# LESSONS LEARNED FROM THE PRODABI PROJECT: SHAPING PERSPECTIVES AT THE INTERSECTION OF DATA, AI, AND EDUCATION TITLES

Rolf Biehler[1] & Carsten Schulte[2]
[1]Paderborn University, Germany, rolf.biehler@uni-paderborn.de
[2]Paderborn University, Germany

*Focus Topics: AI and Data Science Competencies, AI and Data Science Curricula and Implementation in School*

## Abstract

The ProDaBi (Project Data Science and Big Data in Schools) initiative aims to rethink the interplay between data, artificial intelligence (AI), and education and aims to develop educational conceptions, classroom material, professional development courses and accompanying research since its launch in 2017. Since then, the fields of data science and artificial intelligence (AI) have undergone rapid transformation, including profound changes in their societal role and the way they are addressed in education. This talk reflects on key insights, emerging trends, and challenges encountered through the experiences and outcomes of the ProDaBi project

## Overview of the ProDaBi Project

The ProDaBi project was conceived as a response to the growing need for critical engagement with data and AI technologies in school. Its core mission was to bridge the gap between theoretical knowledge and practical application, fostering awareness and literacy that extend beyond technical expertise. The core idea was to bridge computing education, statistics education and societal issues (Biehler et al. 2018).

The project has developed tools, materials, and frameworks that integrate technical concepts with real-world issues, offering students and educators tools and approaches to explore data and AI in a socially aware manner.

Starting with a project course for the grade before the Abitur over one full school year that was developed in close cooperation with a local Gymnasium, the project expanded to develop teaching units to be used in at secondary level in grade 5/6 and 9/10, accompanied in-service teacher education and currently to integrate its teaching concepts and its approach to data projects in an interdisciplinary setting, again with a local school combining the subjects computer science, biology and geography in secondary education at grade 9/10.

These developments also reflect changes in data science and AI and how educational systems respond to these transformations. Before presenting some overarching elements of ProDaBi in some more detail we reflect on these developments from the ProDaBi perspective.

## Data and AI in Society: A Comparative Analysis of selected 2017 to 2024 Trends

The societal landscape for data and AI has evolved dramatically since 2017. What began as a field with limited public awareness has now become a cornerstone of everyday life, influencing governance, healthcare, business, and education.

Regulatory efforts, such as the European Union's AI Act (Council of the European Union 2022; Gstrein et al. 2024), reflect growing recognition of the need for structured oversight of AI technologies. These frameworks aim to balance innovation with accountability, addressing concerns about transparency, fairness, and ethical usage, see e.g. the German Ethikrat (Ethikrat 2023).

In parallel, the discourse around using AI and data science for societal benefit has evolved. The concept of “AI or data science for social good” has gained traction, inspiring initiatives that harness AI to address complex societal challenges, from climate change to public health (Tomašev et al. 2020). Organizations like *Algorithm Watch* and academic journals such as *Big Data in Society* reflect these developments.

Generative AI is transforming the way individuals interact with technology. The growing accessibility of these tools is enabling widespread use in creative, professional, and educational

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.

contexts. However, this also raises critical questions about ethical boundaries, misinformation, and societal dependency on automated systems. The need for a fundamental understanding of AI technologies, including machine learning and the critical role of data, is more relevant than ever.

**Data and AI in Education: Shifts and Opportunities**

These developments significantly influenced educational practices, particularly in terms of curriculum design and pedagogical priorities.

This can be seen in the growing emphasis on AI literacy, which combines technical understanding with critical perspectives on societal impacts, e.g. (Long & Magerko 2020). It aims to equip students with both technical skills and the ability to critically assess AI systems and their societal impact. This is reflected in the increasing inclusion of topics like machine learning and algorithmic decision-making in computing education. This growing emphasis on AI literacy combines technical understanding with critical perspectives on societal impacts (Casal-Otero et al. 2023; Long & Magerko 2020; Ng et al. 2021; Sperling et al. 2023). It aims to equip students with both technical skills and the ability to critically assess AI systems and their societal impact.

