Work Package 2: Diagnosis

D2.1 DIAGNOSIS REPORT ON DATA LITERACY

Work Package 2:

Diagnosis

Deliverable D2.1

Diagnosis report on data literacy

Co-funded by the European Union

Project Title:	Data literacy in the age of AI for education
Project Acronym:	**AI- DL**
Project Number	101195923
Grant Agreement Number	101195923 — AI-DL, ERASMUS-EDU-2024-POL-EXP-DIGITAL
Instrument:	Erasmus+
Deliverable number	D2.1
Work Package	2 - Diagnosis
Work Package Leader	H2 Learning
Work Package Partners	H2 Learning, Nantes University, SCRIPT, INDIRE, CNR, Maribor University, FEI
Dissemination level	Public
Delivery date	2026-01
Status	Final
Version	1.8
Author(s)	Michael Hallissy, Maria Fojk, Harvey Mellar, Jotsna Iyer, Manuel Gentile, Jessica Niewint Gori
List of contributors	Deirdre Butler, Colin de la Higuera, Hoel Le Capitaine, Gerson Sunyé, Sara Mori, Francesca Storai, Salvatore Perna, Giuseppe Città, Mario Allegra, Alenka Lipovec, Barbara Arcet, Kosta Dolenc, Andrej Flogie, Tadej Todorović, Claude Reuter, Lina Rivera, Evelyne Huré, Sophie Gougeon

Disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

ABSTRACT	This diagnosis report presents comprehensive desk research examining data literacy and GenAI adoption in secondary education across seven European countries. Through systematic literature review (59 publications), framework analysis (DigComp, UNESCO, CDC, AILit), European project review (100 Erasmus+ initiatives), and national policy surveys in seven countries, the research reveals that data literacy is a fluid, context-dependent concept and there is minimal connection to GenAI usage in existing frameworks. While recent frameworks increasingly address AI competencies, they focus primarily on "input-side" data practices rather than GenAI-specific skills. The study identifies critical gaps: no framework offers integrated data literacy for GenAI contexts; teachers lack training in pedagogical approaches for developing critical engagement with AI-generated outputs; and approaches to implementing data literacy vary widely across countries. Key findings emphasize the need for flexible, co-constructed frameworks that balance technical competencies with critical thinking, ethical awareness, and democratic participation. The research confirms the need for the AI-DL project and provides clear direction for developing teacher training programmes and an emerging framework that addresses data literacy specifically in the context of GenAI tool usage in upper secondary education.
KEYWORDS	Data literacy, generative AI, teacher training, digital competence frameworks, secondary education

Dissemination level		
PU	**Public**	**X**
PP	Restricted to project partner (including the Commission)	
RE	Restricted to a group defined by the consortium (including the Commission)	
CO	Confidential, only for members of the consortium (including the Commission)	

Executive Summary

Background and Purpose

The AI-DL project addresses the growing need for data literacy in education, particularly in the context of widespread GenAI tool usage in upper secondary schools. This diagnosis report on data literacy (Deliverable 2.1) presents comprehensive desk research examining current practices, frameworks, and policies related to data literacy and GenAI adoption across the seven partner countries (France, Ireland, Italy, Lithuania, Luxembourg, Slovenia, and Spain).

The research was conducted through four parallel work streams: (1) a systematic literature review of academic research; (2) an analysis of existing digital competence frameworks; (3) a review of European Erasmus+ projects; and (4) a survey of national policies and curricula in partner countries.

Key Findings

Literature Review: A literature review of recent peer reviewed publications identified 59 publications related (at school level) to definitions and components of data literacy education and the key strategies, pedagogical principles and tools used for data literacy education. Analysis of these publications showed that data literacy is a fluid, multidimensional concept with no universally accepted definition. The research showed minimal connection between existing data literacy frameworks and GenAI usage, indicating a significant gap that AI-DL would like to address.

The review identified three categories of data literacy components:

- **Disciplinary data literacy components** include understanding data basics, designing investigations, evaluating data quality and sources, creating datasets, sense-making that considers bias and social factors, processing and analysing data, making data-based claims, and publishing analyses ethically.
- **Personal data literacy components**, drawing on Pangrazio and Selwyn's (2023) framework "Critical Data Literacies: Rethinking Data and Everyday Life," which encompasses five key dimensions: *Data identification*: *Data understandings*: *Data reflexivity*: *Data strategies*: *Data tactics.* These critical data literacies framework shifts the emphasis from purely technical skills toward understanding data as a social, political, and cultural phenomenon that shapes power relationships.
- **Teacher-specific components** address handling student data ethically, using data for iterative decision making, and understanding recommended pedagogical approaches for teaching data literacy.

Other key findings were:

- Data literacy definitions are highly contextual and evolve with technological change
- Traditional frameworks focus on "input-side" data practices (collection, management, evaluation) rather than GenAI-specific competencies
- Critical thinking, metacognition, and collaboration are essential aspects of data literacy
- Teachers require pedagogical approaches that encourage curiosity and creativity with emerging technologies

Framework Analysis: Review of major frameworks (DigComp 2.2/3.0, UNESCO AI frameworks, CDC, AILit) for Data Literacy and GenAI coverage demonstrated uneven coverage across six competence domains. For Data Literacy it found that coverage was uneven, reflecting each framework's original purpose and development context. While GenAI coverage, across the six domains, was comparatively broad and more uniform, reflecting the fact that contemporary GenAI-oriented frameworks were developed with an explicit intention to integrate technical understanding with responsible use and classroom-relevant practices.

The analysis revealed the following:

- Core knowledge of AI and data is well-covered in modern frameworks but absent from pre-2022 documents, specifically in the CDC framework.
- “Data Stewardship” and “Model Documentation” remain comparatively less visible. The analysed frameworks prioritise user-facing competences, such as critical evaluation of AI outputs and creative or pedagogical use, over less visible, but consequential, competences related to data preparation, provenance, documentation, and transparency.
- While Data Literacy stays a foundational competence, GenAI Literacy extends it into interaction, creativity, and co-creation with AI systems.
- Data literacy frameworks (e.g., DigComp, or UNESCO) offer partial reference points for proficiency and progression, but the literature does not yet provide a single, formalised definition of “data literacy for generative AI”.
- Data literacy and GenAI literacy are both grounded in critical thinking and ethical use, and both combine technical abilities with social and communicative competences. In practice, this means learners are expected to question sources and outputs, reason about risks and bias, and act responsibly toward others when they use data or AI tools. But the two literacies diverge in scope and emphasis.

European Projects: Analysis of 100 Erasmus+ projects showed growing investment in data literacy education, with most focusing on collaborative institutional partnerships (KA2). Notable projects include DALI (adult data literacy), DALI4US (primary education), TRAIN-DL (teacher training integrating AI and data literacy), and MILES (media literacy and disinformation). These projects provide valuable frameworks and methodologies but rarely address GenAI specifically.

Co-funded by the European Union

National Policies: Survey results showed that five of seven partner countries have data literacy policies, all have curricula, but implementation approaches vary widely - from compulsory named courses (Luxembourg) to cross-curricular integration (Ireland, Italy, Slovenia). GenAI policies exist in only four countries, with usage primarily experimental. This variability presents both challenges and opportunities for the AI-DL project.

Implications for AI-DL

The research confirms the project's core hypothesis: there is an urgent need for practical tools as frameworks and training that specifically address data literacy in the context of GenAI usage. Key implications include:

1. **Framework Development:** AI-DL should not attempt to create rigid definitions but rather develop flexible, co-constructed approaches that adapt to local contexts and evolving technologies.
2. **Teacher Training:** Training should balance technical understanding with critical, ethical, and pedagogical competencies. Teachers need support in developing students' ability to evaluate AI outputs, understand data provenance, and engage in informed decision-making.
3. **Pedagogical Approaches:** The project should emphasize active learning, collaboration, real-world problem-solving, and creative engagement with both data and GenAI tools.
4. **Cross-Curricular Integration:** Given the diverse implementation approaches across countries, AI-DL resources should be flexible enough to support both subject-specific and cross-curricular applications.
5. **Ethical and Democratic Competences:** Data literacy in the GenAI context must explicitly address citizenship, privacy, bias, and the social impacts of AI-driven decision-making.

The desk research provides a foundation for WP3 - Training design and content (teacher training development) and the creation of an emerging AI-DL Framework (D2.3) that will be refined throughout the project lifecycle through co-construction with teachers, students, and school leaders.

Co-funded by the European Union

Table of Contents

Co-funded by the European Union

List of Figures

List of Tables

1. Introduction

1.1 Background and Methodology

What the Proposal Stated

There are number of deliverables contained in Work Package 2 (WP2 - Diagnosis) in order to build a landscape on the current discussion about Data Literacy in the context of GenAI. Within that framework, the first deliverable, 2.1, sets out to conduct comprehensive desk research to understand the current state of data literacies and the adoption of GenAI technologies in education in the context of enhancing democratic competences in our young people.

This research was undertaken to establish current thinking and practices in relation to data literacy in a context where schools are increasingly using GenAI tools, particularly in upper secondary education. The project proposal indicated that the use of these GenAI tools could be the context for developing the data literacies of teachers and students. The purpose of undertaking the desk research was to firstly review current definitions and practices in relation to data literacy in schools, and to see if academics and/or schools are already using the arrival of GenAI tools as an opportunity to develop data literacy competencies. In the proposal we did not have such information, so the first task was to check our hypotheses and to do this we reviewed a range of sources that included the following:

- Academic research
- Existing digital frameworks
- Erasmus+ projects
- Current data literacy and GenAI policies and curricula in each of the partner countries

The proposal stated the purpose of conducting this work was twofold:

- To determine the key competences required for critically using data when employing GenAI tools in educational settings
- To produce a framework to guide discussions on current GenAI practices in participating schools

The project proposal identified the need to create a framework that would support all citizens, but particularly teachers and students in upper secondary education, to engage critically with data when interacting with GenAI tools (see D2. 3 – Framework for teachers and schools). As citizens they need to understand how these technologies operate and how they utilise and generate data. The project views the use of these tools as providing opportunities for teachers and students to further develop their data literacy competencies, so they can make informed decisions around the use of these tools and on the data they generate.

1.2 Objectives of the desk research (Task 2.1)

At the first project meeting in Dublin in May 2025, partners reviewed the proposal in relation to WP2 and considered how best to carry out the desk research task for the deliverable 2.1. This work package was led by H2 Learning with support from INDIRE and the CNR (Italy), Maribor University (Slovenia), Nantes University (France), and SCRIPT (Luxembourg).

Prior to undertaking the desk research tasks, the project partners engaged in discussion around what we hoped to achieve from engaging in this task and we developed the following guiding principles:

The Purpose of the Desk Review:

From the outset our discussions focused on two over-arching issues, the development of data literacy competences and GenAI usage in schools (see Figure 1 below). Regularly the discussions veered more towards GenAI usage and to issues of AI literacy, rather than to focusing on data and data literacy competences in the context of using GenAI. The partners spent quite an amount of time refining and articulating the focus of the project, which is to develop data literacy competences in a very specific context, that of using GenAI tools in schools. This clarification is important in attempting to create manageable boundaries for the project and in managing project outcome expectations. Nevertheless, these two issues, continued to surface regularly during all aspects of WP2, as they are integrally integrated in this project.

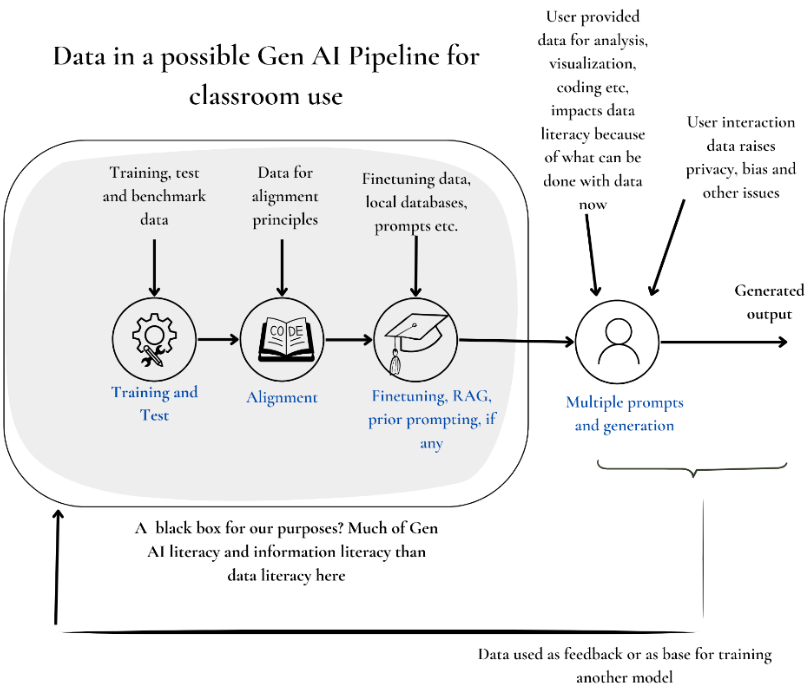


Figure 1 Data in a possible Gen AI Pipeline for classroom use

The proposal described Task 2.1 as conducting "comprehensive desk research to understand the current state of data literacies and the adoption of GenAI technologies in education in the context of enhancing democratic competences in our young people" (T2.1, p.93). Partners made the following decisions on how best to achieve this task.

They decided that:

- The desk research will primarily focus on data literacy in the context of GenAI use.
- GenAI usage in schools, by teachers and students, is the context within which data literacy competences would be developed.
- The project is not, primarily, focused on developing teacher and student AI or GenAI literacy.
- The focus is on developing teacher and student data literacy in the context of using GenAI to enhance their democratic competences.

The above parameters have informed the work of the project to date and specifically the work of WP2.

What is the problem we are trying to solve

In today's world citizens, that includes teachers and students, are using digital tools, such as GenAI tools regularly in their lives, and these tools are using increasing amounts of data. This project wants to better support citizens to understand how these tools use data and to this end we want to develop their data literacy competences. Thus, there are three issues at play in the project: data literacy, GenAI and digital citizenship and democracy (Figure 2), while our focus is data literacy.

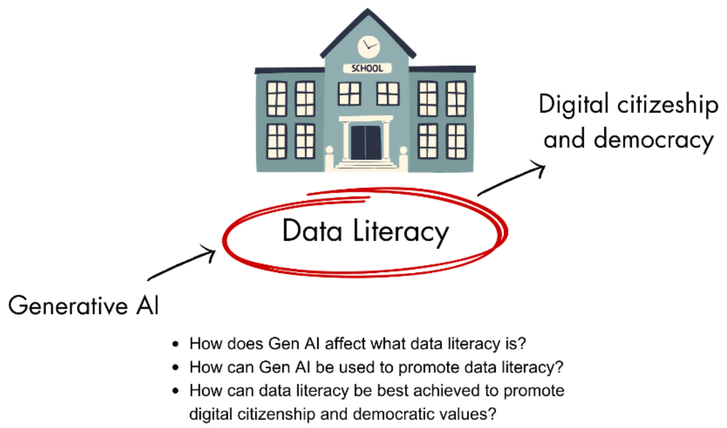


Figure 2 Developing Citizens Data Literacy when using GenAI tools

To help citizens become more data literate when using these tools, we need to find out what is typically meant by data literacy in educational contexts and to see how it has been defined and activated in other educational contexts (in European projects and frameworks), to learn if we can reuse these definitions and approaches or to see if we need to develop some new definitions and approaches for our specific context.

In AI-DL, we will be working with teachers and students in upper secondary school education, and these teachers will come from a range of curricular areas, to consider the use of data more critically. We are describing this ability to use data more critically as data literacy. We recognise that data literacy is a complex area and, in this project, we are keen to support teachers and students interact with data from both a technical and social perspective, by enabling them to engage in practical tasks by interacting with datasets, while also enabling them to discuss and debate the impact of data in the context of GenAI use, in an informed way. We anticipate that the project will develop a range of different use cases, which will be specific to country and subject area contexts, once we begin working in schools.

Thus, before embarking on the next phase of the project, which is to develop a training programme for teachers, we need to find out what is already in place in the participating countries in relation to data literacy curricula, while also reviewing academic research, a selection of existing frameworks and current and recent Erasmus+ projects in the field of data literacy. The findings from these separate reviews would inform how the project approaches the design and activation of the teacher training work package and to the creation of a meaningful data literacy framework.

Review of Existing Definitions and Approaches towards developing data literacy

The proposal stated that the project would conduct a thorough examination of current definitions and methodologies related to data literacy, particularly in the context of GenAI. It also stated that the team would review emerging research on the reconceptualization of human thought in the context of GenAI, by considering how we can reduce its risks to society, and by considering how it offers us a picture of our society and of human thinking. The proposal stated that the project would take a wide lens approach to this issue and not just focus on technical issues. However, after discussion among all the partners a decision was made to lessen the focus on the evolution of GenAI, and to focus more on the evolution of data literacy definitions and approaches in the context of a growing use of GenAI. This decision was taken for several reasons, which included the following:

- The call was focused primarily on data literacy and not on AI literacy and thus the review would focus primarily on data literacy definitions and practices
- That issues such as digital citizenship and democratic competences would be more in the background, rather than the primary focus of the desk research
- Pragmatic reasons related to scope of work, project deadlines and human resources informed this decision

Review of existing (and upcoming frameworks)

The proposal specifically stated that the project would review existing and upcoming frameworks, such as DigComp, UNESCO's AI Framework, Framework of Competences for Democratic Culture

(RFCDC) and consider how we can amalgamate key competencies into a new data literacy framework. This task was completed as outlined in the proposal.

Review of existing European (i.e. Erasmus+ projects) in the field of data literacy

The partners added a new sub task, that of reviewing existing and previous Erasmus+ projects that have, or are, focusing on data literacy. The project team felt the addition of this task would help the project learn from the work of others and inform the creation of the data literacy framework.

Review of Data Literacy Curricula in Partner Countries

The proposal stated that the project would conduct a review of existing data literacy curricula in partner countries to identify best practices and areas for improvement. This action was primarily conducted to establish what policies and implementation practices were in place to develop student data literacy competences in the participating countries, while also capturing the ministries high-level approach to GenAI usage.

