



Data Literacy as a Fundamental Component of AI Education in Schools

Data is at the heart of all AI systems.

Building Artificial Intelligence (AI) systems based on **machine learning** typically involves finding or creating datasets for training, testing, and evaluation [1]. Therefore, 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 [2].

Therefore, I hypothesize that the effective development of AI literacy should include and go beyond the development of **data literacy**, defined as "the ability to collect, manage, evaluate, and apply data, in a critical manner" [3].

RQ 3 + 4

Which teaching approaches could be used to develop data-related AI competencies among school students?

What are engaging case studies for students to effectively build data-related skills when working with AI?

BACKGROUND

According to a recent systematic literature review, AI educators often use the project-based learning approach to educate K-12 students about AI [4]. I decided to take the advantages of this approach and combine it with the method of real-world case studies, which is often used in the education of data scientists and AI engineers in higher education.

A **case study** describes a real-world case or situation [5], usually based on a real-world dataset or including a collection of the own dataset. It includes requirements for a deliverable, such as producing an action report, and is linked to an abstract concept that educators want students to master.

METHODS

In my research, I want to adapt the case study method used in

higher education to school education. In collaboration with experts from other domains, I am investigating domain-specific real-world cases that are solved using AI techniques. To assess whether the case could be transferred to schools context, I collaborate with educators in schools and universities. I follow the method of participatory design research (e.g., by conducting workshops with teachers and students to co-design the case studies) and the design-based approach.

RESULTS

I collaborated with experts from a discipline commonly offered in schools (**history**) and designed a real-world case study to introduce students to data lifecycle and machine learning.

The case study is set in a historical **museum** and can be used in either history or computer science classes.

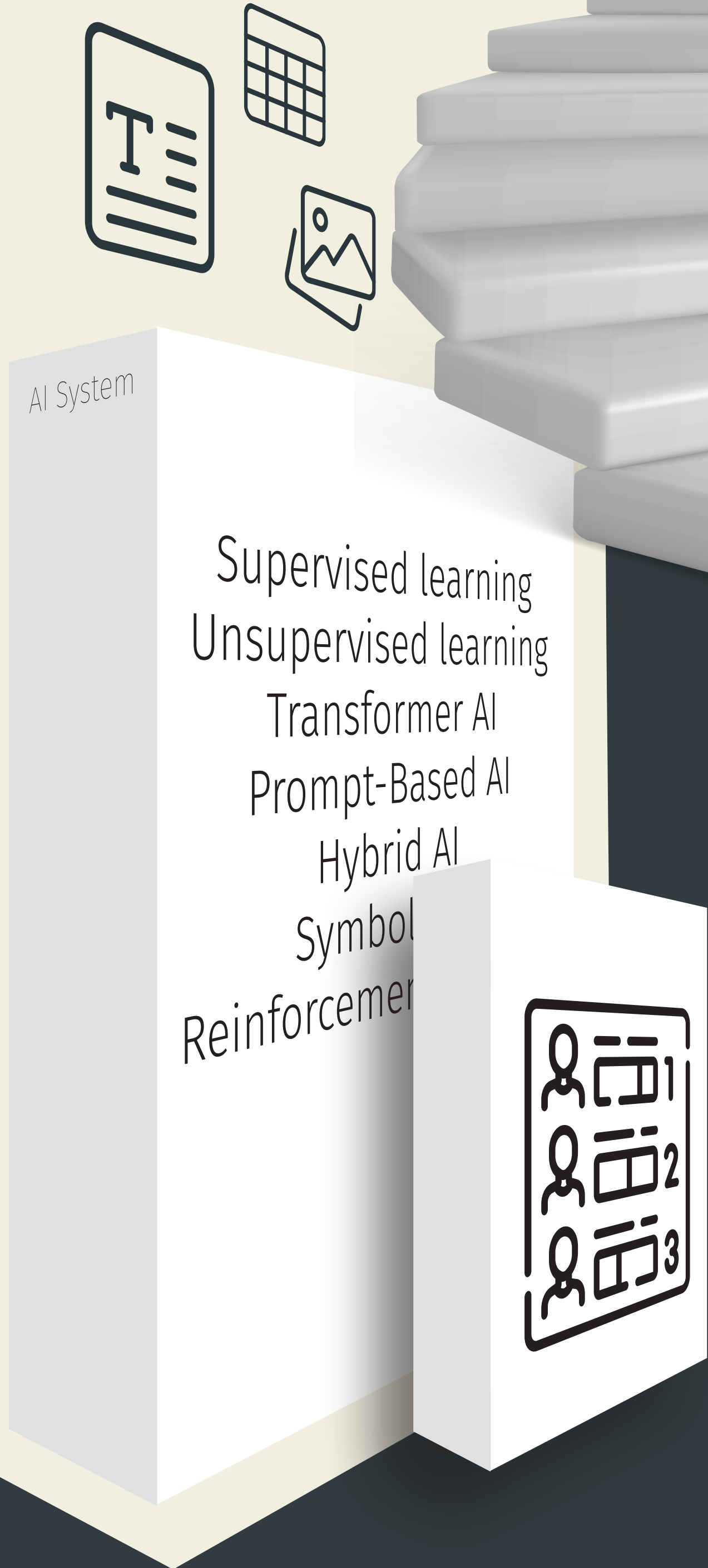
First, students are presented with the problem and develop a solution under the guidance. While working on the case study, they acquire competencies in **data literacy** (related to the data lifecycle incl. modeling, collecting and cleaning data) and **AI literacy** (requirements analysis, training, 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, students are encouraged to develop further ideas and to implement their ideas in a new case study.



