Shared Vocabulary and Pedagogy for AI Education: Data Concepts, Data Practices, and the Data Case Study

Viktoriya Olari
viktoriya.olari@fu-berlin.de
Freie Universität Berlin
Berlin, Germany

Ralf Romeike
ralf.romeike@fu-berlin.de
Freie Universität Berlin
Berlin, Germany

Abstract

Curricula around the world have started to include content related to artificial intelligence (AI) in their agendas. Although this process is timely and important, it is also challenging because the elaboration of the AI field for K–12 remains ongoing. Current efforts often underappreciate the critical role of data literacy for AI education. If the goal is to enable students to understand how AI systems work and what their implications are, they must understand what data underpins these systems and how that data is collected and processed. To advance knowledge about data literacy in AI education, we conducted a comprehensive theoretical analysis of the data science and AI fields, which led to creating a model of key data-related practices and a collection of key data-related concepts. Using a design-based research process, we also developed a pedagogy for educating K-12 students about data: the data case study. The collection and the model equip teachers with a map and a shared vocabulary. The resulting pedagogy provides them with practical ways to help students develop a conceptual understanding and agency.

CCS Concepts

• **Computing methodologies** → **Artificial intelligence**; • **Social and professional topics** → **K-12 education**.

Keywords

Data Literacy, AI Education, Data Concepts, Data Practices, The Data Case Study

ACM Reference Format:
Viktoriya Olari and Ralf Romeike. 2026. Shared Vocabulary and Pedagogy for AI Education: Data Concepts, Data Practices, and the Data Case Study. In *Proceedings of CHI '26 Workshop: Data Literacy for the 21st Century: Perspectives from Visualization, Cognitive Science, Artificial Intelligence, and Education (CHI '26).* ACM, New York, NY, USA, 6 pages. https://doi.org/10.5281/zenodo.19335526

1 Introduction

Due to advances in artificial intelligence (AI), data-driven systems are becoming a focal point of computer science school education [11, 36]. Young people use systems that learn from data and contribute to them as data producers [6]. Accordingly, computer science school education should prepare them for a life with such systems. This task is challenging because little is known about how young people learn about data in the context of data-driven, opaque, and stochastic AI systems [38]. As data is a fundamental component of data-driven systems, data literacy [30, 40], statistical literacy [33], data management education [8], and transformative data agency[37] can provide a stable foundation for AI education, both conceptually and pedagogically.

2 Related Work

Almost all literature in computer science education research has focused on teaching and learning about classical, rule-based systems and computational thinking [38]. Extensive research within and outside computer science education has addressed the computer science concepts to be taught [5, 32], the design of learning environments, the difficulties involved, and the scaffolding methods [9, 10]. However, suggestions related to teaching about data-driven systems are only starting to emerge. Extensive work has been done from the content perspective, namely, the question "What should be learned?" [18, 21, 34, 35]. However, previous work has largely underestimated the role of data literacy [25]. Pedagogical models [14, 38], studies of young people's preconceptions and ideas [19, 20, 22], and reports from the development of teaching materials [2, 15, 17, 39] have provided little insight into *what* should be learned about data and *how*. The difficulties students face when learning about data in context of data-driven systems, the effects of the teaching methods used, the cognitive load, or the learning processes initiated remain largely unknown.

3 Data Concepts and Data Practices as Shared Vocabulary

In our prior work, we contributed to clarifying the data literacy perspective for AI education [24]. Through a comprehensive analysis of the data literacy, data science, and AI fields, we identified 28 data practices and 133 data concepts that are essential for understanding data-driven systems [26, 29].

Knowing data concepts is essential for communicating about data-driven systems. Just as describing a cell and its functions in biology lessons requires students to become familiar with a set of terms and understand their meanings (e.g., nucleus, Golgi apparatus, etc.), describing data-driven systems requires students to understand the types of data-based tasks that systems can perform (e.g., classification, regression), the data formats they use (e.g., tabular, time series, image), how and where data is stored (e.g., datasets and databases), the types of data problems that occur (e.g., outliers, and missing data), how the data is transformed (e.g., cleaning and feature engineering), and how it is used to solve a task (e.g., data flow, training, validation, and testing data). Students must also understand how the model results are tested and interpreted (e.g., performance metrics, underfitting, and overfitting) [28]. Figure 1 provides an overview of the full list of data concepts.

