Impact of a Data-driven Teaching Approach on 9th Graders Conceptual Understanding of Machine Learning

Erik Marx[1], **Thiemo Leonhardt**[2], **Nadine Bergner**[2]

[1]Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI) Dresden/Leipzig, Technische Universität Dresden
[2]RWTH Aachen University, Computer Science Education Teaching and Research Unit

Abstract

This study aimed to investigate the impact of a data-driven teaching approach on students' conceptual understanding of machine learning (ML). To this end, an exemplary intervention was designed and evaluated using a pre- and post-test design and a German-language Concept Inventory on Machine Learning. A total of 83 German ninth-grade students participated in the study. The results revealed significant learning gains related to data handling and the ML workflow. In contrast, conceptions about the inner workings of ML models largely persisted. The effectiveness of the intervention varied depending on context, with greater gains observed in the text generation domain than in facial recognition, highlighting challenges in cross-contextual transfer of understanding. A regression analysis showed no significant influence of students' pre-instructional conceptions on learning outcomes. These findings demonstrate both the potential and the limitations of data-driven teaching approaches and emphasize the need for more explicit engagement with learners' misconceptions to foster deeper conceptual change.

Supplementary files — https://osf.io/94vfn

1 Introduction

Pupils interact with machine learning (ML) applications in their everyday lives. However, due to the opaque nature of these technologies, they develop intuitive explanations on the functionality of these systems. These conceptions often fail to reflect the underlying principles of ML-based applications (Vartiainen et al. 2021; Bewersdorff et al. 2023; Szczuka et al. 2022). In response, there have been increasing efforts to promote foundational AI literacy in school. Importantly, such literacy must extend beyond factual or technical knowledge, encompassing a conceptual understanding of ML processes. This foundational knowledge is considered a prerequisite for students to meaningfully engage with and interpret the wide variety of ML-based technologies they encounter (Schaper et al. 2023; Tedre, Denning, and Toivonen 2021). To support this goal, numerous educational interventions have been developed to introduce core ML concepts to learners at the K-12 level (Morales-Navarro and Kafai 2024). The majority of these introductory approaches can be broadly categorized as **data-driven approaches** (DdA), which aim to make ML systems more transparent by emphasizing the role of data in learning processes (Morales-Navarro and Kafai 2024; Waite et al. 2023). The present study seeks to examine which fundamental ML concepts are addressed through a DdA-based intervention and to what extent this approach supports students in transforming their pre-instructional conceptions about ML systems.

2 Theoretical Background

2.1 Introducing ML via a Data-driven Approach

In their review of instructional approaches to teaching ML, Morales-Navarro and Kafai (2024) identify the DdA as the most prevalent strategy for introducing ML in K-12 settings, accounting for 55% of the studies they analyzed. The DdA is characterized by its emphasis on the influence of data on model behavior while deliberately black-boxing the internal mechanisms of learning algorithms and models in favor of the ML workflow and the role of human decision-making within that process (Morales-Navarro and Kafai 2024; Bilstrup et al. 2022; Waite et al. 2023). This abstraction aims to scaffold foundational ML concepts such as data handling and model evaluation, making them more accessible and tangible for learners (Morales-Navarro and Kafai 2024; Touretzky, Gardner-McCune, and Seehorn 2022). These findings are corroborated by Waite et al. (2023), who outline four abstraction layers in ML teaching. The third layer, "Model," closely aligns with the DdA and includes activities such as data collection, preprocessing, visualization, model training, and evaluation. This layer featured in approximately 60% of the studies they reviewed. Similarly, Casal-Otero et al. (2023), in their literature review identify that a number of learning experiences include the ML workflow and its basic principles. Similarly, Perach and Alexandron (2024) found that the practical components of ML curricula emphasize iterative experimentation, data handling and model evaluation, which are hallmarks of the DdA.

In line with this, Höper and Schulte (2024) argue that conceptual models which are an idealized representation of computational concepts should be treated as learning goals in their own right. These models support learners in interpreting digital artifacts and reasoning about abstract ideas. From this perspective, the ML workflow—encompassing the steps of data acquisition, preprocessing, training, testing, evaluation and application—can be regarded as a conceptual model central to the DdA (Hitron et al. 2019; Tedre, Denning, and Toivonen 2021). To conclude, the DdA seeks to cultivate conceptual understanding by engaging students with the procedural steps of ML and data handling and by black-boxing the inner workings of ML models. Consequently, this raises a central research question:

RQ1: Which ML-related concepts can be effectively conveyed through a data-driven approach, particularly when this approach is structured around the ML workflow as a guiding conceptual model?

2.2 Students' Conceptions of ML

Although various AI literacy frameworks have been proposed, there remains considerable ambiguity regarding which content should be taught and at what level of depth (Broll and Grover 2023; Sanusi et al. 2023; Waite et al. 2023). This uncertainty stems in part from limited understanding of which concepts are actually accessible and comprehensible to learners (Zhang, Perry, and Lee 2024). A productive approach to this challenge lies in examining students' pre-instructional conceptions. According to constructivist learning theory, students do not approach new content as blank slates but bring with them pre-existing ideas, often developed informally or through everyday media exposure (diSessa 2018; Vosniadou 2013). As Gnoth and Novak (2025) highlight, students' preconceptions can present a significant barrier to the development of AI literacy.

Prior research has documented a range of pre-instructional conceptions, particularly in the broader context of AI. The following section synthesizes the most prevalent conceptions specifically related to ML, drawing on the terms introduced by Marx, Witt, and Leonhardt (2024).

