Revisiting the Role of Learning AI Principles in Moving Students Beyond Anthropomorphic thinking

Hagit Gabbay[1,*], Ronit Halstuch[1], Ruth Ston[2], Ido Roll[2], Yael Shwartz[1] and Giora Alexandron[1]

[1] *Weizmann Institute of Science, Rehovot, Israel*
[2] *Technion - Israel Institute of Technology, Haifa, Israel*

Abstract

Research suggests that learning the scientific principles underlying artificial intelligence (AI) can support a conceptual shift among middle and high school students from anthropomorphic interpretations of AI toward data-driven thinking. However, today's students interact extensively with advanced generative AI systems that exhibit highly human-like conversational behavior. This calls for a reexamination of whether and how such conceptual change occurs among students immersed in AI-saturated contexts. In this paper we present preliminary findings from an ongoing field study conducted within an AI literacy curriculum implemented in middle school classes. We analyze pre-questionnaire data from 279 ninth-grade students to examine the prevalence of anthropomorphic versus data-oriented conceptions of AI prior to instruction. The findings indicate a relatively high proportion of students holding anthropomorphic conceptions at the outset, alongside no significant associations with prior background, knowledge, or the diversity of their GenAI use. These initial results suggest that frequent interaction with generative AI tools does not necessarily foster data-driven understanding and may, in some cases, reinforce anthropomorphic reasoning. Our next steps include administering the post-questionnaire and examining the nature of conceptual change (if observed) and considering whether curricular adjustments may be needed to support a transition toward data-oriented understanding of AI.

Keywords

K-12 AI literacy, GenAI knowledge, conceptual change, anthropomorphic thinking

1. Introduction

As Generative Artificial Intelligence (GenAI) systems become an integral part of everyday life, students' conceptions of how these systems function have growing implications for critical thinking and responsible use [1]. Anthropomorphic interpretations of AI such as attributing intention, understanding, emotions or agency to these systems may weaken students' ability to critically evaluate AI outputs. Subsequently, such attributions may encourage unreflective reliance on these tools and hinder human accountability in the design and use of these systems [2, 3, 4, 5]. Supporting a conceptual transitioning toward understanding AI as a data-driven, probabilistic system is therefore a central educational goal [6], enabling students to engage with GenAI tools in a more informed, critical, and responsible manner [1, 2, 7].

Several studies suggest a relationship between learning the scientific principles underlying AI and conceptual change in students' understanding of AI decision-making processes, facilitating a transition from anthropomorphic misconceptions to more data-oriented and critical thinking [8, 1, 6, 2, 3]. However, the rapid advancement of the human-like conversational capabilities of GenAI systems, together with the increasing frequency and intensity of students' interactions with diverse AI tools, may fundamentally reshape how students relate to and reason about AI, underscoring the need to reexamine this relationship. In this paper, we present preliminary findings from an ongoing study addressing this need. The study

Joint Proceedings of LAK 2026 Workshops, co-located with the 16th International Conference on Learning Analytics and Knowledge (LAK 2026), Bergen, Norway, April 27 – May 1, 2026

*Corresponding author.
hagit.gabbay@weizmann.ac.il (H. Gabbay); ronit.halstuch@weizmann.ac.il (R. Halstuch); rut.savir@campus.technion.ac.il (R. Ston); roll@technion.ac.il (I. Roll); yael.shwartz@weizmann.ac.il (Y. Shwartz); giora.alexandron@weizmann.ac.il (G. Alexandron)
0000-0002-2012-1964 (H. Gabbay)

is conducted within an AI literacy curriculum implemented during the current academic year in science and technology middle school classes.

2. RELATED WORK

Trying to promote AI Literacy, research focuses on the following question: how can we advance students' conceptual understanding of AI as statistical inference algorithms that learn from data? Specifically, can such understanding be facilitated by instruction that presents the underlying machinery of AI? In a study conducted in 2022-2023, Dong et al. (2025) [6] implemented a five-lesson module within English Language Arts classes where students iteratively built text classification models. Qualitative analyses of student responses to open questions revealed a clear progression where students moved from anthropomorphic thinking and simplistic keyword matching toward sophisticated data-based thinking.

Similar conceptual shift was demonstrated by Kim et al. (2023) and Klemettilä et al. (2025) [2, 3], utilizing interventions containing both unplugged activities and hand-on tools such as "Teachable Machine". In both studies, participants transitioned from idealizing the capabilities of AI to understanding the relationship between system performance and the quality of the dataset on which it was trained. Dangol et al. (2025) [1] utilized visual pattern grids to probe children's mental models. They identified a statistically significant developmental shift from attributing intelligence as inherent to AI toward conceptualizing AI as a data-based predictive system. It is worth noting that, prior to instruction, no connection was found between the type of AI the participants used (e.g., voice assistants, chatbots) and their mental models of how AI "thinks". As we seek to investigate whether students' intensive interactions with generative AI tools may influence relationship between instructional programs and conceptual change, examining students' baseline knowledge levels as well as the presence of associations between this knowledge and students' misconceptions about AI is of significant importance.

