sustainability

MDPI

Article

Developing Standards for Educational Datasets by School Level: A Framework for Sustainable K-12 Education

In-Seong Jeon [1], Shin-Yu Kim [2] and Seong-Joo Kang [2,*]

[1] Gwangju Hyodong Elementary School, Gwangju 61212, Republic of Korea; jinsung4069@gnue.ac.kr
[2] Department of Chemistry Education, Korea National University of Education, Cheongju 28173, Republic of Korea; 20226043@knue.ac.kr
* Correspondence: sjkang@knue.ac.kr

Abstract: As artificial intelligence (AI) and data science education gain importance in K-12 curricula, there is a growing need for well-designed sustainable educational datasets tailored to different school levels. Sustainable datasets should be reusable, adaptable, and accessible to support long-term AI and data science education goals. However, research on the systematic categorization of difficulty levels in educational datasets is limited. This study aims to address this gap by developing a framework for sustainable educational dataset standards based on learners' developmental stages and data preprocessing requirements. The proposed framework consists of five levels: Level 1 (grades 1–4), where data preprocessing is unnecessary; Level 2 (grades 5–6), involving basic data cleaning; Level 3 (grades 7–9), requiring attribute manipulation; Level 4 (grades 10–12), involving feature merging and advanced preprocessing; and Level 5 (teachers/adults), requiring the entire data science process. An expert validity survey was conducted with 22 elementary and secondary school teachers holding advanced degrees in AI education. The results showed high validity for Levels 1–4 but relatively lower validity for Level 5, suggesting the need for separate training and resources for teachers. Based on the CVR results and expert feedback, the standards for Educational Datasets were revised, particularly for Stage 5, which targets teachers and adult learners. The findings highlight the importance of expert validation, step-by-step experiences, and an interdisciplinary approach in developing educational datasets. This study contributes to the theoretical understanding of educational datasets and provides practical implications for teachers, students, educational institutions, and policymakers in implementing effective and sustainable AI and data science education in K-12 settings, ultimately fostering a more sustainable future.

Keywords: artificial intelligence; data science; sustainable education; AI education; data science education; educational datasets; expert validity; K-12 education

Citation: Jeon, I.-S.; Kim, S.-Y.; Kang, S.-J. Developing Standards for Educational Datasets by School Level: A Framework for Sustainable K-12 Education. *Sustainability* **2024**, *16*, 4954. https://doi.org/10.3390/su16124954

Academic Editors: GuoJun Ji, Jaejong Park and Michael H. Gyamerah

Received: 23 April 2024
Revised: 31 May 2024
Accepted: 5 June 2024
Published: 10 June 2024

1. Introduction

AI is rapidly developing and having a massive impact on our lives [1,2]. AI has the potential to drive significant advancements in sustainability by optimizing resource use, reducing waste, and enhancing efficiency across various sectors [3]. Recognizing its potential, countries worldwide, including the United States, China, and the United Kingdom, are formulating national policies to integrate AI into their educational systems. First, in the United States, AI education utilizing machine learning is emphasized, and the Association for the Advancement of Artificial Intelligence (AAAI) and the Computer Science Teachers Association (CSTA) have formed and are operating a collaborative body called the 'AI4K12 Initiative' [4]. Through this collaborative body, they develop and research national guidelines for AI education for K-12, standard curricula, and teacher training programs. After presenting the necessity of AI education through the "Next Generation AI Development Plan", the Chinese government implemented AI textbook distribution, online platforms, and AI pilot schools through Zhilong X Artificial Intelligence

Education to introduce AI into school education [5]. In the United Kingdom, students equivalent to elementary school students in Korea are taught understanding, writing, debugging, and predicting programs, as well as skills to create digital content. They also teach how to use software to achieve goals using data, and recognizing the need for AI education, they have developed platforms such as Machine Learning for Kids (ML4K) to allow students to directly create programs that solve problems with data such as text, images, and numbers [6]. In line with this global trend, Korea has recently aimed to cultivate digital literacy in the 2022 revised curriculum to meet the competencies required by future society. In terms of information education, the necessity of convergence education was emphasized so that students can understand and utilize cutting-edge digital innovation technologies such as AI and big data. As artificial intelligence education begins to be introduced into elementary and secondary education, the importance of data education as a fundamental field of artificial intelligence is being further emphasized [7].

However, despite the growing recognition of the importance of AI and data science education, there are significant challenges and barriers to incorporating these subjects into K-12 curricula. One major challenge is the lack of teacher training and expertise in these fields, as highlighted by [8]. Many teachers may not have the necessary knowledge and skills to effectively teach AI and data science concepts, as these are relatively new and rapidly evolving areas [9]. Providing adequate training and professional development opportunities for educators is crucial for successful integration of these subjects into the curriculum.

The term 'Big Data' is closely related to the development of AI and is defined as an information technology that goes beyond the simple meaning of large-scale data in various forms to extract valuable information by analyzing data in various ways and predict future changes [10]. As the term 'Big Data' emerged, the term 'Data Science' (DS) naturally became common as well. Data science analyzes and infers the characteristics of data based on data [11]. As the whole world is changing to a data-based society, students' ability to understand and utilize data has been greatly emphasized in the recent educational field. In this respect, students need to cultivate 'data literacy', the ability to read and use data [12]. Ref. [13] presented the concept of science data literacy (SDL), which is the ability to handle data related to science, especially among data literacy, and emphasized the ability to understand and utilize scientific data. Two effects can be expected through SDL education; first, students can have the ability to understand and utilize the data necessary for the inquiry process, and second, students can acquire professional skills related to data management such as data collection, processing, analysis, and evaluation. SDL requires abilities in various aspects of handling data. Specifically, in the data collection process, the ability to understand and design the process of collecting data through observation, experimentation, and investigation is required. In the data processing and analysis process, the ability to understand and utilize techniques for preprocessing data, the ability to find and handle errors or outliers, the ability to understand and apply statistical analysis methods, and the ability to identify patterns or trends using data visualization techniques are required. In addition, the ability to evaluate data necessary to draw conclusions requires the ability to evaluate the reliability and validity of data, and the ability to interpret data and results through statistical inference and significance testing [13].

However, ref. [14] pointed out that in actual educational settings, teachers often face the challenge of needing to provide well-preprocessed data that are suitable for analysis so that students can analyze accurate data, rather than allowing students to experience the entire process of data collection, processing, and analysis.

Another challenge is the limited resources and infrastructure available for implementing AI and data science education [15]. Schools and educational institutions, especially those in underserved areas, may lack the necessary technology, software, and hardware to support these programs [13]. Developing age-appropriate content and ensuring a coherent progression of skills and knowledge across grade levels are also essential for effective implementation, but can be difficult to achieve.

Ethical and privacy concerns surrounding the use of AI and data science in education pose additional challenges. The collection, storage, and usage of student data raise important issues that must be addressed to ensure responsible and trustworthy practices [16]. Efforts must also be made to ensure equal access to AI and data science education for all students, regardless of their socioeconomic background or geographic location, in order to bridge the digital divide and provide opportunities for underrepresented groups [14].