- **Programmed Behaviour**: Students believe that the functionality of ML models are entirely—or at least partially—predefined by human developers (Marx, Witt, and Leonhardt 2024; Kim et al. 2023; Zhang, Perry, and Lee 2024). This view aligns with a general belief that computers operate strictly through explicit programming (Rücker and Pinkwart 2016). Several studies report that students assume AI systems are "intelligently programmed" and do not require training at all (Mertala and Fagerlund 2024; Kim et al. 2023; Kreinsen and Schulz 2021).
- **Storage**: While students are often aware that ML systems rely on data, they frequently lack clarity about how this data is actually used. This leads to conceptions that data is merely stored and later retrieved during model use. For example, students may believe an image recognition model works by comparing new inputs to stored examples (Marx, Witt, and Leonhardt 2024; Mühling and Große-Bölting 2023; Ude, Vo, and Pancratz 2024; Kreinsen and Schulz 2021). More broadly, computers themselves are frequently perceived as all-knowing data repositories (Rücker and Pinkwart 2016).
- **Exactness**: ML is commonly perceived as producing precise, infallible results, with students overlooking its probabilistic and non-deterministic nature (Marx, Witt, and Leonhardt 2024; Gnoth and Novak 2025). This belief is often rooted in traditional computational thinking (CT), where systems are expected to behave deterministically (Tedre, Denning, and Toivonen 2021; Perach and Alexandron 2024). It is also reflected in general beliefs on AI being fair, objective or even omniscient (Kim et al. 2023; Lee et al. 2021; Rücker and Pinkwart 2016; Belghith et al. 2024).
- **Continuous Learning**: Students believe that ML models continue to learn autonomously even after deployment partly by being trained by the user (Marx, Witt, and Leonhardt 2024; Ude, Vo, and Pancratz 2024). This notion aligns with general associations between AI and learning (Mertala and Fagerlund 2024; Vandenberg and Mott 2023; Mühling and Große-Bölting 2023). It is often tied to anthropomorphic conceptions, in which AI is likened to human cognition (Bewersdorff et al. 2023; Belghith et al. 2024).
- **Autonomous Data Acquisition**: Sometimes, students assume that ML models need to search for and select the data they require. This conception is particularly prevalent in relation to systems like chatbots, voice assistants, or search engines (Marx, Witt, and Leonhardt 2024; Ude, Vo, and Pancratz 2024). This is typically paired with the misconception that AI systems can autonomously interpret and understand data (Ude, Vo, and Pancratz 2024).
- **Quality through Quantity**: This misconception reflects the belief that ML systems function effectively simply by being trained on large amounts of data, regardless of its quality (Kim et al. 2023; Yang et al. 2018; Sanusi et al. 2023).

Overall, students' conceptions can be attributed to **two primary factors**. Firstly, a number of conceptions arise from generalized beliefs about AI, where students interpret ML systems through an anthropomorphic understanding of intelligence and autonomy (Bewersdorff et al. 2023; Mertala and Fagerlund 2024; Belghith et al. 2024; Marx, Witt, and Leonhardt 2024). Secondly, other conceptions stem from the transfer of traditional CT to the ML domain. As Hitron et al. (2019) observed, when students struggle to explain the internal logic of ML systems, they tend to revert to familiar CT concepts, assuming, for instance, that ML behavior is fully programmed or that results are exact and objective.

Since ML requires fundamentally different mental models than classical rule-based computation (Tedre, Denning, and Toivonen 2021; Perach and Alexandron 2024), the aforementioned preconceptions can hinder the development of appropriate understanding. We hypothesized that misconceptions concerning the internal functioning of ML models—such as *Storage*—would persist or even be reinforced due to the black-box nature of the DdA. In contrast, we expected a reduction in misconceptions that are more explicitly addressed through the conceptual model embedded in the ML workflow—such as *Continuous Learning*.

RQ2: How do students' conceptions change through a data-driven teaching approach?

2.3 Conceptual Change Theory

Research on students' learning processes is commonly discussed under the umbrella of **Conceptual Change** (CC) theory, which investigates how individuals transform central, organizing concepts into fundamentally different, scientifically normative ones (Posner et al. 1982; Vosniadou 2013). From this perspective, learning entails more than knowledge accumulation—it requires a deep restructuring of prior beliefs.

Within the CC research tradition, various theoretical models have been proposed that differ primarily in their assumptions about the coherence of students' knowledge (diSessa 2013). Framework theory, as developed by Vosniadou (2013), represents a prominent coherence-based account of CC. It posits that children begin with intuitive frameworks that are broad in scope that serve as implicit theories through which learners make sense within a given domain. Learning, in this theory, primarily involves the gradual refinement of these initial frameworks through the incorporation of new information (Vosniadou 2013).

In contrast, Knowledge in Pieces (KiP), introduced by diSessa (2013, 2018), conceptualizes learning as the reorganization of fine-grained, context-sensitive knowledge elements rather than large-scale theoretical structures. Rather than assuming that students possess coherent frameworks, KiP emphasizes the fragmented and emergent nature of intuitive knowledge. It portrays learning as an incremental and contextual process in which small knowledge elements are refined and gradually assembled into more stable conceptual systems.

Despite their differences, all major CC theories share "the constructivist idea that 'old' ideas constrain learning." (diSessa 2013, p.31). Given the characteristics of the DdA we hypothesize that prior CT-conceptions may conflict with this approach as learners may underestimate the autonomy and statistical nature of ML systems. We therefore outline the following research question:

> *RQ3: Is learning of ML-concepts via the data-driven approach influenced by students' pre-instructional conceptions?*

3 Related Work

A growing body of research has explored students' pre-instructional conceptions of ML, as summarized in the previous section. However, fewer studies have systematically investigated how these conceptions change in response to instructional interventions, particularly within introductory learning settings. This section highlights selected studies that address conceptual development in the context of ML and AI education, focusing on hypotheses and findings relevant to our work.