Ultimately these combined sub-tasks or parallel work streams, as captured in Figure 3, were designed to identify a range of competences for critical use of data in the context of GenAI and to produce an "emerging" framework that would continue to be developed over the lifetime of the project.

Figure 3 Overview of the desk review activities in WP2

Table 1 below captures which institutions led the creation of each of the four parallel work streams, and where they received support from another partner.

Task	Lead Partner
Literature Review	Nantes University with support from University of Maribor
Literacy Frameworks Review	INDIRE

EU Projects Review	CNR
Policy and Curriculum review	H2 Learning with support from SCRIPT

Table 1 Assignment of desk review sub-tasks

The next section of the report introduces the literature review report task, highlighting a number of key takeaways and identifying elements of the report that are relevant to the AI-DL project.

2. Literature Review

2.1 Task Overview and Approach

The WP2 team spent considerable time considering how best to conduct a review of the relevant academic research, so that it focused on the topic of data literacies in the context of secondary education. The proposal stated that GenAI technologies have the potential:

- to change how we think of data literacy and
- its role as a tool for activities and resources that facilitate data literacy in schools.

The literature review was led by Nantes University with support from H2 Learning and the University of Maribor. The main author was Jotsna Iyer from Nantes University who led the design of the literature review task, and the write-up of this section of the overall report. The review team decided that in order to answer the above questions we needed to firstly understand what data literacy is and how the term has evolved, and if possible then consider how GenAI tools are impacting on data literacy.

It was decided to use a Systematic Literature Review (SLR) approach. Such an approach has been defined as" a research methodology to collect, identify, and critically analyse the available research studies (e.g., articles, conference proceedings, books, dissertations) through a systematic procedure (Pati & Lorusso, 2018).

The SLR attempted to answer the following two questions:

1) What are the definitions and components of data literacy education at school level and how are they evolving?
2) What are the key strategies, pedagogical principles and tools used for data literacy education at school level?

The aim of this review was to identify and understand all potentially relevant ideas that could inform the creation of a pertinent and up to date data literacy framework and inform the creation of a training programme for teachers that will address their needs and those of students.

2.2 Methodology

Prior to designing the research approach the team undertook an exploratory study by reading books and papers on data literacy and data literacy education. Some of the material was found using online searches on different platforms and search engines, while others were recommended by partners. These materials were subsequently discussed during several workshops to identify relevant ideas and approaches.

Co-funded by the European Union

Arising out of these discussions various search terms and filters were tried on Scopus[1], and the results were checked for breadth and depth of the topics covered. The goal was to generate a list of peer reviewed material that focused on schoolteachers and students which would help us get a good overview of what has been happening in the last ten years in data literacy education. The review did not attempt to cover all the material that is published in this area, but it endeavoured to make the most of what was found in the search. To this end the team collaborated to identify the most appropriate search string that would find the relevant open access peer reviewed research articles.

Inclusion and exclusion information

The search string included all English language articles, conference papers, reviews, book chapters, editorial and notes that:

- mention data literacy, data and information literacy, data and AI literacy, generative AI literacy, gen AI literacy, in the age of AI, in the era of AI, in the age of generative AI or in the era of generative AI in the title, abstract or keywords,
- mention school, teacher or student or K-12 in the title, abstract or keywords but **not** university, higher education, undergraduate, graduate, postgraduate, medical, maritime, engineering or librarian,
- do **not** belong to the subject areas of medicine, engineering, decision sciences, psychology, biochemistry, genetics and molecular biology, business, management and accounting, nursing, neuroscience, pharmacology, toxicology and pharmaceutics, health professions, dentistry, veterinary, economics, econometrics and finance, energy or agricultural and biological Sciences,
- were published in the ten-year period after 2015 and before 2026,
- are open access and
- available on Scopus

The search string returned a list of 113 papers that fitted with our search criteria and their details and abstracts were exported from Scopus as a csv file for further analysis.

The search query:

TITLE-ABS-KEY ("data literacy" OR "Data and information literacy" OR "Data and AI literacy" OR "Generative AI literacy" OR "Gen AI literacy" OR "in the age of AI" OR "in the era of AI" OR "in the age of generative AI" OR "in the era of generative AI") AND TITLE-ABS-KEY (school OR teacher OR student OR k-12 AND NOT university AND NOT "higher education" AND NOT undergraduate AND NOT graduate AND NOT "post graduate" AND NOT postgraduate AND NOT medical AND NOT maritime AND NOT engineering AND NOT librarian) AND PUBYEAR > 2015 AND PUBYEAR < 2026 AND (LIMIT-

[1] Scopus is a scientific abstract and citation database, launched by the academic publisher Elsevier in 2004. https://www.elsevier.com/products/scopus.

Co-funded by the European Union

TO (OA, "all")) AND (EXCLUDE (SUBJAREA,"MEDI") OR EXCLUDE (SUBJAREA,"ENGI") OR EXCLUDE (SUBJAREA,"DECI") OR EXCLUDE (SUBJAREA,"PSYC") OR EXCLUDE (SUBJAREA,"BIOC") OR EXCLUDE (SUBJAREA,"BUSI") OR EXCLUDE (SUBJAREA,"NURS") OR EXCLUDE (SUBJAREA,"NEUR") OR EXCLUDE (SUBJAREA,"PHAR") OR EXCLUDE (SUBJAREA,"HEAL") OR EXCLUDE (SUBJAREA,"DENT") OR EXCLUDE (SUBJAREA,"CENG") OR EXCLUDE (SUBJAREA,"VETE") OR EXCLUDE (SUBJAREA,"ECON") OR EXCLUDE (SUBJAREA,"ENER") OR EXCLUDE (SUBJAREA,"AGRI")) AND (LIMIT-TO (LANGUAGE, "English"))

The title and abstracts of all the selected papers were read and those that met the following criteria were excluded:

- not targeted at school students, pre-service or in-service school teachers
- did not address data literacy and
- not open access.

This generated a final list of 66 papers which was then shared amongst the partners for coding. A Zotero group library[2] was created to facilitate collaboration. The coding process further shortened the list of relevant papers to 59, with 6 papers found to be irrelevant after they were read, while one was excluded as it was not open access.

Coding process

Codes were developed by the review team to help the lead author summarise key aspects of data literacy education, without increasing the cognitive load of partners who agreed to help review the publications.

The coding scheme had 7 tags:

- ***Literacy***: Relevant information on the “literacy” part of DL, including why frame the subject as a literacy or literacies, what are the implications for calling it so, what does it highlight, what are the confusions and pitfalls etc.
- ***Data***: What is data relevant to students and teachers, what is or could be data in general, how we look at data, how we should look at data, its collection and use, how what can be done with data has evolved with available technologies and infrastructures.
- ***Data Literacy (DL):*** Definition and scope of data literacy (ies), DL competencies (KSA), AI literacy, Gen AI and other literacy components that may be relevant to data literacies, knowledge or competencies a teacher or student should have when dealing with data and data-based systems in their personal and professional lives.
- ***Principle***: Useful principles, pedagogical approaches and best practices used and recommended by data literacy educators.

[2] Zotero enables users to create shared group libraries to collaboratively manage research sources and materials, both online and through the Zotero client. https://www.zotero.org/groups/

Co-funded by the European Union

- ***Activity***: Interesting activities or curricular interventions which could inform future interventions.
- ***Problem***: Problems encountered while trying to formulate or implement a successful data literacy education programme. These could be problems in definition, possible tensions and contradictions, difficulties in implementing professional learning on data literacies etc.
- ***Evolution***: How have data literacy definitions and training approaches evolved with time and availability of technology and infrastructure and change in socio-political and economic factors.
- ***Other***: Any other information that the reviewer finds relevant, but which does not fall into any of the above categories.

It was found that a section of a paper could be and often was, highlighted with multiple tags, with additional comments where clarity was required.

Coding Challenges

On reviewing the coded documents, it became clear that each partner had different understandings of which paper was relevant to the study. Some reviewers stuck to the predefined criteria and excluded papers which did not strictly deal with data literacy. Examples are publications which mentioned data literacy, but whose main theme is learning analytics or formative assessment. Others took a broader approach looking for parallels in implementations of statistical literacy and critical digital literacies.

After the coding process was over and the different parts were put together, various themes and sub themes emerged, merged and re-emerged in new forms. Some themes reflected the research questions posed and the tags used while others emerged naturally from the material in the review.

The activity tag was discontinued, since there were too many external references and resources to follow. It was also found that the data literacy tag had too many components and it was found to be prudent to refine the classification scheme as follows:

1) ***Data literacy components*** that apply while learning or teaching a subject or an interdisciplinary topic.
2) ***Personal data literacy components***, as defined and broken down into Data identification, understandings, reflexivity, strategies and tactics in the book “Critical Data Literacies: Rethinking Data and Everyday Life” (Pangrazio & Selwyn, 2023).
3) ***Additional competencies*** that a teacher would need when using student data for pedagogical purposes or while educating them on data literacies.
4) ***Key attitudes*** that are recommended in the literature for a data literate student and teacher.

The categories listed above were useful in grouping and analysing recommended competencies, but they are neither exhaustive nor mutually exclusive. There is not only some overlap in the classification,

Co-funded by the European Union

but competencies acquired in one area can feed into or fortify those identified in other sections. For example, some of the elements in, say, personal data literacy components can help a student understand better the data in a topic related to a subject and vice versa. Another example is personal data that is used for a science project.

All themes and sub themes were first summarised and then analysed to understand how GenAI technologies might already be affecting them, or likely to do so in the future.

2.3 Key findings

Jotsna Iyer, Nantes University, wrote this report, which is reproduced in its entirety here, with additional commentary from H2 Learning, where we highlighted Key Takeaways from her research that have implications for AIDL. In addition, annex 7A, contains all the citations for the documents cited in this section and other publications that were reviewed by the main author.

The changing literacy landscape

Since its conception, literacy has undergone definitional shifts that reflected the prevailing political and cultural ideas of a given period (Chaka, 2019). As technology is also affected by the same currents, they have both evolved together, affecting each other in multiple ways.

During the mid-20th century, literacy was understood to be the ability to read, write, speak, and listen (Chaka, 2019; Lestari & Rosana, 2020). It revolved around one language and its text. Pre-digital age technologies like TV, radio and newspapers had one main function and broadcast knowledge in a one-dimensional way, the literate meant to receive this knowledge passively.

By the early part of the twenty-first century, the discourse came to be focused on skills and on doing as the preferred way of acquiring them. Literacy became a question of actively acquiring a set of measurable skills (reading, writing, speaking and listening skills, and information literacy skills) with emphasis on phonemic awareness, comprehension and fluency. It was the mid-digital age, and digital technology had not only become a part of everyday lives, but it also put in the hands of its users certain affordances when it came to enacting or doing literacies (Chaka, 2019).

In the third digital age, technology became more invisible and started to mediate many facets of life and learning. It brought forth concerns related to privacy, equity and distribution of wealth and power. Critical literacies that went beyond reading, writing, listening and speaking critically became more important. They sought to decode and manipulate text in their varied forms and to understand the power relationships, the social and cultural conditions behind the texts, discourses and genres. The idea was to bring to surface the hidden dynamics, to challenge them and transform them (Chaka, 2019, Ilomäki et al., 2023).

With the availability of super-fast internet, connected devices, instant messaging and livestreaming, the lines between mobile and desktop, online and offline, physical and virtual, instant messaging and writing started relentlessly merging and unmerging. Digital practices and processes got increasingly diverse and complex. Earlier conceptions of literacy were too rigid, conventional reading and writing

Co-funded by the European Union

no longer adequate and new text forms had to be accommodated (Ilomäki et al., 2023). Thus, came about, among others, terms such as new literacies, multiliteracies, critical literacies, digital literacies and data literacies (Chaka, 2019, Ilomäki et al., 2023).

New literacies and multiliteracies recognise literacy as a social practice that changes across time and space, is subjective and contextual, and is expressed using multimodal forms of text where text in its basic form mingles with images, icons, sound, animation and video. Literacies were now viewed as being more collaborative, participatory, distributed, deictic, multilingual and multicultural, going beyond language itself to embrace other emergencies. In these new contexts meaning making is now acknowledged to happen in multiple ways, and with mobile technologies and the internet, literacies occur anywhere and at any time (Chaka, 2019, Ilomäki et al., 2023).

Technologies act as a medium for enacting these literacies, multiple literacies at a time, and change how users experience and practice them. The data-based technologies through which we increasingly receive information about the world, stay in touch with those in our lives and go about the processes at school, work and home, have the power to encode and transform literacies and meaning making practices (Chaka, 2019). Thus, they change not only the literacies that exist already, but render new ones like digital literacies, AI literacies and data literacies indispensable.

Digital literacies involve understanding and using diverse digital devices, expressing ourselves and understanding others using digital text ethically, in socially acceptable and culturally sensitive ways. AI literacies include critically evaluating and using AI technologies and understanding how data are generated, processed and used for different purposes (Kahila et al., 2024). This brings us now to data literacies.

All the papers in the study concur that the power of data-based ideologies and technologies are on the rise. Data is collected practically everywhere, and it ends up almost all the time in its digital form. Technologies and devices which are now indispensable to us depend on data for their functioning, can take exhaustive data from us, behave in certain ways because of that data, and output more data when used. Not surprisingly, the papers reviewed deal with the possibilities and concerns around what it takes to understand data and data-based devices, how to read and speak data and how to go further beyond, and take control of it for our own ends.

Key Takeaway

Comment

- *The need to understand data and data-based devices, how to read and speak data and how to take control of it for our own ends are key to the AI-DL project.*
- *The idea that digital technologies depend on data are key to AI-DL.*

Co-funded by the European Union

Situating and defining data literacies

While data literacy was initially considered a subset of digital literacy, it has more recently become an intrinsic part of digital (Hansen & Wasson, 2016), visual and other literacies (Nwagwu, 2025, Kahila et al., 2024). It intersects and draws on statistical literacy, scientific literacy, computational literacy, media literacy, critical literacy (Vacca et al., 2022), mathematical literacy (Henderson & Corry, 2021), analytical skills and argumentation (Ambarwati et al., 2020).

Data science can be looked at as a commitment to systematically observing and working with data (D'Ignazio & Klein, 2020 as cited in Lee et al., 2022). While data science education prepares learners for careers that require data skills, data literacy education aims to prepare learners in all roles of an increasingly datafied society (Matuk et al., 2022).

In its most basic form, data literacy for everyone is the ability to read, understand, create, and communicate data as information (Lichti et al., 2021, Nwagwu, 2025). Adding context, it could entail evaluating data, making meaning and drawing conclusions from data (Israel-Fishelson et al., 2024, Lestari & Rosana, 2020, Moon et al., 2023). Adding critical thinking and combined with a purpose, data literacy could be seen as making, analysing and interpreting data-based claims critically to foster evidence-based decision-making (Israel-Fishelson et al., 2024).

Key Takeaway

Comment

- *The above finding links data literacy to notions of criticality and to making evidence-based or more informed decisions, which links it to notions of citizenships and democratic behaviours.*

Data literacy could come into play within an inquiry process that calls for the ability to use appropriate data sets ethically, to ask and answer real world questions (Wolff et al., 2017 as cited in Robertson & Tisdall, 2020, Vacca et al., 2022 and Watson & Smith, 2022), especially as part of various disciplines in school (i.e. Physics Education Department, Faculty of Mathematics and Natural Sciences, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia & Dewi, 2024, Rahmita & Rosana, 2020, Suryadi et al., 2024) and as a basis for taking informed action (Ambarwati et al., 2020).

Early definitions of data literacy neglected aspects of power, equity, empowerment, and emancipation and focused more on the potential benefits compared to the critical questions (Dangol & Dasgupta, 2023). However, many papers published in the 2020s combine data processing techniques with the ability to analyse data as a contextual, socio-political and cultural phenomenon, widening the scope of data literacies. They also include personal data literacy components which stress the importance of understanding data-based algorithms and their implications for privacy, equity and social justice (Ilomäki et al., 2023, Kahila et al., 2024, Matuk et al., 2022).

Co-funded by the European Union

Finally, some papers look at data literacies as a collective endeavour (Cowie et al., 2021, Dander & Macgilchrist, 2022, Fagerlund et al., 2025) and as an essential component of civic engagement and social inclusion in a democracy where citizens have a voice in "how and for whom data are used and what they are used for" (Vartiainen et al., 2020), and where they are actively engaged in "critiquing the societal implications of future uses of data" (Robertson & Tisdall, 2020).

In the educational context, teacher's data literacies are used either interchangeably with assessment literacies (Henderson & Corry, 2021, Kim & Yu, 2023), as intersecting with them or as subsuming them (Henderson & Corry, 2021, Kim & Yu, 2023). The most cited definition puts teacher data literacy as a broader concept, extending assessment literacy to include the collection, analysis, and use of data beyond conventional assessment data, such as student satisfaction surveys, along with background information about students disciplinary knowledge and practice, curricular knowledge, pedagogical content knowledge, and an understanding of how children learn, with the intention of making better pedagogical decisions (Mandinach & Gummer, 2013, 2016, as cited in Børte et al., 2023, Dunlap & Piro, 2016, Henderson & Corry, 2021,Jackson, 2022, Mandinach & Jimerson, 2022, Melnikova et al., 2023, Michos et al., 2023, Ndukwe & Daniel, 2020, Schildkamp et al., 2020), and specifically to take instructional action (Cowie et al., 2021, Hoogland et al., 2016, Kippers et al., 2018).

Key Takeaway

Comment

- The above finding shows that the context of data literacy in the educational context most often featured examples related to assessment practices. While this is interesting, it is not the context we are exploring in AI-DL.

Data literacy for educators includes ICT skills like working with digital assessment and student data systems (Schildkamp et al., 2020). One paper stresses multimodal meaning making by defining teacher data literacies as teachers' meaning-making practice involving documenting, noticing/interpreting, visualising, and acting on students' multimodal meaning-making as data (Kim & Yu, 2023).

What emerges from this review is the impression of data literacy as a fluid, ambiguous and multidimensional concept with no universally accepted meaning or skill set. There is no standard definition of data literacy that could cut across contexts and disciplines (Ansyari et al., 2020, Mertala, 2020). At the same time, data literacy is also cross-disciplinary, a fundamental skill that is part of diverse domains (Chaka, 2019, Nwagwu, 2025).