Despite these challenges, the importance and necessity of research on incorporating AI and data science education cannot be overstated. As these technologies become increasingly prevalent in various industries, it is crucial to equip students with the knowledge and skills needed to succeed in a rapidly changing world. Research on effective strategies for incorporating these subjects into education will help ensure that students are well-prepared for future careers and challenges [10]. This study aims to contribute to this important endeavor by developing criteria for educational datasets tailored to different school levels.

Moreover, AI and data science education can help students develop critical thinking, problem solving, and computational thinking, which are essential for fostering innovation and tackling complex problems in various domains. By engaging with real-world applications and data-driven approaches, students can learn to develop solutions and contribute to positive social change [11]. Research on best practices for nurturing these skills through AI and data science education is crucial for harnessing the full potential of these technologies for the betterment of society.

In summary, while there are significant challenges and barriers to incorporating AI and data science education, the importance and necessity of research in this area cannot be understated. By addressing these challenges and leveraging the insights gained from research, we can ensure that students are well-prepared for the future and equipped with the skills and knowledge needed to harness the full potential of AI and data science for the benefit of society. This study focuses on developing standards for educational datasets to support effective data science education across different school levels and subjects.

2. Theoretical Background

2.1. Data Science

The term 'Data Science' is becoming more common along with the term 'Big Data'. Science refers to knowledge obtained through systematic research. Accordingly, data science focuses on data and expands to mean systematic research on statistics [11].

2.1.1. Components of Data Science

According to the data science Venn Diagram shown in Figure 1, the components that make up data science are computer skills, mathematical and statistical knowledge, and domain expertise. If you only have computer skills and mathematical and statistical knowledge without domain expertise, it will be limited to machine learning. The intersection of mathematical and statistical knowledge and domain expertise belongs to the realm of traditional disciplines, where relatively more time is spent on increasing expertise in the discipline than acquiring computer skills. And if you only have computer skills and domain expertise, you can structure data without accurately understanding the meaning of the coefficients, so it is said to be a danger zone [17]. Although you need to have all three areas of skills to perform data science, you do not need to fully possess the skills in all three areas [18].

2.1.2. Data Science Process

Currently, the process of doing data science is not clearly defined, and scholars define it differently.

Ref. [19] presented a comprehensive data science life cycle consisting of seven stages, as shown in Figure 2: business understanding, data mining, data cleaning, data exploration, feature engineering, predictive modeling, and data visualization. This life cycle provides a structured approach to solving problems using data science techniques, ensuring that each stage is given adequate attention and resources. The rationale behind this cycle is to optimize the accuracy and effectiveness of data-driven solutions by carefully considering the problem, collecting relevant data, cleaning and organizing the data, exploring trends, selecting and constructing features, choosing appropriate machine learning algorithms, and effectively communicating insights through visualization.

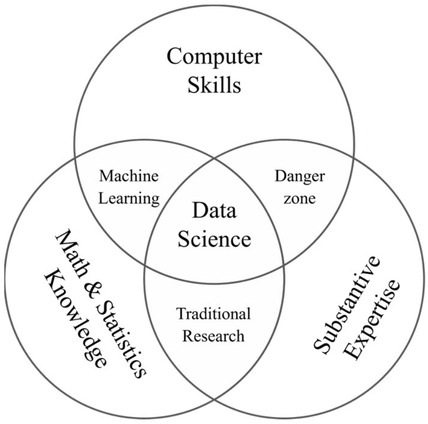

Figure 1. Data science Venn Diagram [17].

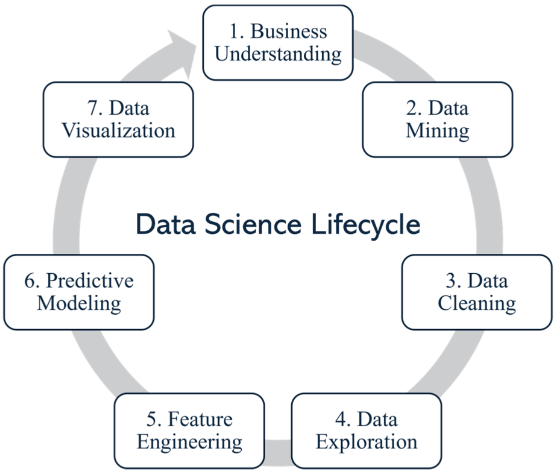

Figure 2. Data science life cycle [19].

Ref. [20] presented a data science process that resembles an algorithm, consisting of six stages, as shown in Figure 3: Business Understanding, Data Collection, Data Understanding, Data Preparation, Data Modeling, Deploy and Iterate. The rationale behind this process is to ensure that the data science project is aligned with the business objectives, the data are collected and prepared appropriately, and the model is developed, validated, and deployed effectively. The process emphasizes the importance of understanding the problem, collecting relevant data, preparing the data for modeling, evaluating the model's accuracy and relevance, and iterating based on feedback from real-world application. This structured approach helps to optimize the outcomes of data science projects and ensures that the solutions developed are both accurate and relevant to the business needs.

Ref. [21] presented the OSEMN model, which represents the data science process as data collection (Obtain data, O), data preprocessing (Scrub data, S), data exploration (Explore data, E), data modeling (Model data, M), and data interpretation (iNterpret data, N). In the O stage, necessary data are collected through various sensors, and in the S stage, preprocessing processes, such as merging individual data into one, cleaning incorrect values, data normalization, and handling extreme values, are performed. In the E stage, data can be explored to identify trends or correlations between data. In the M stage, predictive models can be created using machine learning, and the prediction accuracy can be measured according to how well the model predicts new data. Finally, in the N stage, conclusions are drawn from the data, and the results are evaluated.

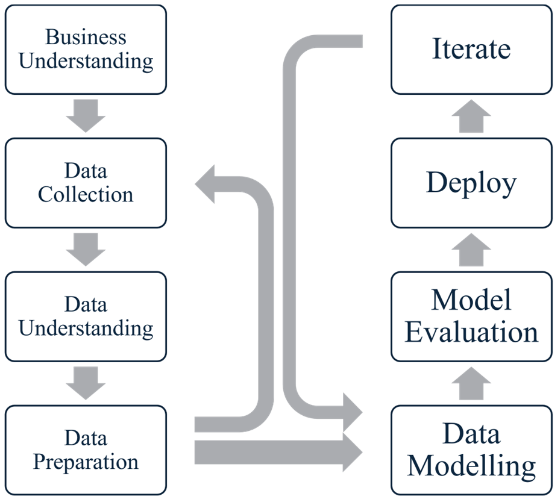

Figure 3. Data science process [20].

Ref. [22] outlines a five-stage approach to data science projects, as shown in Figure 4: Business Understanding, Data Acquisition & Understanding, Modeling, Deployment, and Customer Acceptance. The rationale behind this process is to ensure that the project is aligned with the business objectives, the data are collected and understood, the machine learning models are developed and evaluated, and the final product is deployed and accepted by the customer. This structured approach helps to optimize the outcomes of data science projects by focusing on key variables, feature engineering, model training, and evaluation, while ensuring that the deployed model meets the customer's goals. The TDSP framework provides a clear roadmap for data science teams to follow, ensuring that projects are completed efficiently and effectively.