Vartiainen et al. (2021) conducted an exploratory case study with Finnish sixth-grade students that centered on co-designing ML applications within a constructionist pedagogical framework. The study demonstrated that co-design tasks can initiate productive engagement with ML concepts, although students often struggled to relate these experiences to familiar, real-world applications. Similarly, Mühling and Große-Bölting (2023) investigated secondary students' evolving conceptions of ML after interacting with a reinforcement learning-inspired game (Hexapawn). Even though there were no reflection tasks, some conceptual shifts occurred although moderate. In a design-based experimental study, Hitron et al. (2019) examined how varying degrees of black-boxing influence students' understanding of ML, particularly the process of data labeling and evaluation. Results showed that only participants with access to the full ML workflow demonstrated substantial gains in their ability to explain data handling processes. Complementing these perspectives, Kim et al. (2023) conducted a qualitative study tracing the evolution of naive conceptions about AI of middle school students. The study found that targeted questioning and reflective prompts from educators were instrumental in facilitating CC. While most studies focus on ML-specific interventions, Baldoni et al. (2024) investigated the effects of a robotics-based curriculum grounded in classical CT concepts. No significant differences in AI-related conceptions were found between pre- and post-tests, suggesting that conceptions of CT and AI may develop along independent trajectories. Finally, Kreinsen et al. (2024) explicitly situated their study within the framework of CC. They reported statistically significant improvements related to the *Programmed Behaviour* conception following the use of CC texts, although effect sizes and learning gains were not reported.

Collectively, these studies converge on three key insights. First, students enter learning environments with robust preconceptions about ML. Second, although some conceptual development occurs through interaction alone, these effects tend to be moderate unless explicitly scaffolded through reflection, guided inquiry, or designated materials such as CC texts. Third, interventions that are interactive and directly linked to core ML concepts appear more promising for fostering meaningful change in students' conceptual understanding.

4 Methodology

This study employed a pre- and post-test design. To assess students' conceptual understanding of central ML concepts we employed the Concept Inventory on Machine Learning (CIML). The pre-test questionnaire was administered in participating schools by classroom teachers prior to the intervention. Subsequently, the students attended a three-hour ML workshop held at our institution. Participants were drawn from four 9th-grade classes, along with an additional group of interested students from grades 8 to 10. For consistency, only responses from 9th-grade students were included in the analysis. Students had no prior formal instruction in ML within their school curriculum. The post-test was administered immediately following the workshop. Pseudonyms were used to match pre- and post-test responses, and only fully completed, matched datasets were considered for analysis. The final sample consists of 83 german students (53 male, 29 female, 2 unspecified).

4.1 Research Instrument: Concept Inventory on Machine Learning (CIML)

The CIML is a validated German-language concept inventory specifically designed for use in secondary education and aims to identify students conceptions related to ML (Marx, Leonhardt, and Bergner 2025, a transcribed version is included in the supplementary material). The CIML assesses 7 central ML **concepts**:

1. **Model Generation** (MG): The fundamental approach of ML is to derive a model from a set of data that captures the underlying patterns and relationships within that data. This model can then be used to process new, unseen data. Once the model is generated, the original training data is no longer needed for its application.
2. **Data Selection** (DS): ML requires a sufficient amount of data to identify statistical patterns. Developers either collect the necessary data or, in the case of reinforcement learning or use of webcrawlers, define which data should be gathered and processed. Therefore, selecting and evaluating relevant data is the responsibility of the developers.
3. **Statistical Models** (SM): ML models generate probabilistic predictions, where uncertainty is quantified using probability values. This probabilistic nature is a core feature of how ML-models function.
4. **Implementation** (IM): Implementing a ML-model involves defining its architecture and training process. Developers program the model structure and learning procedure, but the model parameters themselves are learned through optimization during training and are not explicitly specified in the code.
5. **Evaluation** (EV): The parameters optimized during training are not directly interpreted or manually checked by developers. Therefore, appropriate evaluation metrics are used to validate model functionality.
6. **Data Quality** (DQ): The data used for ML must adequately represent the underlying problem and contain meaningful statistical patterns. It should be free of bias and errors. Thus, model performance depends not only on the amount of data but especially on its quality.
7. **Phase Separation** (PS): In ML, a clear distinction is made between the training phase and the application phase. During training, the model is adapted to the data—this is where learning takes place. After training during application, new data is processed based on the learned structure.

Since students have varying conceptions of different ML-based applications (Marx, Witt, and Leonhardt 2024; Ude, Vo, and Pancratz 2024), the questionnaire is divided into two **contexts**: one focusing on **facial recognition** (FR) for smartphone unlocking, and the other on **text generation** (TG) using the example of ChatGPT. Each application includes seven questions, with each question addressing one ML concept. Thus, for each concept, there are two questions in the CIML. All questions except those relating to the concept of *Statistical Models* are multiple-response questions. Each question consists of several statements which are independently rated by the students as true or false. Each question contains one correct statement (further called **concept statement**), pertaining to the previously mentioned concepts, and several **distractors**. These distractors are derived from the misconceptions described in section 2.2. Every correct identified statement scores one point. The questions on *Statistical Models* are multiple-choice-questions where students have to choose one of the statements.

4.2 Workshop Concept

To investigate the impact of the DdA on students' understanding of ML, we designed a three-hour workshop that adopts an interactive, activity-oriented approach within a fictional museum setting. The workshop explicitly centers around the ML workflow, as emphasized by Höper and Schulte (2024), and introduces each phase through hands-on, engaging tasks. The workshop is described in detail in (Marx and Bergner 2025) and included in the supplementary files. The central tool employed is **"Teachable Machine"** (TM), a widely used web-based application that exemplifies the DdA approach (Morales-Navarro and Kafai 2024; Gnoth and Novak 2025). TM enables learners to collect training data using a webcam, train simple image classifiers, and evaluate their models in real time within the browser environment. The workshop's overarching narrative revolves around the development of an intelligent museum app capable of recognizing artworks through image-based classification. To simulate this, students work with custom-designed 3D-printed figurines that vary in shape and appearance, thereby introducing realistic data variability.

