Extended Abstract for the
Symposium on Integrating AI and Data Science into School Education Across Disciplines 2025

EMPOWERING STUDENTS IN A DATA-DRIVEN WORLD: EXPLANATORY MODELS FOR UNDERSTANDING DATA-DRIVEN TECHNOLOGIES FROM EVERYDAY LIFE

Lukas Höper [1] & Yannik Fleischer [2]
[1]Paderborn University, Germany, lukas.hoeper@uni-paderborn.de
[2]Paderborn University, Germany

Focus Topics: AI and Data Science Competencies, Explanatory Models

Introduction

Data-driven technologies and applications, particularly those using artificial intelligence (AI) and especially machine learning (ML), play an increasingly significant role in everyday life. Data-driven practices are essential to these technologies. To empower students as critical and informed citizens in a data-driven world, teaching students about these data-driven technologies and their role in students' everyday lives is an important rationale. This aligns with goals of data science, statistics, and computing education at the school level. To navigate the data-driven world, we argue that students need to understand the inner workings of the applications they encounter in their daily lives and develop capabilities to actively shape and design such technologies. Teaching about data-driven technologies, therefore, combines empowering students to critically engage with these systems from a user perspective and equipping them to adopt a designer perspective in creating data-driven applications.

These perspectives are often described as different continua. For example, from a computer science perspective, Fischer (2002) introduces a consumer-designer continuum, arguing that people should be enabled to take more active roles and participate in technological development rather than remaining passive consumers. Similarly, the concept of computational empowerment emphasizes the combination of consumer and designer perspectives (Dindler et al., 2020). It emphasizes enabling students to analyze and evaluate digital technologies developed by others, while also designing technologies for others, thereby fostering empowerment through active participation in technological development. A similar discussion exists in statistics education research, where the "consumer vs. producer dichotomy" (Weiland, 2017, p. 34) advocates for students to adopt a producer perspective in the classroom. This is intended to support in enabling students to become critical citizens or critical consumers of data-based phenomena (e.g. Engel, 2017, p. 48; Weiland, 2017, p. 38) or, as Louis (2022, p. 8) puts it, for "supporting learners' abilities to either 'read' or 'write' the world with data."

The second aspect that we want to highlight for our following discussion of teaching about data-driven technologies is the idea of explanatory models (see: Höper et al., 2024a; Höper & Schulte, 2024b). Höper and Schulte (2024) argue for the integration and explicit teaching of explanatory models, especially in the context of AI and ML technologies. They characterize explanatory models from two perspectives: they are representations of digital artifacts or socio-technical systems, and they have specific educational purposes (for more detailed explanations, see: Höper and Schulte, 2024). They may represent aspects of data-driven technologies (e.g., basic concepts for large language models), but they always serve specific purposes, highlighting certain aspects and hiding others, and are intended to be used for specific purposes, such as exploring or reasoning about the inner workings of such systems or designing ML applications. For example, teaching about data-driven technologies might include an explanatory model of the role of data in data-driven systems for exploring and evaluating data practices in such systems, or an explanatory model of specific ML techniques for reasoning about the behavior of an ML system and discussing its limitations in different situations.

The rationale of enabling students to understand data-driven technologies in their everyday lives, and empowering them to become critical and informed citizens, and the idea of explanatory models form the foundation of this work. In this paper, we outline two research projects that explore different approaches to teaching about data-driven technologies.

In S. Podworny & S. Schönbrodt (Eds.), *Towards Fostering AI and Data Science Literacy in Schools Across Disciplines. Proceedings of the 1st Symposium on Integrating AI and Data Science into School Education Across Disciplines (AIDEA1 2025)*, Salzburg, Austria.

The role of data in data-driven technologies: developing data awareness

Given the significant role of data-driven technologies in school students' everyday lives, we have previously emphasized the need for teaching students about these technologies and how they work. Data plays an essential role in such technologies. For instance, social media applications, streaming services, or search engines rely on data practices to provide their features, such as displaying personalized recommendations for news feeds or movies or listing search results tailored to the user. In these contexts, students are often in the role of consumers and focus primarily on these features, which are typically the main reasons why they use these data-driven applications. As mentioned earlier, our goal is to support students in understanding these technologies and become informed citizens who can critically reflect on them. In this section, we focus on the role of data in data-driven technologies.