The goal of computer science education has always been to facilitate not only conceptual understanding but also agency. School

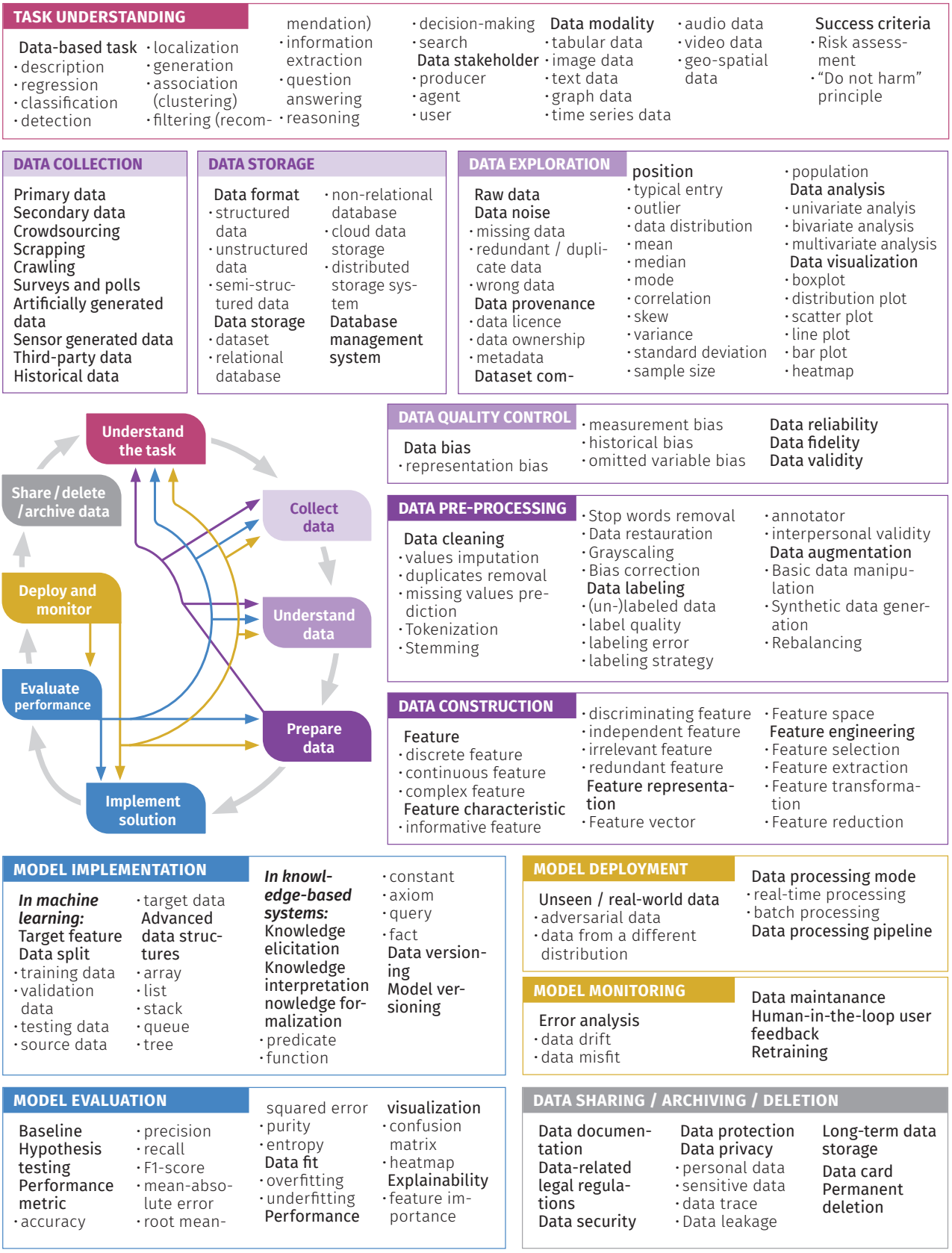

Figure 1: Data concepts form the basis of a shared vocabulary [26].

