**Understanding understanding AI**

Andreas Mühling1 & Lukas Scheppach2

1Leibniz Institute for Science and Mathematics Education, Germany,muehling@leibniz-ipn.de

2 Leibniz Institute for Science and Mathematics Education, Germany

Focus Topics: Explanatory Models, AI and Data Science Competencies

## Introduction

With an increasing demand to teach artificial intelligence and machine learning in K12 settings, there are several gaps that computer science education research needs to address. Determining suitable learning goals for the age group is one of them and has led to first curricula, such as AI4K12 (Touretzky et al., 2019). It is designed around “Big Ideas” and “Key Insights” and follows a very basic idea of “opening up black boxes” as far as the mathematical and CS background of the students will allow to make them understand the underlying principles behind the technology so dominant in our everyday lives (Essinger & Rosen, 2011; Mariescu-Istodor & Jormanainen, 2019; Touretzky et al., 2019). This approach is in line with typical science lessons that also aim to help students understand the natural world. It is, however, not in line with a CT or engineering based approach to computer science lessons in which construction and not understanding is the ultimate goal. In this line of thinking, Tedre et al. (2021) have proposed CT 2.0 as a new variant of computational thinking that is not based on sequential, procedural programs but instead on the notion of learnable machines (mostly neural networks).

Regardless of the chosen approach, developing good teaching materials typically involves understanding how students perceive a topic, what kind of prior knowledge they might bring into lessons and what kind of misconceptions might develop. This – together with instructional strategies – can form the basis for a body of pedagogical content knowledge (Shulman, 1986) that teachers should acquire. While there is work on conceptions of machine learning and artificial intelligence, we currently do not know much about students’ progressions in understanding when learning about AI, neither is there much knowledge about misconceptions or the suitability of teaching approaches.

## A Phenomenographic Model

In recent work (Mühling & Große-Bölting, 2023), we explored how beginners conceptualize machine learning based on students’ responses and identified a phenomenographic outcome space that is structured along two dimensions: The *learning process* itself and the *internal model* of the learning agent. For the dimension of the learning process, four consecutive stages of understanding – None, Unclear, Repetition and Improvement – were identified, whereas for the internal model dimension there are three stages: None, Implicit and Explicit. A detailed description of the stages including anchoring examples are presented along with the model (Mühling & Große-Bölting, 2023).

This outcome space was also used to classify the responses of students from grades 12-13 prior and after a short 90 minute intervention based on an unplugged activity (Gardner & Michie, 1982) centered around reinforcement learning of a simple game. Even this short workshop already had a medium effect on improving learners conceptualization regarding the learning process (W = 999, r = 0.36, p = 0.0002), however only a small and non-significant effect (r = 0.16) was observable regarding the model dimension.

## Operationalizing the Outcome Space

Based on this initial work, we are currently investigating how to operationalize the outcome space into a diagnostic assessment that could be used to determine students’ stages of understanding. In a first attempt, we used actual statements from students together with our understanding from coding to create items that students can agree or disagree with on a 5-point Likert scale.

We piloted a version of such an assessment with 12 items in a three hour workshop on artificial intelligence with students from grades 9 and 10 in a pre-post setting and again could observe an improvement in the learning-process dimension (W = 113.5, r = 0.36, p = 0.02) and a small but non-significant effect on the dimension of the internal model (r = 0.10).

However, the items also show only a weak internal consistency (Cronbach’s alpha = 0.54). As they are combining two dimensions of the original model into one set of items, this is expected to a degree - however it also raises questions about the structure of the items and the overall design of the instrument. A rather large sample might be needed to investigate the internal consistency in such a setting.



Figure 1: The phenomenographic outcome space and improvements between the classifications of pre- and post-test of a workshop (Mühling & Große-Bölting, 2023)

In an alternative approach we are currently developing concept cartoons (Keogh & Naylor, 1999) centered around each of the two dimensions in order to investigate whether this method might be

a suitable way to operationalize a phenomenographic outcome space and diagnose students’ understanding. Concept cartoons allow presenting phenomena and possible explanation from the perspective of peers of the learners and thus better align with the idea of phenomenography in which a normative, expert-like or “correct” understanding is not necessarily in the focus of the model but instead a description of the various ways of sense-making of learners.

We designed an initial set of cartoons based on explanations of recommender systems and text-generative AI and piloted them in a workshop with two classes of grade 10 students. These initial results are now used to refine the cartoons and create a more diverse set for the next round of piloting.

## Discussion and Future Work

The work has some limitations, most prominently the limited scope of the intervention – focusing solely on reinforcement learning – that was used to derive the outcome space. Nevertheless, the two dimensions map on the central aspects of the third (“Learning”) and second (“Representation and Reasoning”) “Big Idea” of the AI4K12 curriculum (Touretzky et al., 2019) and also align well with the modern notion of a learning agent (Russell & Norvig, 2016) that keeps a model of the world it acts in and uses data to improve this model. Both provide some external validity to the structure of the outcome space.