Co-funded by the European Union

Key Takeaway

Comment - *The review is suggesting that this notion of data literacy is a fluid, ambiguous and multidimensional concept with no universally accepted meaning or skill set is very important for AI-DL. The definitions or descriptors typically reflect the context in which the authors were working within.* - *The statement that there is no standard definition of data literacy that could cut across contexts and disciplines is also extremely relevant to AI-DL.* - *As is the suggestion that data literacy is also cross-disciplinary, a fundamental skill that is part of diverse domains*.

As in the case of literacies in general, data literacies impact and are impacted by changing technological, cultural and political factors. As technology changes, data literacies have to necessarily adapt to new changes by not only including new skills but also reviewing old definitions and looking at commonly agreed contexts in new light (Ilomäki et al., 2023, Nwagwu, 2025). This is even more evident in the case of generational advances like Artificial Intelligence (Chaka, 2019, Ilomäki et al., 2023).

Key Takeaway

Comment - *The idea that data literacies impact and are impacted by changing technological, cultural and political factors is very relevant in AI-DL, as GenAI is a new tool in education.* - *The statement that as technology changes, data literacies must adapt to such changes by not only including new skills but also reviewing old definitions and looking at commonly agreed contexts in new light is particularly relevant in this project*.

Perhaps, with such rapid advances in technology, where technology impacts both the composition of literacies and how they are enacted, and where any activity, particularly those that involve technology enacts multiple literacies at the same time, it does not make sense to try and fix the exact boundaries and contents of the various literacies (Chaka, 2019, Ilomäki et al., 2023).

Key Takeaway

Comment - *The AI-DL project has been experiencing challenges in fixing such exact boundaries for a range of emerging literacies, and thus this suggestion to not try and fix exact boundaries is relevant.*

As suggested in the case of critical digital literacies, data literacies can also be looked at as "as an assemblage of meanings and practices and less as a finite and tightly bound concept" (Nichols & Stornaiuolo, 2019 as cited in Ilomäki et al., 2023), which changes with the context and the humans involved, and which changes with time as they and the world around them evolves.

Key Takeaway

Comment - *This idea of critical digital literacies being seen as "as an assemblage of meanings and practices and less as a finite and tightly bound concept" that changes with the contexts and the human involves, is very relevant to AI-DL.* - *In the case of AI-DL we need to consider if the definitions presented here are relevant and applicable to our context.*

It may be better to identify a small set of widely accepted core dimensions that educators can use as a flexible checklist or guide - one that can adapt to emerging phenomena and one that remains mindful of the needs of individuals, groups, and society (Ilomäki et al., 2023).

Factors that have impacted the evolution of data literacies

Some factors emerged as the most impactful ones on the evolution of data literacies and their components. These factors, many of which are interrelated, directly or indirectly affect the competences and attitudes that are recommended as part of data literacy education.

1. *Change in what is considered data*

For a long time, data was perceived to be a set of numbers that were recorded to answer some question. This was because the analytical instruments that were used to treat data were often more conducive to manipulating numbers.

However, as many papers in the study point out, this is not the case anymore. Data is now understood to mean widely different things, including but not limited to swathes of text, audio, video and images

(Dangol & Dasgupta, 2023). It could be geospatial (Juergens, 2020) or biophysical (Carmichael, 2020) and could be the result of activity recorded online, or physical changes captured by sensors.

Data could describe personal, work or school life often cutting across them as mobile technologies, wearable devices, internet and cloud services have become commonplace (Carmichael, 2020, Chaka, 2019, Robertson & Tisdall, 2020).

With the advent of recommender systems and ecommerce sites, digital traces and metadata used by them to describe user behaviour, and the user models created as a result are also data (Juergens, 2020, Mertala, 2020). Finally, data could also mean synthetic data generated by AI systems (Mouta et al., 2023).

For a long time, data was perceived to be a set of numbers that were recorded to answer some question. However, as many papers in the study point out, this is not the case anymore. Data is now understood to mean widely different things, including but not limited to the swathes of text, audio, video and images (Dangol & Dasgupta, 2023). Finally, data could also mean synthetic data generated by AI systems (Mouta et al., 2023).

Key Takeaway

Comment
- *In considering the above section, this idea of the changing perceptions in relation to what is data is very relevant to the AI-DL project.* - *While the literature does not specifically mention GenAI, it mentions AI and there is clearly a strong connection for AI-DL.*

2. *Increased intensity and scale of data collection and storage*

Almost all papers mention increased data capture due to ubiquitous internet, cloud and mobile services and cheap and powerful sensors integrated in everyday appliances (Israel-Fishelson et al., 2024, Mertala, 2020, Vartiainen et al., 2020). Youth create data when engaging in social media, completing school assignments and playing video games with friends (Israel-Fishelson et al., 2024, Mertala, 2020, Moon et al., 2023, Vartiainen et al., 2020).

These data could be and are, put together and as an aggregate determine what aspects of life are accessible to algorithms, and therefore to corporations, school, political and governmental forces. Effects described range from more effective learning to differential distribution of benefits to tracking and surveillance (Carmichael, 2020, Dander & Macgilchrist, 2022, Vartiainen et al., 2020).

In schooling, there is a big change from manually recorded, limited data informing pedagogical decisions at the classroom level, to unprecedented levels of automatically recorded digital data from

Co-funded by the European Union

digital learning environments and student information systems that could be used at the individual, classroom, school level and further (Mertala, 2020, Michos et al., 2023).

Key Takeaway

Comment
- *The above text captures the increasing datafication of education, particularly the use of algorithms and digital platforms in schools. While this is the case the academic literature reviewed for this task, as yet, has not begun considering how AI and GenAI tools might be impacting on such decisions in schools.*

3. *Availability of data recorded invisibly and automatically*

Technologies with seamless interfaces now operate silently in the background and continuously harvest our data (Chaka, 2019). Geo-spatial data that is recorded when we use fitness trackers, smart watches or smart phones are not actively provided by the user. Gestures, facial features and other behavioural signs are recorded by unobtrusive computer vision systems used in some universities (Juergens, 2020, Mertala, 2020, Ndukwe & Daniel, 2020).

Key Takeaway

Comment
- *This text captures the increasing collection and use of our personal data in all aspects of our lives. While the above example does not specifically mention GenAI, this is something teachers and students should be aware of in the lives. This has strong linkages to citizenship and democracy issues.*

4. *Growing tolerance for tracking and loss of privacy*

There is always a tension in using media platforms, online communities and any other data intense environment or application that asks us to trade the benefits they offer against the personal data they capture. The captured data include an obscene amount of data that is unrelated to the functioning of the platform or app, which is becoming normalised and internalised as necessary (Vartiainen et al., 2020).

Often, to use a technology is to forsake some of our rights. Privacy, among other things, has become the cost of doing business. It has also become the cost of schooling since wishing to attend a school often means that we comply with the protocols and data practices of that school (Mertala, 2020). With all the digital traces we leave, we have become 'transparent subjects' of governments, which would

Co-funded by the European Union

have been unacceptable just a few decades back (Dander & Macgilchrist, 2022). This acquiescence is starker in the case of digital natives, who, while aware of privacy issues, are not willing to take measures to protect their data (Israel-Fishelson et al., 2024).

Key Takeaway

Comment
- *Similarly, this issue of loss of privacy has relevance to the use of AI and it impacts on issues such as citizenship and democracy.*

5. *What algorithms can do with the data*

Rich data on people, often cobbled together from different sources, and advances in algorithmic techniques and associated hardware, enable interpretations of data in unprecedented ways (Dangol & Dasgupta, 2023). Algorithms can now draw out insights from invisible patterns in data collected at scale, and use these to educate, profile, infer and predict people's needs, values, and behaviour (Ndukwe & Daniel, 2020, Vartiainen et al., 2020).

This is experienced as better recommendations, engaging media experiences and behavioural nudges. This also means leading tech firms decide which objects and ideas enter society (Vartiainen et al., 2020). Nowadays people construct their personal identities increasingly based on these mediated experiences (Vartiainen et al., 2020).

Within the walls of the school, learning analytics collect, analyse and report multimodal student data, including the digital traces they leave while interacting with digital systems, with the view of helping to understand and optimise learning and learning environments (Michos et al., 2023, Ndukwe & Daniel, 2020). AI technologies promise much more personalised pathways with active, human-like interaction (Kim et al., 2022).

Key Takeaway

Comment
- *The literature speaks about algorithms in the context of AI systems and personalised learning, but not in the context of GenAI. This once again has relevance to citizenship and democracy.*

6. *Learnification, accountability and evidence-based decision making*

In education, transnational policy now demands accountability from teachers and school leaders; where accountability measures are calculated using statistical analyses on educational data which compare one student's past with present, each student against the other, teachers and schools with each other (Mertala, 2020). Questions around education are increasingly rephrased as questions about learning, since learning is more easily broken down into measurable, quantifiable and analysable outcomes when compared to ambiguous and undeterminable educational objectives (Mertala, 2020). Teachers are then urged to investigate their practices based on these analyses since data use is assumed to improve teacher's methods of instruction (Amels et al., 2019).

Key Takeaway

Comment
- *Here again the literature is focused on learning analytics to enhance student learning, which applies to the increasing role of AI technologies in schools.*

7. *In what situations and under what terms do algorithms get to be used*

While accountability concerns for teachers have changed, so has accountability constraints and acceptability of AI. Some tasks traditionally thought of as strictly human are now either executed by AI or with the help of AI. As such, many human experiences are now mediated by AI. Many papers in the study urge integrating AI into classrooms for enjoying all the benefits that the technology can potentially bring about (Kim et al., 2022). They speak of the potential to improve school administration processes, enhance students' learning experiences, simplify teachers' daily tasks, and broaden opportunities for lifelong learning (Mouta et al., 2023). Other papers express concern related to AI use, which includes inscrutability of algorithms and the tech companies' proprietary practices leading to questionable accountability, equity and accessibility, privacy, environmental and other concerns (Mouta et al., 2023, Vartiainen et al., 2020).

Key Takeaway

Comment
- *This factor is very relevant to AI-DL, specifically the need to consider the pros and cons of AI use, and by extension GenAI use, in schools. It also raises a number of concerns that very relevant for all citizens, particularly teachers and students.*

8. *What data and data analysis tools are accessible to citizens*

While sizes of datasets are growing, so are the number of tools that help make sense of data that are available to students and teachers. These advanced tools not only multiply the affordances for data literacy (Chaka, 2019), they can also be used for data justice efforts and data activism (Dander & Macgilchrist, 2022). Children are growing up as a part of online maker communities that build collaborative knowledge where they could use technology to classify, organise, construct and evaluate information that contribute to a common interest (Vartiainen et al., 2020).

Key Takeaway

Comment - *This factor is very relevant to AI-DL as GenAI usage could help teachers and students to classify, organise, construct and evaluate information.*

Components of data literacies for students and teachers

This literature review lists the data related knowledge, skills and attitudes cited, recommended or inferred from the publications reviewed. They have been classified as data literacy components that come into play while studying other disciplines, those that revolve around personal data, its uses and protection, those that have particular relevance for teachers, and data literacy attitudes that multiple papers recommend. There might be overlap in the categories depending on context and component, and mastery in one category might presuppose a good grasp of another.

The rationale for cataloguing data literacy components is to provide a clear picture for analysing which parts are impacted by GenAI technologies, which parts are not and what is missing overall. It is also to help teachers identify areas for teaching and professional learning, those organising data literacy interventions to have a basis for reference, and policy makers to create or update data literacy frameworks.

This area of the review is broken into parts a to d, as follows:

a) Disciplinary data literacy components
b) Personal data literacy components
c) Teacher specific components
d) Data literacy attitudes

a) Disciplinary data literacy components

Part a consists of data literacy components that are necessary for studying a subject or a topic, or those required for doing an interdisciplinary project (Inter, intra and transdisciplinary competences). Not all these competences will come into play in a data-based process, and those that do might pan out in a different order. The whole process might involve multiple iterations and variations. These competencies are listed below:

Understanding data basics

Understand the basics of data, including:

- how could data be defined or described: Examples being “any type of information that is systematically collected, organized, and analysed” (D'Ignazio & Klein, 2020, as cited in Lee et al., 2022); “the raw material produced by abstracting and reducing the world into representative forms” (Kitchin, 2014 as cited in Mertala, 2020).
- what could be considered data: for example, numbers, words, stories (Ambarwati et al., 2020, Lee et al., 2022).
- how data can be captured using sensors (Zhang et al., 2024).
- how could data be both the input and output of a data process (Israel-Fishelson et al., 2024).
- that data can be qualitative and quantitative (Matuk et al., 2022); it can be open data, big data and/or public data (Dangol & Dasgupta, 2023, Pellegrino & Antelmi, 2023).
- how and where could data be stored (Juergens, 2020).
- what are datatypes, data formats and databases (Ambarwati et al., 2020, Israel-Fishelson et al., 2024).
- what could data be used for (Dangol & Dasgupta, 2023).
- how “raw data are collected and transformed into numerical descriptions of the world” (Frankenstein, 2013, as cited in Mendez-Carbajo et al., 2019).
- how data is shaping society (Ambarwati et al., 2020, Dangol & Dasgupta, 2023).

Key Takeaway

Comment

- *This section is particularly relevant to informing what content should be covered in the AI-DL training module and the issue of what is data and understanding the key concepts, such as data vs information etc.*

Asking a question

As part of an inquiry or during a scientific experiment, a question might be posed that requires data and data literacy components. This question could also be posed during data literacy education, since this is seen as an authentic and meaningful way of working with data. Anchoring data work in a real-life scenario helps students understand what kinds of questions can be answered using data (Dangol & Dasgupta, 2023).

Data literacy components here would involve the ability to:

- ask a relevant question, find a problem, write a hypothesis or design an experiment (Lichti et al., 2021, Dangol & Dasgupta, 2023, Suryadi et al., 2021, Watson & Smith, 2022).
- write out methods to answer the question or solve the problem, including limitations of the chosen methods (Lichti et al., 2021).
- understand when data is needed and what kind of data is needed (Rahmita & Rosana, 2020).

Co-funded by the European Union

Finding data

- verify if the data needed has been collected already (Dangol & Dasgupta, 2023).
- Search, copy or download data relevant to the problem (Ambarwati et al., 2020, Irish et al., 2019, Kippers et al., 2018, Lichti et al., 2021, Mendez-Carbajo et al., 2019, Nwagwu, 2025, Physics Education Department, Yogyakarta and Dewi, 2024).
- Evaluate the quality of the source (data creator) and platform (data aggregator) (Ambarwati et al., 2020, Mendez-Carbajo et al., 2019, Rahmita & Rosana, 2020).
- Read metadata to understand the context in which data was collected and the ways in which it was processed and determine data quality and accuracy (Israel-Fishelson et al., 2024, Juergens, 2020, Lee et al., 2022, Mendez-Carbajo et al., 2019, Vacca et al., 2022).
- Determine if other types of data from other sources are required to add to the data found (Irish et al., 2019).
- Combine data correctly or merge quantitative and qualitative data where required, to form a bigger dataset (Israel-Fishelson et al., 2024, (Lichti et al., 2021, Suryadi et al., 2024).

Creating data

- determine what variables, what form and type of data to include in a dataset and how to quantify or categorise it (Lee et al., 2021, Matuk et al., 2022, Moon et al., 2023, Suryadi et al., 2024).
- select appropriate sampling techniques (Lee et al., 2021, Matuk et al., 2022, Watson & Smith, 2022).
- collect data by planning and carrying out investigations, conducting physical experiments, running simulations, measuring personal data, using virtual, remote labs or distributed systems like crowd mapping etc. (Ambarwati et al., 2020, Dangol & Dasgupta, 2023, Ilomäki et al., 2023, Israel-Fishelson et al., 2024, Juergens, 2020 , Lee et al., 2021, Pellegrino & Antelmi, 2023, Rahmita & Rosana, 2020, Suryadi et al., 2021, Watson & Smith, 2022).
- create datasheets for collected data including relevant information like context, metadata etc. that will help with the reuse of data (Ambarwati et al., 2020, Ilomäki et al., 2023, Lee et al., 2022).
- clean data without compromising its integrity (Lee et al., 2021, Moon et al., 2023).

Sense-making

Sense-making happens at every point of a data process and might engender rethinking and redesigning of the whole process multiple times. Visualisation and other techniques might be needed even before a preliminary exploration and sense making is possible, especially when it comes to large datasets.

Sense-making can include and go beyond the ability to:

- gauge if collected data is enough to solve the problem (Rahmita & Rosana, 2020, Schildkamp, 2019).
- identify capabilities and constraints of the datasets in solving the given problem (Lichti et al., 2021, Mendez-Carbajo et al., 2019, Moon et al., 2023, Schildkamp, 2019, Wilkerson et al., 2025).
- reflect on what caused variability, if any, in the data, if unexpected events affected the dataset and what are their implications for inferences made with this data (Irish et al., 2019, Lichti et al., 2021, Watson & Smith, 2022).
- understand how the methodologies used to collect data highlights some aspects and not others of a system or process, especially what was not measurable and therefore not reflected in the data (Dangol & Dasgupta, 2023, Schildkamp, 2019, Watson & Smith, 2022, Wilkerson et al., 2025).
- identify social, political and environmental factors that went into the collection and processing of the data and how they have affected the result (Dangol & Dasgupta, 2023, Lee et al., 2022, Mendez-Carbajo et al., 2019).
- look for evidence of bias in the data (Dangol & Dasgupta, 2023).
- reflect on how else the data could've been collected and how it will change the data (Matuk et al., 2022).
- get inputs from local experts, stakeholders and additional information sources to understand and evaluate the data (Dangol & Dasgupta, 2023, Lee et al., 2022, Schildkamp, 2019).
- design improvements, collect and organise data again if necessary (Schildkamp, 2019).

