(Author's Draft)

Data Literacy as a Fundamental Component of Artificial Intelligence Education in Schools (Doctoral Consortium)

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

ABSTRACT

School students actively contribute to artificial intelligence (AI) technologies as data producers. As future professionals and creators of AI technologies, they will shape the data culture in various domains. Therefore, AI literacy is proposed as a top priority for education systems worldwide. However, when it comes to education about AI in schools, there is little evidence on which data-related skills are fundamental and how to help students acquire them effectively. In this work, I propose considering data literacy as a fundamental component in AI education. The main objective of this research is to discuss this position and present a method for introducing AI through the lens of data literacy using case studies from the real world.

CCS CONCEPTS

• **Computing methodologies** → *Artificial intelligence*; • **Applied computing** → *Education*; • **General and reference** → *Empirical studies*.

KEYWORDS

data literacy, Artificial Intelligence literacy, school education, case study

ACM Reference Format:
Viktoriya Olari. 2023. Data Literacy as a Fundamental Component of Artificial Intelligence Education in Schools (Doctoral Consortium). In *23rd Koli Calling International Conference on Computing Education Research (Koli Calling '23), November 13–18, 2023, Koli, Finland*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3631802.3631839

1 INTRODUCTION

Due to the advancement and ubiquity of AI technologies, AI education is proposed to become a top priority in education systems worldwide [12]. Researchers in the field of AI education have suggested that in order to thrive in a data-driven world, children need to become familiar with AI from an early age [11, 13] and develop *AI literacy* - a set of competencies that enable individuals to critically evaluate AI technologies, communicate and collaborate effectively with AI, and use AI as a tool [4]. In addition to being critical end users, children also need to develop competencies to be creators of AI [5], which requires students to understand the technology behind the scenes. Since building AI systems based on machine learning typically involves finding or creating datasets for training, testing, and evaluation [10], the ability to work critically with data is paramount, both when deploying an AI system and using one that has already been built. Data skills are also important when working with other AI approaches, such as symbolic AI, where data is manually handcrafted when engineering knowledge [9]. Therefore, I hypothesize that the effective development of AI literacy should include the development of *data literacy*, defined as "the ability to collect, manage, evaluate, and apply data, in a critical manner" [7]. Although current research suggests a connection between data literacy and AI literacy [4, 5, 11], empirical work in the field of AI education rarely prioritizes the explicit development of data literacy and a solid theoretical foundation that explores the relationship between these two literacies (including related concepts such as critical big data literacy [8], and statistical literacy [2]) is lacking. Moreover, there is little evidence on pedagogical approaches to effectively include data-related competencies in AI school education.

Koli Calling '23, November 13–18, 2023, Koli, Finland

ACM ISBN 979-8-4007-1653-9/23/11.
https://doi.org/10.1145/3631802.3631839

2 METHOD, FEASIBILITY STUDY AND FUTURE WORK

Recent literature review investigated how AI literacy and data literacy are related and currently taught in schools [6]. It found that several empirical studies included one or more stages of the data lifecycle when introducing AI to K-12 students. Pedagogical activities often addressed artificially created, pre-structured tasks, and students were not expected to develop their knowledge and apply it to new contexts. Based on these findings, I investigated approaches to teaching data science and AI in higher education, where there is a longer tradition of teaching these subjects. I found that a recommended approach is to use real-world case studies [1], which refers to the application of AI technology to a real-world problem. To assess whether this could be feasible in schools context, I collaborated with experts and educators in a discipline commonly offered in schools (history) and designed a real-world case study to introduce students to data lifecycles and machine learning (Table 1). The case study is set in a historical museum and can be used in either history or computer science classes. First, students are presented with the problem as described in the problem statement and develop a solution under the guidance. While working on the case study, they acquire competencies in data literacy (related to the data lifecycle incl. modeling, collecting and cleaning data) and AI literacy (requirements analysis, training, fine-tuning and deployment of classification models) as well as historical literacy (learning about the objects and their historical contexts such as place and context of discovery, time of origin, material and design style). Second, after understanding the workflow, students are encouraged to develop further ideas about how classification models might be used in historical research in a museum and to implement their ideas in a new case study. I tested the case study with two groups of adolescents (n=30, 14-16 years old) in two 4h workshops. Initial experiments have shown that the approach is feasible, although adjustments need to be made, e.g. to introduce more technical details of classification models according to student feedback. Following the design-based approach, I aim to further develop the real-world case studies for different contexts by involving computer science teachers and teachers from other subjects. I also aim to use qualitative evidence synthesis [3] to identify specific data-related skills and knowledge domains relevant to the development of AI-based systems. Finally, I plan to conduct a study with a large group of students to understand whether the explicit development of data literacy positively correlates with the development of AI literacy.