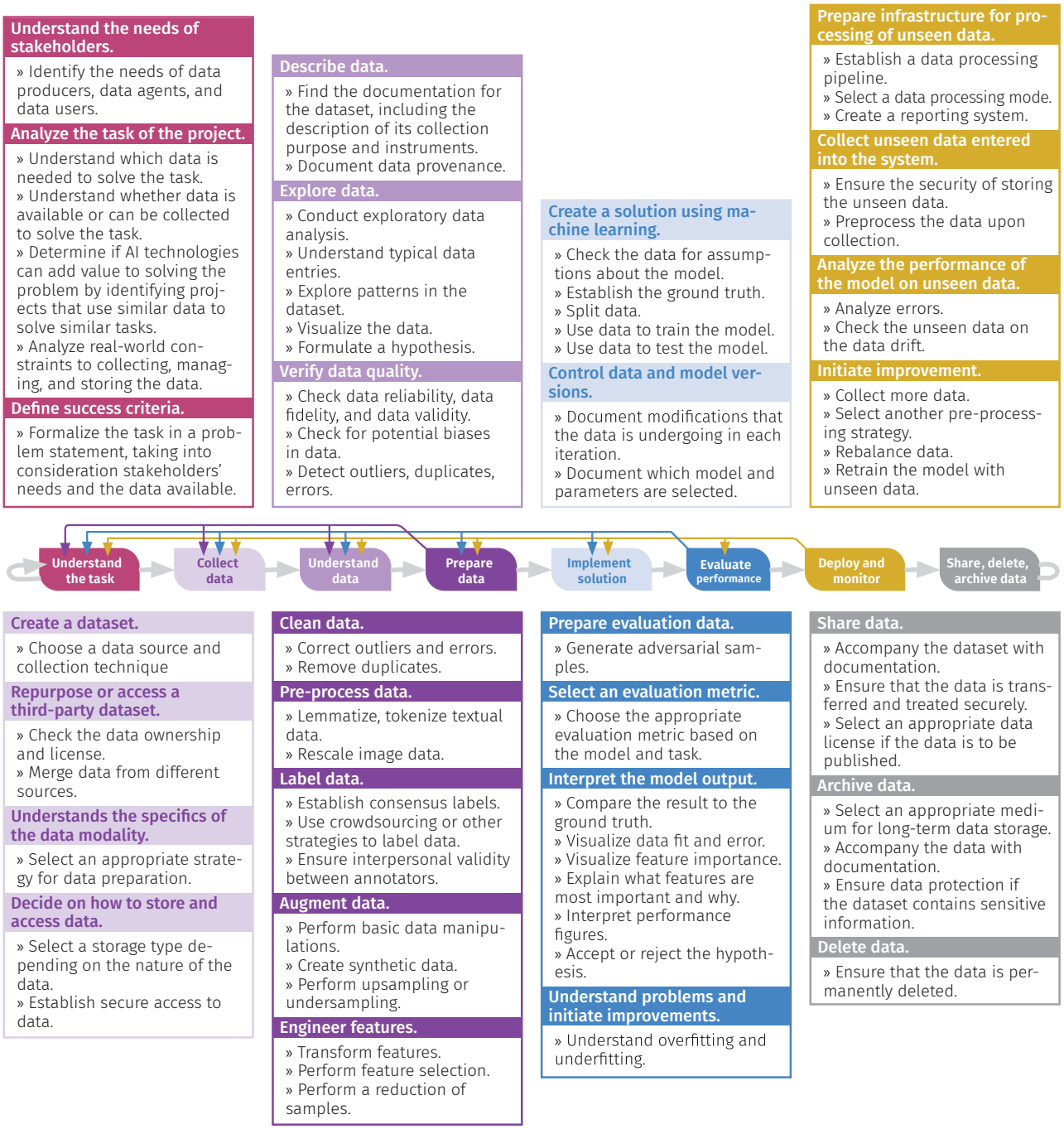

Figure 2: Data practices form the basis for defining competencies [27].

students should learn to design computational systems—not to develop market-ready software products, but to understand how such systems work. Therefore, knowing and understanding the concepts is not sufficient; engaging in the practices is essential. Figure 2 illustrates 28 data practices (and 69 subpractices) that materialize the data-related concepts for teachers in practice and can be used to specify skills.