3. METHODOLOGY

3.1. Research context - our conceptual framework and curriculum

The examination of the impact of teaching the scientific and mathematical principles of AI on students' conceptions is conducted within the context of instruction in an AI literacy curriculum (integrated with programming), targeting ninth-grade middle school students enrolled in science and technology tracks that include programming instruction. The conceptual framework underlying the curriculum builds on UNESCO's AI Competency Framework and AI4K12's "Five Big Ideas" [9, 10]. It includes four main aspects: AI foundations, skills, ethics, and attitudes [11]. The curriculum is organized into ten comprehensive learning modules, each aligned with the conceptual framework dimensions. These modules focus on key AI applications, combining both plugged and unplugged activities. The program includes instructional units addressing classification, Markov models and LLMs, and reinforcement learning. Students are encouraged to explore meaningful concepts such as data representation, abstraction, and algorithmic bias.

The program is being implemented during the current academic year in several dozen schools, with variations of the pace of progress. At the time of writing, about half of the students having begun the instruction and completed the pre-questionnaire.

3.2. Data collection

Data for the present study were collected through students' responses to an online questionnaire administered before starting the AI literacy program. The questionnaire was administered to students in 18 ninth-grade science-technology classes across different schools. A total of N = 279 students completed the questionnaire. Of the participants, 55% identified as boys (n = 154), 41% as girls (n = 114) and the remaining participants preferred not to disclose their gender.

The questionnaire was originally developed to measure different types and domains of AI knowledge among adults [12] and was subsequently adapted for use with middle and high school students. It was administered in the participants' native language, and the examples presented here are translated for reporting purposes. In the present study, three selected sections of this instrument were employed. In the first one, respondents are asked to provide demographic information (e.g., age, gender, and class affiliation), information about their background in programming and AI, and the purposes for which they use GenAI technologies in their everyday lives by selecting all applicable options from a list of six, with an additional "other" option. Examples for common usage purposes are "For searching for information and facts (e.g., answering questions across a wide range of topics)" and "As a conversational partner (e.g., chatting with a chatbot for fun)".

The second section includes a knowledge test assessing understanding of scientific principles of AI, consisting of 13 statements with three response options: true, false, or "don't know". Examples of knowledge items are "Generative artificial intelligence may infringe on artists' copyright", "Using GenAI requires prior knowledge of programming" and "The quality of the responses generated by AI is similar across all languages".

The third section includes measures of students' attitudes and perceptions regarding AI, assessed through their level of agreement with 15 statements on a 1–5 Likert scale, where higher values indicate more positive attitudes. For example: "AI has uses that are relevant to my everyday life", "I can successfully use a new tool that is based on AI". In addition, this section includes an item assessing agreement with the statement "Artificial intelligence is moral" (on a 1–5 Likert scale with an additional "don't know" option) and an open-ended question in which respondents are asked to justify their position regarding the morality of artificial intelligence.

3.3. Data Analysis

The internal reliability of the questionnaire items assessing knowledge and attitudes toward AI was found to be good (Cronbach's $\alpha = 0.73$ and 0.83, respectively). Data analysis was conducted as follows. First, the relevant variables were computed:

- Diversity of GenAI use – the number of selected usage purposes (0-7).
- Knowledge level – the proportion of correct responses out of all responses to the knowledge test (range: 0-1).
- Attitude toward AI - the mean level of agreement across all attitude items (1-5).

Responses to the open-ended question were analyzed manually by an expert member of the research team to identify anthropomorphic interpretations. Texts that included references to human-like agency such as intentions, emotions, or desires attributed to AI were coded as anthropomorphic, whereas texts indicating an understanding of AI as data-driven software or algorithms or referring user (i.e. human) responsibility of its "behavior", were coded as non-anthropomorphic. Neutral responses or responses that did not provide sufficient information for classification were coded as neutral with respect to anthropomorphism. Relationships among the study variables were examined using appropriate statistical tests (Spearman correlation or the Wilcoxon rank-sum test for independent samples). The specific test used for each analysis is reported in the Findings section.

4. INITIAL FINDING AND DISCUSSION

The study is ongoing, and the findings reported below refer to the pre-questionnaire completed by the students who have begun the program. Descriptive statistics for the pre-questionnaire measures are presented in Table 1. Overall, students demonstrated moderate diversity in GenAI use, limited prior practical experience with programming or AI and mid-high range knowledge and attitudes toward AI. Responses to the AI morality item indicated moderate agreement overall, with a notable proportion of students selecting "don't know".