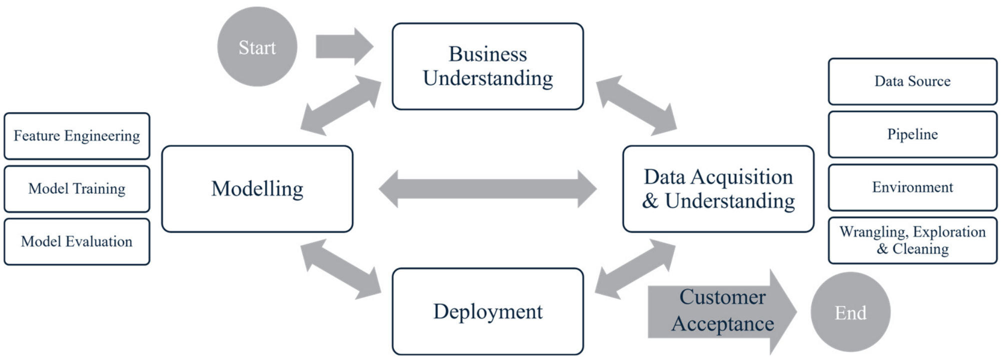

Figure 4. Data science process in TDSP [22].

Based on the previous studies reviewed, it can be confirmed that the data science process goes through common stages, albeit with different expressions. Table 1 summarizes the data science process discussed so far, divided into the stages of problem recognition, data collection, data preprocessing, data exploration, modeling, and value creation.

Among the stages of data science, many scholars have further subdivided and studied the data preprocessing stage in particular. According to [23], the data preprocessing process consists of data integration, data transformation, data cleaning, data reduction, and data discretization. Ref. [24] composed the data preprocessing processes of data cleaning, data reduction, data scaling, data transformation, and data partitioning. Based on previous studies, Ref. [25] summarized the contents of data preprocessing, as shown in Table 2.

Table 1. Comparison of data science process.

Process	[19]	[20]	[21]	[22]
Business Understanding	Business understanding	Business Understanding		Business Understanding
Data Collection	Data mining	Data Collection	Obtain data	Data Acquisition
Data Preprocessing	Data Cleaning Feature Engineering	Data Preparation	Scrub data	
Data Exploration	Data Exploration	Data understanding	Explore data	Data Understanding
Modeling	Predictive Modeling	Data Modeling	Model data	Modeling
Deployment	Data Visualization	Deployment and Iteration	Interpret data	Deployment Customer Acceptance

Table 2. Data preprocessing process and activities [25].

Category	Process	Detailed Activity
Data Cleaning	Filling in Missing Values	Removing or replacing missing or empty values with other values (e.g., mean, mode)
	Clarifying and Removing Outliers	Removing or changing values that significantly deviate from the normal range
Data Transformation	Grouping/Calculating Summary Statistics	Extracting specific information from the dataset and utilizing representative attributes such as mean or median
	Creating Attributes	Creating new data using existing data to add and utilize attributes
	Normalization	A method of expressing learning data with different units in a unified system, utilizing methods such as min–max normalization and z-score normalization
Data Discretization	Converting Continuous Data to Categorical Data	Converting continuous data to categorical data for cases that only target categorical data or to increase the accuracy of analysis and measurement
Data Integration	Integrating Multiple Databases or Files	Combining two or more datasets to utilize as learning data
Data Reduction	Dimensionality Reduction	A method of reducing the attributes of a dataset to improve learning efficiency by extracting and utilizing essential attributes
	Sampling	A method of reducing the number of data (records) by selecting and utilizing arbitrary data

2.2. Data Science in School Education

To better understand the purpose and characteristics of educational datasets, it is important to define what constitutes a "dataset" in this context. In this study, an educational dataset refers to a structured collection of data that can be used to support teaching and learning activities in K-12 education. The purpose of using educational datasets is to provide students with opportunities to engage in authentic data analysis, develop data literacy skills, and apply their knowledge to real-world problems.

Data science has significant potential for cross-curricular integration in K-12 education. The skills and knowledge involved in data science, such as data literacy, computational thinking, and statistical reasoning, are applicable across various school subjects [26]. For example, data analysis techniques can be used in science classes to interpret experimental results, in social studies to examine demographic trends, and in language arts to analyze text data. Incorporating data science principles into different subject areas allows students to develop transferable skills and see the real-world relevance of data in diverse contexts.

However, despite the growing recognition of the importance of data science education, there is limited research on effective strategies for integrating data science into K-12 curricula. Existing studies have primarily focused on specific aspects of data science education, such as teaching data literacy, developing computational thinking skills, or using data visualization tools [27,28]. There is a need for more comprehensive research that examines the challenges and opportunities of implementing data science across different subject areas and grade levels.

In the field of subject didactics, some studies have explored the use of data in teaching and learning specific disciplines. For example, in science education, researchers have investigated how students engage in data collection, analysis, and interpretation as part of scientific inquiry [29]. In mathematics education, studies have examined how students develop statistical reasoning skills through working with datasets [30]. These studies provide valuable insights into the role of data in disciplinary learning, but they do not explicitly address the broader concept of data science as an interdisciplinary field.

To enable students to experience the data science process discussed earlier in the school education setting, instructors can utilize two types of datasets and must select the data needed based on the purpose. The data selection process varies depending on whether students work with primary or secondary data, and through this, students can have diverse experiences related to data science [14]. In the case of primary data, instructors increase complexity by not providing protocols or guidelines for data collection. This way, students take responsibility for collecting or using data on their own. In this activity, students must decide for themselves what data to collect and the methods and tools needed to collect those data, requiring knowledge of how to design protocols for data collection. On the other hand, in the case of secondary data, instructors start with the experience of providing students with only the data needed to solve a specific scientific problem, and students select variables from the provided dataset and explore independently. Therefore, it builds students' ability to select appropriate data from larger datasets later on.

Students can experience inquiry-based learning depending on the type of dataset provided by the teacher. The inquiry level is a measure of how deeply students engage in inquiry and problem solving during the inquiry learning process. Inquiry levels can be divided into four categories according to students' learning objectives [31]. First, confirmation inquiry, a low-level inquiry, is when the instructor provides students with the learning procedure and informs them of the results in advance. This is useful when the learning objective is to confirm and verify already given facts or concepts, and it is mainly conducted by confirming results through simple experiments or observations. Second, in structured inquiry, a mid-level inquiry, the instructor similarly provides the inquiry procedure but does not mention any facts or concepts. Students collect evidence on their own and generate hypotheses. This allows students to gradually develop the ability to perform more open-ended inquiries. Third, in guided inquiry, a high-level inquiry, the instructor only provides students with questions, and students directly design the procedures and explanations to answer the questions. This allows students to learn how to plan experiments and record data on their own. It also helps students engage more actively in learning, build knowledge in a self-directed manner, and improve problem-solving skills. Finally, in open inquiry, a creative inquiry, students act like scientists, have the opportunity to derive questions, design and conduct experiments, and draw conclusions.