Beyond computer science, another important development is the integration of data science education into interdisciplinary contexts (NCTM et al. 2024). By extending beyond traditional technical domains, subjects such as mathematics, social sciences, and humanities are incorporating data science to address real-world challenges. Critical data literacy (Weiland 2017)—teaching students to analyze and interpret data with an understanding of its societal and cultural contexts—has become a focal point. It encourages learners to question how data is produced, analyzed, and used, as well as its implications for decision-making and social structures (Ridgway 2022).

At the same time, the adoption of AI tools in education has accelerated. Generative AI technologies are increasingly integrated into teaching and learning practices. Tools like ChatGPT are being explored not only for their utility in assisting with academic tasks but also for their potential to reshape teaching and learning practices.

Another noticeable trend is the shift toward tool-centric (or: tool supported) pedagogy (see e.g. (Gresse von Wangenheim et al. 2021; Höper & Schulte 2024a) ). By embedding tools in the curriculum, educators help students connect abstract computational concepts with everyday experiences. This approach promotes intuitive understanding and equips learners with practical skills to engage with AI and data-driven systems effectively.

**ProDaBi's Contributions: Concepts, Materials, and Innovations**

The talk will conclude with an introduction and overview of the ProDaBi project, which provides a specific approach to integrating data science and AI into education.

We present five core aspects and initiatives; focusing on the approach to teaching data science at school – somewhat leaving out our parallel efforts in in-service teacher education and workshops with teachers. Note, our approaches were developed and tested in close cooperation witch teachers (our network of teachers comprises roughly 300 teachers, mostly from computer science), and also in connection and collaboration with the community and our virtual ProDaBi colloquium series.

Our teaching approaches focus on four aspects in two areas. First regarding AI and data literacy, navigating the digital world is one area. The second area is about analyzing and exploring data, where fundamental ideas of statistics education must be revisited (Biehler 2022).

*Data Awareness*

A first focus of the project is on supporting students' development of data awareness (Höper, Schulte & Mühling 2024; Höper & Schulte 2024b). By combining practical exercises with reflective activities, ProDaBi helps students understand the role and significance of data in shaping both individual experiences and collective structures. Our interventions address data-driven technologies through a combination of exploring their inner workings and critically reflecting on their role in everyday life. To do this, ProDaBi provides an explanatory model (see for the idea of explanatory models: Höper & Schulte 2024a) to equip learners with the skills and perspectives needed to navigate a data-driven world.

*Machine Learning*

The project's tools and materials help educators illustrate complex topics like machine learning, data transparency, and the ethical dimensions of algorithmic decision-making through accessible, real-world examples. The project introduces learners to foundational concepts in machine learning. It enables students to engage critically with algorithmic systems, understanding not only how they work but also their potential biases and limitations. We use different tools such as data cards, CODAP with its plug-in Arbor and Python-based Jupyter Notebooks with different degrees of necessary coding knowledge (Biehler et al. 2021; Podworny et al. 2021; Fleischer et al. 2022, Fleischer et al. 2024)

*Data exploration*

ProDaBi also emphasizes innovative approaches to data exploration with multivariate data (Podworny et al. 2022), such as storytelling, which encourages students to combine technical analysis with creative and interpretive skills. One aspect is to design data exploration processes that move beyond traditional frameworks like the PPDAC (Problem, Plan, Data, Analysis, Conclusion) (Wild & Pfannkuch 1999) cycle to emphasize more dynamic and engaging approaches such as the CRISP-DM (Shearer 2000) cycle for data science. The project also addresses the evolving role of visual and coding and tools in data exploration, but also hands-on materials for enactive data exploration. Data exploration is embedded within meaningful narratives that highlight societal relevance (Ridgway 2022).

*Epistemic Programming*

ProDaBi also emphasizes the importance of epistemic programming—teaching programming as a means for exploring personally meaningful phenomena or questions, for example in the context of data explorations (Hüsing, Schulte, & Winkelnkemper 2023). This (learning) process can be supported by a guiding worked example that presents a worked out and documented inquiry process on a similar topic (Hüsing et al. 2024). For documentation, they could develop a computational essay that is suitable for capturing their programming and inquiry process in a comprehensible way (Hüsing & Podworny 2022). Overall, students are encouraged to reflect on the epistemological implications of programming and to use it as a tool for exploring personal interests.