Recent research indicate that many students lack a clear understanding of the role of data in everyday applications (e.g., Dowthwaite et al., 2020; Goray & Schoenebeck, 2022; Höper & Schulte, 2024c; Pangrazio & Selwyn, 2019). In addition, studies suggest that many people have passive and powerless roles regarding data practices in such technologies, as can be seen in stances and feelings of powerlessness, resignation, and apathy (e.g., Draper & Turow, 2019; Hoffmann et al., 2024; Keen, 2020). Moreover, a study with middle school students indicates that many students have only moderate motivation and intention to engage with the data-driven practices of everyday applications that collect and process data about them (Höper et al., 2024b). Taken together, these research findings underscore the need for teaching students about data-driven technologies in a way that empowers them to engage with the inner workings of the data-driven applications they interact with in everyday life and to critically reflect on their own roles in these interactions.

Based on this rationale, we have developed the *data awareness* framework (Höper et al., 2024a; Höper & Schulte, 2024c, 2024a). Its primary goals are to enable students to understand and reflect on data-driven technologies and applications from their everyday lives, to form their own opinions about these technologies, to participate in societal discussions, and to contribute to shaping the data-driven world. Thus, data awareness seeks to empower students to navigate the data-driven world.

To teach data awareness in middle schools, we developed an approach for teaching an explanatory model of data-driven technologies. The model, designed from the perspective of individual interactions between a user and a data-driven digital artifact, characterizes the role of data. This includes various aspects of data collection, data processing (e.g., via ML methods), the construction and use of data models (e.g., ML model and user models), and the purposes underlying these data practices.

We implemented this approach in two teaching units for secondary school computing education. Students learn the explanatory model and apply it to analyze and evaluate data-driven technologies they encounter in their everyday lives. These interventions build on students' daily experiences and are designed to enable them to move beyond passive consumer roles. Students engage with, explore, analyze, and evaluate such applications, enabling them to adopt more informed and active roles in their daily interactions with data-driven applications. For example, one of these interventions focuses on recommendation systems, which targets students in grades 8 to 10. Students first subsequently learn the explanatory model. According to the different aspects of the model, the students reconstruct the inner workings of recommendation systems. In particular, this includes data-driven practices of collecting user data, processing it with an ML technique (k-nearest neighbor) to generate predictions, refining data models about users (e.g., with inferred data), and creating movie recommendations. Students then critically discuss and evaluate these practices, guided by the "purposes" aspect integrated in the explanatory model. Having learned the different aspects of the model, students identify other examples of data-driven applications in their daily lives. They select an example and deconstruct the role of data in this data-driven application, using the explanatory model as an analytical lens.

In our design-based research project for developing and evaluating the data awareness framework, we have empirically examined students' development of data awareness (Höper et al., 2024a; Höper & Schulte, 2024c, 2024a). The findings suggest that students from grade 6 onwards can learn the explanatory model, apply it to everyday data-driven applications, and relate insights from the teaching units to their everyday experiences. Learning and using the model to analyze and evaluate the role of data helps students develop a more nuanced understanding of data practices. In addition, learning about data awareness encourages students to reflect on their everyday interactions with data-driven

applications and empowers them to take a more informed and empowered role in their interactions with such digital artifacts.

In this workshop, we present an overview of the data awareness framework, detail the educational approach designed to support students' development of data awareness, and share findings from our studies.