Processing, exploring and analysing data and its representations

- use appropriate tools to convert and format data to facilitate analysis (Ambarwati et al., 2020, Israel-Fishelson et al., 2024, Mendez-Carbajo et al., 201, Rahmita & Rosana, 20209).
- visualise data in multiple ways, including graphs (Dangol & Dasgupta, 2023, Irish et al., 2019, Lee et al., 2021, Lee et al., 2022, Lichti et al., 2021, Mendez-Carbajo et al., 2019, Moon et al., 2023, Pellegrino & Antelmi, 2023, Physics Education Department, Yogyakarta and Dewi, 2024, Rahmita & Rosana, 2020, Watson & Smith, 2022, Wilkerson et al., 2025, Woods et al., 2024).
- label and format graphs and other visualisations appropriately (Lichti et al., 2021, Mendez-Carbajo et al., 2019).
- identify trends using data visualisations and representations, created by self or others (Irish et al., 2019, Lee et al., 2021, Pellegrino & Antelmi, 2023, Physics Education Department, Yogyakarta, and Dewi, 2024, Watson & Smith, 2022).

Co-funded by the European Union

- reflect on what caused these patterns in the visualisation (Wilkerson et al., 2025).
- find personal connections with a visualisation in order to understand the data in context (Lee et al., 2022, Wilkerson et al., 2025).
- compare a given visualisation and those from other sources to discover new patterns (Wilkerson et al., 2025).
- critically analyse and interpret visualisations and assess how mathematical, contextual and social factors shaped them (Irish et al., 2019, Woods et al., 2024, Wilkerson et al., 2025).
- analyse data appropriately with numerical methods, with or without relevant software or programming languages to test the research hypothesis (Lee et al., 2021, Moon et al., 2023, Nwagwu, 2025, Watson & Smith, 2022, Physics Education Department, Yogyakarta, and Dewi, 2024, Suryadi et al., 2021).

Making and evaluating data-based claims

- make a decision, claim or explanation, answer questions, support or reject a hypothesis based on data (Ambarwati et al., 2020, Dangol & Dasgupta, 2023, Irish et al., 2019, Lestari and Rosana, 2020, Lichti et al., 2021, Mendez-Carbajo et al., 2019, Moon et al., 2023, Physics Education Department, Yogyakarta, and Dewi, 2024, Rahmita & Rosana, 2020, Suryadi et al., 2021, Watson & Smith, 2022, Wilkerson et al., 2025).
- identify, evaluate, combine, compare and critique data-based claims within a given context (Ambarwati et al., 2020, Irish et al., 2019, Israel-Fishelson et al., 2024, Gal, 2022, Lestari and Rosana, 2020, Rahmita & Rosana, 2020, Watson & Smith, 2022).
- make the evidence explicit (Vacca et al., 2022).
- calculate, verify or gauge the probability of and the uncertainty in a claim (Gal, 2022, Matuk et al., 2022, Nwagwu, 2025, Watson & Smith, 2022, Vacca et al., 2022).
- look for assumptions, bias, reasoning errors and incidences of cherrypicked data in understanding a claim (Dangol & Dasgupta, 2023, Gal, 2022, Irish et al., 2019, Nwagwu, 2025, Watson & Smith, 2022).
- look for biases introduced by the tools, techniques and algorithms used for inference (Dangol & Dasgupta, 2023, Lee et al., 2022).
- infer what this claim could mean for student's personal and school life, the communities that may be impacted and how they could reinforce power structures (Dangol & Dasgupta, 2023, Gal, 2022, Lee et al., 2022, Pellegrino & Antelmi, 2023, Rahmita & Rosana, 2020, Wilkerson et al., 2025).
- look for other interpretations of the same data and how other parameters could affect the data (Gal, 2022, Irish et al., 2019, Lestari and Rosana, 2020, Lichti et al., 2021, Mandinach & Jimerson, 2022, Wilkerson et al., 2025).

Extrapolating and predicting

- extrapolate data and trends in data to make predictions in a given context based on a scientific model (Gal, 2022, Lichti et al., 2021, Mendez-Carbajo et al., 2019, Physics Education Department, Yogyakarta, and Dewi, 2024, Wilkerson et al., 2025).

- design machine learning apps and other artefacts of the new generation of technologies to help in making predictions (Vartiainen et al., 2020).

Publishing data-based analyses and reports

- verify that data has been handled ethically in all parts of the data process and they are not harmful to privacy, copyrights or safety (Ambarwati et al., 2020, Ilomäki et al., 2023, Nwagwu, 2025, Wilkerson et al., 2025).
- publish datasets with datasheets that include metadata and appropriate sharing and reuse rights (Ambarwati et al., 2020).
- present data in a way that supports existing or new narratives, along with sources, assumptions, limitations, evidence, clear arguments and alternative explanations, to diverse audiences (Ambarwati et al., 2020, Dangol & Dasgupta, 2023, Irish et al., 2019, Pellegrino & Antelmi, 2023, Nwagwu, 2025, Pellegrino & Antelmi, 2023, Vacca et al., 2022).
- recommend an evidence-based course of action (Irish et al., 2019).
- be aware of the role design and presentation of data can have in the perception and reception of a narrative (Juergens, 2020).
- create new ways to express data to reflect personal meanings, values and relations (Dangol & Dasgupta, 2023, Lee et al., 2022).
- envision futures for data use to expand the investigation and pose questions for further research (Mendez-Carbajo et al., 2019, Wilkerson et al., 2025).

Key Takeaway

Comment - *Much of what is covered above is not directly related to the context of using GenAI in schools. This shows how data literacy is extremely contextual and that AI-DL needs to consider if any of these components are relevant to our context.*

b) Personal data literacy components

Part B is concerned with data literacy components that involve personal data organised according to the lead author's understanding of the framework given in "Critical Data Literacies: Rethinking Data and Everyday Life" (Pangrazio & Selwyn, 2023). As indicated earlier, the lines between disciplinary data literacies and personal data literacies might not be that clearcut in practice. Particularly, a lot of personal data literacy components listed here come into play during an ethical and safe disciplinary data practice and therefore cannot and should not be ignored.

Data identification

This section is about identifying personal data and its types, be it information voluntarily given to digital systems, that extracted with and without a person's knowledge or the data processed on behalf of people, whether they are sold to companies and institutions or provided to the users themselves (Pangrazio & Selwyn, 2023). To identify data, we need to have a basic idea of how, when and where data is collected and how it can be used.

Key Takeaway

Comment - *The above statement regarding the need to have a basic idea of how, when and where data is collected and how it can be used is extremely relevant to AI-DL.*

The components of data literacy involved is identifying personal data include:

- online data recorded by search engines, mail providers, web sites, gaming and media platforms, ecommerce systems etc. (Israel-Fishelson et al., 2024, Moon et al., 2023, Ndukwe & Daniel, 2020).
- rich personal data on social media (Moon et al., 2023, Ndukwe & Daniel, 2020).
- learning data captured by educational technologies. Examples: Time spent on the learning management system, time spent on individual tasks, assignment submissions and participation in discussions (Israel-Fishelson et al., 2024, Melnikova et al., 2023, Ndukwe & Daniel, 2020).
- geo-spatial data recorded in mobile phone apps and laptops (Juergens, 2020).
- images and videos captured on computer cameras, smartphones, stand-alone video cameras etc. (Ndukwe & Daniel, 2020).
- facial data captured by facial recognition systems. Examples: Student attention data captured using computer vision systems in universities (Ndukwe & Daniel, 2020).
- physical, chemical and biological data captured by sensors, sensor networks and wearables like fitness trackers. Examples: Data recorded by the sensors in cars, household appliances and lighting (Moon et al., 2023, Ndukwe & Daniel, 2020, Vartiainen et al., 2020).
- demographical data on a topic of public interest. Example: Data and statistics on Covid (Ilomäki et al., 2023, Watson & Smith, 2022).
- AI generated data (Mouta et al., 2023).

Key Takeaway

Comment - *It is interesting that only one source mentioned AI generated data, again showing that the connections between data literacy and GenAI usage are minimal at best.*

Data understandings

Data understandings is about looking at the context in which data was collected and the methods that were used for its collection. It also looks at how this data is likely to be processed, used and reused by others. Data understandings also cover interpretation of processed data; however, it may be represented, but stops short of connecting this to personal choices, experiences, action and reform. Finally, data understandings could also involve use of tools that help us in gaining these insights (Pangrazio & Selwyn, 2023).

The data literacy components related to data understandings found in the studied papers include understanding:

Personal data collection:

- how all technology-mediated actions generate data (Mertala, 2020).
- how data is non-neutral, subjective and depends on context (Carmichael, 2020, Dangol & Dasgupta, 2023, Lee et al., 2022).
- how personal data could be shared or given by the user (Ilomäki et al., 2023, Vartiainen et al., 2020).
- how personal data can be collected or captured with or without the knowledge of the user (Ilomäki et al., 2023, Vartiainen et al., 2020). This could be done, among others, through the use of GPS to record location data (Juergens, 2020), by accessing metadata of photos and videos, or through the use of cookies and trackers to harvest all sorts of user data (Mertala, 2020).
- how personal data could be extracted (Ilomäki et al., 2023, Vartiainen et al., 2020). For example, data extracted by teachers and administration from learning management systems (Israel-Fishelson et al., 2024), that extracted by platforms from sports and health apps (Israel-Fishelson et al., 2024).
- how personal data of minors can be generated by other actors on their behalf, including parents, teachers and care givers (Vartiainen et al., 2020).
- how long and how much data could apps retain (Lee et al., 2022).
- who are the stakeholders that collect personal data and for what purposes (Israel-Fishelson et al., 2024, Vartiainen et al., 2020). Examples here are technology companies, financial organizations, educational institutions, and governmental agencies (Israel-Fishelson et al., 2024).
- how data is sold or exchanged amongst stakeholders (Robertson & Tisdall, 2020, Vartiainen et al., 2020).

Co-funded by
the European Union

Key Takeaway

Comment

- *The above citations are all extremely relevant to the issues related to citizenship and democracy, but once again they don't reference GenAI.*

Personal data processing:

- How personal data could be visualised to understand what is collected and how they are processed (Juergens, 2020, Mertala, 2020). One example is to use the dedicated apps and websites that come with health and performance trackers which provide visualisation of some of the data collected (Mertala, 2020).
- How personal data, collected from a lot of individuals, can be processed through algorithms to create models or representations to identify people, places, objects, emotions etc. (Juergens, 2020, Mertala, 2020, Robertson & Tisdall, 2020, Vartiainen et al., 2020, Zhang et al., 2024).
- How "sensitive information can be inferred from seemingly innocuous data" (Mouta et al., 2023).
- How these representations are used to understand, predict, and influence people. Examples here are recommendation systems that track and combine data from multiple similar users to predict their interests (Mertala, 2020, Vartiainen et al., 2020) and learning analytics that model past learning processes to facilitate decisions regarding future learning (Mertala, 2020, Robertson & Tisdall, 2020).
- How personal data is used at school by data-based technology to track fees, attendance, dropout rates, student behaviour and teacher performance and to take a range of actions from facilitating an individual student's learning to informing institution wide processes (Kim et al., 2022, Mertala, 2020, Mouta et al., 2023, Vartiainen et al., 2020).
- How data representations depend on proxies. For example, learning analytics depend on proxies that indicate student attention instead of attention itself. These proxies can be clicks on links, duration of activities, measurements of electrodermal activity etc, which have been identified as meaningful for learning (Kim et al., 2022, Mertala, 2020).
- How AI models are trained and tested using labelled data and how they can reflect the bias in the data and algorithmic processes (Dangol & Dasgupta, 2023, Israel-Fishelson et al., 2024, Mouta et al., 2023, Zhang et al., 2024).
- How uncertainty in AI's decisions and suggestions should be considered before carrying out an action based on it (Kim et al., 2022).
- How these technologies are proprietary, often lack transparency, and can harm interpersonal, institutional, and commercial privacy (Israel-Fishelson et al., 2024, Vacca et al., 2022).

Key Takeaway

Comment
- *The above citations are all extremely relevant to the issues related to citizenship and democracy, but once again they don't reference GenAI. However, there is mention of AI models and the issue of lack of transparency, which is relevant to AI-DL.*

Data reflexivity

Data reflexivity is about bringing "consciousness and thought to when and how (personal) data might be processed and used" (Pangrazio & Selwyn, 2023). It is about connecting how one's interactions with data-based technologies change the future experiences with those technologies for us and enable profiling and predictions for others. Data reflexivity is also about figuring out the reasons behind privacy breaches or infringement of rights, about sorting out the societal norms that shape our data-sharing practices and about how these practices shape society in turn (Pangrazio & Selwyn, 2023). As part of data reflexivity are understanding:

how humans and systems affect data:

- how complex sociopolitical and cultural practices and perspectives determine what data matters, how it is used, how it is categorised and labelled, what visualisations are used to represent it and what shape data based tools take (Carmichael, 2020, Ilomäki et al., 2023, Kim et al., 2022, Lee et al., 2022, Matuk et al., 2022, Vartiainen et al., 2020, Woods et al., 2024).
- how available mathematical, statistical and algorithmic techniques dictate how different forms of data are produced and used. Example: Automated image processing techniques enable extraction of emotions from image data sets (Juergens, 2020, Woods et al., 2024).

how humans and systems are affected by data:

- what aspects of lives are affected by data (Israel-Fishelson et al., 2024, Lee et al., 2021).
- what are the social, political, and economic implications of data-driven design and automation (Dangol & Dasgupta, 2023, Israel-Fishelson et al., 2024, Moon et al., 2023, Mouta et al., 2023, Vartiainen et al., 2020).
- what practices data-based methods support and what practices lie beyond them and therefore are ignored, discouraged or suppressed (Carmichael, 2020, Vartiainen et al., 2020).
- how data mediates and changes interactions between individuals and the broader social world (Carmichael, 2020, Dangol & Dasgupta, 2023, Israel-Fishelson et al., 2024, Lee et al., 2022, Kim et al., 2022, Vartiainen et al., 2020, Woods et al., 2024). Data based technologies bring people together and give them platforms to discuss and take communal actions, but they can also fragment, isolate and polarise (Vartiainen et al., 2020).

Co-funded by the European Union

- how what is measured and highlighted by data-based technologies changes our practices and sense of identity, liberty, privacy and agency. For example, how using measurable, quantifiable indicators for learning might make us align our behaviour in line to modify these indicators to the exclusion of other indeterminable qualities (Lee et al., 2022, Mertala, 2020, Vartiainen et al., 2020, Woods et al., 2024).
- how judgements or decisions regarding what to watch or what to buy is based on data-based technologies and how they change what we access, buy and value (Israel-Fishelson et al., 2024, Vartiainen et al., 2020).
- how biases and sampling decisions affect the people subjected to decision making systems (Israel-Fishelson et al., 2024, Lee et al., 2021, Lee et al., 2022, Matuk et al., 2022).
- who data represents and who is ignored, who has power over collected data, who can access it and make sense of massive amounts of information of multitudes (Dangol & Dasgupta, 2023, Lee et al., 2021, Lee et al., 2022, Mertala, 2020, Robertson & Tisdall, 2020, Vartiainen et al., 2020).
- how data visualisations and representations reinforce the notion that data is objective and naturalise data's position as an inherent phenomenon (Woods et al., 2024).
- how effects of data based technology at schools affect a student and their class, including flagging attention, anxiety because of constant monitoring, lesser sense of agency and human interactions and inevitability of less than ethical choices, oversimplification of complex educational practices and social settings (Carmichael, 2020, Israel-Fishelson et al., 2024, Kahila et al., 2024, Kim et al., 2022, Mertala, 2020, Mouta et al., 2023, Vartiainen et al., 2020).

Key Takeaway

Comment

- *This section is very relevant to issues of citizenship and democracy and in particular how aspects of our lives are impacted by data, and it mentions issues such as bias, power relationships and the wider social impact of data on our lives. Again, it does not specifically mention GenAI, but these are wider issues that AI-DL needs to consider.*

Data strategies

This is about taking data understandings and reflexivity as a basis for action. An example would be to gauge the tensions between data privacy and visibility and choosing an appropriate course of action within a system or device. The authors (Pangrazio & Selwyn, 2023) stress the importance of being critically reflexive when it comes to dealing with data-based systems, especially when prompted for action, like clicking on terms of agreement or filling an online survey—*why am I being asked to do this? Who might be interested in the information I give and how might it be used?* Data strategies that people can use to manage, access and protect data would ask them to:

- weigh the consequences before sharing personal data (Israel-Fishelson et al., 2024, Robertson & Tisdall, 2020, Vartiainen et al., 2020). Being data literate and exercising agency would imply choosing between participation and abstention from activities that involve collecting personal data (Vartiainen et al., 2020).
- read user agreements and verify the policies of data platforms before engaging with them (Israel-Fishelson et al., 2024, Vartiainen et al., 2020).
- select privacy settings and the visibility of our online content and presence (Israel-Fishelson et al., 2024, Vartiainen et al., 2020).

Data tactics

Data strategies are about flexing one's agency and putting humanness back in data processes while playing by the rules of the system. Data tactics takes this further, towards "subverting and resisting the dominance of Big Tech", by not relying solely on official strategies. Here, we develop and engage in alternative or even disruptive ways of producing, processing, and using data, especially as a collective (Pangrazio & Selwyn, 2023).

Data tactics could have users:

- access and use personal data to infer and reveal personal and social implications (Dander & Macgilchrist, 2022, Matuk et al., 2022).
- communicate those implications to engage audiences emotionally and intellectually (Dander & Macgilchrist, 2022, Matuk et al., 2022) This might include exploring and creating new methods of expression with readily available materials (Dangol & Dasgupta, 2023) or using data-art (Matuk et al., 2022, Woods et al., 2024).
- obfuscate data traces (Brunton & Nissenbaum, 2015 cited in Dander & Macgilchrist, 2022). This could be by use of encryption mechanisms (Dander & Macgilchrist, 2022).
- engage in data activism and use data to counter discrimination, and to change what and who is visible, who is heard and is therefore part of politics (Carmichael, 2020, Dander & Macgilchrist, 2022, Ilomäki et al., 2023). One of the ways this could be done is by receiving, collating and anonymising crowdsourced data, in order to counter mainstream and authoritative sources that might embed biases (Dander & Macgilchrist, 2022).
- use data in arguments for change, to envision and bring about a better world (Carmichael, 2020, Dander & Macgilchrist, 2022, Israel-Fishelson et al., 2024, Lee et al., 2022, Vartiainen et al., 2020, Woods et al., 2024).

