Exploring Students' Preinstructional Mental Models of Machine Learning: Preliminary Findings

Erik Marx[1,2][0000−0002−5918−804X], Thiemo Leonhardt[1,2][0000−0003−4725−9776], Nadine Bergner[3][0000−0003−3527−3204], and Clemens Witt[1][0009−0005−8160−4029]

[1] TUD Dresden University of Technology, Germany
[2] Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI) Dresden/Leipzig, Germany
[3] RWTH Aachen University, Germany

Abstract. This poster proposal provides insights into research on school students' conceptions of machine learning (ML) and highlights the effectiveness of interviews in exploring their mental models. In our study, we use semi-structured interviews to explore the mental models students develop prior to instruction on ML, outlining the advantages of this method for obtaining detailed insight. Eight interviews with German school students were conducted, revealing different perspectives. The preliminary findings indicate that some students imbue artificial intelligence (AI) with anthropomorphic qualities. Traditional concepts of computational thinking are also referenced, but often do not match the realities of ML. These findings contribute to computer science education by providing a nuanced understanding of school students' conceptions of AI and ML and highlighting the need for accurate education in the evolving landscape of AI and ML.

Keywords: Machine Learning · Mental Models · School Students' Conceptions · Preconceptions · Interview Study

1 Introduction

Considering the growing integration of artificial intelligence (AI) into students' daily lives, it becomes crucial for them to acquire a fundamental understanding of the functioning of this technology. Based on constructivist learning theories the assimilation of new knowledge occurs through the adaptation of existing knowledge structures [10, 2]. One concept that explains how individuals explain the world, solve problems, or form hypotheses is that of mental models [1, 2]. Mental models are cognitive representations of situations or domains that help in understanding, learning, reasoning, or predicting [1, 2]. In educational research, other terms are used that address students' preinstructional understanding such as preconceptions, alternative conceptions, or p-prims but with little consistency as to how these terms relate to each other [10]. We understand student conceptions and related terms as the observable phenomena of the use of internal mental

J.-P. Pellet and G. Parriaux (Eds.): ISSEP 2023 Local Proceedings, pp. 233–236, 2023.
https://doi.org/10.5281/zenodo.8432012

models. The study of mental models is necessary because they provide us with information about the learning process of students, for example, by explaining typical errors and thus enabling us to design better learning experiences [1].

Students' view on AI is influenced by a diverse array of sources. Media and in particular science fiction wield substantial influence over students' conceptualizations of AI, often leading to anthropomorphism of AI, thereby conflicting with accurate computational concepts [7]. Furthermore, in the field of computer science education (CSE) machine learning (ML) presents distinctive characteristics compared to conventional subjects within CSE. This contrast is elucidated by Tedre et al., who argue that ML introduces a paradigm shift, challenging established notions of computational thinking (CT). For instance, in the context of imperative programming, correctness predominantly revolves around syntactic accuracy or the generation of precise outputs. Conversely, within the domain of ML an element of probability is introduced, indicating the likelihood of correctness for each generated output [11].

The primary objective of this interview study is to delve into the preinstructional mental models that school students employ in their attempts to elucidate phenomena within ML. This investigation seeks to assess the potential implications of these mental models on the learning process and their potential conflicts with correct computational concepts of ML.

2 Related Works

Some works already exist that delve into school students' preconceptions about AI. Mertala et al. used a qualitative questionnaire to inquire middle-school students about their perceptions of how AI functions. The analysis primarily focuses on discerning how children conceptualize the type of technology AI represents and where AI finds application [9]. Kim et al. assessed middle school students' preconceptions about AI through video observations and learning artifacts, tracking their development across a summer camp [4]. Kreinsen & Schultz conducted interviews with students on their understanding of AI [5]. These endeavors share the commonality of encompassing a general assessment of school students' conceptions of AI, without concentrating on specific technologies like ML, which holds particular relevance for CSE research, given the fundamental differences between some concepts of ML and traditional CS concepts. Additionally, we identify a lack of utilization of interview-based data collection methods, despite their capacity to pose follow-up questions, thereby faciliting deeper understanding of students' mental models. With our investigation, we aim to narrow the focus onto the realm of ML, thereby exploring school students' preinstructional mental models concerning ML via an interview study.

3 Methodology

To assess students mental models, we opted for the traditional semi-structured interview, commonly used in mental model research [1]. A total of eight inter-

views with German school students were conducted with a duration between 43 and 69 minutes. Of the children interviewed, 6 were in eighth grade, 1 was in fifth grade, and 1 was in third grade. As Jones et al. found, interviewees' responses are more detailed and in-depth when they are directly exposed to the phenomena about which they are being questioned [3]. Thus, the students were introduced to two technologies to interact with during the interview: facial recognition on smartphones and ChatGPT. After each of these interactions, the participants were asked questions about how these technologies work. A semi-structured interview guide was created, outlining topics of interest derived from literature (such as correctness of AI [11]). Further detailed questions were formulated to facilitate deeper discussions. In the interview the interviewer also had flexibility to ask unlisted follow-up questions, aligning with the semi-structured interview approach.