4 The Data Case Study as Pedagogy

How can data-related concepts and practices be introduced in computer science school education on AI? In academic education in AI and data science, where the teaching of data concepts and practices naturally occurs, an established pedagogical method is the *data case study* (also known as a "case study" or "lab"). This method

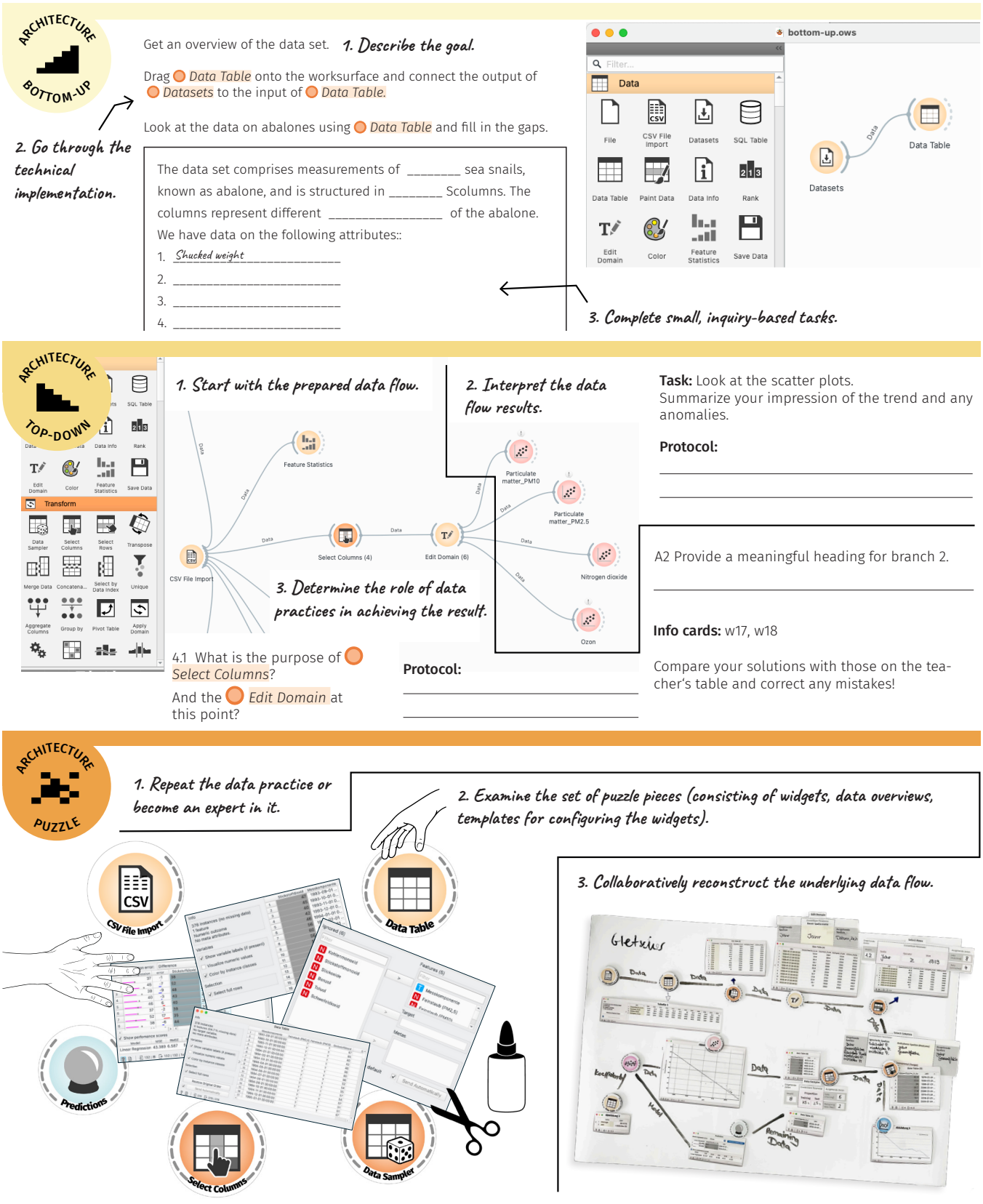

Figure 3: Three architectures of the data case study method: bottom-up, top-down and puzzle-like [28, 29]. The examples above are implemented in the open-source, flow-based data mining environment *Orange3* [4].