Based on these insights, we hypothesize that integrating data science education through a structured framework of educational datasets will enhance students' data literacy, computational thinking, and problem-solving skills across various subjects. By providing students with datasets prepared by teachers that are appropriate for their school level and proficiency, students can receive data science education that is tailored to their developmental stage. The objectives of this study are to develop and validate a framework for educational datasets tailored to different school levels and to evaluate its effectiveness in improving students' data science competencies.

3. Research Methods

3.1. Description and Characteristics of Educational Dataset for K-12

Based on the insights gained from the comprehensive literature review on data science processes, methods of utilizing data in school education, and existing educational dataset standards, we propose a set of criteria for educational datasets tailored to different school levels. The proposed criteria align the data preprocessing stages with the developmental stages of learners, ensuring that the datasets are appropriate for their cognitive abilities and learning objectives.

To apply this educational dataset in the school setting, instructors must decide whether to use primary or secondary data according to the learning objectives. Instructors can divide the way they provide secondary data to students according to difficulty. The difficulty of data is low in the order of level 1, level 2, level 3 and level 4. Level 1 data are complete data that learners do not need to preprocess, level 2 allows students to experience preprocessing within a single dataset, level 3 data allow students to experience preprocessing processes, and level 4 data allow students to experience all preprocessing processes, including data integration. Specifically, level 1 is the level of data through which learners can most easily experience data science, and it is the final data that have undergone all preprocessing processes. If level 1 data are provided to learners, learners can analyze data using well-selected data without experiencing the complex data preprocessing process. This allows learners to draw more accurate and cleaner conclusions from the dataset and focus more on data analysis. Table 3 presents the suggested description and characteristics of educational datasets for each school level.

Table 3. Description and characteristics of educational dataset for K-12.

Level (Appropriate Grade)	Description	Data Preprocessing Process Experienced by Learners	Characteristics
Level 1 (K-1 to 4)	Stage where data preprocessing is unnecessary		Unstructured data Data are completed by the instructor or a completed dataset is used
Level 2 (K-5 to 6)	Stage where only the cleaning process to remove unnecessary data is required	Data Cleaning	Dataset is provided by the instructor, and unnecessary data are removed according to the instructor's guidance as needed
Level 3 (K-7 to 9)	Stage where attributes (e.g., labels) are added or unnecessary data are removed	Data Cleaning, Data Discretization, Data Transformation	Dataset is provided by the instructor, and data are removed or necessary data such as sum/average/labels are added by the learner as needed
Level 4 (K-10 to 12)	Stage where necessary features are merged and subsequent cleaning processes must be performed	Data Integration, Data Cleaning, Data Discretization, Data Transformation	Data sources and collection methods, preprocessing methods are provided Outliers, missing values, etc., must be identified and cleaned, and necessary data such as sum/average/labels are added by learner
Level 5 (Teacher (Adult))	Stage where learners must devise the entire process from data collection to preprocessing on their own	Entire Process	Dataset is not provided Data collection and preprocessing methods are not provided (including the task of identifying what features should be included in the data and finding where data with such features are located)

Educational datasets for each school level are divided into five stages according to the learners' developmental stage.

In Stage 1, corresponding to elementary school grades 1–4, learners handle structured data and utilize datasets that have been fully preprocessed by the instructor. This allows learners to focus on basic data analysis tasks without the need for data preprocessing, as they are still developing foundational data literacy skills.

In Stage 2, corresponding to elementary school grades 5–6, learners at this level are introduced to basic data cleaning techniques, such as removing unnecessary data, under the guidance of the instructor. This helps them understand the importance of data quality and the need for data preprocessing, without overwhelming them with more advanced techniques.

In Stage 3, corresponding to middle school grades 1–3, learners experience more advanced data preprocessing techniques, such as data discretization and transformation, in addition to data cleaning. They learn to add attributes or remove unnecessary data from datasets provided by the instructor, developing a deeper understanding of how data preprocessing impacts data analysis.

In Stage 4, corresponding to high school grades 1–3, learners engage in various preprocessing processes, including data integration, cleaning, discretization, and transformation. They are provided with data sources, collection methods, and preprocessing methods, but are required to identify and handle outliers and missing values independently, fostering critical thinking and problem-solving skills.

Finally, in Stage 5, for teachers (adults), learners are expected to devise the entire process from data collection to preprocessing independently. This level is designed for teachers and adult learners who require a comprehensive understanding of the data science process to effectively guide their students or apply data science techniques in their professional lives.

In this way, educational datasets for each school level are systematically structured according to the learners' developmental stage. Learners experience various data preprocessing processes at each stage and can gradually develop their data processing competency.

3.2. Research Methods

First, a comprehensive literature review is conducted on data science processes, methods of utilizing data in school education, and existing educational dataset standards, as aforementioned in the theoretical background. The insights gained from the literature review are used to develop draft criteria for educational datasets tailored to different school levels (K-1 to 4, K-5 to 6, K-7 to 9, K-10 to 12, and teachers/adults).

The size and complexity of educational datasets can vary depending on the intended learning objectives and the developmental stage of the students. For younger students (K-1 to 4), datasets are typically smaller in size and less complex, focusing on simple data structures and basic analysis tasks. As students' progress through the grade levels, the size and complexity of the datasets increase, allowing for more advanced analysis and problem-solving activities.

Next, an expert Delphi survey and focus group interviews are conducted with a sample of 22 elementary and secondary school teachers holding a master's degree or higher in artificial intelligence convergence education to assess the validity of the developed criteria [32]. The Delphi method was chosen for its ability to gather consensus from a panel of experts through multiple rounds of questionnaires, ensuring a comprehensive evaluation of the criteria. The collected survey data are analyzed using descriptive statistics and Content Validity Ratio (*CVR*) values to determine the overall validity of the criteria [33].

For this purpose, a questionnaire was developed to respond to the appropriateness of school level, appropriateness of stage-by-stage description, data provision format, suitability of data preprocessing process, and appropriateness of stage-by-stage characteristics of the dataset. The survey was composed to respond to the degree of validity on a 5-point Likert scale, and if the response was 'not suitable (3 points or less)', it was structured to describe the reason or direction for revision. The composition of the items in the developed questionnaire is as shown in Table 4. The Content Validity Ratio (*CVR*) was calculated to determine the overall validity of the criteria. The *CVR* is a statistical measure used to quantify the degree of agreement among experts on the relevance of specific items [33]. The *CVR* is a widely used method for quantifying the degree of agreement among experts, with values ranging from -1 to +1. A *CVR* of 0 indicates that 50% of the experts agree an item is essential, while values closer to +1 indicate higher levels of agreement. The scale for calculating *CVR* generally uses a 5-point scale, and the formula is as follows. *N* is the total number of panels that responded, and *Ne* is the number of cases that responded 'valid'. The minimum *CVR* value according to the number of expert panels of 22 is approximately 0.40.

$$CVR = (N_e - N/2)/(N/2) \quad (1)$$

Table 4. Characteristics and validity questions of educational dataset for K-12.