Credit: Staatliche Museen zu Berlin, Vorderasiatisches Museum / Olaf M. TesmerCC BY-SA 4.0

Step 3: Map the data competencies from Step 1 to the data lifecycle from Step 2.

Rationale: The knowledge extracted in step 2 helps uncover gaps in current AI education frameworks.



RQ 1 + 2

What school students need to know about data in order to create and responsibly use AI systems?

Which data-related competencies are important for school students, in order to create and responsibly use AI systems?

METHODS

Step 1: Find relevant K-12 AI education frameworks through reviewing the AI education research. Gather data-related competencies from these frameworks.

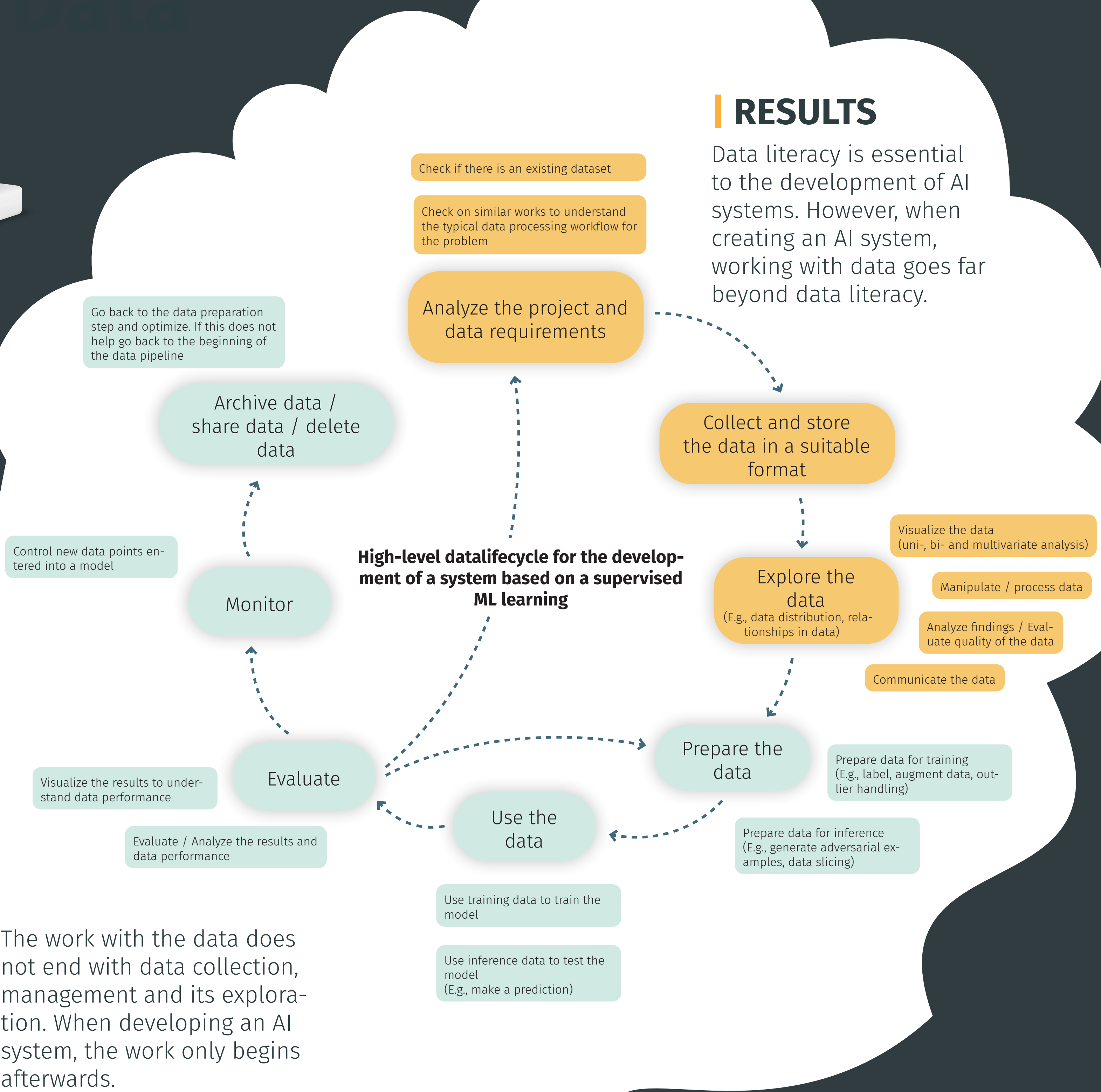
Rationale: The frameworks are based on expert knowledge and are the most comprehensive source of what professionals think students need to know about AI.

Creator of an AI-based system
Student collects, analyzes, models, evaluates, and designs his/her own AI-based system.

Student produces data while interacting with AI-based system (e.g., likes, attention span).

Producer of data

Consumer of data
Student consumes data produced by AI-based system (e.g., video recommendations)