Key Takeaway

Comment - *This section is again relevant to the wider issue of citizenship and democracy and in particular the line "communicate those implications to engage audiences emotionally and intellectually" and the notion of "data activism".*

c) Teacher specific components

This section covers data literacy components that deal with handling and using student data and data based technologies safely and correctly in order to bring about better processes and outcomes, pedagogical components for teaching data literacy to students, and being aware of both the effects of data use in the classroom and the forces that push for such an use, with the view of choosing the best courses of action for all the students.

Handling data ethically

- Approach data use ethically. Be thoughtful about what to do with data and how to go about it (Mandinach & Jimerson, 2022). Know when to refrain from doing something just because it is doable and when to probe further and do one's utmost to handle a problem correctly and take valid action (Mandinach & Jimerson, 2022).
- Use diverse data sources to address the whole child, including personal and contextual factors, interests and strengths and not just on academic performance (Mandinach & Jimerson, 2022)
- Try to get to the root causes of a problem (Mandinach & Jimerson, 2022).
- Consider multiple interpretations, viewpoints and pathways, the diverse needs of all learners and collaborate to avoid cognitive fallacies and confirmation bias, subconsciously cherry-picking data and datapoints to improve metrics (Mandinach & Jimerson, 2022, Schildkamp, 2019).
- Centre data use on long term benefit to all stakeholders and communities involved (Mandinach & Jimerson, 2022).
- Weigh ethical issues in potential actions based on data and ensure equitable access to the rewards of data work (Mandinach & Jimerson, 2022).
- Be aware of personal biases, institutional and political pressures to choose a particular pathway, certain inequities and barriers to change (Mandinach & Jimerson, 2022).

Co-funded by the European Union

Key Takeaway

Comment
- *The references above to knowing "when to refrain from doing something just because it is doable or when to probe further and do one's utmost to handle a problem correctly and take valid action" seems to fit very nicely with the use of GenAI tools, as does the notion of considering "multiple interpretations", weighing the "ethical issues" and being aware of a range of "biases". These are all relevant to the context of GenAI usage.*

Using data for iterative data-based decision making (DBDM)

- Set a clear educational purpose for collecting and analysing student data (Amels et al., 2019, Ansyari et al., 2020, Dunlap & Piro, 2016, Hoogland et al., 2016, Kim & Yu, 2023, Kippers et al., 2018, Mandinach & Jimerson, 2022, Mandinach and Schildkamp, 2021, Suryadi et al., 2021). The purpose could be, among other things, to pinpoint a specific learning gap and hypothesise why it occurred (Bolhuis et al., 2019, Kim & Yu, 2023, Nwagwu, 2025), or simply assess student progress (Kim & Yu, 2023)
- Collect appropriate multimodal student data (including classroom assessments, structured observations, student information and analytics systems, research data and big data) from different sources towards validating or negating this hypothesis (Amels et al., 2019, Ansyari et al., 2020, Bolhuis et al., 2019, Dunlap & Piro, 2016, Hoogland et al., 2016, Jackson, 2022, Kim & Yu, 2023, Kippers et al., 2018, Mandinach & Jimerson, 2022, Ndukwe & Daniel, 2020, Nwagwu, 2025, Schildkamp, 2019, Schildkamp et al., 2020).
- Evaluate quality of collected data with respect to the hypothesis, and assess if more or better data needs to be collected (Ansyari et al., 2020, Kippers et al., 2018, Mandinach & Jimerson, 2022)
- Document and handle collected data appropriately and ethically (Hansen & Wasson, 2016, Kim & Yu, 2023, Mandinach & Jimerson, 2022).
- Visualise, make sense of, analyse and interpret collected data individually or collaboratively (Amels et al., 2019, Ansyari et al., 2020, Bolhuis et al., 2019, Cowie et al., 2021, Dunlap & Piro, 2016, Hoogland et al., 2016, Jackson, 2022, Kim & Yu, 2023, Kippers et al., 2018, Mandinach & Jimerson, 2022, Mandinach and Schildkamp, 2021, Melnikova et al., 2023, Schildkamp et al., 2020). This might ask for knowledge of basic statistics (Dunlap & Piro, 2016).
- Provide targeted interventions based on connecting this analysis to student understanding and teaching practice (Amels et al., 2019, Ansyari et al., 2020, Bolhuis et al., 2019, Cowie et al., 2021, Hansen & Wasson, 2016, Hoogland et al., 2016, Jackson, 2022, Kim & Yu, 2023, Kippers et al., 2018, Melnikova et al., 2023, Nwagwu, 2025, Schildkamp, 2019, Schildkamp et al., 2020, Suryadi et al., 2021)

Co-funded by the European Union

- Validate or refute current hypothesis using multiple outcome measures (Ansyari et al., 2020, Bolhuis et al., 2019, Cowie et al., 2021, Hoogland et al., 2016, Kim & Yu, 2023, Kipper2018, Mandinach, Mandinach and Schildkamp, 2021).
- Restart the next cycle or iteration of this process by either reformulating current hypothesis or coming up with the next one or choosing a different problem to study (Ansyari et al., 2020, Bolhuis et al., 2019, Hoogland et al., 2016, Kim & Yu, 2023, Kippers et al., 2018).
- Report and discuss findings with stakeholders (Amels et al., 2019, Ansyari et al., 2020, Mandinach & Jimerson, 2022).
- Use pedagogical and subject matter knowledge in each step of data use from posing valid questions and deciding what data to collect to interpreting results and connecting this to student understanding, gaps in curriculum or teaching methodologies (Cowie et al., 2021).

Key Takeaway

Comment - *While much of the above references refer to data collection, which is not the context for AI-DL, it does reference the need to "validate or refute current hypothesis using multiple outcome measures". Again, this is something relevant to the use of GenAI from a data literacy perspective.*

Note: All references cite one of Mandinach & Gummer, 2013, 2016 and 2017 as basis for these steps in data-based decision making.

Using data-based technology appropriately

- Use digital assessment and data systems appropriately (Hoogland et al., 2016, Ndukwe & Daniel, 2020, Schildkamp et al., 2020)
- Understand how different data is used to model learning processes in learning and teaching analytics (Melnikova et al., 2023, Selwyn, 2019 as cited in Mertala, 2020, Ndukwe & Daniel, 2020)
- Understand how these systems use these models to make decisions regarding future learning (Melnikova et al., 2023, Selwyn, 2019 as cited in Mertala, 2020, Ndukwe & Daniel, 2020)
- Critically assess, evaluate and interpret single and multiple user information from dashboard and data visualisation tools (Melnikova et al., 2023, Mertala, 2020, Ndukwe & Daniel, 2020)
- Evaluate machine-based recommendations and decide appropriate course of action (Henderson & Corry, 2021, Kim et al., 2022)
- Use these insights, pedagogical and content knowledge and professional discernment to make improvements in the classroom (Melnikova et al., 2023, Ndukwe & Daniel, 2020)
- Note when school datafication reduces student agency and give students voice in the decisions that affect them (Carmichael, 2020, Woods et al., 2024).

Co-funded by the European Union

- Be aware that a student's identity as projected by data-based systems changes how the student looks at themselves and others look at them (Carmichael, 2020, Woods et al., 2024).
- Watch out for the possible effects of data-based technology in school including surveillance causing anxiety, change in human interactions, student's freedom of choice, deskilling etc. (Carmichael, 2020, Mouta et al., 2023)
- How the stress on student's learning outcomes and accountability changes classroom dynamics, teacher's professional identity and the role of school and schooling in the community and society (Carmichael, 2020)

Key Takeaway

Comment

- *Several of the above citations focus on the use of digital tools to collect student assessment data and to make suggestions or predictions on student future performance, in what is commonly referred to as data analytics. These citations highlight that context in relation to data literacy is very important. While some schools may use GenAI tools in assessment tasks, it will most likely differ from the citations in this section.*

Understanding recommended approaches for teaching data literacy

- Guide data inquiries and design of experiments that are anchored in learner's context and an authentic real world problem (Dangol & Dasgupta, 2023, Gal, 2022, Suryadi et al., 2024), and progress steadily in level of difficulty, size and complexity of data sets and real life situations (Wolff et al., 2019 as cited in Pellegrino & Antelmi, 2023).
- Select appropriate datasets and data-based tools after weighing their advantages, disadvantages and how they would shape the data work and learning (Lee et al., 2021).
- Encourage sharing of lived experiences and personal connections with data and visualisations (Dangol & Dasgupta, 2023, Israel-Fishelson et al., 2024, Lee et al., 2021, Suryadi et al., 2024, Wilkerson et al., 2025, Woods et al., 2024). This could be done by selecting topics of interests to the class (Dangol & Dasgupta, 2023, Lee et al., 2022), news items with lots of statistics like Covid was (Watson & Smith, 2022), a project that involves collection and analysis of personal data etc. (Dangol & Dasgupta, 2023, Woods et al., 2024, Wolff et al., 2019 as cited in Pellegrino & Antelmi, 2023) Personal impressions could include immediate experiences, interests and past stories (Lee et al., 2021, Suryadi et al., 2024).
- Support students in coming up with ideas and communicating them, and in envisioning and arguing for change, while maintaining a critical perspective (Dangol & Dasgupta, 2023, Lee et al., 2021, Woods et al., 2024).
- Encourage multiple perspectives and multiple ways of knowing and doing data work (Dangol & Dasgupta, 2023, Israel-Fishelson et al., 2024, Lee et al., 2021, Lee et al., 2022, Woods et al., 2024). This includes encouraging new methods that emerge in the classroom as part of discussions (Dangol & Dasgupta, 2023, Lee et al., 2021, Woods et al., 2024).

- Encourage students to examine context, exclusion, labels, classes and hierarchies and imbalances of power in the datasets and data processes (Gal, 2022, Israel-Fishelson et al., 2024, Lee et al., 2022).
- Make labour behind construction of data visible (Lee et al., 2022).
- Highlight cultural, political and sociotechnical values that are implicit in dataset collection and use (Israel-Fishelson et al., 2024, Lee et al., 2021). This could include methods and instruments within science and statistics communities, computational tools from the data science community which shape what can be done with data. This could also be cultural practices and local knowledge that impact meaning making and sharing. These determine "whose knowledge and approaches are validated during data work, and what sources of data may be considered trustworthy sources of evidence by students" (Lee et al., 2021).
- Scaffold data tasks with questions, hints and other material to help students critically engage with the steps (Gal, 2022).
- Focus on foundational competencies rather than practical skills (Wolff et al., 2019 as cited in Pellegrino & Antelmi, 2023).
- When dealing with new technologies like AI, engage in process-oriented assessments that check student understanding of content and processing of data rather than technical skills (Kim et al., 2022).
- Take a STEAM approach to data literacy, working with colleagues from multiple departments (Ilomäki et al., 2023, Nwagwu, 2025, Wolff et al., 2019 as cited in Pellegrino & Antelmi, 2023).
- Ensure that the learning environment are designed so that "learners can (i) connect and reason with data (ii) participate in projects that enable the construction of knowledge about the varied purposes of data, and (iii) develop new methods of expression that reflect what they know, personally value and relate to" (Dangol & Dasgupta, 2023).

Key Takeaway

Comment

- *The references in this section relate to approaches that schools typically use in developing data literacy. While only one citation refers specifically to AI, the suggestions point towards learning environments that support active learning approaches that encourage critical engagement with data.*

d) Data literate attitudes

Curiosity, courage, problem-solving and debugging mindset, being at ease with uncertainty are all recommended attitudes when it comes to data literacy. However, the three that are mentioned extensively across all papers studied are critical thinking, learning and acting; self-reflection and metacognition; and a collaborative spirit.

Co-funded by the European Union

Key Takeaway

Comment
- *The development of data literate attitudes will be key in AI-DL.*

Criticality

Criticality is a thread that weaves through all aspects of data literacy, and perhaps literacies in general. Given the temperamental outputs and complex social, political, cultural and economic dependencies of the latest breed of data-based technologies, this is to be expected. All papers, even the ones most optimistic about data prospects, recommend a critical approach to data assemblages. Criticality here is about "socially situated reflection and evaluation" (Ilomäki et al., 2023) and involves looking at a product, problem or process from multiple perspectives, including those that "involve self-critique" (Banegas & Villacañas, 2016 as cited in Ilomäki et al., 2023). Systems, subjectivities, practice and language can all be put to scrutiny and a phenomenon along with its attendant theories, methodologies and practices are to be continually re-examined and reevaluated (Carmichael, 2020).

Criticality in data literacy includes, but is not limited to,

- Understanding and critiquing the political and complex nature of data, including its hidden aspects (Dangol & Dasgupta, 2023).
- Investigating the origins of the data and the perspectives of those who handle it and use it (Dangol & Dasgupta, 2023, Israel-Fishelson et al., 2024, Moon et al., 2023), to gauge its suitability for the purpose at hand (Mendez-Carbajo et al., 2019).
- Questioning messages and claims purportedly based on data or quantitative evidence as an internalised habit of mind (Gal, 2022, Irish et al., 2019).
- Evaluating possible alternative interpretations of presented conclusions (Gal, 2022).
- Identifying, analysing and publishing data in ways that either verifies and supports established ideas or puts forth new narratives (Dander & Macgilchrist, 2022, Dangol & Dasgupta, 2023).
- Understanding the implications of data and data-based processes and outlook on society, including social, economic, cultural and political issues and movements for social change (Israel-Fishelson et al., 2024, Lee et al., 2022, Mertala, 2020, Watson & Smith, 2022).
- Being aware of social injustices around datafication and questioning and challenging "dominant ideologies, beliefs and practices" (Pangrazio & Selwyn, 2020 as cited in Mertala, 2020).

Taking a critical approach to data would involve reflection, self-reflection, awareness of the background and aims of data sources and data actors. This is fruitless without a sound knowledge of content, context and processes (Ilomäki et al., 2023).

Co-funded by the European Union

Key Takeaway

Comment
- *The statement that, "criticality here is about "socially situated reflection and evaluation" (Ilomäki et al., 2023) and involves looking at a product, problem or process from multiple perspectives, including those that "involve self-critique" is particularly relevant to AI-DL and something we will need to develop in teachers and students in the context of GenAI usage.*

Metacognition and self-reflection

Data based technologies pervade all aspects of life, blurring lines between the professional and personal, online and offline, while classifying, categorising, predicting and nudging individuals and collectives. Metacognition and self-reflection have become indispensable especially in the context of learning, since these systems personalise learning, leading to both better learning and loss of agency in the form of reduced choices and pathways. They offer help to visualise concepts and processes, interpret, translate and simulate, which can make concepts easier to grasp but, at the same time, lead to dependency and deskilling (Carmichael, 2020).

Metacognition and self-reflection imply constant self-inquiry into what skills and practices are enhanced by using data based systems and which ones are ignored, discouraged or suppressed (Carmichael, 2020, Vartiainen et al., 2020), how these systems change how we learn and what we learn (Mertala, 2020, Vartiainen et al., 2020, Woods et al., 2024), how they change how we look at ourselves, how we feel about ourselves and collaborate with others (Dander & Macgilchrist, 2022) and how they affect our physical and emotional well-being(Mouta et al., 2023). This introspection should lead to self-regulation, self-efficacy, and adaptation of their usage to help attain our long-term goals while reducing harmful side effects. Effective self-reflection and self-regulation could only stem from the strong belief that datafication is not inevitable and can be "made and unmade in everyday data-based interactions" by "relatively autonomous, reflective, deliberative, and intentional" individuals (Dander & Macgilchrist, 2022).

Key Takeaway

Comment
- *Both metacognition and self-reflection should be to the fore in AI-DL, so that teachers and learners can reflect on how these tools impact their lives, particularly student learning.*

Collaboration

Different collaborations, exchange of feedback, and collective sense making and action are recommended by the papers with respect to data literacy:

Co-funded by the European Union

Collaboration between students as a means of meaning making by including multiple perspectives, personal histories and background to ask better questions, uncover missing or bad data and understanding limitations of data and data collection practices (Dangol & Dasgupta, 2023, Lee et al., 2022).

Collaboration with multiple stakeholders of the data project and local and indigenous experts from different backgrounds as a check to top down, industrial, political, limited and distorted views of the world (Lee et al., 2022).

Collaboration amongst teachers and students in discussing what data is collected for instructional improvement, how they are recorded and interpreted and how they are translated into learning requirements, methodology and curriculum changes (Schildkamp et al., 2020).

Collaboration amongst teachers to discuss and develop plans for action based on student data (Ansyari et al., 2020, Cowie et al., 2021, Henderson & Corry, 2021, Kim and Yu, 2023, Schildkamp et al., 2020), to exchange ideas, resources and best practices within and across subjects(Kim and Yu, 2023, Nwagwu, 2025), understand legal and ethical protections (Henderson & Corry, 2021), and undertake interdisciplinary projects (Kim and Yu, 2023, Nwagwu, 2025).

Collaboration between teachers and school leaders to plan infrastructure and resources, align values with policy and for emotional support (Amels et al., 2019, Ansyari et al., 2020, Schildkamp, 2019).

Collaboration between teachers and subject experts on both technology and pedagogy so that teachers can make sense of new innovations, new data literacy competencies and pedagogies relevant to incorporating these developments into the classroom (Ilomäki et al., 2023).

Going further, some advocate for data literacy itself to be explicitly framed as collective, relational undertaking (Dander & Macgilchrist, 2022, Fagerlund et al., 2025).

Key Takeaway

Comment
- *The AI-DL proposal stated that collaboration would be to the fore in the project, and this section captures a range of examples of what this might look like. It also highlights examples of where teachers and school leaders worked together, and such practices will need to be embedded throughout the AI-DL project.*

Tools for Data literacy education

Data literacy education is shaped and mediated by the tools used for data practices. New data analysis tools, with simple interfaces and functionalities to automate complex tasks make it possible for students to actively engage in collecting, sorting, extracting and analysing data (Moon et al., 2023).

Co-funded by the European Union

Data gathering tools enable students to collect their own data. Many data science projects use survey tools like SurveyMonkey, Google Forms and Application Programming Interfaces offered by sites like Spotify, Google Maps, Twitter, YouTube etc, often through Python (Moon et al., 2023).