Characteristics	Question
Appropriateness of school level	Are the criteria for educational datasets by school level appropriate for the stages of each school level?
Appropriateness of stage-by-stage description	Are the stage-by-stage descriptions of the criteria for educational datasets by school level appropriate?
Data provision format	Is the data provision format according to the criteria for educational datasets by school level suitable for the developmental stage of each school level?
Suitability of data preprocessing process	Is the learner's data preprocessing process according to the criteria for educational datasets by school level suitable for the developmental stage and data preprocessing process of each school level?
Appropriateness of stage-by-stage characteristics of the dataset	Are the characteristics of each school level for educational datasets by school level appropriate?

Educational datasets can support a wide range of learning activities, such as data visualization, statistical analysis, machine learning, and data-driven problem solving. These activities are designed to help students develop critical thinking, computational thinking, and data literacy skills, which are essential for success in the 21st century workforce.

Finally, based on the quantitative and qualitative findings, the educational dataset criteria are refined and finalized, incorporating expert recommendations to enhance their validity and practical applicability in K-12 AI and data science education.

4. Research Results

4.1. Expert Validity Results

The expert validity results of the educational dataset are as follows. First, an expert validity survey was conducted on 22 elementary and secondary school teachers with a master's degree or higher in artificial intelligence integrated education. The results of the expert's validity review on the draft criteria for educational datasets for elementary and secondary artificial intelligence education are summarized in Table 5. As a result of the review, high validity was confirmed in all items, such as appropriateness of school level (*CVR* = 1.00), appropriateness of stage-by-stage description (*CVR* = 0.82), data provision format (*CVR* = 0.82), suitability of data preprocessing process (*CVR* = 0.82), and appropriateness of stage-by-stage characteristics of the dataset (*CVR* = 0.82). Respondents emphasized the importance of education that can judge the suitability of data when providing datasets or teaching data science, and suggested approaching it as a spectrum rather than stage-by-stage classification as students' levels vary. They suggested that it is good to provide scaffolding (support system) densely between stages, and that it is necessary to subdivide the curriculum by grade level in middle and high school courses.

The results of the expert review on the validity of the step-by-step description and characteristics of the educational dataset criteria for developed elementary and secondary artificial intelligence education are as follows. Stage 1 (grades 1–4, *CVR* = 0.82), Stage 2 (grades 5–6, *CVR* = 0.82), Stage 3 (middle school grades 1–3, *CVR* = 1.00), and Stage 4 (high school grades 1–3, *CVR* = 0.82) showed high validity, but Stage 5 (teachers (adults), *CVR* = 0.45) showed relatively low validity, as shown in Table 6.

Table 5. Expert validity results.

Characteristic	Question	M	SD	CVR
Appropriateness of school level	Are the criteria for educational datasets by school level appropriate for the stages of each school level?	4.50	0.50	1.00
Appropriateness of stage-by-stage description	Are the stage-by-stage descriptions of the criteria for educational datasets by school level appropriate?	4.50	0.50	0.82
Data provision format	Is the data provision format according to the criteria for educational datasets by school level suitable for the developmental stage of each school level?	4.45	0.65	0.82
Suitability of data preprocessing process	Is the learner's data preprocessing process according to the criteria for educational datasets by school level suitable for the developmental stage and data preprocessing process of each school level?	4.36	0.65	0.82
Appropriateness of stage-by-stage characteristics of the dataset	Are the characteristics of each school level for educational datasets by school level appropriate?	4.45	0.64	0.82

Table 6. Expert validity results.

Level (Appropriate Grade)	Description	Characteristics	M	SD	CVR
Level 1 (K-1 to 4)	Stage where data preprocessing is unnecessary	Unstructured data Data are completed by the instructor or a completed dataset is used	4.50	0.50	1.00
Level 2 (K-5 to 6)	Stage where only the cleaning process to remove unnecessary data is required	Dataset is provided by the instructor, and unnecessary data are removed according to the instructor's guidance as needed	4.50	0.50	0.82
Level 3 (K-7 to 9)	Stage where attributes (e.g., labels) are added or unnecessary data are removed	Dataset is provided by the instructor, and data are removed or necessary data such as sum/average/labels are added by the learner as needed	4.45	0.65	0.82
Level 4 (K-10 to 12)	Stage where necessary features are merged and subsequent cleaning processes must be performed	Data sources and collection methods, preprocessing methods are provided Outliers, missing values, etc., must be identified and cleaned, and necessary data such as sum/average/labels are added by the learner	4.36	0.65	0.82
Level 5 (Teacher (Adult))	Stage where learners must devise the entire process from data collection to preprocessing on their own	Dataset is not provided Data collection and preprocessing methods are not provided (including the task of identifying what features should be included in the data and finding where data with such features are located)	4.45	0.64	0.82

4.2. Modified Description and Characteristics of Educational Dataset for K-12

Based on the CVR results, the standards for Educational Datasets were revised, as shown in Table 7. For adults, three levels were defined and included from Level 3 to Level 5, as follows. Level 3 (Adults—Beginner) is for adults who are new to artificial intelligence education. The datasets at this level should cover basic concepts and terminology of AI, and provide simple hands-on exercises to gain practical experience. The content should be presented in an easy-to-understand manner suitable for beginners. Level 4 (Adults—Intermediate) is targeted towards adults who have some prior knowledge of AI and want to deepen their understanding. Level 5 (Adults—Advanced) is for adults who want to gain expertise in artificial intelligence. The datasets at this level should cover state-of-the-art techniques and research in AI, and provide opportunities to work on real-world projects.

Table 7. Modified description and characteristics of educational dataset for K-12.

Level (Appropriate Grade)	Description	Data Preprocessing Process Experienced by Learners	Characteristics
Level 1 (K-1 to 4)	Stage where data preprocessing is unnecessary		Unstructured data Data are completed by the instructor or a completed dataset is used
Level 2 (K-5 to 6)	Stage where only the cleaning process to remove unnecessary data is required	Data Cleaning	Dataset is provided by the instructor, and unnecessary data are removed according to the instructor's guidance by learner as needed
Level 3 (K-7 to 9, Adult 1)	Stage where attributes (e.g., labels) are added or unnecessary data are removed	Data Cleaning, Data Discretization, Data Transformation	Dataset is provided by the instructor, and data are removed or necessary data such as sum/average/labels is added by the learner as needed
Level 4 (K-10 to 12, Adult 2)	Stage where necessary features are merged and subsequent cleaning processes must be performed	Data Integration, Data Cleaning, Data Discretization, Data Transformation	Data sources and collection methods, preprocessing methods are provided Outliers, missing values, etc., must be identified and cleaned, and necessary data such as sum/average/labels are added by the learner
Level 5 (Adult 3)	Stage where learners must devise the entire process from data collection to preprocessing on their own	Entire Process	Dataset is not provided Data collection and preprocessing methods are not provided (including the task of identifying what features should be included in the data and finding where data with such features are located)

5. Discussion

The expert validity results demonstrated high validity for Stages 1 through 4, indicating that the educational datasets were well-suited to the developmental stages of K-12 students. However, Stage 5, which is targeted at teachers and adults, displayed significantly lower validity scores. This finding aligns with [34], who observed a similar trend where the data science education programs were highly valid for elementary, middle, and high school levels but showed decreased validity for adult education, particularly for teachers. These results suggest a gap in the effectiveness of current data science education programs for teachers, underlining the need for specifically tailored educational datasets and separate training programs that can effectively enhance data science competencies among educators. This tailored approach would ensure that teachers are not only consumers of data science education but are also proficient in applying data science concepts effectively in their teaching practices.