Teaching data-based decision trees in secondary classroom to exemplify ML and data-based algorithmic models

As mentioned above, AI and ML affect the modern life. For instance, AI-driven recommender systems give online platforms substantial influence, especially on adolescents. In adolescents various misconceptions about AI are found: some students' perception of AI is to think it can solve "problems 'magically' through its intelligence" (Kim et al., 2023, S. 9835) or "that AI is flawless and complete without human interventions and input from the data" (Kim et al., 2023, S. 9838). This underscores the need for education that demystifies AI and equips students with foundational knowledge about data-based modeling, and the roles of data and human in the process. The concept of AI literacy can give an orientation of what to teach. Casal-Otero et al. (2023) identified subareas of AI literacy, three of which are: Learning how AI works (e. g., teaching ML algorithms), Learning tools for AI (using AI and ML techniques and tools), and Learning for life with AI (critically assessing AI and responsible use of ML). A problem is that despite its relevance, ML is perceived as not accessible by students (Sulmont et al., 2019). So, there is a need for an elementarized version of ML that appears authentic and teachable. Data-based decision trees offer such an accessible entry point to ML, as advocated by researchers (e. g. Erickson & Engel, 2023; Martignon et al., 2022) and taken up by curricula for middle and high school (IDSSP Curriculum team, 2019; MSB NRW, 2023). Teaching about trees is furthermore suspected to support statistical literacy and critical thinking (Erickson & Engel, 2023) but all these hypotheses have to be investigated empirically.

Because these topics are quite new to school teaching, it is not obvious what exactly should be taught and how. To address this, we designed four teaching modules on ML with decision trees for grades 6, 9, 12 basic, and 12 advanced. We developed (or adapted) digital and unplugged tools for teaching and designed activities for students to acquire skills and knowledge about data-based construction of decision trees, broader ML phenomena like overfitting and bias, and detailed evaluation of ML models in a certain context. As outlined above the students take both perspectives on ML created AI-systems: the consumer perspective and the producer perspective. We started investigating in a grade 12 advanced course where the producer perspective was focused. We then went down from there with an orientation towards a spiral curricula approach of teaching decision trees and a stronger focus on students taking the critical consumer perspective. We conducted qualitative studies to characterize what students have taken from class: one study in a grade 12 advanced class (Fleischer et al., 2022), two studies in a grade 12 basic class (Fleischer & Biehler, 2025; Fleischer & Biehler, accepted), and one study in grade 6 (Fleischer et al., 2024). The results of the studies show, on the one hand, that young students can successfully learn and apply fundamental concepts of data-based decision trees and, on the other hand, provide insights into challenges that are important for further developing the instructional design. In the presentation, we will detail the findings and present the design ideas of the teaching.

References

Casal-Otero, L., Catala, A., Fernández-Morante, C., Taboada, M., Cebreiro, B., & Barro, S. (2023). AI literacy in K-12: A systematic literature review. *International Journal of STEM Education*, 10(1), 29. https://doi.org/10.1186/s40594-023-00418-7

Dindler, C., Smith, R., & Iversen, O. S. (2020). Computational empowerment: Participatory design in education. *CoDesign*, *16*(1), 66–80. https://doi.org/10.1080/15710882.2020.1722173

Dowthwaite, L., Creswick, H., Portillo, V., Zhao, J., Patel, M., Vallejos, E. P., Koene, A., & Jirotka, M. (2020). „It's your private information. It's your life.": Young people's views of personal data use by online technologies. *Proceedings of the Interaction Design and Children Conference*, 121–134. https://doi.org/10.1145/3392063.3394410

Draper, N. A., & Turow, J. (2019). The corporate cultivation of digital resignation. *New Media & Society*, *21*(8), 1824–1839. https://doi.org/10.1177/1461444819833331

Engel, J. (2017). Statistical Literacy for Active Citizenship: A Call for Data Science Education. *Statistics Education Research Journal*, 16(1), 44–49. https://doi.org/10.52041/serj.v16i1.213

Erickson, T., & Engel, J. (2023). What goes before the CART? Introducing classification trees with Arbor and CODAP. *Teaching Statistics*, 45(S1). https://doi.org/10.1111/test.12347

Fischer, G. (2002). Beyond „Couch Potatoes“: From Consumers to Designers and Active Contributors. *First Monday*, *7*(12). https://doi.org/10.5210/fm.v7i12.1010

Fleischer, Y., & Biehler, R. (2025). Exploring students' constructions of data-based decision trees after an introductory teaching unit on machine learning. *ZDM – Mathematics Education*, 57(1), 153–173. https://doi.org/10.1007/s11858-025-01663-6

Fleischer, Y., & Biehler, R. (accepted). Analyzing students' informal approaches to creating decision trees in the classroom.

