for many students.

Further reasons that emerged from the data about the third core category were difficulties with certain concepts, such as hyperparameters or over- and underfitting, as illustrated by the following two statements:

> *The concept of hyperparameter value was not tangible for the students.* [T0, protocol 4]
> *It is also difficult to abstract overfitting and underfitting to decision trees.* [T3, protocol 6]

A company wants to develop an AI system that decides whether a person should be invited to an interview based on their age, education, and work experience. The following data should be used to learn a decision tree. Identify the first and second questions that can be used to generate the best possible decision tree.

age	degree	years of work experience	invitation
30	Master	3	yes
32	Bachelor	1	no
35	Master	1	yes
28	Master	2	no
25	Master	8	yes
23	Bachelor	5	no
32	Master	6	yes
27	Master	4	yes

age < 27
Bachelor or Master
work experience > 4
work experience > 5
age > 28
work experience > 2

a) first question
b) second question

Fig. 8. Item 11 of the learning assessment on applying decision tree learning

Another reason identified was the lack of a suitable tool for evaluating AI model results. For example, the teachers reported that Orange Data Mining is too difficult for students to get started with and, therefore, takes a lot of time:

> *Working with Orange was difficult. [...] Working with Orange is not easy, more detailed guidance is needed. There is not yet a deeper understanding of Orange's processes.* [T3, protocol 5]

Question and Hypothesis. Based on the difficulty of evaluating AI models and the possible reasons for it, one hypothesis and one question were derived. First, it seems that students do not design meaningful AI systems within the lessons. Second, the question was raised whether there is a suitable tool for evaluating AI models that provides an easy start (low floors) and allows students to design different and meaningful systems (high ceilings, wide walls).

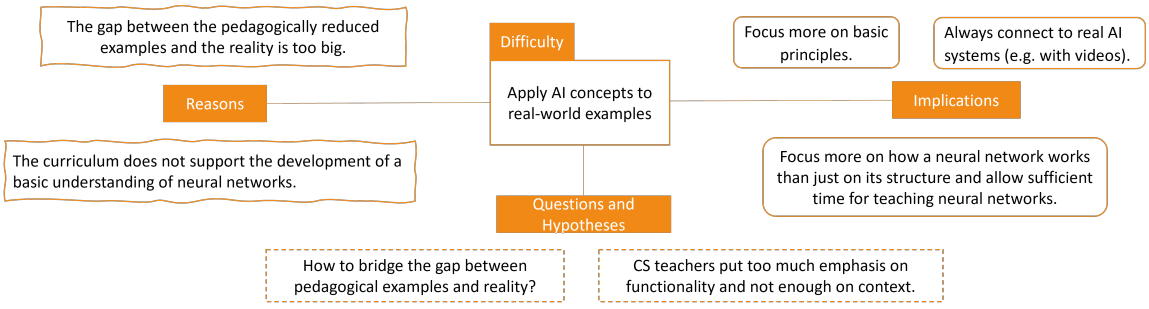


Fig. 9. Core category 4: Difficulty in applying AI concepts to real-world examples

Core Category 4. The fourth core category concerns the difficulty of **applying AI concepts from the classroom to real-world examples** (see Fig. 9). During data analysis, this issue was especially evident in the context of artificial neural networks (ANNs). For example, the low mean item score for the assessment item 15 on designing an ANN for facial recognition (0.16, see Tab. 2) and the analysis of the selected responses, indicates that students held different incorrect ideas about network design (see Fig. 10a). In the final workshop, teachers emphasised that this difficulty extends beyond ANNs to all taught AI concepts. This is consistent with the data, as following interview quotes illustrate:

> *But then you get to the point, so quickly, where the students say, OK, I know the basic principle, but how does it really work [...], that's the step that's difficult for the students.* [T6, interview at 17:36]

The relatively low mean scores for the first two parts of the learning assessment item 4, which ask students to assign an ML approach to real-world tasks (0.36 and 0.35, see Tab. 2), support the interpretation that applying AI concepts is generally difficult for students (see Fig. 10b).