In Stage 1, the data preprocessing is intentionally omitted to align with the cognitive abilities of elementary school learners, who typically concentrate on discrete data values rather than on complex data manipulations. According to [35], at this early educational stage, students are better suited to engage with unstructured data that do not require preprocessing, which simplifies their interaction with the information. This approach allows learners to directly observe and interpret individual data points without the complexities of data cleaning or transformation. Such a method is beneficial, as it nurtures initial curiosity and foundational understanding of data, accommodating the students' developmental stage, where the focus is on tangible and immediate data observation, fostering a more intuitive grasp of basic data science concepts.

In Stage 2 of the educational dataset framework, learners are introduced to basic data cleaning tasks, which involves the removal of unnecessary data. This stage serves as an initial step into the more technical aspects of data science but remains accessible to younger students, such as those in elementary school grades 5 to 6. The datasets used are curated by the instructor to ensure relevance and manageability. This controlled introduction mirrors the design principles set forth by [36], where the use of binary values for variables and the limitation of independent variables simplify the learning process. This simplification helps to build a foundational understanding of how data quality directly influences analytical outcomes, preparing students for more complex tasks.

Stage 3 advances the learners' engagement with datasets by introducing them to the process of adding or modifying attributes and further refining the data through cleaning, discretization, and transformation. This stage is pivotal as it encourages middle school students to take a more active role in preparing their data for analysis. Ref. [37] highlights the importance of such data science tools that enable transformation and visualization, which are integral to this stage. By allowing students to manipulate data and see the results of their adjustments, they gain a deeper understanding of how data preparation impacts the exploratory and eventual analytical results, fostering a hands-on comprehension of data manipulation.

In Stage 4, the educational tasks become more complex and aligned with high school curricular goals, involving data integration and advanced processing techniques like cleaning, discretization, and transformation. Students at this level are expected to identify and correct outliers and missing values, thus ensuring the integrity and usability of the data. The authors of [38] emphasize the significance of data authoring, which is critical in this stage. By engaging students in these higher-level tasks, they not only learn the technical skills necessary for effective data management but also develop critical thinking skills as they make decisions on how to best prepare and represent data for specific analytical purposes.

Stage 5 of the educational dataset framework poses considerable challenges, as it necessitates that learners, specifically teachers and adults, independently manage the entire data science process, from data collection through to preprocessing. This stage received notably lower validity scores from experts, which may reflect a general deficiency in readiness or existing expertise among adult learners in mastering comprehensive data science practices. The evaluative feedback underscores a critical need for enhanced structured support and educational resources tailored to empower adults to efficiently design and execute data science workflows. The inadequacy highlighted by this stage accentuates the imperative for specialized training programs that are equipped to furnish educators with both the theoretical underpinnings and practical acumen necessary for effective dissemination of data science knowledge within educational settings.

This observation aligns with concerns previously noted by [8], regarding the observable shortfall in data science competency among educators. Survey respondents advocated for initiating educational interventions with novice teachers by providing them access to data and subsequent training. This approach underscores the necessity of a gradual, experiential learning process in data science for adults, as reflected in the feedback from the educational field. Additionally, experts highlighted the importance of curating datasets that are subject-specific yet versatile enough to be integrated across various disciplines. This strategy suggests that artificial intelligence education should transcend traditional subject boundaries, promoting an interdisciplinary methodology in teaching data science, as advocated by [39]. This approach not only enhances educational inclusivity but also broadens the applicability and relevance of data science in diverse academic contexts.

Based on the CVR results and expert feedback, the standards for Educational Datasets were revised, particularly for Stage 5, which targets teachers and adult learners. Three new levels were defined for adults and included. These revisions aim to address the gap in the effectiveness of current data science education programs for teachers and adult learners. By providing tailored educational datasets and separate training programs, the goal is to effectively enhance data science competencies among educators and ensure they are proficient in applying data science concepts in their teaching practices. The revised standards also emphasize the importance of curating datasets that are subject-specific yet versatile enough to be integrated across various disciplines. This promotes an interdisciplinary methodology in teaching data science, enhancing educational inclusivity and broadening the applicability and relevance of data science in diverse academic contexts.

The proposed educational dataset framework shares some similarities with existing research and guidelines in data science education, such as the emphasis on incorporating real-world data and hands-on experiences and the importance of introducing data science concepts in high school [25–27]. However, our framework extends beyond these principles to provide a comprehensive, developmentally appropriate approach tailored to the specific needs of learners at each grade level from K-12. While existing research highlights the importance of understanding data in K-12 data science education [28], our framework goes further by proposing a structured progression of data preprocessing skills and techniques across different developmental stages. By aligning the complexity of data preprocessing tasks with students' cognitive abilities at each stage, our framework ensures that students can gradually build their data literacy and computational thinking skills in a manner appropriate to their grade level.

6. Conclusions

6.1. Theoretical Contribution

This study makes a significant theoretical contribution to the field of educational datasets and materials by exploring the under-researched area of systematic categorization of difficulty levels in educational resources. Prior research has predominantly concentrated on elementary aspects such as data structuring and the use of unstructured data [13,14,40]. However, a comprehensive framework for classifying the complexity levels appropriate for different educational stages has been notably absent. This research addresses this gap by introducing a structured framework that categorizes educational datasets based on the developmental stages of learners and the required complexity of data preprocessing tasks. The framework promotes hands-on learning and active learning pedagogies to foster sustainable education. Furthermore, the study highlights the importance of effective bridge programs to support learners' transitions between different educational levels.

One of the key findings of this research is the critical role of expert validation in developing educational materials, particularly within science and data-based education sectors. Previous studies have not adequately explored the systematic incorporation of expert opinions in the development of educational resources, a gap that is evident in the literature concerning the use of technologies and data sources in educational settings [8]. This study enriches the existing body of knowledge by conducting an expert validity survey that assesses draft criteria for educational datasets in elementary and secondary AI education, thus underscoring the importance of expert involvement as suggested by [9].

Additionally, the validation results from this study show a high degree of acceptance for the proposed dataset criteria among elementary and secondary education levels. This aligns with Lee, Wilkerson, 2018, who stressed the necessity of age-appropriate data handling and educational scaffolding [38]. Conversely, the lower validation scores observed for datasets intended for adult educators highlight a critical need for tailored dataset designs and educational strategies to improve data science competencies among teachers, reinforcing the findings of [8], and echoing [36], who advocate for gradual, experiential learning in adult education.