The workshop unfolds in multiple phases, each designed to scaffold students' conceptual understanding of ML through guided experimentation and reflection. It begins with a short group discussion aimed at activating prior conceptions about AI. Students are presented with everyday AI applications—including *facial recognition* and *text generation*—to facilitate connections with familiar technologies. These are then revisited in a debrief, where the term "machine learning" is introduced and both applications are explicitly framed as ML-based technologies.

The core ML portion begins with an introduction to TM, followed by a walkthrough of the initial workflow steps: data collection and model evaluation. The focus then shifts to a more in-depth engagement with the data collection process. In pairs, students train the model for classifying the 3D-printed figurines for the intelligent museum guide. Using USB webcams, students capture images from multiple angles and are encouraged to exchange figurines between groups to increase dataset diversity. Crucially, students then evaluate their models using prepared test data that includes varied backgrounds and camera angles. They document their results and calculate standard performance metrics including accuracy, recall, and precision. Building on these insights, students are guided to reflect on characteristics of good training data. They examine example images and discuss factors such as blurriness or distracting backgrounds. Based on their findings, the teacher then adds the "Data Preparation" step to the conceptual ML workflow model. Students then refine

Group	Mean	SD	Median	Min	Max	SE
pre	26.41	5.66	25	16	43	0.62
post	31.04	7.01	30	16	51	0.77

Table 1: Descriptive statistics for questionnaire scores in pre- and post-test

their models by retraining and retesting them, thereby experiencing the iterative nature of ML development, which is again added to the ML workflow conceptual model. In the final phase, students deploy their trained models into a simulated museum app. This activity introduces the "Model Application" to the conceptual model and introduces the distinction between model training and deployment.

At the end of the workshop, the teacher gives an insight into how ML models and learning algorithms work using the example of decision trees, adapted from educational materials by Fleischer, Podworny, and Biehler (2024). Students then engage in a learning game designed by Witt et al. (2025) that illustrates the functionality of decision trees through an interactive experience. In summary, while the intervention follows a DdA at its core, it deliberately incorporates insights on the functionality of ML models and learning algorithms in the form of demonstrations and simulations which Morales-Navarro and Kafai (2024) refer to as a DdA "with sprinkles."

5 Results

In this chapter, the results of the pre-post study are presented. The aim was to examine the impact of the intervention on learners' conceptual understanding of ML. Additionally, the influence of learners conception on learning gain was investigated. All analyses were conducted using R, version 4.5.1 (complete R code and anonymized dataset included in the supplementary files).

The mean total score for participants in the post-test ($M = 31.04$, $SD = 7.01$) was significantly higher than their mean score in the pre-test ($M = 26.41$, $SD = 5.66$, see Table 1). A paired-samples t-test was conducted to compare the total scores before and after the intervention. The analysis revealed a statistically significant difference, $t(82) = 8.74$, $p < 0.001$. The assumption of normality of the difference in scores was tested using the Shapiro-Wilk test ($W = 0.988$, $p = 0.542$) and visualization via histogram and QQ plots, which did not indicate a violation (Field 2024).

To further investigate the results, changes in individual question scores between the pre- and post-test were examined. As the assumption of normality was violated for all 14 questions, as indicated by significant Shapiro-Wilk tests, Wilcoxon signed-rank tests were conducted for polytomous question scores, corresponding to multiple-response items, and McNemar's tests for dichotomous question scores, representing multiple-choice items (Field 2024). To account for the multiple comparisons problem, which increases the likelihood of a Type I error (falsely rejecting the null hypothesis), the Bonferroni-Holm correction was applied to the 14 tests. This method adjusts the p-values sequentially, offering greater statistical power than the more conservative Bonferroni correction while still controlling the family-wise error rate (Holm 1979). The adjusted p-values (p_{adj}) are reported in Table 2. Regardless of the context, significant improvements were observed for the concepts *Evaluation* and *Phase Separation*. For the concept *Data Selection*, a significant change was found only in the context of text generation.

Context	Concept	Test	p_{adj}	Gain G
FR	MG	Wilcoxon	1.000	-
FR	DS	Wilcoxon	0.279	-
FR	SM	McNemar	0.633	-
FR	IM	Wilcoxon	0.591	-
FR	EV	Wilcoxon	< 0.001	0.305
FR	DQ	Wilcoxon	0.279	-
FR	PS	Wilcoxon	< 0.001	0.283
TG	MG	Wilcoxon	1.000	-
TG	DS	Wilcoxon	< 0.001	0.517
TG	SM	McNemar	0.279	-
TG	IM	Wilcoxon	0.313	-
TG	EV	Wilcoxon	< 0.001	0.383
TG	DQ	Wilcoxon	0.168	-
TG	PS	Wilcoxon	< 0.001	0.359

Table 2: Question-wise results of statistical significance tests and normalized gain (G)

To assess the effectiveness of the intervention, we used the normalized gain (G). This measure is the standard approach for analyzing learning gains when using concept inventories. It is calculated as the ratio of the actual gain to the maximum possible gain. This method addresses the challenge of comparing learning outcomes from different initial knowledge levels, as it accounts for the pre-test scores S_{pre} (Coletta and Steinert 2020; Hake 1998). The formula for normalized gain is: $G = \frac{S_{\text{post}} - S_{\text{pre}}}{100 - S_{\text{pre}}}$.

The normalized gain for the total score was $G_{\text{Total Score}} = 0.188$, which constitutes a low learning effect (typically defined as $G < 0.3$, Hake 1998). However, a more granular analysis of the concepts with a significant difference between pre- and post-test scores shows gains that predominantly fall into the range of a medium learning effect ($0.3 \leq G < 0.7$), as detailed in Table 2.