An AI system for facial recognition is to be developed using an artificial neural network. Which of the following statements are true? The developers...	*Responses (%)*
...consider which edges occur in a face. The network is then built with one artificial neuron per edge.	36%
...divides the face into geometric shapes (e.g. circles for the eyes). The network contains one artificial neuron for each shape.	52%
...divides the face into different areas (e.g. mouth area, eye area). The network contains one layer of artificial neurons per area.	68%
None of the statements are true.	15%

(a) Item 15: Designing a neural network

Assign a machine learning approach (supervised, unsupervised, reinforcement learning) to each of the following applications.	*Correct answer*
a) A robot explores an unfamiliar environment on its own.	**reinforcement learning**
b) Customers are clustered in groups.	**unsupervised learning**
c) Diabetes is predicted based on certain characteristics.	**supervised learning**

(b) Item 4: Assign a ML approach to an application

Fig. 10. Two learning assessment items related to the application of AI concepts to real-world examples (correct answers are bold)

Reasons. The main reason for this difficulty, identified in the data and confirmed by the teacher in the final workshop, is that the gap between pedagogically reduced examples taught in the classroom and real-world examples is too big, as the following interview statement illustrates:

> *I think the biggest challenge is that there is such a huge gap between theory and practice. So you try to make incredibly simple examples [...]. The end result [real-world AI systems; authors' annotation] is so incredibly complex that I think that gap is too big.* [T9, interview at 02:15]

Another reason, particularly regarding ANNs, is that the curriculum does not support the development of a basic understanding of ANNs, as the related learning objective focuses more on their structure than on their functionality.

Question and Hypothesis. From the difficulty and its reason, one question and one hypothesis were derived. The question of how to close the gap between pedagogically reduced examples and complex reality could not be answered in the final workshop. The hypothesis that CS teachers put too much focus on functionality and not enough on context was neither confirmed nor denied by the teachers in the final workshop.

4.2 RQ2: Addressing the Difficulties

Based on the implications derived from the data analysis, the teachers developed strategies in the final workshop, to address the four identified difficulties. Regarding the difficulty, **evaluating whether real-world systems use AI approaches**, the teachers formulated the implication that the question of what constitutes an AI problem should be considered in detail in the classroom. To this end, they proposed discussing indicators of AI systems in future lessons (e.g., dealing with uncertainty, handling large problems). In addition, they suggest analysing a wide range of real-world computing systems with students to examine whether AI approaches are used. The teachers emphasised the importance of also analysing non-AI systems and justifying why AI approaches are not used. Intending to develop a basis for a classroom activity, the teachers collected systems and attempted to assess whether and which AI approaches are used (Tab. 3). In doing so, they realised that they lack the technical expertise in some areas to properly evaluate real systems.

Regarding the difficulty of **distinguishing between knowledge-based and data-based approaches**, the teachers formulated the implication that real, currently used knowledge-based AI systems should be demonstrated in the classroom. At the same time, they noted the challenge of identifying such systems that relate to students' everyday lives. In addition, the teachers identified some key differences between the two approaches that should be explicitly discussed in the classroom (e.g., knowledge-based systems represent knowledge and draw conclusions from it, while data-based systems generate new knowledge). Teachers further suggested using decision trees in both approaches

computing system	teachers' thoughts on the use of AI approaches
video toll	object recognition on a taken photo uses AI approaches; matching the license tag in the database does not need AI
dynamic traffic light control	everything is feasible: reinforcement learning to train the traffic light control, as well as an expert system or a classical algorithm to program the control depending on time, day, etc.
booking system for a restaurant	no AI approach is necessary
parking lot entrance with barrier	no AI approach is necessary, but when working with automatic recognition of the license tag, AI approaches are used
navigation system	a combination of different AI approaches and classical algorithmic approaches is possibly used

Table 3. Teachers' attempt to create a list of systems with and without the use of AI approaches

to highlight the differences. Additionally, the workshop revealed some inconsistencies in teachers' understanding of knowledge-based systems, prompting them to refine their own definitions.

To address the difficulty of **evaluating AI models**, the teachers formulate the implication to introduce as few abstract terms as possible while using familiar terms instead (e.g., four-field table instead of confusion matrix in the context of quality metrics for the evaluation of AI models). They also suggest using small, reduced datasets as well as real-world data. As an intermediate step, they propose using large datasets that are prepared in such a way that changing specific hyperparameters has a clear and easily recognisable effect on the results. However, they noted that such datasets have yet to be created. Regarding the use of Orange Data Mining, teachers agree with the researchers that sufficient scaffolding is needed to support students, especially when using the tool for the first time. One teacher who had already used such scaffolding reported fewer problems with the tool.

To address the difficulty, **applying AI concepts to real-world examples**, the teachers formulated the implication to focus more on basic principles in future lessons and explicitly connect them to real-world AI systems. For this purpose, they shared and discussed videos that explain real-world examples and their underlying concepts. Regarding ANNs, the teachers noted that more time should be spent on this topic in the classroom, even though the curriculum does not intend to. However, due to the high complexity of AI systems, also the teachers sometimes struggle to identify the exact underlying concepts. The question of how to bridge the gap between pedagogically reduced examples and reality, as well as how to connect lessons to the complex AI systems in everyday life, could not be fully answered.