Moreover, this study extends the theoretical discussions on the interdisciplinary potential of AI education. It highlights the significance of creating versatile datasets that can be adapted for various subjects and integrated into diverse curricular areas. The importance of a subject-transcendent approach to AI education has been supported by previous frameworks such as the GAISE II guidelines for precollege statistics and data science education and further discussed in [39]. This suggests a move towards an integrative curriculum model that embraces the multifaceted nature of data science, allowing for enhanced applicability across educational disciplines.

The implications of this study are profound for scholars seeking to further explore the integration of AI and data science into educational curricula. It invites further research into the effective implementation of interdisciplinary approaches and the development of educational resources that are accessible and relevant to both students and educators across all stages of learning. By continuing to focus on these areas, educational researchers can contribute to the development of more effective, inclusive, and comprehensive educational technologies and methodologies.

Moreover, the interdisciplinary nature of artificial intelligence education, as emphasized by the respondents in this study, aligns with the principles of Education for Sustainable Development (ESD). ESD aims to empower learners to make informed decisions and take responsible actions for environmental integrity, economic viability, and a just society [41]. By developing datasets that are applicable to various subjects and integrating them into the curriculum, this study promotes an interdisciplinary approach to AI education, which can foster the development of critical thinking, problem-solving, and decision-making skills necessary for sustainable development [42].

6.2. Practical Implications

The findings of this study have significant practical implications for various stakeholders in the field of artificial intelligence education, including teachers, students, educational institutions, and government.

For teachers, the study highlights the need for separate education and dataset design to enhance their data science competency. The relatively low validity score for the adult teacher stage in the expert review suggests that teachers require specialized training and resources to effectively teach artificial intelligence and data science concepts. Educational institutions and government should prioritize the development and implementation of professional development programs that provide teachers with the necessary knowledge and skills to integrate artificial intelligence education into their teaching practices. These programs should focus on step-by-step experiences with datasets and emphasize the practical application of data science principles in the classroom setting.

Students stand to benefit greatly from the findings of this study. The high validity scores for the stage-by-stage description and characteristics of the educational dataset criteria for elementary and secondary students indicate that well-designed datasets and materials can effectively support student learning in artificial intelligence education. Educational institutions should invest in the development of age-appropriate datasets and materials that align with the curriculum and provide students with hands-on experiences in data analysis, visualization, and interpretation. By engaging with these resources, students can develop critical thinking, problem-solving, and data literacy skills that are essential for success in the 21st century.

Educational institutions play a crucial role in the implementation of artificial intelligence education. The study emphasizes the importance of developing datasets that are applicable to various subjects and integrating them into the curriculum. Educational institutions should adopt an interdisciplinary approach to artificial intelligence education, ensuring that datasets and materials are not limited to a specific subject but rather integrated across different subjects. This approach can foster a more comprehensive understanding of artificial intelligence and its applications among students. Additionally, educational institutions should collaborate with industry partners and experts in the field to ensure that the datasets and materials used in the classroom are relevant, up-to-date, and aligned with real-world applications.

Government has a vital role in supporting the implementation of artificial intelligence education. The findings of this study underscore the need for government to allocate resources and funding towards the development of educational datasets and materials, as well as the training of teachers in data science and artificial intelligence. Government should also establish policies and guidelines that promote the integration of artificial intelligence education into the curriculum and ensure that all students have access to high-quality resources and instruction. By investing in artificial intelligence education, government can help prepare the next generation of students for the challenges and opportunities of the digital age.

6.3. Limitations and Future Research Directions

While this study provides valuable insights into the development of educational datasets and materials for artificial intelligence education, it is not without limitations. The sample size of the expert validity survey is relatively small, and future research should consider expanding the sample to include a larger and more diverse group of experts. Additionally, the study focuses primarily on the South Korean context, and further research is needed to examine the generalizability of the findings to other educational settings and cultural contexts.

Future research should also explore the effectiveness of the proposed dataset criteria in practice, by implementing and evaluating the use of datasets and materials developed based on these criteria in real classroom settings. This can provide valuable feedback and insights into the practical application of the findings and inform further refinements and improvements to the criteria.

Author Contributions: Conceptualization, I.-S.J. and S.-J.K.; methodology, S.-J.K.; validation, I.-S.J.; formal analysis, S.-Y.K.; investigation, I.-S.J.; resources, S.-Y.K.; data curation, I.-S.J.; writing—original draft preparation, I.-S.J.; writing—review and editing, S.-J.K.; visualization, I.-S.J.; supervision, S.-J.K.; project administration, S.-J.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data available on request due to privacy. The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Paek, S.; Kim, N. Analysis of worldwide research trends on the impact of artificial intelligence in education. *Sustainability* **2021**, *13*, 7941. [CrossRef]
2. Lee, B.H. International Conference on Artificial Intelligence and Education. In *Planning Education in the AI Era: Lead the Leap (16–18 May 2019)*; Opening Remarks HE Ambassador LEE Byong-Hyun Chairperson of the Executive Board; Beijing Hotel: Beijing, China, 2019.
3. Taghikhah, F.; Erfani, E.; Bakhshayeshi, I.; Tayari, S.; Karatopouzis, A.; Hanna, B. Artificial intelligence and sustainability: Solutions to social and environmental challenges. In *Artificial Intelligence and Data Science in Environmental Sensing*; Elsevier: Amsterdam, The Netherlands, 2022; pp. 93–108.
4. Touretzky, D.; Martin, F.; Seehorn, D.; Breazeal, C.; Posner, T. Special session: AI for K-12 guidelines initiative. In Proceedings of the 50th ACM Technical Symposium on Computer Science Education, Minneapolis, MN, USA, 27 February–2 March 2019; pp. 492–493.
5. Wu, F.; Lu, C.; Zhu, M.; Chen, H.; Zhu, J.; Yu, K.; Li, L.; Li, M.; Chen, Q.; Li, X. Towards a new generation of artificial intelligence in China. *Nat. Mach. Intell.* **2020**, *2*, 312–316. [CrossRef]
6. Lane, D. *Machine Learning for Kids: A Project-Based Introduction to Artificial Intelligence*; No Starch Press: San Francisco, CA, USA, 2021.
7. Ministry of Education. *A Better Future, Education for All 2022 Revised Curriculum Overview Main Points*; Ministry of Education: Sejong, Republic of Korea, 2021.
8. Rosenberg, J.M.; Schultheis, E.H.; Kjelvik, M.K.; Reedy, A.; Sultana, O. Big data, big changes? The technologies and sources of data used in science classrooms. *Br. J. Educ. Technol.* **2022**, *53*, 1179–1201. [CrossRef]
9. Touretzky, D.; Gardner-McCune, C.; Martin, F.; Seehorn, D. Envisioning AI for K-12: What should every child know about AI? In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; pp. 9795–9799.
10. Mariani, M.; Baggio, R. Big data and analytics in hospitality and tourism: A systematic literature review. *Int. J. Contemp. Hosp. Manag.* **2022**, *34*, 231–278. [CrossRef]
11. Dhar, V. Data science and prediction. *Commun. ACM* **2013**, *56*, 64–73. [CrossRef]
12. Organisation for Economic Co-operation and Development. The Future of Education and Skills: Education 2030. In *The Future We Want*; Organisation for Economic Co-operation and Development: Château de la Muette in Paris, France, 2018.
13. Qin, J.; D'Ignazio, J. *Lessons Learned from a Two-Year Experience in Science Data Literacy Education*; Purdue University: West Lafayette, IN, USA, 2010.
14. Kjelvik, M.K.; Schultheis, E.H. Getting Messy with Authentic Data: Exploring the Potential of Using Data from Scientific Research to Support Student Data Literacy. *CBE—Life Sci. Educ.* **2019**, *18*, es2. [CrossRef] [PubMed]