Table 1
Descriptive Statistics of Study Variables

Measure	N	Statistic	Value
Background in programming or AI	213	Reported any background	61% (n = 129)
		No background	39% (n = 84)
	129	Practical experience	55% (n = 71)
Diversity of GenAI use	279	Mean (SD)	3.31 (1.68)
		Median	3
		No GenAI use	6.1% (n = 17)
		Use beyond listed options	1.1% (n = 3)
AI knowledge	279	Mean correct (out of 13)	9.23 (2.92)
		Knowledge proportion (SD)	0.71 (0.23)
		Range	0-13
Attitudes toward AI	279	Mean (SD)	4.30 (0.78)
AI morality (single item)	236	"Don't know" responses	17.4% (n = 41)
	195	Mean agreement (SD)	3.09 (0.94)
		Median	3

The qualitative analysis of the justifications for the selected levels of agreement, and their classification as indicating the presence or absence of anthropomorphic thinking, revealed that out of n = 236 respondents, 37% exhibited anthropomorphic thinking, while 27% exhibited data-oriented thinking and/or an understanding that AI lacks human-like agency. The remaining 35.6% of respondents were classified as neutral with respect to anthropomorphism, primarily because they did not provide any justification. After excluding these neutral responses, 57.9% of respondents were found to exhibit anthropomorphic thinking, whereas 42.1% exhibited non-anthropomorphic thinking. An additional insight emerging from the qualitative analysis of the open-ended justifications is that students interpreted the agreement item concerning AI morality in different ways. For example, some respondents who selected "disagree" intended to express the view that AI is not moral, whereas others who selected "disagree" rejected the premise of the question itself, arguing that AI is not an entity capable of morality (thereby demonstrating a more advanced, non-anthropomorphic thinking model). Accordingly, the numerical agreement scores were interpreted only in conjunction with the qualitative analysis of the accompanying justifications.

Examination of the relationships between respondents' background and their level of knowledge revealed a significant difference, where learners who reported on previous practical experience in programming or previous formal learning about AI showed higher knowledge level compared to those without such background (Wilcoxon rank-sum test, $W=1380.5$, $p<.001$). It is worth noting that most of the prior experience reported was in programming, while only a few students had learned about AI in advance, suggesting that computational experience in general may support AI-related knowledge. In contrast, no significant associations were found between students' background and their attitudes toward AI or the presence (or absence) of anthropomorphic thinking ($p>.05$, Spearman correlations and Wilcoxon rank-sum tests). Similarly, knowledge level was not significantly correlated with attitudes toward AI ($p>.05$, Spearman correlation). Finally, anthropomorphic thinking was not significantly associated with either the diversity of GenAI use or knowledge level ($p>.05$, Wilcoxon rank-sum tests).

These findings are in somewhat disagreement with previous findings [6] as a higher proportion of students exhibited anthropomorphic thinking prior to instruction. Possible explanation for this difference lies in the contemporary learning context, in which students are routinely exposed to and interact with generative AI systems. Unlike [6], the present study also examined the relationship between students' background and their initial thinking models and found no significant associations. This result is consistent with findings reported by [1, 7], who similarly observed no association between prior knowledge and students' initial conceptions of AI.

The findings are surprising in several respects. First, one might have expected to find an association between students' knowledge and the presence of anthropomorphic conceptions, even though previous studies have similarly reported no such relationship [1, 7]. One possible explanation is that conceptual knowledge of AI mechanisms does not necessarily reshape deeper conceptual or intuitive mental models about intelligence and agency. Students may acquire procedural knowledge about AI without fundamentally revising their conceptual interpretations of what AI "is," or advancing toward a deeper epistemic understanding of how knowledge is generated and represented in AI systems as opposed to human cognition. Consequently, they may continue to attribute human-like understanding or intentionality to AI systems. Another unexpected finding is that the proportion of students demonstrating anthropomorphic thinking was relatively high compared to previous studies (e.g.,[6]). This result may suggest that experience in using GenAI tools, as reported by most respondents, does not necessarily lead to an understanding of how AI models operate. Moreover, frequent interaction with generative AI tools that exhibit advanced human-like conversational capabilities may even foster the attribution of human-like agency to these systems.

5. NEXT STEPS

The study is currently being conducted in the field, and data collection from participating classrooms is ongoing. The next steps will involve administering the questionnaire at the conclusion of the program and examining whether students demonstrate a conceptual shift from anthropomorphic thinking toward data-driven reasoning and an understanding that AI lacks human-like agency. In line with the study's goal of revisiting prior findings, we will compare the nature of conceptual change (if observed in the present context) with that reported in earlier studies, particularly the work of [6]. This comparison will allow us to examine whether students' extensive and routine interaction with generative AI tools influences the relationship between learning AI principles and changes in students' thinking, and consequently whether refinements to the curriculum are needed to more effectively support a transition toward a data-oriented understanding of AI.