5 Discussion

Investigating students' difficulties is a common strategy for exploring emerging or insufficiently understood topics, within (e.g., [42, 44]) and beyond CS education (e.g., [50]). In line with this tradition, the present study addresses an underexplored dimension of AI education by systematically examining students' difficulties in learning about AI in compulsory CS lessons. Although our study is situated within a specific national context with factors that might limit its generalisability – such as a state-mandated curriculum and comparably well-educated teachers – we argue that our results provide an important starting point for further empirical research and theory-building in formal AI education. Our PAR approach constitutes a particular strength in this regard. By intentionally conducting our study in the field, we were able to examine authentic teaching and learning processes and maximise the practical implications of our results.

Our findings on **non-AI-specific difficulties**, such as insufficient programming skills or mathematical problems, align with the barriers identified in AI education at the university level by Sulmont et al. [49]. We extend this line of work by showing that these difficulties also persist in K-12 AI education and by highlighting the role of students' heterogeneity in prior knowledge. As these difficulties are also known from CS education research in other topics, it can be argued that certain challenges in AI education may stem from general CS difficulties. For example, programming

remains challenging for students when implementing ML algorithms. Therefore, our findings reinforce the call to build AI education on established insights from CS education research instead of "reinventing the wheel" [13].

The identified AI-specific difficulties point to interesting aspects related to fundamental characteristics of AI systems, such as their focus on probabilistic predictions and their ability to handle uncertainty. For example, our investigations indicate that the difficulty of **evaluating whether real-world systems use AI approaches** partly stems from students' limited awareness that AI systems can deal with uncertainty. This aligns with prior research showing that students struggle to identify key features of AI technologies, particularly prediction [63]. However, our findings indicate a broader pattern: students in our sample seem to struggle to develop a coherent understanding of what constitutes AI. In this context, recognising AI systems' ability to handle uncertainty may point to a potential threshold concept, as this requires a fundamental shift in how computing systems are understood. This is particularly important as the ability to identify technologies that use AI approaches is explicitly listed in several frameworks that define what to teach about AI [31, 37] and constitutes a prerequisite for critically evaluating AI technologies and understanding their social impact.

The difficulty of **applying AI concepts to real-world examples** concerns a fundamental goal of (CS) education. As application and transfer are also issues in other CS topics, the question arises whether AI systems present a unique challenge in this regard due to their high complexity and limited transparency. At the same time, this difficulty may also reflect teachers' still-developing PCK for teaching about AI [1]. However, given that all our participating teachers were well qualified and had attended AI-specific PD programs, we argue that application and transfer within AI education is indeed particularly challenging. In this context, our findings suggest that the limited transparency of AI systems may affect students' ability to form stable mental models of AI concepts, thereby hindering transfer to real-world examples. Moreover, the question of what level black boxes should be opened [45] gains a new dimension in AI education.

The identified difficulty of **evaluating AI models** can be interpreted in light of Tedre et al.'s CT 2.0 framework, which documents several aspects that distinguish traditional programming education from education aimed at creating ML models [51]. For example, Tedre et al. argue that new strategies for testing are necessary in the context of ML systems. In contrast to deterministic programs, ML systems require evaluating models' predictions and validating them against datasets. Our findings empirically support this argument by showing that students encounter specific difficulties when engaging with model evaluation. Beyond the reasons identified in our data, this difficulty may also be grounded in students' (and teachers') limited familiarity with such testing strategies. Research in this area remains in an exploratory phase with limited guidance for classroom practice [40]. Our findings therefore provide an important starting point for investigating the challenges that students face in this context. This is particularly relevant because designing and evaluating computing systems — in this case, AI systems — constitutes a core objective of CS education [18].

Regarding the **distinguishing between knowledge- and data-based approaches**, our results indicate that this difficulty is closely related to students' perceptions of AI. While prior research has identified fragmented preconceptions of AI, often shaped by everyday experiences [36], our findings extend this perspective to a more conceptual level by showing that knowledge-based approaches are largely absent in students' understanding of AI. This may reflect the widespread tendency that AI is largely seen as ML, but may also be influenced by the limited attention given to knowledge-based approaches in current AI education initiatives [3]. Addressing this difficulty, therefore, requires targeted conceptual change approaches as well as a more explicit integration of learning objectives related to knowledge-based systems in AI education. At the same time, it remains an open question to what extent the instructional approach itself — in our case, the use of decision tree learning — may exacerbate this difficulty.