Visual analysis tools can be a simple spreadsheet that helps in collecting data and performing simple manipulations like filtering by value, deleting records, and displaying basic statistics. GeoGebra, CODAP, Tinkerplots, Tableau etc. have graphical user interfaces with drag and drop features that allow datasets to be represented in tables, graphs and other visual representations. Online map services enable visualising geo-spatial data. Most visual analysis tools are intuitive, interactive and allow easy exploration of patterns in the data. They could also give a first taste of programming, an example being the formula bar in Excel, but they limit the ability to manipulate data and reflect on more sophisticated data practices. They also preclude exploring artistic visualisation methods (Lee et al., 2021, Lee et al., 2022, Moon et al., 2023, Saddiqa et al., 2019).

Programming tools like R, Python, and Pyret, are suitable for doing advanced operations on large datasets but have a steep learning curve. With friendly environments like Colab and RStudio and using different levels of scaffolding in the form of pre-written but customisable scripts (Use -> Modify -> Create pedagogical sequence common to introductory programming contexts), these can be rendered more accessible. However, these environments are less interactive than in the case of visual analysis tools and scaffolding reduces student agency in choosing what type of analysis to pursue and leads to same or similar outcomes. EduBlocks and Scratch are graphical block-based programming environments that can bridge this gap (Moon et al., 2023).

Both visual analysis tools and programming environments often have pre-loaded datasets available for ready analysis. Working with simple and complex datasets, students can gain experience in data work (Moon et al., 2023).

Data analysis tools should be chosen considering interactivity, ease of use, functionality, time available, programming knowledge and cost (Moon et al., 2023, Saddiqa et al., 2019). More than the tools themselves, "examining the context in which they are embedded is just as important" (Moon et al., 2023).

Datasets for DLE

It is recommended to use different types of datasets for data literacy education. The datasets should be about diverse topics, of different sizes and contain multiple datatypes. The publications talk about using data created by students themselves, or data from public databases, open or otherwise, collected online or from physical sources, and data made up for pedagogical needs (Moon et al., 2023, Saddiqa et al., 2019). The last could also be AI-generated (Mouta et al., 2023).

Datasets used are characterised by their recency, how relatable they are to the learners and how many observations they contained (Moon et al., 2023). Multiple authors recommend choosing relatable datasets either created by the students themselves, or those that are situated in their lived

Co-funded by the European Union

experiences within or outside academic contexts. This is believed to increase student involvement, ownership and agency (Irish et al., 2019, Israel-Fishelson et al., 2024, Lee et al., 2021, Moon et al., 2023, Wilkerson et al., 2025).

Use of authentic data sets is stressed, messy data seen as being closer to what is dealt with in real life and is believed to foster critical thinking skills and drive home the fact that variability exists must be reckoned with while making claims based on data (Lichti et al., 2021). Along the same lines, getting some experience with large datasets will also reflect real life conditions and highlight the importance of programming tools in data work and could be engaging, provided they are also freshly collected and relatable (Moon et al., 2023).

Key Takeaway

Comment
- *The above raises very interesting questions for WP3 in relation to how and where we might use datasets in teacher training activities and how teachers might use datasets with students in their contexts. The creation and collection of such classroom examples within AIDL will be extremely valuable in the future for all teachers and students.*

Creating datasets

Students can collect and analyse their own personal data using wearable activity trackers, conducting surveys in the classrooms or the larger community, or by recording events or feelings in a journal (Lee et al., 2022, Moon et al., 2023). They could conduct science experiments, use sensors and input devices to capture data or use geographic information systems, including online map services and other community data visualisation tools (Israel-Fishelson et al., 2024, Lee et al., 2022, Moon et al., 2023).

Such datasets are personally relevant, more likely to engage learner’s interest and help them see for themselves how context is essential to understand a dataset and how personal experiences and perspectives shape how datasets are created and understood (Dangol & Dasgupta, 2023, Lee et al., 2022, Moon et al., 2023, Wilkerson et al., 2025). While capturing datasets by themselves, they also need to reflect on methodologies of collection, sampling and modelling techniques, how their own mobility and instruments might highlight some aspects and not others and help them to think critically on data’s origins and the nonneutral aspect of data (Dangol & Dasgupta, 2023, Lee et al., 2022). Such experience also prepares them better to explore statistical patterns in data and make inferences based on context of collection. They also understand better the rules and rationales behind conventional treatments of data and are more open to explore novel data forms and patterns (Lehrer & Schauble, 2004 as cited in Lee et al., 2021).

However, creating datasets might not always lead to productive outcomes. It might skew the students' take of the collected data more towards their own personal views rather than the broader patterns that might exist in it. They could also hesitate to look at self-created datasets as authoritative since they are aware of its limitations, without realising that external datasets are subjected to the same errors as their own (Dangol & Dasgupta, 2023, Lee et al., 2021, Lichti et al., 2021). Sharing their work, receiving feedback, iterating and inventing may help students overcome this by helping them to create better datasets and by developing richer understandings and inferences (Dangol & Dasgupta, 2023).

To overcome the problem of size, students can also create data as part of a larger community effort, like citizen science projects. This way, they could analyse their own data while also using large long-term data sets. Citizen science projects also help them make the connections between their findings and scientific topics (Lichti et al., 2021).

Using public datasets

Despite the advantages of self-created data, the data currently used in professional practice often requires making sense of second-hand data, and so do data-based decisions involved in daily life (Lee et al., 2021). This is where accessing and using public datasets could add value. Such datasets include data published by official sources for public access like historical and civic data, weather, air and water quality data etc. and data generated algorithmically from online environments and accessible through APIs (Dangol & Dasgupta, 2023, Israel-Fishelson et al., 2024, Moon et al., 2023). These public datasets may or may not be found with a selection of visualisation and analysis aids (Wilkerson et al., 2025).

Open data are public data with additional characteristics like completeness, non-propriety nature and availability of documentation which make it better suited for educational purposes (Pellegrino & Antelmi, 2023, Saddiqa et al., 2019). But open data connecting to local contexts might not be available everywhere (Dangol & Dasgupta, 2023).

Learners can find themselves a part of many of these datasets (Dangol & Dasgupta, 2023), even if they cannot directly engage with matters of data construction thus pushing them to a more reactive stance rather than taking the position of designers and authors of data (Lee et al., 2021).

Publicly available datasets often come with the problem of missing or poorly organized data, though it might be possible to compensate for this by other means from students' local communities (Dangol & Dasgupta, 2023). But such messy data can be argued to be more advantageous since they demand certain understanding and skills. In addition, "The multivariate relationships that may be present in a public dataset allow learners to have choices in deciding the type of questions that can be answered with data" (Dangol & Dasgupta, 2023). The sociopolitical contexts of existing public data sets are one of these relationships that could contribute to a deeper understanding of data science and the world around them (Lee et al., 2021, Saddiqa et al., 2019).

What could make public datasets more suitable for educational projects is to maintain a locally curated collection of suitable datasets with support for integration of new data, along the lines of the CORGIS project (Dangol & Dasgupta, 2023).

Co-funded by the European Union

Strategies used for data literacy education of teachers

Co-construct DLE

As touched upon earlier, literacies are collaborative, participatory, distributed and deictic (Chaka, 2019, Ilomäki et al., 2023). Data literacy is a fluid concept that changes with contexts, technologies and humans involved and with time as their experiences and needs evolve. A data literacy intervention, like that described for digital literacy ones by Ilomäki et al., 2023, "are compromises between users of various backgrounds and competencies" and should prioritise topics according to their current goals, knowledge, attitudes and concerns about data (Cowie et al., 2021, Matuk et al., 2022, Robertson & Tisdall, 2020). The definitions of data literacies will have to be revisited and reformulated multiple times during Data Literacy Education (Cowie et al., 2021).

Those performing data literacy interventions should choose data literacy definitions, competences, inquiry processes, medium and technology along with subject-specific goals in discussions with teachers (Cowie et al., 2021, Matuk et al., 2022). Researchers can then create lesson materials and have teachers giving feedback based on their and their student's interests and challenges and on logistical considerations (Matuk et al., 2022). This approach directly impacts not only the relevance of the initiative, but also the commitment and ownership that teachers feel (Ansyari et al., 2020).

Key Takeaway

Comment

- *The suggestions in this section are very applicable to WP3 and in particular ideas such as "that data literacy is a fluid concept" and that definitions will need to be "revisited and reformulated" and that they are co-constructed by those involved in the project.*

Tie DLE to teacher's data use

One way to ensure data literacy education is not isolated from teacher's daily practice, is to integrate data use into it (Dunlap & Piro, 2016, Kim & Yu, 2023). Groups of teachers can come together to make sense of real assessment data, discuss and develop instruction plans and strategies, including curricular changes (Hoogland et al., 2016). Shared resources like protocols for data use and reflection, tools for data interpretation and analysis and external experts can add value to this discussion (Ansyari et al., 2020, Cowie et al., 2021, Dunlap & Piro, 2016).

Co-funded by the European Union

Key Takeaway

Comment
- *The AI-DL project is attempting to link data literacy to their daily practice, which in this case is their use of GenAI tools. Once again it suggests taking a whole of school approach, that might include guidelines and even policies.*

Provide avenues for collaboration

Collaboration among teachers is often advocated for developing and supporting teacher data literacy. Collaboration can help address problems that individuals face in various steps of data use (Dunlap & Piro, 2016, Henderson & Corry, 2021, Hoogland et al., 2016, Cowie et al., 2021).

A productive collaboration will need motivation and ownership of all the teachers, a shared understanding of the goals, and an environment of trust and mutual respect. Teachers need to feel free to experiment, make mistakes, help each other and learn (Bolhuis et al., 2019, Cowie et al., 2021). Further, interdisciplinary groups can bring in richness in the exchange of ideas, tools and best practices, provided the discussions build on synergies between the domains (Matuk et al., 2022, Nwagwu, 2025). Coaches and mentors can mediate these exchanges and bring in clarity with respect to pedagogy and data analysis (Ansyari et al., 2020, Cowie et al., 2021, Dunlap & Piro, 2016).

Integrate art teachers

Data art inquiry empowers students to express their impressions and interpretations of a data set through stories, artwork and artefacts constructed from readily available material (Dangol & Dasgupta, 2023, Woods et al., 2024). Creating artefacts forces students to pause and reflect critically on data processes (Dangol & Dasgupta, 2023), ask personally relevant questions to locate themselves within datasets, revealing how context shapes the meaning of data and making visible the personal and social relevance of data (Matuk et al., 2022, Woods et al., 2024).

Data art may or may not include technology and digital media (Lee et al., 2022) and makes it easy to communicate data analyses to a broader audience (Matuk et al., 2022). It also helps bring in teachers and non-technical learners who may not traditionally identify as doing data work, bringing in different perspectives and practices (Lee et al., 2022).

However, data art inquiry might be difficult to implement in traditional classrooms "given the tendency for schools to structure learning with disciplinary silos, and to unequally prioritize mathematics and the arts" (Matuk et al., 2022). However, establishing shared, flexible goals allowing in-the-moment adaptations by teachers could help with smoother scheduling and lesser hitches (Matuk et al., 2022).

Co-funded by the European Union

Key Takeaway

Comment
- *While the above section refers to Art teachers, the ideas here could be transferred to other curriculum areas. The notion of creating artefacts seems particularly important here and relates to the use of GenAI, as does the suggestion of taking a cross-curricular approach where teachers, from different disciplines, work together.*

Provide time to learn, experiment and discuss

Teacher's adoption and implementation of any novel content or approach is gradual and iterative. Having teachers test out a method in small groups and giving time for discussion and critique after, may shorten the process (Zhang et al., 2024). A DLE conducted over a long period would allow them time to experiment in their own class settings, come back, discuss and learn more (Hoogland et al., 2016). Such interventions can last from one and half to four years (Ansyari et al., 2020). Shorter programmes do not result in student achievement (Ansyari et al., 2020). Online and hybrid courses can add flexibility and convenience but, in this case, collaboration should be actively facilitated (Pellegrino & Antelmi, 2023).

Provide scaffolding and guides

To make interventions productive, it helps to have scaffolding in the form of thematic questions along with datasets or visualisations, which the teachers can use later for other activities. These can be questions on data or on subject-matter (Wilkerson et al., 2025). Such discussion guides can avoid certain aspects of data literacy getting downplayed (Wilkerson et al., 2025) and help teachers to slow down the data process for a deeper understanding (Cowie et al., 2021). Teachers seem to appreciate additional supportive material that they can consult in their own time, to help with content from other subjects or to enact social data discussions in class. These could be resources, multilingual where necessary, on graph-reading skills, language/literacy supports for argumentation and statistical vocabulary, support for interdisciplinary connections etc. (Wilkerson et al., 2025).

Key Takeaway

Comment
- *The idea of providing scaffolding and guides seems very apt for AI-DL and such supportive materials in AI-DL might include the AI-DL Framework and other resources.*

Make use of appropriate technology

As technology changes, data literacies must adapt by including new competencies and reviewing old ones. This is more so, with the new data-based technologies which mediate and enact data literacies and without which a huge part of data analysis will remain the terrain of experts. These technologies could be used during a data literacy intervention to develop skills and practice in working with data, and to understand how these they impact data literacies, how they act on data and what their limitations could be (Henderson & Corry, 2021).

Some papers point out that using classroom technology the teachers are already familiar with can serve to anchor data literacy concepts and lead to a better use of such technology (Ansyari et al., 2020). Others argue for the use of smartphones (Banowati et al., 2019), and artificial intelligence technologies, which, “on the one hand, are portable, pervasive and ubiquitous; on the other hand, are not temporally, spatially and geographically bound.” (Chaka, 2019).

On top of that, using them could also give teachers the opportunity to practice weighing their advantages against the ethical issues when it comes to classroom use. The intervention could drive home the point of starting *with when, how and to what pedagogical end,* when it comes to technology in education (Mouta et al., 2023).

Key Takeaway

Comment

- *This idea of “using classroom technology the teachers are already familiar with can serve to anchor data literacy concepts and lead to a better use of such technology”, seems particularly relevant to AI-DL. We are expecting to work with teachers already familiar with GenAI tools and in the process deepen their data literacy competences.*

Discuss pedagogies and implementations

Multiple projects referenced elements of constructivism and social constructivism, with frequent mention of different active learning and design-oriented pedagogies. Data literacy education is moving towards “far-reaching, power-related, collective, organisational, relational and distributed data practices” (Dander & Macgilchrist, 2022). Such pedagogies would involve setting up trusting classroom communities which encourage open student discussion and collaboration (Wilkerson et al., 2025).

Therefore, giving examples of different activities, time and space for discussing what they would mean for the teacher should be part of any data literacy intervention for teachers. Topics discussed should include setting up constructivist and collaborative learning spaces (Dangol & Dasgupta, 2023, Kahila et al., 2024), creating open ended but structured tasks (Ansyari et al., 2020, Dangol & Dasgupta, 2023, Kahila et al., 2024), negotiating scaffolding and student agency (Dangol & Dasgupta, 2023, Lee et al., 2021, Moon et al., 2023, Vartiainen et al., 2020), planning resources and technology aids (Ansyari et al., 2020, Kahila et al., 2024,Pellegrino & Antelmi, 2023, Vartiainen et al., 2020), implementing assessment for learning (Kahila et al., 2024, Kim et al., 2022), facilitating projects and mediating sensitive discussions (Ansyari et al., 2020, Dangol & Dasgupta, 2023, Kahila et al., 2024, Wilkerson et al., 2025), all within the schedules, routines and goals of traditional schooling.

Key Takeaway

Comment
- *The idea of "trusting classroom communities" again seems very relevant for AI-DL, as we are keen to create spaces where teachers and students can engage in robust discussion around data literacy issues in the context of GenAI use.*

Problems encountered during data literacy education

Defining data and data literacies

There is a fundamental problem in approaching data literacy programs methodically since there are a lot of differences in how different people perceive data and what constitutes data (Dangol & Dasgupta, 2023). Changing definitions and scope of data literacies with time and context, a lack of a clear list of constituent knowledge and skills, and their optimal integration into current systems, hamper curriculum designers (Henderson & Corry, 2021, Nwagwu, 2025). It is not clear if a standard instrument is possible for assessing data literacies either (Ansyari et al., 2020, Nwagwu, 2025). Educational actors need new flexible approaches for working with a shifting concept that will not be pinned down. At the same time, new approaches should be firmly grounded in education research (Wang and Lester, 2023). Co-constructing data literacy programs, discussed earlier, could be a good start towards solving this problem.

Key Takeaway

Comment
- *This challenge, re defining data and data literacies, has surfaced many times in this paper. The suggestion of co-constructing data literacy programs, such as AI-DL, with a range of stakeholders might be a good way forward in our context, as little has been done in this specific context to date re data literacy.*

Time, infrastructure and Institutional support

Change in data literacies and the technologies that demand them and enact them, calls for changes in various levels of the educational system. As with the case of digital literacies, at the macro level, curriculum should be updated and professional learning programs designed and implemented (Ilomäki et al., 2023, Kahila et al., 2024, Kim et al., 2022, Zhang et al., 2024).

At the meso level, school wide changes of educational practices, priorities and attitudes, classroom setups and scheduling, changes in available infrastructure and consistent collaboration between

Co-funded by the European Union

teachers needs to take place (Ilomäki et al., 2023, Kahila et al., 2024, Kim et al., 2022, Matuk et al., 2022, Schildkamp, 2019, Zhang et al., 2024). Classroom set ups are particularly important for constructionist approaches recommended for data literacy education (Dangol & Dasgupta, 2023, Kahila et al., 2024, Kim et al., 2022).

At the micro level, the teacher level, update in knowledge, skills and attitudes, practice in trying out and iterating new pedagogical approaches and technological tools are needed (Ilomäki et al., 2023, Zhang et al., 2024). This is particularly true for AI technology since the output and interactions differ from student to student (Kim et al., 2022).

When it comes to technology, there are also unclarified legal issues, ethical dilemmas and non-availability of those tools in the local language (Kim et al., 2022, Mouta et al., 2023, Saddiqa et al., 2019). An overall environment of trust in all levels is needed to implement a successful data literacy education (Amels et al., 2019, Schildkamp, 2019).