*GeoMINT: An interdisciplinary data science project course for grades 9/10*

In project-based learning environments, such as the GeoMint interdisciplinary courses, which is currently under development and the upper-secondary (Sek II) project course offerings (Biehler et al 2018; Heinemann et al. 2018), ProDaBi has demonstrated the effectiveness of collaborative, hands-on approaches. These courses emphasize the integration of computational tools with domain-specific knowledge, providing students with a holistic understanding of data science and AI applications. We combine units with systematic learning of science topics (such as climate change) or topics from data science (such as data exploration and machine learning methods) with phases of self-directed project work by the students themselves.

Through courses such as advanced secondary-level classes and the GeoMint program, students engage in collaborative, real-world problem-solving that integrates perspectives from computer science, mathematics, and societal challenges. These courses demonstrate how data science and AI education can foster both technical proficiency and broader critical awareness.

## Future Directions and Challenges

One pressing issue is the need for sustained investment in teacher education, ensuring that educators are equipped to deliver AI and data literacy content effectively. ProDaBi will move in the direction to offer on-line professional development courses with a geographically speaking wider range. Moreover, we are restructuring our website www.prodabi.de in the direction of a resource center for teachers with classroom material, tools, data sets, and background information of AI and data science topics. It will offer many self-learning opportunities for students and teachers, such as instructional videos. Collaborative research will also be crucial for refining pedagogical strategies and identifying best practices.

Another challenge is addressing disparities in access to AI and data literacy education. As these topics become increasingly central to societal participation, ensuring equitable access will be critical to

avoiding digital divides. ProDaBi's inclusive approach offers a model for addressing these challenges, emphasizing the importance of contextualized and culturally relevant teaching materials.
ProDaBi also aims to continue adapting its frameworks to the evolving technological landscape. This includes exploring new applications of generative AI within the current units ProDaBi has developed, fostering interdisciplinary collaborations, and deepening its engagement with ethical and societal issues. By remaining responsive to these trends, ProDaBi seeks to ensure that its contributions remain relevant and impactful in the years to come.

**References**

Biehler, R., Budde, L., Frischemeier, D., Heinemann, B., Podworny, S., Schulte, C., & Wassong, T. (Eds.). (2018). *Paderborn Symposium on Data Science Education at School Level 2017: The Collected Extended Abstracts*. Universitätsbibliothek Paderborn. https://doi.org/10.17619/UNIPB/1-374.

Biehler, R., Frischemeier, D., Podworny, S., Wassong, T., Budde, L., Heinemann, B., & Schulte, C. (2018). Data Science and Big Data in Upper Secondary Schools: A Module to Build up First Components of Statistical Thinking in a Data Science Curriculum. *Archives of Data Science, Series A (Online First)*, *5*(1), 28. https://doi.org/10.5445/KSP/1000087327/28

Biehler, R., & Fleischer, Y. (2021). Introducing students to machine learning with decision trees using CODAP and Jupyter Notebooks. *Teaching Statistics*, *43*(S1), S133-S142. https://doi.org/10.1111/test.12279

Biehler, R. (2022). Revisiting Fundamental Ideas for Statistics Education From the Perspective of Machine Learning and Its Applications. In S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), *Bridging the Gap: Empowering and Educating Today's Learners in Statistics. Proceedings of the Eleventh International Conference on Teaching Statistics*. https://doi.org/10.52041/iase.icots11.T1A2

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). https://doi.org/10.1186/s40594-023-00418-7

Council of the European Union (2022). Artificial Intelligence Act: Council Calls for Promoting Safe AI That Respects Fundamental Rights. 6 December 2022. https://europa.eu/!7kcT3x.

Ethikrat, Deutscher (2023). Mensch und Maschine - Herausforderungen durch Künstliche Intelligenz https://mediatum.ub.tum.de/1753389.