While the identification of students' difficulties was mainly researcher-led, the development of possible strategies to address them (RQ2) was highly participatory, revealing a particular strength of the chosen PAR approach: Teachers were

able to reflect on the difficulties in relation to their individual lessons and, despite their different ways of teaching, found agreement on certain strategies to address them. We therefore argue that the difficulties reflect fundamental challenges in learning about AI rather than mere artefacts of specific classroom implementations. Furthermore, the proposed strategies can be interpreted as emerging design principles for AI education, providing a basis for future research within and beyond this study. For example, teachers' suggestion to include real-world knowledge-based systems together with their consideration of using decision trees to contrast data- and knowledge-based approaches, indicates the need for targeted conceptual change interventions addressing students' predominantly ML-centric views of AI. Furthermore, teachers' expressed desire to allocate more time to teaching about neural networks may reflect not only curricular constraints but also research suggesting that ANNs may involve threshold concepts [2]. Finally, teachers' emphasis on more explicitly addressing what constitutes an AI problem points to the importance of an ontological foundation when teaching about AI. It also reflects the epistemological shift highlighted in Tedre et al.'s CT 2.0 framework [51].

Limitations. Given that this first PAR cycle intentionally **foregrounded teachers' perspectives**, it is important to reflect on how teachers' perceptions of students' difficulties are shaped by their own CK and PCK. When teaching a topic for the first time, teachers themselves may face uncertainties, for example, when selecting appropriate examples, reducing complexity, or working with unfamiliar tools. Such uncertainties may influence how student learning difficulties are perceived and articulated. At the same time, teachers' diagnostic expertise and their observations across multiple student groups and lessons provide valuable insights, particularly in the early stages of implementing a new curriculum. One limitation concerns the **sample of teachers**, as it must be assumed that the lessons of these particularly motivated teachers are not representative. Nevertheless, it is likely that the difficulties identified here also occur in less elaborate teaching situations, perhaps even more. In addition, only four teachers provided **teaching materials**, limiting the scope of material-based conclusions. Additional insights into the lessons came from existing textbooks used by the teachers. However, teaching materials only served as supplementary data, while interviews and learning assessments formed the core of the analysis. Regarding the data analysis, we did not apply an explicit **test of coding reliability**. While this is consistent with GT, where coding is understood as an iterative and interpretative process, we enhance transparency and traceability by systematically documenting the process through memos and by conducting peer debriefing at different stages within the research group. The final results were validated with the participating teachers in the workshop.

When conducting research in the field, it is important to consider the findings' **generalisability**. Our study is situated within a state-mandated curriculum that defines a specific selection of AI learning objectives, which, for example, does not explicitly address the current widely discussed topics of generative AI and LLMs. While this limits direct transfer to teaching approaches focusing on generative AI, the investigated curriculum includes foundational AI concepts that serve as a basis for other approaches, such as LLMs. We therefore consider our difficulties to be relevant beyond the specific curriculum. Taken together, we are aware that this study cannot be reproduced precisely. However, by applying a grounded theory-informed analysis and triangulating data from quantitative and qualitative instruments, we achieve a higher level of abstraction following the different coding phases. We therefore argue that our findings could be investigated in other teaching environments covering similar AI concepts. Moreover, they can lead to further research in AI education, moving beyond the exploratory phase by systematically investigating specific learning difficulties.

6 Conclusion and Future Work

This study systematically investigated the first year of teaching the AI-related part of a compulsory Year 11 CS curriculum. Using participatory action research, we identified four AI-specific learning difficulties, primarily derived from teachers'

perspectives, complemented by a student learning assessment: (1) evaluating whether real-world systems use AI approaches, (2) distinguishing between knowledge-based and data-based approaches, (3) evaluating AI models, and (4) applying AI concepts to real-world examples. Furthermore, our results revealed possible reasons for these difficulties, including students' preconceptions, the lack of adequate tools, and the high complexity of AI systems. To address these difficulties, implications for practice were derived from the data and elaborated collaboratively with teachers, such as discussing indicators of AI usage, analysing AI and non-AI systems, and working with varied pre-structured datasets.

Overall, our results provide a foundation for further empirical research and theory-building on teaching and learning about AI in K-12 education. They also contribute to the evidence-based advancement of teaching about AI, both in the local context of this study and by informing the development of curricula and teaching resources in other contexts. Another contribution of this study is the diagnostic value of the identified difficulties. For example, our findings can inform the development of assessment instruments like concept inventories, or can be used in professional development programs to raise teachers' awareness and better prepare them to teach about AI. Finally, this study demonstrates how PAR can be applied to systematically investigate the teaching and learning of newly introduced CS topics, enabling immediate classroom improvement while simultaneously laying a basis for theory-building. This is especially relevant when these topics have already been integrated into curricula, but evidence from research is limited.