Key Takeaway

Comment

- *These ideas appear relevant to AI-DL in relation to implementing change and introducing curriculum change, an issue that was reviewed elsewhere in this report. As is the point that new pedagogical approaches and tools are needed, particularly in the case of AI.*

Roadblocks to interdisciplinary instruction

The main obstacle to interdisciplinary approaches, particularly those that integrate art, is the current organisation of schooling into "disciplinary silos" (Matuk et al., 2022). Apart from each separate discipline having its own subject goals, assessment and accountability structures (Lee et al., 2022, Matuk et al., 2022), some subjects are often prioritised over others especially with respect to instructional time (Matuk et al., 2022). Insisting interdisciplinary instruction is one of the key components of implementing data literacy education further adds to the time taken for a successful implementation and would ask for setting up of spaces and projects, gathering technological resources, and a synching of classroom norms and routines of different teachers (Matuk et al., 2022, Wilkerson et al., 2025). Before asking for the time and effort from teachers for such an approach, they must be convinced of the pedagogical purpose of data literacy and the value of interdisciplinary approach as a means to achieve it (Cowie et al., 2021, Matuk et al., 2022, Wilkerson et al., 2025).

Key Takeaway

Comment

- *"Before asking for the time and effort from teachers … they must be convinced of the purpose of data literacy and the value of an interdisciplinary approach", seems extremely relevant to AI-*

DL. The project must convince teachers of the value of developing student data literacies, using a cross-curricular approach, as being hugely important.

Roadblocks to constructionist approaches

The structure of schooling around subjects, curriculum, textbooks and exams also impede constructionist approaches to data literacies where students have the agency to engage in open ended activities by which they explore and make meaning with data (Dangol & Dasgupta, 2023, Saddiqa et al., 2019). But there are ongoing reforms in schooling worldwide in this direction (Wilkerson et al., 2025).

Activities that ask students to use data for their own ends and advocate for change in their schools and communities, encounter the usual resistances and obstacles to such activism. The time and sustained effort to make real changes happen might be too much to ask from classroom settings without external support (Lee et al., 2022).

On top of these constraints, curiosity and creativity, which are important not only to constructionist approaches but are also essential for adapting and coping up with changing technologies, seem to be routinely underplayed in school education (Ilomäki et al., 2023).

Key Takeaway

Comment

- *This section suggests finding pedagogical approaches that encourage curiosity and creativity, when using changing technologies and this is something we are going to attempt in AI-DL.*

Navigating ethical and legal issues associated with technology

Problems in using artificial intelligence and other data-based technologies, include, but are not limited to issues related to accessibility, privacy, surveillance, transparency, and environmental footprint (Lee et al., 2024). For generative AI chatbots and agents, there is the danger of unreliable outputs ("hallucinations" in the case of chatbots), unequal representation of languages and cultures and concerns about copyright and intellectual property (Lee et al., 2024). The policies and practices of the companies that churn out these technologies are inscrutable and hidden behind proprietary and commercial restrictions.

All this makes the use of such technologies complicated and data literacy competencies that deal with understanding these systems and visualising their operations impractical (Vartiainen et al., 2020). There are very few open alternatives when working with big datasets making commercial intermediaries indispensable in making raw data accessible to teachers and students (Dander & Macgilchrist, 2022).

Co-funded by the European Union

Before implementing a data literacy curriculum, the policies for using such systems in the classroom must be laid down clearly at the governmental and institutional level and regulation like GDPR decoded since it has been shown to be too complex when used in education (Kitto & Knight, 2019 as cited in Mouta et al., 2023).

Key Takeaway

Comment
- *The issue of using AI/GenAI tools in schools raises many challenges for schools, as cited above. The AI-DL project needs to be aware of these issues and ensure that our approaches are compliant with existing policies.*

Normalising datafication and standardisation

There are some inherent pitfalls in implementing data literacy education. Focusing excessively on data could establish data-based analyses as "undisputed cognitive authority" (Mertala, 2020) leading students to overestimate the accuracy and validity of such results. Care should be taken to present data analysis in context, with all its limitations and avoid the rush to embrace data (Lee et al., 2021, Mertala, 2020). It is to be borne in mind that the very nature of these technologies, due to how they train and optimise their outputs, is to minimise pluralities and herd its users towards homogenisation and standardisation (Lee et al., 2021, Mertala, 2020).

Also, it is important that the school policies and approach to technology usage are coherent with the values and attitudes promoted by the data literacies approach, since "the everyday data-related practices of contemporary education can be approached as functional forms of data literacy education" (Mertala, 2020). They have the power to naturalise and drive home the inevitability of data collection practices, despite a data literacy programme advocating for student privacy and agency (Mertala, 2020, Vartiainen et al., 2020). Indeed, some studies already show that students are sceptical of personal data collection but do not necessarily take steps to protect it (Israel-Fishelson et al., 2024).

Key Takeaway

Comment
- *The idea of students overestimating the accuracy and validity of their outputs is key to AI-DL, as is the suggestion that "school policies and approach to technology usage are coherent with the values and attitudes promoted" by a data literacies approach.*

Co-funded by the European Union

Limitations

This review included only open access papers found on Scopus using a limited set of key words. Other publications, including those that are open access but not available on this platform and those that refer to data literacy using other terms were not included. Also, projects that did not have associated papers are not part of this study. University papers which could have included interesting approaches to data literacy education that could have applied to schooling were not included either, owing to time and resource constraints. The same goes for papers written in languages other than English, which is unfortunate given that this is a European effort. These restrictions have limited the ideas and approaches that were considered in this study.

Conclusions

This review has catalogued and grouped competencies and common themes and sub themes associated with data literacies and data literacy education. In doing this, some meta elements that go into making of these competencies have become apparent. Whether dealing with disciplinary, personal or teacher specific competencies, a good DLE would include:

- Hands-on and interdisciplinary data analysis which would equip teacher and student to understand the elements, processes and tools that go into such analyses, both at small scale and large.
- Understanding mathematical and statistical tools and the working of data based technologies,
- Probing into the technical, socio cultural, economic and political context and implications of data processes, their ethics and environmental impact
- Taking a critical attitude, posing questions and looking for alternative datasets, processes and interpretations, often in collaboration with diverse stakeholders and experts from different disciplines
- Making personal connections and foregrounding personal and collaborative meaning making practices
- Engaging in self-reflective and metacognitive practices to both identify data based processes and tools are at work in daily life and to monitor their effects on self and others.
- Exercising agency in all stages of data literacy education starting with deciding the topic and medium to deciding against indulging in data based practices or rerouting them towards different ends

This work was done with the hope of preparing the terrain for the analysis of the impacts of Generative artificial technologies on data literacies and the design of data literacy initiatives to accommodate this impact. The possible lines along which further investigation could take place are:

Co-funded by the European Union

1. *How GenAI technologies and their usage can change the character of literacies?*

For example, it might explore:

- the lines between some literacies seem to have blurred further after the integration of GenAI into many digital technologies like search engines.
- the usage of GenAI technologies seemed to have impacted the relative importance of different literacies and how they are activated. Conventionally, mathematics and programming have mediated literacies linked to science, data, statistics, computation etc. Now, these could be mediated through text in the form of prompts. Information and media literacies play a role in parsing and understanding the outputs thus generated.
- Versatility, effectiveness and the ease of use of GenAI applications can disincentivise their users from directly engaging with data processes, predicting a certain loss of agency. On the other hand, GenAI technologies also seem to make many of these literacies more accessible to laypeople.
- After the internet era helped put in place collaborative and socialised literacies, GenAI, with its chatbots and personalisation capabilities, seem to be a step towards either hyper-individualisation, or, in the form of filter bubbles, creation of islands of similar individuals.
- Multilingual and multicultural aspects of literacies should also be investigated in the light of GenAI's tendencies to standardise and homogenise using a distorted sample of one language and culture. Its multimodality, however, makes more types of expressions possible.

2. *What are the affordances provided by GenAI technologies for developing data literacy competencies?*

GenAI technologies can make data less abstract and facilitate how we "read" and understand it. This affects how we can use GenAI applications to find, generate or aggregate data, to make sense of it, process, explore or visualise it. Being multimodal and adaptable for different tasks, GenAI tools can also help us "write" and "express" with data. This affects how we can use it to make extrapolations and predictions, publish claims with visualisations. What was accessible to programmers and technicians is now available to novices, since data is doable by text.

The use of GenAI technologies for facilitating various parts of the inquiry process like finding and merging data, running simulations, creating documentation for datasets, preparing data for analysis, exploring and analysing data, creating creative visualisations, searching for alternative interpretations, making apps for prediction and presenting data analyses to a diverse audience should be investigated.

GenAI technologies could make working with programming languages like Python and R easier, by using GenAI augmented notebooks and programming environments. Such environments can also integrate multiple functionalities, avoiding the need for learning multiple languages and environments. Being versatile, Gen AI technologies also permit artistic visualisations of datasets.

Co-funded by the European Union

Being interactive, they help in bypassing the drawbacks of other programming and visual analysis tools.

3. *How GenAI technologies impact data literacy competencies?*

For one, data literacy would now involve a general understanding of GenAI technology, whether this technology is directly involved with schooling or is embedded in other commonly used technology. This would be more so if GenAI agents are used to automate decisions in DBDM systems and learning analytics. Different aspects of GenAI might have to be included in a data literacy curriculum depending on the context.

GenAI is trained, tested and fine-tuned based on huge amounts of data. Accessing them often include some multimodal input, multiple textual exchanges in the case of chatbots, and possible uploading of supplementary data. This again leads to output data or, in the case of agents, metadata of the resulting action. This whole process is also valuable data, which, along with user personal data like name, location, IP addresses etc. could be used as feedback.

The data that can be fed into these systems could take any form. A piece of someone's recorded voice, for example, could be used to read output text. This has extended what is usable and valuable data to include any content or process that is digitisable. Understanding all the ways these data can be used by present and future systems will be part of data literacies.

Regarding personal data literacy competencies, GenAI technologies facilitate analysis of different forms of data that were difficult to analyse before. Identifying such data is tied to its possible uses and reuses by LLMs and other programs, apps and agents that integrate LLMs. A basic understanding of how these systems work and are put together to make larger more sophisticated systems, especially with respect to what they can do with user provided data and with user personal data and how to evaluate and interpret their output is indispensable. Their effects on classroom dynamics, interpersonal relationships, personal well-being, security and agency should also be analysed, especially given the possibility of human-like interactions. How to go about scaffolding tasks that involve GenAI technologies and how to verify individual outputs and conversations should also be looked into. So are setting up other pathways for families who opt out of using GenAI.

4. *How GenAI technologies impact data literacies attitudes?*

To read between the data and read beyond data, critical thinking and collaboration with diverse stakeholders are necessary. The data literacy attitudes of collaboration as a way to access alternative perspectives and keeping up to date with respect to technology and pedagogy, and metacognition and self-reflection as a way to counteract deskilling and over dependence have become crucial in the context of GenAI.

As for criticality, it has been a chief motif in literacies, education and democracies. But how does it play out in a classroom using GenAI, especially one that is using it to work towards data literacy?

There are a few discrepancies to work out here. How to aid individual critical thinking in a constructive way inside of the classroom which depends on rules and tight structures for functioning? What is the role of the teacher in facilitating constructive criticism, sharing of doubts and questioning in the classroom while covering all the learning objectives in limited time? Given critical thinking depends on a good grasp of content, context and processes, how does it happen with generative AI technologies which produce information hiding the very processes the data literacy learner is supposed to engage in (Costa & Murphy, 2025) ? How to focus on teasing out the processes when it would involve a lot more work than abiding to LLMs and knowledge summaries as epistemic authorities (Chen, 2025)? Is critical thinking possible if there is only limited agency to follow up thinking with action?

5. *How to maintain focus on data literacies with all that GenAI technologies bring into the equation?*

The affordances of GenAI tools for data literacy are non-negligible, making real data work with large datasets possible without the corresponding training and cognitive leap. However, if the teacher or student is not already initiated to data work using simpler datasets, the associated techniques, processes and analyses that are supposed to be understood as part of DLE would only get more obscure. In addition, this knowledge is a basic requirement for critically analysing the generated output and engaging in follow up prompting to get correct results. Data literacy is attained only when the techniques behind data work are understood. If not, we are speaking of media and information literacies, not data literacies.

6. *How to resolve ethical and environmental issues with GenAI technologies:*

We are becoming more and more inured to ethical and environmental issues associated with technological use. The problems of accessibility, inequity, lack of transparency and accountability associated with GenAI are not trivial issues. The environmental costs of training and using GenAI are starting to be felt even in developed countries in the form of water scarcity and rising electricity costs (Bashir et al., 2024).

Any intervention using GenAI tools should take care not to normalise trivial use of such power hungry tools. Models trained by governmental agencies with open curated datasets and full documentation of associated processes would go some way in addressing lack of transparency and corporate influence. Models fine-tuned for pedagogical use could facilitate safer educational use.

Co-funded by the European Union

Next Steps

Data literacy is an evolving concept that is very closely linked to the context in which teachers and students are interacting with data and tools, such as GenAI. The literature reviewed identified a wide range of definitions and approaches to developing data literacies, in a range of contexts. However, there was minimal reference to AI and or GenAI and it appears that this area requires further exploration, and the AI-DL project can assist in furthering our understanding in relation to this specific context. This review paper has identified a number of questions that could be further explored with teachers, school leaders and students.

These and other questions should be further explored using a co-construction approach, as suggested in the project proposal and by many of the papers reviewed in this paper. The ideas garnered from reviewing the selected academic literature should be considered alongside the findings from the review of frameworks, European projects and existing curricular policies and practices, to inform the teacher training activities and the emerging AI-DL Framework.

Co-funded by the European Union

3. Review of existing frameworks

3.1 Task Overview and Approach

The second element of the desk review was to analyse existing and upcoming frameworks, and this task was led by INDIRE, with support from H2 Learning.

The proposal stated that:

> *"In parallel we will also review other frameworks, such as DigComp 2.2, UNESCO's AI Framework, Framework of Competences for Democratic Culture (RFCDC) and consider how we can amalgamate key competencies into a new data literacy framework."*

The review set out to:

- Capture a range of existing data literacy definitions
- Capture the competences that are listed to data literacy and GenAI.
- Compare and contrast the competences in relation to key similarities and differences

The output from this analysis was then fed into wider discussions on developing an enhanced framework that will empower school leaders, teachers, and students to engage in critical discussions about data literacy within the context of GenAI.

This work commenced in June and was completed by mid-July 2025. It should be noted that this task was completed before the publication the DigComp 3.0 Framework (Cosgrove & Cachia, 2025).

This task reviewed the following frameworks:

- DigComp 2.2 (Vuorikari et al., 2022) and available drafts of the emerging DigComp 3.0 Framework (Cosgrove & Cachia, 2025)
- UNESCO AI Frameworks for Teacher & Student (UNESCO, 2024)
- Framework of Competences for Democratic Culture (Council of Europe, 2016)
- The AILit Framework (OECD, 2025)

Terminology and scope note: The reviewed frameworks use different labels and scopes. Some address 'AI' broadly (including non-generative approaches), while others refer explicitly to GenAI. In this analysis, we treat 'AI' as the umbrella category and interpret framework statements about AI as applicable to AI-enabled practices more generally. Where the framework does not explicitly specify GenAI, we extend the interpretation to include GenAI as a prominent current instantiation of AI in educational practice.

Co-funded by the European Union

3.2 Key findings

This section outlines the work carried out and the implications of this work for the wider project. This report was developed by INDIRE, and the lead author is Jessica Niewint-Gori.

A. Capture the data literacy definitions

The initial task was to review how each of the referenced frameworks defines literacy. Table 2 below shows that their treatment of literacy varies. In some frameworks they offer a precise, operational definition (e.g., DigComp 2.2), others embed data literacy within broader competence areas (e.g., UNESCO), and earlier frameworks (e.g., CDC, 2016) refer to adjacent ideas without naming data or AI literacy explicitly. The review team extracted the exact wording from each framework where a definition was provided and, where it was not, derived a concise formulation from the framework's descriptors and scope notes. For consistency, the team normalised terms around recurring components, such as articulating information needs, searching and navigating data, organising and managing it, and critically evaluating quality and relevance.

Framework (year)	Data literacy	AI / Gen-AI literacy
Competences for Democratic Culture (CDC) (2016)	No standalone definition. Data-related abilities are summarized under *media- and digital-media literacy* and are part of the broader "Knowledge and critical understanding of the media" competence.	No AI-literacy definition. The model keeps a focus on democratic culture.
DigComp 2.2 (2022)	Explicit definition in "Information & Data Literacy": articulate information needs; search, access, navigate, organise, store, manage and *critically evaluate* data, information and content.	No definition. AI appears as a *transversal* theme across knowledge, skill and attitude descriptors.
UNESCO *AI Competency Framework for Teachers* (2024)	Does not define the term, but states that *AI data literacy is a cross-cutting, progressive component* that grows across the three proficiency levels (Acquire, Deepen, Create).	Structures AI-literacy through five interconnected strands (human-centred mindset, ethics, AI foundations & applications, AI pedagogy, AI for professional development) with the three proficiency levels (Acquire, Deepen, Create).
UNESCO *AI Competency Framework for Students* (2024)	Assumes basic data literacy is already "acquired and fully integrated"; therefore, it is not defined again.	Aims for students to become *effective, ethically aware users and co-creators* of AI through four strands (mindset, ethics,

Co-funded by the European Union

		techniques, system design) across the levels (Understand–Apply–Create)
DigComp 3.0 draft (2025)	DigComp 2.2 data-literacy def. and frames it as *inter-related* with AI competence, media literacy, programming.	AI competence definition: "informed, confident, critical and responsible use of, and engagement with, AI systems ... for learning, work and participation," seen as a blend of knowledge, skills and attitudes.
***AI-Lit* draft (2025)**	Implicit Definition: knowledge of data provenance, bias, ethics and environmental impact; skills to collect, analyse, evaluate and communicate data; attitudes to work with data "responsible, inclusive and future proof".	Formal definition: "technical knowledge, durable skills and future-ready attitudes required to thrive in a world influenced by AI," enabling learners to engage with, create, manage and design AI while critically judging its benefits, risks and ethics.