Acknowledgments

This research was conducted in collaboration with the Eddie and Jules Trump Family Foundation.

Declaration on Generative AI

During the preparation of this work, the authors used Chat-GPT-5.2 in order to: Grammar and spelling check. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] A. Dangol, R. Wolfe, R. Zhao, T. Ramanan, JaeWon Kim, K. Davis, J. A. Kientz, Children's Mental Models of AI Reasoning: Implications for AI Literacy Education, in: In Proceedings of the 24th Interaction Design and Children (IDC '25), Association for Computing Machinery, 2025, pp. 106–123. URL: https://dl.acm.org/doi/epdf/10.1145/3713043.3728856. doi:10.1145/3713043.3728856.

[2] K. Kim, K. Kwon, A. Ottenbreit-Leftwich, H. Bae, K. Glazewski, Exploring middle school students' common naive conceptions of Artificial Intelligence concepts, and the evolution of these ideas, Education and Information Technologies 28 (2023) 9827–9854. URL: https://doi.org/10.1007/s10639-023-11600-3. doi:10.1007/s10639-023-11600-3.

[3] P. A. Klemettilä, S. Sharma, F. Mochiyama, N. Iivari, M. Iwata, J. Koivisto, "It's Just a Machine that Predicts" - Demystifying Artificial Intelligence / Machine Learning with Teenagers, in: Proceedings of the 24th Interaction Design and Children, IDC '25, Association for Computing Machinery, New York, NY, USA, 2025, pp. 168–182. URL: https://dl.acm.org/doi/10.1145/3713043.3728853. doi:10.1145/3713043.3728853.

[4] P. Mertala, J. Fagerlund, Finnish 5th and 6th graders' misconceptions about artificial intelligence, International Journal of Child-Computer Interaction 39 (2024) 100630. URL: https://www.sciencedirect.com/science/article/pii/S2212868923000673. doi:10.1016/j.ijcci.2023.100630.

[5] J. Solyst, E. Yang, S. Xie, J. Hammer, A. Ogan, M. Eslami, Children's Overtrust and Shifting Perspectives of Generative AI, 2024. URL: http://arxiv.org/abs/2404.14511. doi:10.48550/arXiv.2404.14511, arXiv:2404.14511 [cs].

[6] L. Dong, S. Jiang, F. Xu, S. Duan, C.-J. Hsieh, From simplification to sophistication: Secondary students' conceptual change in understanding machine learning model decision making, British Journal of Educational Technology n/a (2025) 1–20. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/bjet.70024. doi:10.1111/bjet.70024, _eprint: https://bera-journals.onlinelibrary.wiley.com/doi/pdf/10.1111/bjet.70024.

[7] L. Morales-Navarro, D. J. Noh, Y. B. Kafai, High school students building babyGPTs: Engaging in data practices and addressing ethical issues through the construction of generative language models, International Journal of Child-Computer Interaction (2025) 100769. URL: https://www.sciencedirect.com/science/article/pii/S2212868925000509. doi:10.1016/j.ijcci.2025.100769.

[8] Y. Dai, Z. Lin, A. Liu, W. Wang, An embodied, analogical and disruptive approach of AI pedagogy in upper elementary education: An experimental study, British Journal of Educational Technology 55 (2024) 417–434. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/bjet.13371. doi:10.1111/bjet.13371, _eprint: https://bera-journals.onlinelibrary.wiley.com/doi/pdf/10.1111/bjet.13371.

[9] D. Touretzky, C. Gardner-McCune, D. Seehorn, Machine Learning and the Five Big Ideas in AI, International Journal of Artificial Intelligence in Education 33 (2023) 233–266. URL: https://link.springer.com/article/10.1007/s40593-022-00314-1. doi:10.1007/S40593-022-00314-1/TABLES/4.

[10] Miao, Fengchun [author] [77], Shiohira, Kelly [author] [10], Lao, Natalie [author] [6], K. Shiohira, N. Lao, AI competency framework for students, UNESCO, 2024. URL: https://doi.org/10.54675/JKJB9835. doi:10.54675/jkjb9835, publication Title: AI competency framework for students.

[11] R. Halstuch, G. Alexandron, Y. Shwartz, Literature Review on AI Education Programs and Guidelines from Leading Countries and International Organizations, Technical Report, 2024. URL: https://shorturl.at/plyqh.

[12] I. Klein-Avraham, R. Ston, O. Atias, I. Roll, A. Baram-Tsabari, Measuring different types and domains of AI knowledge: Developing and validating a performance-based scale, Computers & Education 247 (2026) 105573. URL: https://www.sciencedirect.com/science/article/pii/S0360131526000114. doi:10.1016/j.compedu.2026.105573.