Implications for Future Research. Our findings point to several directions for future research. For example, it should be investigated how education influences students' preconceptions of AI and how teaching can foster the development of explicit criteria for identifying AI systems. Research is also needed on low-floor/high-ceilings tools that empower students to create their own AI systems, as well as on strategies for evaluating AI systems in the classroom. Finally, future work should examine whether decision trees as an introductory ML approach support or hinder students' ability to distinguish between knowledge- and data-based AI approaches. Regarding the next PAR cycle, we will additionally investigate the proposed implications for addressing the identified difficulties with a particular focus on students.

Acknowledgments

While preparing this work, the authors used ChatGPT to improve readability and language. They reviewed and edited its output as needed and take full responsibility for the publication's content..

References

[1] Salomey Afua Addo, Forster D. Ntow, and Sue Sentance. 2025. Exploring Ghanaian Computing Teachers' AI Pedagogical Content Knowledge (PCK) in Junior High Schools. In *Proceedings of the ACM Global Computing Education Conference 2025 - Volume 1* (Gaborone, Botswana) *(CompEd 2025)*. Association for Computing Machinery, New York, NY, USA, 29–35. doi:10.1145/3736181.3747124

[2] Becky Allen, Marie Devlin, and A.Stephen McGough. 2021. Using the One Minute Paper to Gain Insight into Potential Threshold Concepts in Artificial Intelligence Courses. In *Proceedings of the 5th Conference on Computing Education Practice* (Durham, United Kingdom) *(CEP '21)*. Association for Computing Machinery, New York, NY, USA, 21–24. doi:10.1145/3437914.3437974

[3] Omaima Almatrafi, Aditya Johri, and Hyuna Lee. 2024. A systematic review of AI literacy conceptualization, constructs, and implementation and assessment efforts (2019–2023). *Computers and Education Open* 6 (2024), 100173. doi:10.1016/j.caeo.2024.100173

[4] Arne Bewersdorff, Xiaoming Zhai, Jessica Roberts, and Claudia Nerdel. 2023. Myths, mis-and preconceptions of artificial intelligence: A review of the literature. *Computers and Education: Artificial Intelligence* 4 (2023), 100143.

[5] Astrid Carolus, Martin J Koch, Samantha Straka, Marc Erich Latoschik, and Carolin Wienrich. 2023. MAILS-Meta AI literacy scale: Development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change-and meta-competencies. *Computers in Human Behavior: Artificial Humans* 1, 2 (2023), 100014.

[6] John W Creswell and Cheryl N Poth. 2016. *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications, Thousand Oaks, CA.

[7] Yun Dai, Ang Liu, Jianjun Qin, Yanmei Guo, Morris Siu-Yung Jong, Ching-Sing Chai, and Ziyan Lin. 2023. Collaborative construction of artificial intelligence curriculum in primary schools. *Journal of engineering education* 112, 1 (2023), 23–42.

[8] Anna Eckerdal, Robert McCartney, Jan Erik Moström, Mark Ratcliffe, Kate Sanders, and Carol Zander. 2006. Putting threshold concepts into context in computer science education. In *Proceedings of the 11th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education* (Bologna, Italy) *(ITICSE '06)*. Association for Computing Machinery, New York, NY, USA, 103–107. doi:10.1145/1140124.1140154

[9] Ingo Eilks and Bernd Ralle. 2002. Participatory Action Research within chemical education. In *Research in chemical education-What does this mean.* Shaker, Aachen, 87–98.

[10] Yannik Fleischer, Susanne Podworny, and Rolf Biehler. 2024. Teaching and learning to construct data-based decision trees using data cards as the first introduction to machine learning in middle school. *Statistics Education Research Journal* 23, 1 (2024), 3–3.

[11] William Gelder, Xueru Yu, David Touretzky, Christina Gardner-McCune, and Judith Uchidiuno. 2025. From Lecture Hall to Homeroom: Co-Designing an AI Elective with Middle School CS Teachers. *International Journal of Artificial Intelligence in Education* 35 (2025), 1749–1792.

[12] Christiane Gresse von Wangenheim, Jean CR Hauck, Fernando S Pacheco, and Matheus F Bertonceli Bueno. 2021. Visual tools for teaching machine learning in K-12: A ten-year systematic mapping. *Education and Information Technologies* 26, 5 (2021), 5733–5778.