After analyzing the overall and concept-level score changes, we examined how response behavior changed at the item level to address Research Question 2 concerning how students' conceptions changed through the data-driven teaching approach. To analyze the difference in item-level performance between the pre- and post-tests, item scores were scored dichotomously as either correct (1) or incorrect (0) and McNemar's test was applied (Field 2024). Once more, Bonferroni-Holm correction was applied for the 51 item tests (Holm 1979). The test results are shown in Figure 1, which also displays mean pre-post differences for each item. Notably, the majority of conceptions that exhibited change were associated with the context of text generation with most relating to the concepts *Data Selection*, *Evaluation*, *Data Quality* and *Phase Separation*. Furthermore, a

R	adjusted R^2	RMSE	F-Statistic	p-value
0.228	0.028	2.311	2.193	0.118

Table 3: Regression model summary

significant decline was observed in students' understanding of the *correct* concept of *Model Generation* within this context.

Finally, to address Research Question 3, we examined how the two overarching factors of students' conceptions—CT and AI conceptions—influenced the learning effect. To quantify these influences, a score for each predictor (p_{CT} and p_{AI}) was calculated for every participant by counting how often *distractors* reflecting the respective category were selected in the pre-test. The learning effect was operationalized as the difference in the number of correctly identified *concept statements* between the pre- and post-tests. A multiple linear regression was then performed to predict the learning effect from the two predictor scores (p_{CT} and p_{AI}) (Field 2024). In summary, the regression model explains approximately 3% of the variance (adjusted $R^2 = 0.028$, see Table 3). Moreover, the F-statistic is not significant, indicating that the model does not provide a statistically significant explanation of learning outcomes (Field 2024). Pearson correlations of the individual factors with the learning effect are also (negative but) close to zero. Therefore, the data shows no signs that preconceptions had a significant influence on the learning of the *correct* concepts.

6 Discussion

Concerning the first research question pertaining to the efficacy of an intervention grounded in the DdA framework, the findings suggest that this approach can effectively facilitate the development of understanding regarding specific ML concepts. While the overall change in scores was statistically significant, the observed gain of $G = 0.188$ was moderate. This value is comparable to gains typically seen in traditional classroom settings that do not employ interactive engagement methods (Hake 1998). However, a more detailed analysis at the conceptual level reveals a more nuanced picture. Significant gains, ranging from $G = 0.283$ to $G = 0.517$, were observed for the concepts of *Data Selection*, *Evaluation*, and *Phase Separation*. This finding is particularly notable as these concepts were addressed through interactive engagement activities within the workshop. The higher gains observed for these concepts align with the work of Hake (1998), which demonstrates that such interactive methods are more effective in promoting CC than traditional approaches. Moreover, this result is consistent with related research that also highlights the effectiveness of interactive elements in facilitating CC in the context of ML. Furthermore, these concepts are most closely aligned with the ML workflow and are specifically emphasized within our DdA-aligned workshop. In contrast, our analysis revealed no significant differences in students' understanding of ML models' inner workings, despite the DdA "with sprinkles" workshop. We therefore conclude that the DdA is particularly well-suited for facilitating CC in areas related to data handling and the developer's role in shaping model outcomes. However, its effect on a holistic CC appears to be limited. This finding is consistent with related studies that also reported only moderate learning effects for similar interventions not directly designed for fostering CC (Vartiainen et al. 2021; Mühling and Große-Bölting 2023). However, one possible contributing factor may be the duration of the intervention itself. CC often unfolds gradually (Vosniadou 2013), and the three-hour workshop may not have provided sufficient time to fully realize the learning gains intended by the DdA approach.

Mertala and Fagerlund (2024) hypothesized that some conceptions of students are easy to change. With respect to the second research question—how student conceptions change as a result of the intervention—our findings reveal a nuanced picture. On the one hand, the intervention led to a significant reduction in misconceptions, especially for concepts for which learning gains were observed. Specifically, the intervention addressed conceptions concerning *Continuous Learning* (change in six related items) as well as *Autonomous Data Acquisition* (one item). As hypothesized, these are the very conceptions that are directly targeted by the DdA and the conceptual model of the ML workflow presented in our intervention. However, there remains a substantial number of conceptions that were not affected by the intervention. In particular, misconceptions concerning the internal functioning of ML models—such as the "Storage" conception—persisted post-intervention. While it is encouraging that there was no evidence indicating a reinforcement of such misconceptions, as we had hypothesized, their persistence suggests that the DdA is insufficient to effectively challenge these misconceptions. These findings contribute to the ongoing discourse on AI literacy regarding the limits of black-box approaches. Specifically, they lend empirical support to the argument put forth by several researchers that instruction must move beyond interaction with opaque systems and include explicit exploration of model internals and learning algorithms (Ude, Vo, and Pancratz 2024; Morales-Navarro and Kafai 2024; Touretzky, Gardner-McCune, and Seehorn 2022; Michaeli, Romeike, and Seegerer 2022).

A particularly noteworthy finding concerns the context of *Text Generation*, in which recognition of the correct concept, *Model Generation* (*"the model generates text using rules it has learned itself"*), decreased significantly following the intervention. In other words, students were less likely to select the scientifically accurate statement post-intervention. One explanation for this lies in the strong emphasis during the workshop on the role of developers in shaping the ML process. While this is a key feature of the DdA, it may have inadvertently led students to attribute less autonomy to the system itself. This phenomenon, wherein students perceive AI systems as less autonomous after intervention has been observed in previous studies (Große-Bölting and Mühling 2020; Kim et al. 2023). For the DdA, this underscores a critical design implication: while emphasizing the influence of developers is important, it must be balanced with a clear articulation of the model's autonomous learning capabilities to avoid introducing conceptual inaccuracies.