However, this only applied to the “final” stages of each dimension, i.e. the ones that would also be considered “correct” from a normative point of view. From a phenomenographic perspective, it is important to note, that the intermediate stages should not be considered incorrect. They all serve the purpose of explaining a phenomenon subjectively based on the experiences that one encountered so far (Marton & Booth, 2013).

This poses a rather fundamental question of how best to operationalize such a model. Concept cartoons, for example, usually work by combining correct answers with distractors that are derived from known misconceptions. In our case, the stages of understanding do not necessarily represent useful misconceptions however. If a student does understand that a model may be necessary within a learning agent, but does not yet have the capabilities of explicating parts of this model, the student does not hold a misconception. Therefore, designing a distractor that indicates an “implicit” understanding of the model in contrast to an “explicit” understanding – that would be considered correct from a normative perspective, becomes a difficult and eventually maybe even impossible task.

Based on these considerations and the results of our initial piloting, we are therefore currently investigating ways of designing these cartoons in a way that focusses only on some stages of each dimension, for example by leaving out the implicit stage of the model dimension as this allows to more clearly design sample answers in the cartoons that sound “reasonable” and sufficiently distinct from one another.

The final instrument, regardless of its format, can then be used to investigate the effectiveness of teaching interventions and – in particular – how to address the model dimension that currently appears to be not as affected as the dimension of the learning process. Therefore, in another line of future work we aim to look at the validity of the construct, in particular regarding the dimension of internal model that may be aligned with a more general understanding of modelling and models, as described in literature (Upmeier zu Belzen et al. 2010).

### Using the Phenomenographic Model in Teaching

Finally, an interesting future aspect to consider is the suitability of the phenomengrpahic model as an explanatory model (cf. Höper et al. 2024) in (K12) teaching. Teaching could then follow along the stages of the model to iteratively deepen students’ understanding. Since the finale stages align well with curricula and experts ideas, the model could present a suitable an empirically derived series of reductions that can be effective as a teaching device. For example, such a progression in teaching could mean that lessons first leave the internal model implicit, while presenting repetition as the core idea of machine learning algorithms and then make the model explicit to also pinpoint what the purpose of repetition is: optimization of model parameters.

## References

Essinger, S. D., & Rosen, G. L. (2011). An Introduction to Machine Learning for Students in Secondary Education. In *Ieee Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop (DSP/SPE), 2011: 4 - 7 Jan. 2011, Enchantment Resort, Sedona, Arizona, USA* (pp. 243–248). IEEE.

Gardner, M., & Michie, D. (1982). *Logic Machines and Diagrams* (2. ed.). Univ. of Chicago Press.

Höper, Lukas; Schulte, Carsten; Mühling, Andreas (2024): Learning an Explanatory Model of Data-Driven Technologies can Lead to Empowered Behavior: A Mixed-Methods Study in K-12 Computing Education. In: *Proceedings of the 2024 ACM Conference on International Computing Education Research - Volume 1, 2024/08/12*: ACM (10), S. 326–342.

Keogh, B., & Naylor, S. (1999). Concept cartoons, teaching and learning in science: an evaluation. *International Journal of Science Education*, *21*(4), 431–446.

Marton, Ference; Booth, Shirley (2013): *Learning and Awareness*. Mahwah: Taylor and Francis (Educational Psychology Series).

Mariescu-Istodor, R., & Jormanainen, I. (2019). Machine Learning for High School Students. In P. Ihantola & N. Falkner (Eds.), *Proceedings of the 19th Koli Calling International Conference on Computing Education Research - Koli Calling '19* (pp. 1–9). ACM Press.

Mühling, A., & Große-Bölting, G. (2023). Novices’ conceptions of machine learning. *Computers and Education: Artificial Intelligence*, *4*(4), 100142.

Russell, S. J., & Norvig, P. (2016). *Artificial Intelligence: A Modern Approach* (3rd ed.). *Prentice Hall series in artificial intelligence*. Prentice Hall.

Shulman, L. S. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, *15*(2), 4–14.

Tedre, M., Denning, P., & Toivonen, T. (2021). CT 2.0. In O. Seppälä & A. Petersen (Eds.), *21st Koli Calling International Conference on Computing Education Research* (pp. 1–8). ACM.

Touretzky, D., Gardner-McCune, C., Martin, F., & Seehorn, D. (2019). Envisioning AI for K-12: What Should Every Child Know about AI? *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*, 9795–9799.

Upmeier zu Belzen, Annette; Krüger, Dirk (2010): Modellkompetenz im Biologieunterricht. In: Zeitschrift für Didaktik der Naturwissenschaften 16, S. 41–57.