[13] Shuchi Grover. 2024. Teaching AI to K-12 Learners: Lessons, Issues, and Guidance. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1* (Portland, OR, USA) *(SIGCSE 2024)*. Association for Computing Machinery, New York, NY, USA, 422–428. doi:10.1145/3626252.3630937

[14] Sara Guerreiro-Santalla, Francisco Bellas, and Alma Mallo. 2022. Introducing high school students in natural interaction through the Robobo educational robot. In *Iberian Robotics conference*. Springer, Springer, Cham, 500–512.

[15] Mark Guzdial. 2013. Exploring hypotheses about media computation. In *Proceedings of the Ninth Annual International ACM Conference on International Computing Education Research* (San Diego, San California, USA) *(ICER '13)*. Association for Computing Machinery, New York, NY, USA, 19–26.

[16] Marie Hornberger, Arne Bewersdorff, and Claudia Nerdel. 2023. What do university students know about Artificial Intelligence? Development and validation of an AI literacy test. *Computers and Education: Artificial Intelligence* 5 (2023), 100165.

[17] Franz Jetzinger, Sven Baumer, and Tilman Michaeli. 2024. Artificial intelligence in compulsory K-12 computer science classrooms: A scalable professional development offer for computer science teachers. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*. Association for Computing Machinery, New York, NY, USA, 590–596.

[18] Yasmin Kafai, R Benjamin Shapiro, Franz Jetzinger, Tilman Michaeli, Matti Tedre, Henriikka Vartiainen, Netta Iivari, Line Have Musaeus, Ole Sejer Iversen, Safinah Ali, et al. 2025. Youth as designers of artificial intelligence and machine learning technologies: What do we know about the opportunities and challenges of K-12 students creating their own applications?. In *Proceedings of the 19th International Conference of the Learning Sciences-ICLS 2025, pp. 2260-2268*. International Society of the Learning Sciences, Helsinki, Finland, 2260–2268.

[19] Keunjae Kim, Kyungbin Kwon, Anne Ottenbreit-Leftwich, Haesol Bae, and Krista Glazewski. 2023. Exploring middle school students' common naive conceptions of Artificial Intelligence concepts, and the evolution of these ideas. *Education and Information Technologies* 28, 8 (2023), 9827–9854.

[20] Seonghun Kim, Yeonju Jang, Woojin Kim, Seongyune Choi, Heeseok Jung, Soohwan Kim, and Hyeoncheol Kim. 2021. Why and what to teach: AI curriculum for elementary school. In *proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. AAAI Press, Palo Alto, California USA, 15569–15576.

[21] Siu-Cheung Kong, Man-Yin William Cheung, and Olson Tsang. 2024. Developing an artificial intelligence literacy framework: Evaluation of a literacy course for senior secondary students using a project-based learning approach. *Computers and Education: Artificial Intelligence* 6 (2024), 100214.

[22] Moritz Kreinsen, Anna Rabe, Finja Grospietsch, and Sandra Schulz. 2024. Leveraging Conceptual Change regarding Artificial Intelligence in Computer Science Education. In *Proceedings of the 24th Koli Calling International Conference on Computing Education Research*. ACM, New York, NY, USA, 12–17.

[23] Matthias Carl Laupichler, Alexandra Aster, and Tobias Raupach. 2023. Delphi study for the development and preliminary validation of an item set for the assessment of non-experts' AI literacy. *Computers and Education: Artificial Intelligence* 4 (2023), 100126.

[24] Irene Lee, Safinah Ali, Helen Zhang, Daniella DiPaola, and Cynthia Breazeal. 2021. Developing Middle School Students' AI Literacy. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education* (Virtual Event, USA) *(SIGCSE '21)*. Association for Computing Machinery, New York, NY, USA, 191–197. doi:10.1145/3408877.3432513

[25] Seung Lee, Bradford Mott, Anne Ottenbreit-Leftwich, Adam Scribner, Sandra Taylor, Kyungjin Park, Jonathan Rowe, Krista Glazewski, Cindy E Hmelo-Silver, and James Lester. 2021. AI-infused collaborative inquiry in upper elementary school: A game-based learning approach. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35. AAAI Press, Palo Alto, California USA, 15591–15599.

[26] Sang Joon Lee and Kyungbin Kwon. 2024. A systematic review of AI education in K-12 classrooms from 2018 to 2023: Topics, strategies, and learning outcomes. *Computers and Education: Artificial Intelligence* 6 (2024), 100211.

[27] Laura C Leviton and Mathew D Trujillo. 2017. Interaction of theory and practice to assess external validity. *Evaluation Review* 41, 5 (2017), 436–471.