is grounded in the tradition of constructivist, active, and situated learning [3, 7, 12, 13, 23, 41] and requires students to solve a *data case*: an authentic problematic situation accompanied by a dataset. In the process, students apply the data concepts and practices taught in lectures, thereby developing data-based judgment and problem-solving skills [16]. Although several researchers in school education on data literacy and AI have used data cases for teaching (e.g., [1]), they have not adapted them to school specific requirements. The cases are code- and text-heavy and require a significant amount of time and prior knowledge to complete. School teaching, however, requires the consideration of school-specific demands such as rigid time constraints and heterogeneous knowledge of computational concepts among school students. Thus, it can be assumed that school teachers will reject the data case study that has not been explicitly adapted as a teaching and learning method for AI school education due to its impracticability.

After clarifying the technical, practical, and didactic challenges, we adapted the *data case study* method from academic data science education for computer science school classes [28]. In a three-cycle design-based research project, we tested and further refined the method in cooperation with computer science teachers, experts in the field of data science, and 44 students in grades 9 and 10 in Germany. During the research process, which included quantitative analyses of teaching quality, student knowledge, and motivation, as well as qualitative evaluation of audio and video data from the classroom, we identified three types—or *architectures*—of the data case study that address subject-specific and didactic requirements, while taking school challenges into account: top-down, bottom-up, and puzzle-like. Figure 3 summarizes three architectures.

Our work has demonstrated that students exhibit moderate to high levels of motivation when working with the data case study. However, their understanding of data concepts and practices, communicative participation, perceived difficulty, and evaluation of teaching quality vary significantly depending on the architecture [29].

5 Current Work

In our current work, we are developing empirically based architecture profiles regarding the objectives of computer science education on data. Based on previous research, we are following the *conjecture mapping* approach outlined by Sandoval [31] to further develop local learning theories on teaching and learning about data in context of computer science school education on AI. Based on our prior work, we assume the following:

(1) The puzzle-like architecture contributes to a deep understanding of data concepts and data practices, as well as to a high level of communicative participation, but is associated with high cognitive load and low motivation.
(2) Bottom-up and top-down architectures are associated with low communicative participation and contribute to the ability to evaluate and design data-driven systems. Motivation is higher and cognitive load is lower with bottom-up architecture than with top-down architecture.
(3) Students can only evaluate and design data-driven systems if they have a comprehensive understanding of data concepts and data practices throughout the data lifecycle.

During the workshop, we will discuss the ongoing research project and share insights from our previous work.