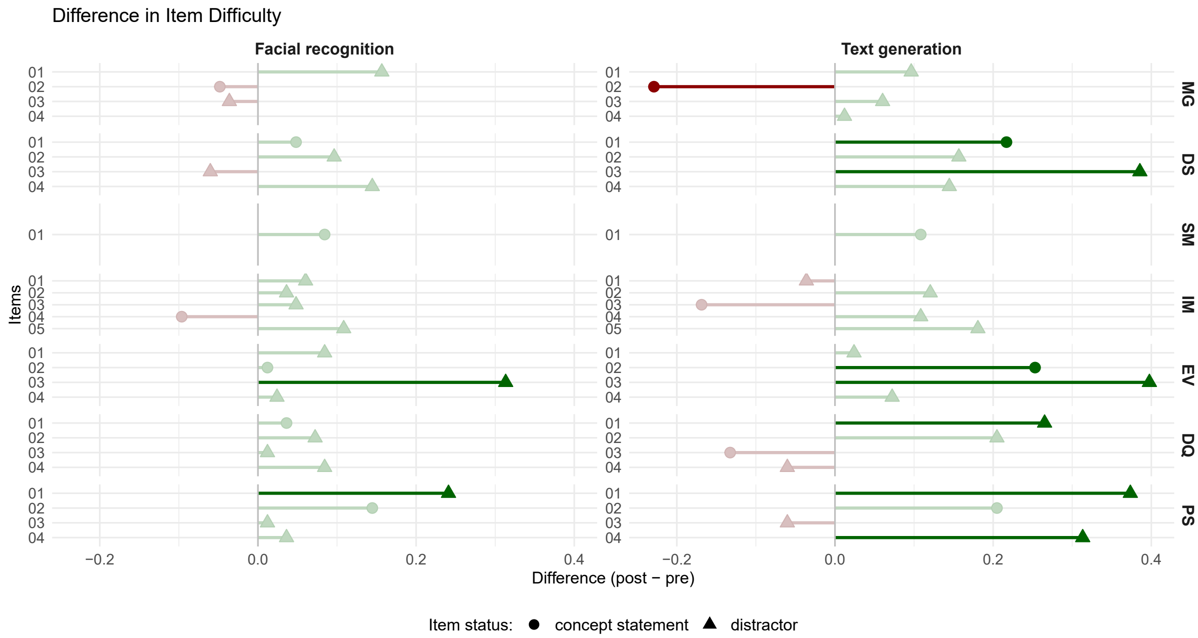

Figure 1: Mean item differences divided into the contexts of *facial recognition* and *text generation*. Non-significant differences are displayed transparently. Items are indicated as either *distractors* or *concept statements*

Another interesting finding emerges from comparing changes in students' conceptions across different application contexts. Most significant changes occurred within the context of text generation, while the changes related to facial recognition were largely non-significant. This is noteworthy because the central tool of the workshop—*Teachable Machine*—is based on image recognition and therefore has more similarities to the context of facial recognition. This finding is in line with previous research by Vartiainen et al. (2021) who also found that students had difficulty to transfer their learnings from TM to ML applications of their everyday life. This also underscores the context-dependence of students' conceptions (diSessa 2013, 2018; Vosniadou 2013). To support effective CC, learning interventions should therefore expose students to a range of ML applications and explicitly reveal the underlying ML components, particularly in domains where the ML processes are not immediately visible. This follows recommendations from CC theory that students must learn to interpret concepts in a variety of contexts (diSessa 2018).

The third research question addressed whether students' pre-instructional conceptions influenced the learning effect achieved through the DdA. Our findings do not indicate a significant relationship between the strength of students' pre-existing conceptions and their subsequent learning gains. In other words, the extent to which a student held a particular preconception prior to the intervention did not statistically affect their ability to acquire scientifically accurate concepts during the learning process. This suggest that within the DdA, it may not be necessary to tailor learning materials specifically to students' pre-instructional conceptions when the goal is to foster understanding of core ML concepts. However, addressing and reconstructing students' misconceptions explicitly remains essential for deeper conceptual understanding. Future learning experiences should therefore include elements that directly engage with prevalent misconceptions. Approaches such as CC texts, introduced by Kreinsen et al. (2024), offer a promising direction.

7 Limitations and Outlook

The most substantial limitation of this study lies in the scope of its conclusions. While we developed and evaluated an intervention grounded in the DdA, our findings are ultimately confined to this specific implementation. As such, no general claims can be made about the overall efficacy of DdA-based instruction without comparative studies involving a range of interventions rooted in the same paradigm. Moreover, the study relied on convenience sampling in the recruitment of participants. Consequently, we cannot determine whether the students included in our sample are representative of the broader target population. This limitation restricts the generalizability of our results.

The instrument used in this study may serve as a valuable tool for evaluating future interventions. Its application could facilitate more direct cross-study comparisons and contribute to a more robust understanding of effective teaching approaches in AI education. In addition, qualitative micro-analytic studies (diSessa 2018) could usefully complement our findings by examining learning processes at the individual level. Such approaches may offer deeper insights into the development and transformation of learners' conceptions over time, thereby enriching the empirical foundation for AI literacy research.

References

Baldoni, M.; Baroglio, C.; Bucciarelli, M.; Capecchi, S.; Gandolfi, E.; Gena, C.; Ianì, F.; Marengo, E.; Micalizio, R.; Rapp, A.; and Ras, I. N. 2024. Does Any AI-based Activity Contribute to Develop AI Conception? A Case Study with Italian Fifth and Sixth Grade Classes. In *Proceedings of the Fourteenth Symposium on Educational Advances in Artificial Intelligence*, EAAI'24, 23060–23068. AAAI Press.