[28] Annabel Lindner, Stefan Seegerer, and Ralf Romeike. 2019. Unplugged Activities in the Context of AI. In *Informatics in Schools. New Ideas in School Informatics*. Springer International Publishing, Cham, 123–135.

[29] Xiaofan Liu and Baichang Zhong. 2024. A systematic review on how educators teach AI in K-12 education. *Educational Research Review* 45 (2024), 100642. doi:10.1016/j.edurev.2024.100642

[30] Marguerite G Lodico, Dean T Spaulding, and Katherine H Voegtle. 2010. *Methods in educational research: From theory to practice.* John Wiley & Sons, San Francisco, USA.

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

[32] Radu Mariescu-Istodor and Ilkka Jormanainen. 2019. Machine learning for high school students. In *Proceedings of the 19th Koli Calling International Conference on Computing Education Research* (Koli, Finland) *(Koli Calling '19)*. Association for Computing Machinery, New York, NY, USA, Article 10, 9 pages. doi:10.1145/3364510.3364520

[33] Ramon Mayor Martins and Christiane Gresse Von Wangenheim. 2023. Findings on teaching machine learning in high school: A ten-year systematic literature review. *Informatics in Education* 22, 3 (2023), 421–440.

[34] Ramon Mayor Martins, Christiane Gresse von Wangenheim, Marcelo Fernando Rauber, and Jean Carlo Hauck. 2024. Machine learning for all!—introducing machine learning in middle and high school. *International Journal of Artificial Intelligence in Education* 34, 2 (2024), 185–223.

[35] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 72 (Nov. 2019), 23 pages. doi:10.1145/3359174

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

[37] Fengchun Miao, Kelly Shiohira, et al. 2024. *AI competency framework for students*. UNESCO Publishing, France. doi:10.54675/JKJB9835

[38] Tilman Michaeli, Ralf Romeike, and Stefan Seegerer. 2022. What students can learn about artificial intelligence–recommendations for K-12 computing education. In *IFIP World Conference on Computers in Education*. Springer, Cham, 196–208.

[39] Tilman Michaeli, Stefan Seegerer, Lennard Kerber, and Ralf Romeike. 2023. Data, trees, and forests-decision tree learning in K-12 education. In *The Third Teaching Machine Learning and Artificial Intelligence Workshop*. PMLR, Grenoble, France, 37–41.

[40] Luis Morales-Navarro, Meghan Shah, and Yasmin B Kafai. 2024. Not Just Training, Also Testing: High School Youths' Perspective-Taking through Peer Testing Machine Learning-Powered Applications. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1* (Portland, OR, USA) *(SIGCSE 2024)*. Association for Computing Machinery, New York, NY, USA, 881–887. doi:10.1145/3626252.3630899

[41] Andreas Mühling and Gregor Große-Bölting. 2023. Novices' conceptions of machine learning. *Computers and Education: Artificial Intelligence* 4 (2023), 100142.

[42] Alannah Oleson, Meron Solomon, and Amy J. Ko. 2020. Computing Students' Learning Difficulties in HCI Education. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3313831.3376149

[43] Nicolas Pope, Juho Kahila, Henriikka Vartiainen, and Matti Tedre. 2025. Children's AI Design Platform for Making and Deploying ML-Driven Apps: Design, Testing, and Development. *IEEE Trans. Learn. Technol.* 18 (Jan. 2025), 130–144. doi:10.1109/TLT.2025.3529994

[44] Yizhou Qian and James Lehman. 2017. Students' Misconceptions and Other Difficulties in Introductory Programming: A Literature Review. *ACM Trans. Comput. Educ.* 18, 1, Article 1 (Oct. 2017), 24 pages. doi:10.1145/3077618

[45] Mitchel Resnick and Brian Silverman. 2005. Some reflections on designing construction kits for kids. In *Proceedings of the 2005 Conference on Interaction Design and Children* (Boulder, Colorado) *(IDC '05)*. Association for Computing Machinery, New York, NY, USA, 117–122. doi:10.1145/1109540.1109556

[46] Saman Rizvi, Jane Waite, and Sue Sentance. 2023. Artificial Intelligence teaching and learning in K-12 from 2019 to 2022: A systematic literature review. *Computers and Education: Artificial Intelligence* 4 (2023), 100145.

[47] Ismaila Temitayo Sanusi, Kissinger Sunday, Solomon Sunday Oyelere, Jarkko Suhonen, Henriikka Vartiainen, and Markku Tukiainen. 2024. Learning machine learning with young children: Exploring informal settings in an African context. *Computer Science Education* 34, 2 (2024), 161–192.