Fleischer, Y., Biehler, R., & Schulte, C. (2022). Teaching and Learning Data-Driven Machine Learning with Educationally Designed Jupyter Notebooks. *Statistics Education Research Journal*, 21(2), 7. https://doi.org/10/gqv59w

Fleischer, Y., Podworny, S., & 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. https://doi.org/10.52041/serj.v23i1.450

Goray, C., & Schoenebeck, S. (2022). Youths' Perceptions of Data Collection in Online Advertising and Social Media. *Proceedings of the ACM on Human-Computer Interaction*, *6*(CSCW2), 1–27. https://doi.org/10.1145/3555576

Hoffmann, C. P., Lutz, C., & Ranzini, G. (2024). Inequalities in privacy cynicism: An intersectional analysis of agency constraints. *Big Data & Society*, *11*(1), 1-13. https://doi.org/10.1177/20539517241232629

Höper, L., & Schulte, C. (2024a). Empowering Students for the Data-Driven World: A Qualitative Study of the Relevance of Learning about Data-Driven Technologies. *Informatics in Education*, *23*(3), 593–624. https://doi.org/10.15388/infedu.2024.19

Höper, L., & Schulte, C. (2024b). New Perspectives on the Future of Computing Education: Teaching and Learning Explanatory Models. *24th Koli Calling International Conference on Computing Education Research*, 1–8. https://doi.org/10.1145/3699538.3699558

Höper, L., & Schulte, C. (2024c). The data awareness framework as part of data literacies in K-12 education. *Information and Learning Sciences*, *125*(7/8), 491–512. https://doi.org/10.1108/ILS-06-2023-0075

Höper, L., Schulte, C., & Mühling, A. (2024a). Learning an Explanatory Model of Data-Driven Technologies can Lead to Empowered Behavior: A Mixed-Methods Study in K-12 Computing Education. *Proceedings of the 2024 ACM Conference on International Computing Education Research - Volume 1*, 326–342. https://doi.org/10.1145/3632620.3671118

Höper, L., Schulte, C., & Mühling, A. (2024b). Students' Motivation and Intention to Engage with Data-Driven Technologies from a CS Perspective in Everyday Life. *Proceedings of the 2024 Innovation and Technology in Computer Science Education V. 1 (ITiCSE 2024)*, 1-7. https://doi.org/10.1145/3649217.3653625

IDSSP Curriculum team. (2019). *Curriculum Frameworks for Introductory Data Science*. http://idssp.org/files/IDSSP_Frameworks_1.0.pdf

Keen, C. (2020). Apathy, convenience or irrelevance? Identifying conceptual barriers to safeguarding children's data privacy. *New Media & Society*, *24*(1), 1–20. https://doi.org/10.1177/1461444820960068

Kim, K., Kwon, K., Ottenbreit-Leftwich, A., Bae, H., & 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. https://doi.org/10.1007/s10639-023-11600-3

Martignon, L., Erickson, T., & Viale, R. (2022). Transparent, simple and robust fast-and-frugal trees and their construction. *Frontiers in Human Dynamics*, *4*, 1-13. https://doi.org/10.3389/fhumd.2022.790033

MSB NRW. (2023). *Kernlehrplan für die Sekundarstufe I*. https://www.schulentwicklung.nrw.de/lehrplaene/lehrplan/340/gesk_wpif_klp_2023_06_01.pdf

Pangrazio, L., & Selwyn, N. (2019). 'Personal data literacies': A critical literacies approach to enhancing understandings of personal digital data. *New Media & Society*, *21*(2), 419–437. https://doi.org/10.1177/1461444818799523

Sulmont, E., Patitsas, E., & Cooperstock, J. R. (2019). Can You Teach Me To Machine Learn? In E. K. Hawthorne, M. A. Pérez-Quiñones, S. Heckman, & J. Zhang (Hrsg.), *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*. ACM. 948–954. https://doi.org/10/ghnpcs

Weiland, T. (2017). Problematizing statistical literacy: An intersection of critical and statistical literacies. *Educational Studies in Mathematics*, *96*(1), 33–47. https://doi.org/10.1007/s10649-017-9764-5