Table 1: Case study "kAIa Expedition at Pergamonmuseum", developed in cooperation with experts from the Pergamonmuseum

Element	Description
Problem statement	You are a student intern working in the depot of the Pergamonmuseum in Berlin - one of the most famous museums in the world for housing objects from ancient times. Your job is to classify fragments from the latest archaeological excavation as vases, plates, or vessels. You have heard that AI is increasingly being used to reconstruct fragments from archaeological finds, analyze ancient inscriptions and fill in the missing vocabulary, and automatically translate texts from ancient languages, so you wonder if there are any techniques you could try to speed up the process: After all, there are still thousands of fragments waiting to be classified! Your mentor suggests that you could use supervised machine learning to train a classification model that would help you with your problem.
Dataset	*Option 1:* You can use an existing dataset. It consists of thousands of images of fragments of objects found during recent excavations. These may include fragments of vases, plates, and vessels. There are many good images, but some of them are very dark, so you cannot easily see the color of the fragment. *Option 2:* You can also collect your own data from the museum to enhance the existing dataset or to create a completely new dataset that meets your needs.
Guiding questions	Is there a visible difference between fragments of vases, plates, and vessels (material, design, size)? How many examples do you need to train a model that can accurately predict the type of fragment? Does the color of the image or the shape of the fragment affect the accuracy of your model?

REFERENCES

[1] Becky Allen, Andrew Stephen McGough, and Marie Devlin. 2021. Toward a Framework for Teaching Artificial Intelligence to a Higher Education Audience. *ACM Transactions on Computing Education* 22, 2 (Nov 2021), 15:1–15:29. https://doi.org/10.1145/3485062

[2] Robert Gould. 2017. Data Literacy is statistical literacy. *Statistics Education Research Journal* 16, 1 (may 2017), 22–25. https://doi.org/10.52041/serj.v16i1.209

[3] Maria J. Grant and Andrew Booth. 2009. A typology of reviews: an analysis of 14 review types and associated methodologies: A typology of reviews, *Maria J. Grant & Andrew Booth. Health Information & Libraries Journal* 26, 2 (June 2009), 91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

[4] Duri Long and Brian Magerko. 2020. What is AI Literacy? Competencies and Design Considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–16. https://doi.org/10.1145/3313831.3376727

[5] Tilman Michaeli, Ralf Romeike, and Stefan Seegerer. 2023. What Students Can Learn About Artificial Intelligence – Recommendations for K-12 Computing Education. In *Towards a Collaborative Society Through Creative Learning*, Therese Keane, Cathy Lewin, Torsten Brinda, and Rosa Bottino (Eds.). Springer Nature Switzerland, Cham, 196–208. https://doi.org/10.1007/978-3-031-43393-1_19

[6] Viktoriya Olari, Kamilla Tenório, and Ralf Romeike. 2023. Introducing Artificial Intelligence Literacy in Schools: A Review of Competence Areas, Pedagogical Approaches, Contexts and Formats. In *Towards a Collaborative Society Through Creative Learning*, Therese Keane, Cathy Lewin, Torsten Brinda, and Rosa Bottino (Eds.). Springer Nature Switzerland, Cham, 221–232. https://doi.org/10.1007/978-3-031-43393-1_21

[7] Chantel Ridsdale, James Rothwell, Mike Smit, Michael Bliemel, Dean Irvine, Dan Kelley, Stan Matwin, Brad Wuetherick, and Hossam Ali-Hassan. 2015. Strategies and Best Practices for Data Literacy Education Knowledge Synthesis Report. https://doi.org/10.13140/RG.2.1.1922.5044

[8] Ina Sander. 2020. What is critical big data literacy and how can it be implemented? *Internet Policy Review* 9, 2 (May 2020). https://doi.org/10.14763/2020.2.1479

[9] Rudi Studer, V.Richard Benjamins, and Dieter Fensel. 1998. Knowledge engineering: Principles and methods. *Data & Knowledge Engineering* 25, 1-2 (March 1998), 161–197. https://doi.org/10.1016/S0169-023X(97)00056-6

[10] Thordis Sveinsdottir, Pinelopi Troullinou, Stergios Aidlinis, Alexandra Delipalta, Rachel Finn, Panagiotis Loukinas, Julia Muraszkiewicz, Ryan O'Connor, Katrina Petersen, Michael Rovatsos, Nicole Santiago, Diana Sisu, Mistale Taylor, and Peter Wieltschnig. 2020. The Role of Data in AI. https://doi.org/10.5281/zenodo.4312907

[11] David Touretzky, Christina Gardner-McCune, Fred Martin, and Deborah Seehorn. 2019. Envisioning AI for K-12: What Should Every Child Know about AI? *Proceedings of the AAAI Conference on Artificial Intelligence* 33, 9795–9799. https://doi.org/10.1609/aaai.v33i01.33019795

[12] UNESCO. 2021. Beijing Consensus on Artificial Intelligence and Education. (Jul 2021). https://unesdoc.unesco.org/ark:/48223/pf0000368303

[13] Randi Williams, Hae Won Park, Lauren Oh, and Cynthia Breazeal. 2019. PopBots: Designing an Artificial Intelligence Curriculum for Early Childhood Education. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 9729–9736. https://doi.org/10.1609/aaai.v33i01.33019729

Received 7 July 2023; revised 25 October 2023; accepted 05 September 2023