For the analysis, interviews were transcribed from audio recordings and augmented with video footage to incorporate students' gestures and facial expressions, enhancing contextual understanding. Subsequently, the transcripts are to be analyzed using qualitative content analysis according to Kuckartz [6], taking into account Mayring & Fenzl's recommendations [8]. In accordance with [6], the analysis started with reading each interview and summarizing initial findings. The preliminary findings presented in section 4 are derived from this step. An inductive-deductive approach is planned post-initial text work to build the coding framework using the main categories from the interview guide. Open coding will refine the category system, followed by coding all interviews and analyzing thematic summaries.

4 Preliminary Findings and Implications

The initial analysis of the data demonstrates that the interviewed students provided a diverse range of explanations. Some students exhibited familiarity with ML concepts, such as neural networks, or possessed a basic comprehension of how facial recognition operates. In contrast, however, there were also more naive perspectives, like the notion that ChatGPT was simply some form of sophisticated search engine. Particularly when dealing with less familiar technical intricacies, students tend to anthropomorphize the technology, attributing human-like traits to it. For instance, one student characterizes AI as striving for self-improvement and perpetual learning. Another student draws a direct parallel between ChatGPT and humans, asserting that both make random decisions because humans are inherently "random" as well.

However, it's important to acknowledge that explanatory approaches rooted in traditional CT concepts also emerge. For instance, students are well aware of the pivotal role that data play and acknowledge its significance. However, a noteworthy observation permeating all the interviews, which has yet to be addressed in related work, is that most students assume data is stored and that AI makes decisions based on a comparison of input data with stored "training data"; this notion directly contradicts the actual functioning of ML systems. In

another instance, a student combines both anthropomorphic ideas and CT concepts. He speculates that ChatGPT was coded by developers with an initial set of handcrafted grammatical rules and vocabulary and subsequently independently scoured the Internet for new data to enhance its vocabulary.

The preliminary results presented herein bear relevance to research in CSE by offering an additional perspective on school students' conceptions of AI. In summary, it can be deduced that the conducted interview study proves effective as a methodology for assessing the various explanations put forth by students. The partially novel findings underscore the value of semi-structured interviews, as they allow to ask specific follow-up questions and thus to render students' mental models on the functionality of ML systems visible.

References

1. Gentner, D.: Mental Models, Psychology of. In: International Encyclopedia of the Social & Behavioral Sciences, pp. 9683–9687. Elsevier (2001). https://doi.org/10.1016/B0-08-043076-7/01487-X
2. Greca, I.M., Moreira, M.A.: Mental models, conceptual models, and modelling. International Journal of Science Education **22**(1), 1–11 (Jan 2000). https://doi.org/10/btgzg6
3. Jones, N.A., Ross, H., Lynam, T., Perez, P.: Eliciting Mental Models: A Comparison of Interview Procedures in the Context of Natural Resource Management. Ecology and Society **19**(1) (2014)
4. Kim, K., Kwon, K., Ottenbreit-Leftwich, A., Bae, H., Glazewski, K.: Exploring middle school students' common naive conceptions of Artificial Intelligence concepts, and the evolution of these ideas. Education and Information Technologies (Jan 2023). https://doi.org/10.1007/s10639-023-11600-3
5. Kreinsen, M., Schulz, S.: Students' Conceptions of Artificial Intelligence. In: The 16th Workshop in Primary and Secondary Computing Education. pp. 1–2. ACM, Virtual Event Germany (Oct 2021). https://doi.org/10/gqjsqq
6. Kuckartz, U., Rädiker, S.: Qualitative Content Analysis: Methods, Practice and Software. SAGE Publications, Thousand Oaks, second edn. (2023)
7. Marx, E., Leonhardt, T., Bergner, N.: Brief Summary of Existing Research on Students' Conceptions of AI. In: Proceedings of the 17th Workshop in Primary and Secondary Computing Education. pp. 1–2. WiPSCE '22, Association for Computing Machinery, New York, NY, USA (Oct 2022). https://doi.org/10/gq2p2n
8. Mayring, P.: Qualitative Content Analysis: Theoretical Foundation, Basic Procedures and Software Solution. Klagenfurt (2014)
9. Mertala, P., Fagerlund, J., Calderon, O.: Finnish 5th and 6th grade students' preinstructional conceptions of artificial intelligence (AI) and their implications for AI literacy education. Computers and Education: Artificial Intelligence **3**, 100095 (Jan 2022). https://doi.org/10.1016/j.caeai.2022.100095
10. Taber, K.S.: The Nature of Student Conceptions in Science. In: Taber, K.S., Akpan, B. (eds.) Science Education, pp. 119–131. New Directions in Mathematics and Science Education, SensePublishers, Rotterdam (2017). https://doi.org/10.1007/978-94-6300-749-8_9
11. Tedre, M., Denning, P., Toivonen, T.: CT 2.0. In: 21st Koli Calling International Conference on Computing Education Research. pp. 1–8. ACM, Joensuu Finland (Nov 2021). https://doi.org/10/gnqv9f