Belghith, Y.; Mahdavi Goloujeh, A.; Magerko, B.; Long, D.; Mcklin, T.; and Roberts, J. 2024. Testing, Socializing, Exploring: Characterizing Middle Schoolers' Approaches to and Conceptions of ChatGPT. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24, 1–17. New York, NY, USA: ACM.

Bewersdorff, A.; Zhai, X.; Roberts, J.; and Nerdel, C. 2023. Myths, Mis- and Preconceptions of Artificial Intelligence: A Review of the Literature. *Computers and Education: Artificial Intelligence*, 4: 100143.

Bilstrup, K.-E. K.; Kaspersen, M. H.; Lunding, M. S.; Schaper, M.-M.; Van Mechelen, M.; Tamashiro, M. A.; Smith, R. C.; Iversen, O. S.; and Petersen, M. G. 2022. Supporting Critical Data Literacy in K-9 Education: Three Principles for Enriching Pupils' Relationship to Data. In *Proceedings of the 21st Annual ACM Interaction Design and Children Conference*, IDC '22, 225–236. New York, NY, USA: ACM.

Broll, B.; and Grover, S. 2023. Beyond Black-Boxes: Teaching Complex Machine Learning Ideas through Scaffolded Interactive Activities. In *Proceedings of the Thirteenth Symposium on Educational Advances in Artificial Intelligence*, EAAI'23, 15990–15998.

Casal-Otero, L.; Catala, A.; Fernández-Morante, C.; Taboada, M.; Cebreiro, B.; and Barro, S. 2023. AI Literacy in K-12: A Systematic Literature Review. *International Journal of STEM Education*, 10(1): 29.

Coletta, V. P.; and Steinert, J. J. 2020. Why Normalized Gain Should Continue to Be Used in Analyzing Preinstruction and Postinstruction Scores on Concept Inventories. *Physical Review Physics Education Research*, 16(1): 010108.

diSessa, A. A. 2013. A Bird's-Eye View of the "Pieces" vs. "Coherence" Controversy (From the "Pieces" Side of the Fence). In *International Handbook of Research on Conceptual Change*, 31–48. New York: Routledge, 2 edition.

diSessa, A. A. 2018. A Friendly Introduction to "Knowledge in Pieces": Modeling Types of Knowledge and Their Roles in Learning. In Kaiser, G.; Forgasz, H.; Graven, M.; Kuzniak, A.; Simmt, E.; and Xu, B., eds., *Invited Lectures from the 13th International Congress on Mathematical Education*, 65–84. Cham: Springer International Publishing.

Field, A. 2024. *Discovering Statistics Using IBM SPSS Statistics*. Los Angeles London New Delhi Singapore Washington DC Melbourne: Sage, 6th edition edition.

Fleischer, Y.; Podworny, S.; and Biehler, R. 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): 3–3.

Gnoth, S.; and Novak, J. 2025. Supporting AI Literacy Through Experiential Learning: An Exploratory Study. In Smith, B. K.; and Borge, M., eds., *Learning and Collaboration Technologies*, 233–251. Cham: Springer Nature Switzerland.

Große-Bölting, G.; and Mühling, A. 2020. Students Perception of the Inner Workings of Learning Machines. In *2020 International Conference on Learning and Teaching in Computing and Engineering (LaTICE)*, 5. Ho Chi Minh City: IEEE Computer Society.

Hake, R. R. 1998. Interactive-Engagement versus Traditional Methods: A Six-Thousand-Student Survey of Mechanics Test Data for Introductory Physics Courses. *American Journal of Physics*, 66(1): 64–74.

Hitron, T.; Orlev, Y.; Wald, I.; Shamir, A.; Erel, H.; and Zuckerman, O. 2019. Can Children Understand Machine Learning Concepts? The Effect of Uncovering Black Boxes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–11. New York, NY, USA: ACM.

Holm, S. 1979. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, 6(2): 65–70.

Höper, L.; and Schulte, C. 2024. New Perspectives on the Future of Computing Education: Teaching and Learning Explanatory Models. In *Proceedings of the 24th Koli Calling International Conference on Computing Education Research*, Koli Calling '24, 1–8. New York, NY, USA: ACM.

Kim, K.; Kwon, K.; Ottenbreit-Leftwich, A.; Bae, H.; and Glazewski, K. 2023. Exploring Middle School Students' Common Naive Conceptions of Artificial Intelligence Concepts, and the Evolution of These Ideas. *Education and Information Technologies*, 28(8): 9827–9854.

Kreinsen, M.; Rabe, A.; Grospietsch, F.; and Schulz, S. 2024. Leveraging Conceptual Change Regarding Artificial Intelligence in Computer Science Education. In *Proceedings of the 24th Koli Calling International Conference on Computing Education Research*, Koli Calling '24, 1–7. New York, NY, USA: ACM.

Kreinsen, M.; and Schulz, S. 2021. Students' Conceptions of Artificial Intelligence. In *The 16th Workshop in Primary and Secondary Computing Education*, 1–2. Virtual Event Germany: ACM.

Lee, I.; Ali, S.; Zhang, H.; DiPaola, D.; and Breazeal, C. 2021. Developing Middle School Students' AI Literacy. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, SIGCSE '21, 191–197. New York, NY, USA: ACM.

Marx, E.; and Bergner, N. 2025. Art Interactive - The Intelligent Museum Guide: A Hands-On Approach to Introducing Key Machine Learning Concepts. In *ISSEP 2025 Local Proceedings*. Zenodo.

Marx, E.; Leonhardt, T.; and Bergner, N. 2025. Concept Inventory Zum Thema Maschinelles Lernen (CIML) - Konzeption, Entwicklung Und Evaluation. In *INFOS 2025 - Grenzen Überwinden - Voneinander Lernen*. Gesellschaft für Informatik e.V.