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). https://doi.org/10.52041/serj.v21i2.61

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). https://doi.org/10.52041/serj.v23i1.450

Gresse von Wangenheim, C., Hauck, J. C. R., Pacheco, F. S., & Bertonceli Bueno, M. F. (2021). Visual tools for teaching machine learning in K-12: A ten-year systematic mapping. *Education and Information Technologies*, *26*(5), 5733–5778. https://doi.org/10/gqv592

Gstrein, O. J., Haleem, N., & Zwitter, A. (2024). General-purpose AI regulation and the European Union AI Act. *Internet Policy Review*, *13*(3), 1–26. https://doi.org/10.14763/2024.3.1790

Heinemann, B., Budde, L., Schulte, C., Biehler, R., Frischemeier, D., Podworny, S., & Wassong, T. (2018). Data Science and Big Data in Upper Secondary Schools: What Should Be Discussed From a Perspective of Computer Science Education? *Archives of Data Science, Series A*, *5*(1), 1-18. https://doi.org/10.5445/ksp/1000087327/26

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

Höper, L., & Schulte, C. (2024b). 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. (2024). 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

Hüsing, S., & Podworny, S. (2022). Computational Essays as an Approach for Reproducible Data Analysis in lower Secondary School. Proceedings of the IASE 2021 Satellite Conference. IASE 2021 Satellite Conference: Statistics Education in the Era of Data Science. https://doi.org/10/gqv58h

Hüsing, S., Schulte, C., Sparmann, S., & Bolte, M. (2024). Using Worked Examples for Engaging in Epistemic Programming Projects. *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, 443–449. https://doi.org/10.1145/3626252.3630961

Hüsing, S., Schulte, C., & Winkelnkemper, F. (2023). Epistemic Programming. In S. Sentance, E. Barendsen, N. R. Howard, & C. Schulte (Eds.), *Computer Science Education: Perspectives on Teaching and Learning in School* (2nd ed., pp. 291–304). Bloomsbury Academic; Bloomsbury Collections. http://dx.doi.org/10.5040/9781350296947.ch-022

Long, D., & Magerko, B. (2020). What is AI Literacy? Competencies and Design Considerations. In R. Bernhaupt, F. "Floyd" Mueller, D. Verweij, J. Andres, J. McGrenere, A. Cockburn, I. Avellino, A. Goguey, P. Bjørn, S. (Shen) Zhao, B. P. Samson, & R. Kocielnik (Eds.), *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). ACM. https://doi.org/10/ghbz2q

NCTM, NSTA, ASA, NCSS, & CSTA. (2024). Data Science A Joint Position of NCTM, NSTA, ASA, NCSS, and CSTA. https://www.nctm.org/Standards-and-Positions/Position-Statements/Data-Science/

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence, 2*, 100041. https://doi.org/10/gqv59z

Podworny, S., Fleischer, Y., Hüsing, S., Biehler, R., Frischemeier, D., Höper, L., & Schulte, C. (2021). Using data cards for teaching data based decision trees in middle school. In *21st Koli Calling International Conference on Computing Education Research (Koli Calling '21), November 18-21, 2021, Joensuu, Finland*. ACM. https://doi.org/10.1145/3488042.3489966

Podworny, S., Fleischer, Y., Stroop, D., & Biehler, R. (2022). An example of reach, real and multivariate survey data for use in school. In J. Jodgen, E. Geraniou, G. Bolondi, & F. Ferretti (Eds.), *Proceedings of the Twelfth Congress of the European Society for Research in Mathematics Education (CERME12)* (pp. 940-947). Free University of Bozen-Bolzano and ERME. https://hal.science/CERME12/hal-03751842v1

Ridgway, J. (2022) (Ed.). *Statistics for empowerment and social engagement. Teaching civic statistics to develop informed citizens*. Springer. https://doi.org/10.1007/978-3-031-20748-8

Shearer, C. (2000). The CRISP-DM Model—The New Blueprint for Data Mining. *Journal of Data Warehousing*, 5(4), 15.

Sperling, K., Stenberg, C.-J., McGrath, C., Åkerfeldt, A., Heintz, F., & Stenliden, L. (2024). In search of artificial intelligence (AI) literacy in teacher education: A scoping review. *Computers and Education Open*, *6*, 100169. https://doi.org/https://doi.org/10.1016/j.caeo.2024.100169

Tomašev, Nenad, Julien Cornebise, Frank Hutter, Shakir Mohamed, Angela Picciariello, Bec Connelly, Danielle C. M. Belgrave, et al. (2020). 'AI for Social Good: Unlocking the Opportunity for Positive Impact'. *Nature Communications* 11 (1): 2468. https://doi.org/10.1038/s41467-020-15871-z

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

Wild, C., & Pfannkuch, M. (1999). Statistical Thinking in Empirical Enquiry. *International Statistical Review*, *67*(3), 223-265.