References

[1] Rolf Biehler and Yannik Fleischer. Introducing students to machine learning with decision trees using CODAP and Jupyter Notebooks. *Teaching Statistics*, 43:S133–S142, 2021.
[2] Jessica Van Brummelen, Viktoriya Tabunshchyk, and Tommy Heng. "Alexa, Can I Program You?": Student Perceptions of Conversational Artificial Intelligence Before and After Programming Alexa. In *Interaction Design and Children*, pages 305–313, Athens, Greece, 2021. Association for Computing Machinery.
[3] Valentina Chkoniya. Success Factors for Using Case Method in Teaching Applied Data Science Education. *European Journal of Education*, 4(1):77–86, April 2021.
[4] Janez Demšar, Tomaž Curk, Aleš Erjavec, Črt Gorup, Tomaž Hočevar, Mitar Milutinovič, Martin Možina, Matija Polajnar, Marko Toplak, Anže Starič, Miha Štajdohar, Lan Umek, Lan Žagar, Jure Žbontar, Marinka Žitnik, and Blaž Zupan. Orange: Data mining toolbox in python. *Journal of Machine Learning Research*, 14(71):2349–2353, 2013.
[5] Peter J. Denning. Great principles of computing. *Communications of the ACM*, 46(11):15–20, November 2003.
[6] Feierabend, Sabine, Rathgeb, Thomas, Gerigk, Yvonne, and Glöckler, Stephan. JIM study 2025: Youth, information, media – a baseline study on media use among 12- to 19-year-olds, published in German; Original title: JIM-Studie 2025: Jugend, Information, Medien – Basisuntersuchung zum Medienumgang 12- bis 19-Jähriger. Technical report, Medienpädagogischer Forschungsverbund Südwest.
[7] Grandon T. Gill. *Informing with the Case Method: A Guide to Case Method Research, Writing, & Facilitation.* Informing Science Press, Santa Rosa, 2011.
[8] Andreas Grillenberger and Ralf Romeike. Key concepts of data management: An empirical approach. In *Proceedings of the 17th Koli Calling International Conference on Computing Education Research*, Koli Calling '17, pages 30–39, New York, NY, USA, November 2017. Association for Computing Machinery.
[9] Shuchi Grover, editor. *Computer Science in K-12: An A to Z Handbook on Teaching Programming*. Edfinity, Palo Alto, CA, 2020.
[10] Mark Guzdial and Benedict du Boulay. The history of computing education research. In Sally Fincher and Anthony Robins, editors, *The Cambridge Handbook of Computing Education Research*. Cambridge University Press, Cambridge, 2019.
[11] Lutz Hellmig, Steffen Burk, André Greubel, Martin Hennecke, Henry Herper, André Hilbig, Tilman Michaeli, Alexander Mittag, Arno Pasternak, Hermann Puhlmann, Gerhard Röhner, Michael Rücker, Thomas Schmalfeldt, Wolf Spalteholz, and Peer Stechert. Educational Standards for Computer Science in Lower Secondary Education – Recommendations of the German Informatics Society (GI), published in German; Original title: Bildungsstandards Informatik für die Sekundarstufe I – Empfehlungen der Gesellschaft für Informatik e. V. (GI). 2025.
[12] Clyde Freeman Herreid. Case study teaching. *New Directions for Teaching and Learning*, 2011(128):31–40, December 2011.
[13] Erin Rae Hoffer. Case-based Teaching: Using Stories for Engagement and Inclusion. 2(2):75–80, 2020.
[14] Lukas Höper, Carsten Schulte, and Andreas Mühling. 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*, pages 326–342, Melbourne VIC Australia, August 2024. ACM.
[15] Sven Jatzlau, Tilman Michaeli, Stefan Seegerer, and Ralf Romeike. It s not Magic After All@ Machine Learning in Snap! using Reinforcement Learning. *2019 IEEE Blocks and Beyond Workshop (B&B)*, pages 37–41, 2019.
[16] Jana Lasser, Debsankha Manik, Alexander Silbersdorff, Benjamin Säfken, and Thomas Kneib. Introductory data science across disciplines, using Python, case studies, and industry consulting projects. *Teaching Statistics*, 43(S1), July 2021.
[17] Phoebe Lin, Jessica Van Brummelen, Galit Lukin, Randi Williams, and Cynthia Breazeal. Zhorai: Designing a Conversational Agent for Children to Explore Machine Learning Concepts. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 13381–13388, April 2020.
[18] Duri Long and Brian Magerko. What is AI Literacy? Competencies and Design Considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pages 1–16, New York, NY, USA, April 2020. Association for Computing Machinery.
[19] Erik Marx, Thiemo Leonhardt, and Nadine Bergner. Secondary school students' mental models and attitudes regarding artificial intelligence - A scoping review. *Computers and Education: Artificial Intelligence*, 5:100169, 2023.
[20] Erik Marx, Clemens Witt, and Thiemo Leonhardt. 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*, pages 1–10, Munich Germany, September 2024. ACM.
[21] Tilman Michaeli, Ralf Romeike, and Stefan Seegerer. What Students Can Learn About Artificial Intelligence – Recommendations for K-12 Computing Education. In Keane, Therese, Lewin, Cathy, Brinda, Torsten, and Bottino, Rosa, editors,

Towards a Collaborative Society Through Creative Learning, pages 196–208, Cham, 2022. Springer Nature Switzerland.

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

[23] Deborah Ann Nolan and Terry Speed. *Stat Labs: Mathematical Statistics through Applications*. Springer Texts in Statistics. Springer, New York Berlin Heidelberg, corr. 2. print edition, 2001.

[24] Viktoriya Olari. Data Literacy as a Fundamental Component of Artificial Intelligence Education in Schools (Doctoral Consortium). In *Proceedings of the 23rd Koli Calling International Conference on Computing Education Research*, pages 1–2, Koli Finland, November 2023. ACM.