Marx, E.; Witt, C.; and Leonhardt, T. 2024. Identifying Secondary School Students' Misconceptions about Machine Learning: An Interview Study. In *Proceedings of the 19th WiPSCE Conference on Primary and Secondary Computing Education Research*, 1–10. Munich Germany: ACM.

Mertala, P.; and Fagerlund, J. 2024. Finnish 5th and 6th Graders' Misconceptions about Artificial Intelligence. *International Journal of Child-Computer Interaction*, 39: 100630.

Michaeli, T.; Romeike, R.; and Seegerer, S. 2022. What Students Can Learn about Artificial Intelligence - Recommendations for K-12 Computing Education. In *Proceedings of IFIP WCCE 2022: World Conference on Computers in Education*, 196–208. Hiroshima.

Morales-Navarro, L.; and Kafai, Y. B. 2024. Unpacking Approaches to Learning and Teaching Machine Learning in K-12 Education: Transparency, Ethics, and Design Activities. In *Proceedings of the 19th WiPSCE Conference on Primary and Secondary Computing Education Research*, WiPSCE '24, 1–10. New York, NY, USA: ACM.

Mühling, A.; and Große-Bölting, G. 2023. Novices' Conceptions of Machine Learning. *Computers and Education: Artificial Intelligence*, 4(100142): 1–11.

Perach, S.; and Alexandron, G. 2024. Mind the Gap: Confronting the Vast Divide Between CS Teaching and Machine Learning Pedagogy. In Ferreira Mello, R.; Rummel, N.; Jivet, I.; Pishtari, G.; and Ruipérez Valiente, J. A., eds., *Technology Enhanced Learning for Inclusive and Equitable Quality Education*, 344–358. Cham: Springer Nature Switzerland.

Posner, G. J.; Strike, K. A.; Hewson, P. W.; and Gertzog, W. A. 1982. Accommodation of a Scientific Conception: Toward a Theory of Conceptual Change. *Science Education*, 66(2): 211–227.

Rücker, M. T.; and Pinkwart, N. 2016. Review and Discussion of Children's Conceptions of Computers. *Journal of Science Education and Technology*, 25(2): 274–283.

Sanusi, I. T.; Oyelere, S. S.; Vartiainen, H.; Suhonen, J.; and Tukiainen, M. 2023. Developing Middle School Students' Understanding of Machine Learning in an African School. *Computers and Education: Artificial Intelligence*, 5: 100155.

Schaper, M.-M.; Tamashiro, M. A.; Smith, R. C.; Van Mechelen, M.; and Iversen, O. S. 2023. Five Design Recommendations for Teaching Teenagers' about Artificial Intelligence and Machine Learning. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference*, 298–309. Chicago IL USA: ACM.

Szczuka, J.; Strathmann, C.; Szymczyk, N.; Mavrina, L.; and Krämer, N. 2022. How Do Children Acquire Knowledge about Voice Assistants? A Longitudinal Field Study on Children's Knowledge about How Voice Assistants Store and Process Data. *International Journal of Child-Computer Interaction*, 33: 100460.

Tedre, M.; Denning, P.; and Toivonen, T. 2021. CT 2.0. In *21st Koli Calling International Conference on Computing Education Research*, Koli Calling '21, 1–8. New York, NY, USA: ACM.

Touretzky, D.; Gardner-McCune, C.; and Seehorn, D. 2022. Machine Learning and the Five Big Ideas in AI. *International Journal of Artificial Intelligence in Education*, 33(2): 1–34.

Ude, N.; Vo, G. M.; and Pancratz, N. 2024. "Like a Real Human with Much More Knowledge" - A Phenomenon-oriented Investigation of Pre-instructional Conceptions in the Context of Artificial Intelligence. In *Proceedings of the 24th Koli Calling International Conference on Computing Education Research*, Koli Calling '24, 1–9. New York, NY, USA: ACM.

Vandenberg, J.; and Mott, B. 2023. "AI Teaches Itself": Exploring Young Learners' Perspectives on Artificial Intelligence for Instrument Development. In *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1*, 485–490. Turku Finland: ACM.

Vartiainen, H.; Toivonen, T.; Jormanainen, I.; Kahila, J.; Tedre, M.; and Valtonen, T. 2021. Machine Learning for Middle Schoolers: Learning through Data-Driven Design. *International Journal of Child-Computer Interaction*, 29: 100281.

Vosniadou, S. 2013. Conceptual Change in Learning and Instruction: The Framework Theory Approach. In *International Handbook of Research on Conceptual Change*. Routledge, 2 edition.

Waite, J.; Tshukudu, E.; Cucuiat, V.; Whyte, R.; and Sentance, S. 2023. Towards a Framework for Learning Content Analysis in K-12 AI/ML Education. In *2023 IEEE Frontiers in Education Conference (FIE)*, 1–5. College Station, TX, USA: IEEE.

Witt, C.; Leonhardt, T.; Marx, E.; and Bergner, N. 2025. From Guesswork to Game Plan: Exploring Problem-Solving-Strategies in a Machine Learning Game. In Pluhár, Z.; and Gaál, B., eds., *Informatics in Schools. Innovative Approaches to Computer Science Teaching and Learning*, volume 15228, 73–84. Cham: Springer Nature Switzerland.

Yang, Q.; Suh, J.; Chen, N.-C.; and Ramos, G. 2018. Grounding Interactive Machine Learning Tool Design in How Non-Experts Actually Build Models. In *Proceedings of the 2018 Designing Interactive Systems Conference*, DIS '18, 573–584. New York, NY, USA: ACM.

Zhang, H.; Perry, A.; and Lee, I. 2024. Developing and Validating the Artificial Intelligence Literacy Concept Inventory: An Instrument to Assess Artificial Intelligence Literacy among Middle School Students. *International Journal of Artificial Intelligence in Education*, 35(1).