15. Kim, S.-Y.; Jeon, I.; Kang, S.-J. Integrating Data Science and Machine Learning to Chemistry Education: Predicting Classification and Boiling Point of Compounds. *J. Chem. Educ.* **2024**, *101*, 1771–1776. [CrossRef]
16. Nguyen, K.A.; Borrego, M.; Finelli, C.J.; DeMonbrun, M.; Crockett, C.; Tharayil, S.; Shekhar, P.; Waters, C.; Rosenberg, R. Instructor strategies to aid implementation of active learning: A systematic literature review. *Int. J. STEM Educ.* **2021**, *8*, 9. [CrossRef]
17. Conway, D. The Data Science Venn Diagram. Available online: http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram (accessed on 9 April 2024).
18. Emmert-Streib, F.; Moutari, S.; Dehmer, M. The Process of Analyzing Data is the Emergent Feature of Data Science. *Front. Genet.* **2016**, *7*, 12. [CrossRef] [PubMed]
19. Agarwal, S. Understanding the Data Science Lifecycle. Available online: https://www.sudeep.co/data-science/2018/02/09/Understanding-the-Data-Science-Lifecycle.html (accessed on 24 February 2024).
20. Sidhu, R. Life Cycle of a Data Science Project. Available online: https://towardsdatascience.com/life-cycle-of-a-data-science-project-3962b9670e5b (accessed on 13 February 2024).
21. Dineva, K.; Atanasova, T. OSEMN process for working over data acquired by IoT devices mounted in beehives. *Curr. Trends Nat. Sci.* **2018**, *7*, 47–53.
22. Ericson, G.; Rohm, W.A.; Martens, J.; Sharkey, K.; Casey, C.; Harvey, B.; Schonning, N. Team Data Science Process Documentation. Available online: https://learn.microsoft.com/en-us/azure/architecture/data-science-process/overview (accessed on 24 February 2024).
23. Zhang, S.; Zhang, C.; Yang, Q. Data preparation for data mining. *Appl. Artif. Intell.* **2003**, *17*, 375–381. [CrossRef]
24. Fan, C.; Chen, M.; Wang, X.; Wang, J.; Huang, B. A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data. *Front. Energy Res.* **2021**, *9*, 652801. [CrossRef]
25. Kim, S.; Jeon, Y.; Kim, T. A Study on Data Preprocessing Content Knowledges According to School Level for Artificial Intelligence Education. *J. Korean Assoc. Comput. Educ.* **2021**, *24*, 1–12.
26. De Veaux, R.D.; Agarwal, M.; Averett, M.; Baumer, B.S.; Bray, A.; Bressoud, T.C.; Bryant, L.; Cheng, L.Z.; Francis, A.; Gould, R. Curriculum guidelines for undergraduate programs in data science. *Annu. Rev. Stat. Its Appl.* **2017**, *4*, 15–30. [CrossRef]
27. National Council of Teachers of Mathematics. NCTM Releases New Position Statement on Teaching Data Science in High School. Available online: https://www.nctm.org/News-and-Calendar/News/NCTM-News-Releases/NCTM-Releases-New-Position-Statement-on-Teaching-Data-Science-in-High-School/ (accessed on 13 March 2024).
28. Israel-Fishelson, R.; Moon, P.; Tabak, R.; Weintrop, D. Understanding the data in K-12 data science. *Harv. Data Sci. Rev.* **2024**, *6*, 2. [CrossRef]
29. Wickman, P.-O.; Hamza, K.; Lundegård, I. Didactics and didactic models in science education. In *Methodological Approaches to STEM Education Research: Volume 1*; White, P.J., Tytler, R., Ferguson, J., Cripps Clarke, J., Eds.; Cambridge Scholars Publishing: Newcastle upon Tyne, UK, 2020; pp. 34–49.
30. Zawacki-Richter, O.; Marín, V.I.; Bond, M.; Gouverneur, F. Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *Int. J. Educ. Technol. High. Educ.* **2019**, *16*, 39. [CrossRef]
31. Banchi, H.; Bell, R. The many levels of inquiry. *Sci. Child.* **2008**, *46*, 26.
32. Green, R.A. The Delphi technique in educational research. *Sage Open* **2014**, *4*, 2158244014529773. [CrossRef]
33. Lawshe, C.H. A quantitative approach to content validity. *Pers. Psychol.* **1975**, *28*, 563–575. [CrossRef]
34. Lee, V.R.; Wilkerson, M.H.; Lanouette, K. A call for a humanistic stance toward K–12 data science education. *Educ. Res.* **2021**, *50*, 664–672. [CrossRef]
35. Konold, C.; Higgins, T.; Russell, S.J.; Khalil, K. Data seen through different lenses. *Educ. Stud. Math.* **2015**, *88*, 305–325. [CrossRef]
36. Srikant, S.; Aggarwal, V. Introducing data science to school kids. In Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education, Seattle, WA, USA, 8–11 March 2017; pp. 561–566.
37. Soto, C.; Gutiérrez de Blume, A.P.; Jacovina, M.; McNamara, D.; Benson, N.; Riffo, B. Reading comprehension and metacognition: The importance of inferential skills. *Cogent Educ.* **2019**, *6*, 1565067. [CrossRef]
38. Lee, V.R.; Wilkerson, M.H. *Data Use by Middle and Secondary Students in the Digital Age: A Status Report and Future Prospects*; National Academy of Sciences Engineering, and Medicine, Board on Science Education, Committee on Science Investigations and Engineering Design for Grades 6-12: Washington, DC, USA, 2018.
39. Franklin, C.; Bargagliotti, A. Introducing GAISE II: A guideline for precollege statistics and data science education. *Harv. Data Sci. Rev.* **2020**, *2*, 1–9.
40. Ramírez-Montoya, M.S.; Andrade-Vargas, L.; Rivera-Rogel, D.; Portuguez-Castro, M. Trends for the future of education programs for professional development. *Sustainability* **2021**, *13*, 7244. [CrossRef]
41. Bucea-Manea-Țoniş, R.; Kuleto, V.; Gudei, S.C.D.; Lianu, C.; Lianu, C.; Ilić, M.P.; Păun, D. Artificial intelligence potential in higher education institutions enhanced learning environment in Romania and Serbia. *Sustainability* **2022**, *14*, 5842. [CrossRef]
42. Chiu, T.K.; Chai, C.-S. Sustainable curriculum planning for artificial intelligence education: A self-determination theory perspective. *Sustainability* **2020**, *12*, 5568. [CrossRef]