The work with the data does not end with data collection, management and its exploration. When developing an AI system, the work only begins afterwards.

[1] Thordis Sveinsdottir, Pinelopi Troullinou, Stergios Aidinis, Alexandra Delipalta, Rachel Finn, Panagiotis Loukinas, Julia Muraszewicz, Ryan O'Connor, Katrina Petersen, Michael Rovatsos, Nicole Santiago, Diana Sis, Mistale Taylor, and Peter Wietchnig. 2020. The Role of Data in AI.

[2] Rudi Studer, V.Richard Benjamins, and Dieter Fensel. 1998. Knowledge engineering: Principles and methods. Data & Knowledge Engineering 25, 1-2 (March 1998), 161-197.

[3] Chantel Ridsdale, James Rothwell, Mike Smit, Michael Bliemel, Dean Irvine, Dan Kel-

ley, Stan Matwin, Brad Wuetherick, and Hossam Ali-Hassan. 2015. Strategies and Best Practices for Data Literacy Education Knowledge Synthesis Report.

[4] Saman Rizvi, Jane Waite, and Sue Sentence. 2023. Artificial Intelligence teaching and learning in K-12 from 2019 to 2022: A systematic literature review. Computers and Education: Artificial Intelligence (June 2023), 100145.

[5] Erin Rae Hoffer. 2022. Case-Based Teaching: Using Stories for Engagement and Inclusion. International Journal on Social and Education Sciences, v2 n2 p15-30 2020