[25] Viktoriya Olari and Ralf Romeike. Addressing AI and Data Literacy in Teacher Education: A Review of Existing Educational Frameworks. In *The 16th Workshop in Primary and Secondary Computing Education*, page Article 17, Virtual Event, Germany, 2021. Association for Computing Machinery.

[26] Viktoriya Olari and Ralf Romeike. Data-related concepts for artificial intelligence education in K-12. *Computers and Education Open*, 7:100196, December 2024.

[27] Viktoriya Olari and Ralf Romeike. Data-related practices for creating Artificial Intelligence systems in K-12. In *Proceedings of the 19th WiPSCE Conference on Primary and Secondary Computing Education Research*, Munich, Germany, 2024. Association for Computing Machinery.

[28] Viktoriya Olari and Ralf Romeike. The data case study - a teaching and learning method for computer science education in schools. In *Proceedings of the ACM Global on Computing Education Conference 2025 Vol 1*, CompEd 2025, pages 225–232, New York, NY, USA, 2025. Association for Computing Machinery.

[29] Viktoriya Olari and Ralf Romeike. Teaching data concepts and practices in secondary school education on artificial intelligence: Approaches, mechanisms, and emerging local theories. In *Proceedings of the 25th Koli Calling International Conference on Computing Education Research*, Koli Calling '25, New York, NY, USA, 2025. Association for Computing Machinery.

[30] Chantel Ridsdale, James Rothwell, Mike Smit, Michael Bliemel, Dean Irvine, Daniel Kelley, Stan Matwin, Brad Wuetherick, and Hossam Ali-Hassan. Strategies and Best Practices for Data Literacy Education Knowledge Synthesis Report. 2015.

[31] William Sandoval. Conjecture Mapping: An Approach to Systematic Educational Design Research. *Journal of the Learning Sciences*, 23(1):18–36, January 2014.

[32] Andreas Schwill. Fundamental Ideas of Computer Science. 1994.

[33] Milo Shields. Information Literacy, Statistical Literacy, Data Literacy. *IASSIST Quarterly*, 28(2):6, August 2005.

[34] Matti Tedre, Peter Denning, and Tapani Toivonen. CT 2.0. In *Proceedings of the 21st Koli Calling International Conference on Computing Education Research*, Koli Calling '21, pages 1–8, New York, NY, USA, November 2021. Association for Computing Machinery.

[35] David Touretzky, Christina Gardner-McCune, Fred Martin, and Deborah Seehorn. Envisioning AI for K-12: What Should Every Child Know about AI? In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 9795–9799, July 2019.

[36] UNESCO. K-12 AI curricula: A mapping of government-endorsed AI curricula. Technical Report ED-2022/FLI-ICT/K-12, UNESCO, Paris, 2022.

[37] Henriikka Vartiainen, Lotta Pellas, Juho Kahila, Teemu Valtonen, and Matti Tedre. Pre-service teachers' insights on data agency. *New Media & Society*, page 146144482210796, February 2022.

[38] Henriikka Vartiainen and Matti Tedre. The CEDE Model: A Learning-Sciences Based Approach for Critical and Transformative K–12 AI Education. In *Proceedings of the 25th Koli Calling International Conference on Computing Education Research*, pages 1–11, Koli Finland, November 2025. ACM.

[39] Xiaoyu Wan, Xiaofei Zhou, Zaiqiao Ye, Chase K. Mortensen, and Zhengyan Bai. SmileyCluster: Supporting accessible machine learning in K-12 scientific discovery. *Proceedings of the Interaction Design and Children Conference*, 2020.

[40] Annika Wolff, Daniel Gooch, Jose J. Cavero Montaner, Umar Rashid, and Gerd Kortuem. Creating an Understanding of Data Literacy for a Data-driven Society. *The Journal of Community Informatics*, 12(3), August 2016.

[41] Carrie Wright, Qier Meng, Michael R. Breshock, Lyla Atta, Margaret A. Taub, Leah R. Jager, John Muschelli, and Stephanie C. Hicks. Open Case Studies: Statistics and Data Science Education through Real-World Applications. *Journal of Statistics and Data Science Education*, pages 1–30, 2024.

Received 11 February 2026; revised 29 March 2026; accepted 26 February 2026