[48] Anselm Strauss and Juliet Corbin. 1998. Basics of qualitative research techniques.

[49] Elisabeth Sulmont, Elizabeth Patitsas, and Jeremy R. Cooperstock. 2019. Can You Teach Me To Machine Learn?. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education* (Minneapolis, MN, USA) *(SIGCSE '19)*. Association for Computing Machinery, New York, NY, USA, 948–954. doi:10.1145/3287324.3287392

[50] Tarzimah Tambychik and Thamby Subahan Mohd Meerah. 2010. Students' Difficulties in Mathematics Problem-Solving: What do they Say? *Procedia - Social and Behavioral Sciences* 8 (2010), 142–151. doi:10.1016/j.sbspro.2010.12.020 International Conference on Mathematics Education Research 2010 (ICMER 2010).

[51] Matti Tedre, Peter Denning, and Tapani Toivonen. 2021. CT 2.0. In *Proceedings of the 21st Koli Calling International Conference on Computing Education Research* (Joensuu, Finland) *(Koli Calling '21)*. Association for Computing Machinery, New York, NY, USA, Article 3, 8 pages. doi:10.1145/3488042.3488053

[52] David Touretzky, Christina Gardner-McCune, Fred Martin, and Deborah Seehorn. 2019. Envisioning AI for K-12: What should every child know about AI?. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. AAAI Press, Palo Alto, California USA, 9795–9799.

[53] UNESCO. 20022. AI Curricula: A Mapping of Government-Endorsed AI Curricula.

[54] Gia Minh Vo, Moritz Kreinsen, Rina Ferdinand, and Nils Pancratz. 2024. Draw, Find, and Describe AI for Me: Investigating Learners' Conceptions of Artificial Intelligence. In *2024 IEEE Global Engineering Education Conference (EDUCON)*. IEEE, Kos Island, Greece, 01–05. doi:10.1109/EDUCON60312.2024.10578628

[55] Stella Vosniadou. 1994. Capturing and modeling the process of conceptual change. *Learning and instruction* 4, 1 (1994), 45–69.

[56] Yuchen Wang, Shangxin Guo, Lin Ling, and Chee Wei Tan. 2024. Nemobot: Crafting Strategic Gaming LLM Agents for K-12 AI Education. In *Proceedings of the Eleventh ACM Conference on Learning @ Scale* (Atlanta, GA, USA) *(L@S '24)*. Association for Computing Machinery, New York,

NY, USA, 393–397. doi:10.1145/3657604.3664671

[57] Patrick Weber, Marc Pinski, and Lorenz Baum. 2023. Toward an objective measurement of AI literacy. In *Pacific Asia Conference on Information Systems*. Association for Information Systems, Nanchang, China, 17 pages.

[58] Franz Emanuel Weinert and Werner Zielinski. 1977. Lernschwierigkeiten–schwierigkeiten des schülers oder der schule. *Unterrichtswissenschaft* 5, 4 (1977), 292–304.

[59] Robert Whyte, Diana Kirby, and Sue Sentance. 2024. Secondary Students' Emerging Conceptions of AI: Understanding AI Applications, Models, Engines and Implications. In *Proceedings of the 2024 Conference on United Kingdom & Ireland Computing Education Research*. ACM, New York, NY, USA, 1–7.

[60] Randi Williams, Safinah Ali, Nisha Devasia, Daniella DiPaola, Jenna Hong, Stephen P Kaputsos, Brian Jordan, and Cynthia Breazeal. 2023. AI+ ethics curricula for middle school youth: Lessons learned from three project-based curricula. *Journal of AI in Education* 33, 2 (2023), 325–383.

[61] Benjamin Xie, Parth Sarin, Jacob Wolf, Raycelle CC Garcia, Victoria Delaney, Isabel Sieh, Anika Fuloria, Deepak Varuvel Dennison, Christine Bywater, and Victor R Lee. 2024. Co-designing AI education curriculum with cross-disciplinary high school teachers. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. AAAI Press, Washington, DC, USA, 23146–23154.

[62] Miao Yue, Morris Siu-Yung Jong, and Yun Dai. 2022. Pedagogical design of K-12 artificial intelligence education: A systematic review. *Sustainability* 14, 23 (2022), 15620.

[63] Helen Zhang, Irene Lee, Safinah Ali, Daniella DiPaola, Yihong Cheng, and Cynthia Breazeal. 2023. Integrating ethics and career futures with technical learning to promote AI literacy for middle school students: An exploratory study. *Journal of AI in Education* 33, 2 (2023), 290–324.