Table 2, Frameworks and definitions of Data Literacy and AI/GenAI Literacy

Over the last decade, the conversation around "what counts" as literacy has evolved from tidy lists of skills to richer, more contextual understandings. In 2022, DigComp 2.2 set a clear anchor for data literacy with a concise, operational definition: "To articulate information needs, to search for data, information and content in digital environments, to access and navigate between them. To organise, store, manage and critically evaluate them." In practice, this frames data literacy as a disciplined cycle of inquiry - knowing what you're looking for, finding it efficiently, keeping it in order, and making sound judgments about quality and relevance. It is the craft of the input side: sources, structure, storage, and scrutiny. By 2025, attention has switched focus decisively towards artificial intelligence (AI), and especially GenAI, as systems that not only process information but co-produce it with us. The AI-Lit draft framework captures this shift with an expansive definition: "AI literacy represents the technical knowledge, durable skills, and future-ready attitudes required to thrive in a world influenced by AI. It enables learners to engage, create with, manage, and design AI, while critically evaluating its benefits, risks, and ethical implications." Here the emphasis moves beyond handling data toward interacting with AI, which includes GenAI systems, co-creating outputs, and weighing social and ethical trade-offs in real time.

It should be noted that not all frameworks state their terms with the same precision. Only DigComp 2.2, the DigComp 3.0 draft (for AI), and AI-Lit provide explicit definitions. Other frameworks, such as theCDC and UNESCO frameworks, embed the literacies inside broader competence areas or use them as transversal themes. DigComp and AI-Lit pin down definitions for measurement and design, UNESCO and CDC integrated the capacities into larger areas such as teaching, learning, and

Co-funded by the European Union

citizenship. Alongside definitions comes a shift in emphasis. Ethical judgment, bias mitigation, and human-centred design appear only at the margins in classic data-literacy definitions, which focus on information quality and management. In the AI-focused frameworks, these elements move to the foreground. The conversation has matured from "Is the data correct and well kept?" to "How do these systems shape decisions, power, and participation - and what responsibilities follow?"

Progression models tell a similar story. UNESCO's Teacher and Student frameworks and DigComp include progression levels (e.g., Acquire–Deepen–Create; Understand–Apply–Create), inviting curricula to scaffold from awareness to design. AI-Lit and CDC, by contrast, present strong single-point statements without explicit level progressions, which can be useful for vision and alignment, and less prescriptive for sequencing. There are also differences in naming. DigComp 3.0 opts for "AI competence," signalling alignment with a competence-based approach, while AI-Lit and UNESCO retain "literacy," preserving links to broader cultural and civic aims and literature still avoid the narrower label "GenAI literacy," implying it rather than declaring it.

Taken together, these trends chart a movement: from discrete literacies to integrated competence models; from tool-focused checklists to human-centred practice; and from static definitions to progressive, contextual frameworks that can flex with technology and setting. However, a gap remains as none of the reviewed frameworks yet offers a fully integrated definition that spans the continuum from foundational data work to generative-AI design and deployment. A pragmatic path forward might be to combine the strengths of exiting framework definitions: by for example pairing DigComp 3.0's articulation of AI competence with AI-Lit's sustained attention to the data thread across the AI lifecycle.

B. Capture the competences that are listed for data literacy and GenAI.

To compare the single skills and competencies defined in the single frameworks it was necessary to define new domains, and these are captured in Table 4 below.

Domain	CDC	AILit	DigComp 2.2	DigComp 3.0	UNESCO Teacher	UNESCO Students
A. Core Knowledge of AI & Data	✗	✓	✗	✗	✓	✓
B. Data Stewardship & Model Documentation	✗	✓	✓	✓	✗	✗
C. Cognitive & Analytical Domain	✓	✓	✓	✓	✓	✗
D. Creative & Innovative Domain	✗	✓	✗	✗	✓	✓

Co-funded by the European Union

E. Ethical, Self & Social Domain	✗	✓	✗	✗	✓	✓
F. Interaction, Collaboration & Communication	✗	✓	✗	✗	✓	✗

Table 2 Domains and competences listed for data literacy and GenAI

C. Framework Coverage Analysis

In this section we map how each framework covers the six competence domains for Data Literacy and GenAI Literacy, translating the comparison tables into concise prose. The analysis, in Table 4 above, highlights where coverage is comprehensive, partial, or absent across CDC (2016), DigComp 2.2/3.0, UNESCO Teacher/Student, and AILit, identifying consistent strengths and recurring gaps and showing the shift from pre-GenAI to post-2022 integrated approach or the shift from a time when Gen-AI was not in popular use (i.e. November 2022) to a time afterwards (post November 2022).

Data Literacy

A review of the frameworks, see Table 5 below, found that coverage of data literacy competencies description across the six domains is uneven, reflecting each framework's original purpose and development context.

The AI-Lit framework is the only one that explicitly addresses all six domains. The others cluster differently: DigComp 2.2/3.0 focuses on "input-side" data practices - particularly data stewardship, documentation, and critical evaluation - while the UNESCO Teacher and Student frameworks prioritise creative-pedagogical use and ethical responsibility, treating data literacy as a cross-cutting prerequisite rather than a distinct competence area.

At the domain level, Core Knowledge of AI & Data (A) appears mainly in AI- and pedagogy-focused frameworks such as AI-Lit and the UNESCO Teacher/Student frameworks, and is weak or absent in CDC and DigComp's data strand, which centres on information handling rather than AI foundations. Data Stewardship & Model Documentation (B) is strongest where data management and documentation are explicit priorities, notably in DigComp and AI-Lit, but receives little attention in UNESCO's competency strands. Cognitive & Analytical (C) is the most consistently represented domain across all frameworks, and CDC appears primarily here, reflecting its pre-generative AI focus on analytical reasoning.

The domains most connected to classroom transformation - Creative & Innovative (D), Ethical/Self/Social (E), and Interaction/Communication (F) - are most visible in UNESCO and AI-Lit. This suggests that data literacy is increasingly understood as a practice embedded in learning design, responsibility, and communication, rather than simply as data handling

Domain	CDC	AILit	DigComp 2.2	DigComp 3.0	UNESCO Teacher	UNESCO Students
A. Core Knowledge of AI & Data	✗	✓	✓	✓	✓	✓
B. Data Stewardship & Model Documentation	✗	✗	✗	✓	✓	✓
C. Cognitive & Analytical Domain	✗	✓	✓	✓	✓	✓
D. Creative & Innovative Domain	✗	✓	✓	✓	✓	✓
E. Ethical, Self & Social Domain	✗	✓	✓	✓	✓	✓
F. Interaction, Collaboration & Communication	✗	✓	✓	✓	✓	✓

Table 3 Framework Coverage Analysis: Data Literacy

A. Core Knowledge of AI & Data

Core Knowledge of AI & Data is covered in AILit and both UNESCO frameworks; it is absent in CDC and in the data-literacy strands of DigComp 2.2 and 3.0. This indicates that "core knowledge" for data/AI is articulated mainly in the AI-oriented or pedagogy-oriented frameworks, while DigComp's data-literacy focus remains on information handling rather than conceptual AI knowledge. The evidence in the Annex 7B shows that UNESCO treats 'AI foundations & applications' (Teacher) and 'techniques & applications' (Student) as the key indicators of foundational AI/data understanding; by contrast, DigComp's data-literacy strand foregrounds information/data handling rather than conceptual AI knowledge.

B. Data Stewardship & Model Documentation

Data Stewardship & Model Documentation is present in AILit and both DigComp versions; not covered in CDC or the UNESCO Teacher/Student frameworks. Stewardship (searching, managing data; documenting model purpose/limits) is therefore strongest where data management is explicit (DigComp) or where the AI lifecycle is spelled out (AILit), but it is not foregrounded in UNESCO's competency strands. The appendix confirms DigComp 1.1/1.3 descriptors and AILit's documentation skills as the core references here.

C. Cognitive & Analytical Domain

Cognitive & Analytical is covered by CDC, AILit, DigComp 2.2, DigComp 3.0 and UNESCO Teacher; it is not covered for UNESCO Students. CDC (2016) maps primarily to this domain, reflecting its focus on

Co-funded by the European Union

analytical and critical thinking rather than on AI- and data-specific practices. In other words, CDC shows concentrated coverage in C but limited or no coverage across the other data-literacy domains.

D. Creative & Innovative Domain.

Creative & Innovative Domain are covered in AILit and both UNESCO frameworks; not present in CDC or DigComp's data-literacy strands. Creativity and innovation with (or enabled by) AI show up strongly where pedagogy and co-creation are emphasised (UNESCO) or where AI practice is holistic (AILit). UNESCO items on AI-enhanced pedagogy, tool customisation, and student design/prototyping substantiate this coverage.

E. Ethical, Self & Social Domain.

Ethical, Self & Social Domain are covered in AILit and both UNESCO frameworks; absent in CDC and DigComp's data-literacy strands. Modern frameworks typically frame data practice within ethics, human agency and social responsibility, whereas DigComp's data-literacy focus is more operational here. The report's narrative underscores the stronger ethical emphasis in newer AI-linked frameworks.

F. Interaction, Collaboration & Communication.

Interaction, Collaboration & Communication are covered in AILit and UNESCO Teacher; not covered in CDC, DigComp's data-literacy strands, or UNESCO Students. This suggests interactional competences (working with others and with AI; explaining AI's role) are recognised where teaching practice and professional collaboration are central. Related descriptors on collaboration/communication with AI appear in the evidence base.

GenAI Literacy

Coverage of AI Literacy across the six domains (see Table 6 below) is comparatively broad and more uniform than in Data Literacy, reflecting the fact that contemporary GenAI-oriented frameworks were developed with an explicit intention to integrate technical understanding with responsible use and classroom-relevant practices. With the exception of CDC, which reflects a pre-generative-AI context and therefore provides only limited alignment, the remaining frameworks cover most domains in a balanced way, pairing foundational knowledge and analytical judgement with creativity, ethics, and interactional competences. Even if the frameworks itself show now further distinction between AI-Literacy and GenAI-Literacy. At the domain level, Core Knowledge (A), Cognitive & Analytical (C), and Ethical/Self/Social (E) are the most consistently present: users should understand what GenAI is, evaluate outputs critically, and act with attention to human agency, fairness, and responsibility. Creative & Innovative (D) and Interaction/Collaboration/Communication (F) are also widely represented, consistent with a shift from "tool use" to co-creation, dialogue, and communicative transparency about AI-supported work. The main gap concerns Data Stewardship & Model Documentation (B), which remains less consistently specified than the other domains. This suggests that even where GenAI literacy is framed comprehensively, lifecycle-oriented practices, such as

Co-funded by the European Union

documentation of inputs, provenance, limitations, and model-related constraints - are more often treated implicitly or distributed across adjacent areas rather than articulated as a stable competence set.

Domain	CDC	AILit	DigComp 2.2	DigComp 3.0	UNESCO Teacher	UNESCO Students
A- Core Knowledge of AI & Data	✗	✓	✓	✓	✓	✓
B-Data Stewardship & Model Documentation	✗	✗	✗	✓	✓	✓
C-Cognitive & Analytical Domain	✗	✓	✓	✓	✓	✓
D-Creative & Innovative Domain	✗	✓	✓	✓	✓	✓
E-Ethical, Self & Social Domain	✗	✓	✓	✓	✓	✓
F-Interaction, Collaboration & Communication	✗	✓	✓	✓	✓	✓

Table 4 Framework Coverage Analysis: GenAI Literacy

A. Core Knowledge of AI & Data

All modern AI and digital frameworks, such as AILit, DigComp 2.2, DigComp 3.0, and UNESCO Teacher/Student, cover foundational knowledge of AI concepts and applications. CDC is the exception in this case, reflecting its development before the arrival of GenAI. Overall, Domain A, "core knowledge" is a shared baseline across all the contemporary frameworks.

B. Data Stewardship & Model Documentation

Data Stewardship & Model Documentation coverage is concentrated in DigComp 3.0 and the UNESCO frameworks; AILit and DigComp 2.2 do not foreground stewardship/documentation for GenAI, and it is absent in the CDC framework. This pattern indicates that, for GenAI, stewardship competences are more explicitly articulated where lifecycle, governance, and educational implementation are emphasized.

C. Cognitive & Analytical Domain

The AILit framework, both DigComp versions, and UNESCO Teacher/Student all include critical evaluation, problem decomposition, and decision-making in relation to AI. The CDC framework does not reference these areas. Thus, these analytical competences are widely recognised as being essential for judging GenAI outputs and integrating them into tasks.

D. Creative & Innovative Domain

All modern frameworks address creative use and design with GenAI by highlighting aspects such as co-creation, prototyping, and experimentation, while the CDC framework does not. Creativity is treated as a central characteristic of GenAI literacy rather than a peripheral skill.

E. Ethical, Self & Social Domain

Ethical use, human agency, responsibility, and attention to social impact are represented in AILit, the UNESCO Teacher/Student frameworks, and though framed more broadly as digital citizenship and responsible digital practice also in DigComp 2.2/3.0. By contrast, CDC does not provide AI/GenAI-specific ethical guidance, which is consistent with its earlier, non-AI scope. Overall, the contemporary frameworks tend to align around a rights- and responsibility-oriented understanding of (Gen)AI practice, while differing in emphasis (e.g., explicit AI ethics and governance versus wider ethical participation, safety, and agency in digital environments).

F. Interaction, Collaboration & Communication

AILit, DigComp 2.2/3.0, and UNESCO Teacher/Student all recognise the following elements: collaboration with AI and with peers, transparency about AI use, and communication of AI's role. Once again, this this is absent from the CDC framework. Interactional competences are framed as integral to effective and accountable GenAI use.

Key Observations

Unlike Data Literacy, where coverage develops unevenly across frameworks with different original purposes, the more recent AI- and digital-competence frameworks that have been updated or published in the 2022–2025 period show a clear pattern of incorporating competences that are relevant to GenAI use. The Framework of Competences for Democratic Culture (CDC, 2016) does not mention AI or GenAI, which is consistent with its pre-GenAI development context and illustrates how quickly the field has shifted over less than a decade. A distinction is important: not even all newer frameworks reference "GenAI" explicitly. Some frameworks (e.g., AI-Lit and the UNESCO teacher/student documents) include AI content that is readily applicable to GenAI practices and, depending on the document wording, may also reference generative systems more directly. Others (e.g., DigComp 2.2/3.0) primarily reference "AI" and responsible digital practice rather than naming GenAI as a category. In this analysis, we therefore treat explicit "AI" competence statements as covering GenAI where the descriptor concerns capabilities, risks, and practices that are characteristic of generative systems (e.g., content generation, evaluation of outputs, transparency about AI-assisted work), while acknowledging that the level of GenAI-specificity varies across sources. Across the reviewed set (excluding CDC), coverage is broad across competence domains; however, topics as "Data Stewardship" and "Model Documentation" remain comparatively less visible. The analysed

Co-funded by the European Union

frameworks prioritise user-facing competences, such as critical evaluation of AI outputs and creative or pedagogical use, over less visible, but consequential, competences related to data preparation, provenance, documentation, and transparency.

Additionally, frameworks, such as AI-Lit and UNESCO, often address competencies in a holistic transversal way instead of specific necessary skills, linking technical abilities with ethical awareness and social responsibility. The UNESCO teacher and student frameworks, in particular, focus on pedagogical integration, positioning AI-enabled tools (including generative tools) as components of teaching, learning, and professional development. Another recurrent feature is an emphasis on human agency: GenAI literacy is framed not only as operating tools, but also as adapting and contextualising their use to specific purposes, which aligns with a shift from passive use toward active appropriation and co-creation. While Data Literacy stays a foundational competence, GenAI Literacy extends it into interaction, creativity, and co-creation with AI systems. Across contemporary frameworks there is a strong emphasis on ethical responsibility and human oversight, supporting an interpretation of GenAI literacy as both a technical and a civic competence - provided that the AI-to-GenAI extension is made explicit and applied consistently in the analysis.

Compare and contrast what they say in terms of common elements and differences

This section compares and contrasts the reviewed frameworks by identifying their shared elements and their distinctive emphases, highlighting where they converge on core competence areas and where they diverge due to scope, target audience, and development context.

Table 7 and 8 summarise this comparison: the first presents the main points of convergence across frameworks, while the second highlights the most salient differences in focus, coverage, and domain-level emphasis.

Missing both DL **AND** GenAI	Only one element (DL **OR** GenAI)	DL **AND** GenAI represented

Domain	CDC	AILit	DigComp 2.2	DigComp 3.0	UNESCO Teacher	UNESCO Students
A-Core Knowledge of AI & Data			only GenAI	only GenAI		
B-Data Stewardship &Model Documentation		only DL	only DL		only GenAI	only GenAI
C-Cognitive & Analytical Domain	only DL					only GenAI
D-Creative & Innovative Domain			only GenAI	only GenAI		

E-Ethical, Self & Social Domain			only GenAI	only GenAI		
F-Interaction, Collaboration & Communication			only GenAI	only GenAI		only GenAI

Table 5 Comparison of DL and GenAI Competence Coverage

Domain	Data Literacy Coverage	GenAI Literacy Coverage	Gap
A. Core Knowledge	3/6 frameworks (50%)	5/6 frameworks (83%)	+33% for GenAI
B. Data Stewardship	3/6 frameworks (50%)	3/6 frameworks (50%)	Equal coverage
C. Cognitive & Analytical	5/6 frameworks (83%)	5/6 frameworks (83%)	Equal coverage
D. Creative & Innovative	3/6 frameworks (50%)	5/6 frameworks (83%)	+33% for GenAI
E. Ethical & Social	3/6 frameworks (50%)	5/6 frameworks (83%)	+33% for GenAI
F. Interaction & Communication	2/6 frameworks (33%)	5/6 frameworks (83%)	+50% for GenAI

Table 6 Data and GenAI Literacy Coverage

Data literacy and GenAI literacy are both grounded in critical thinking and ethical use, and both combine technical abilities with social and communicative competences. In practice, this means learners are expected to question sources and outputs, reason about risks and bias, and act responsibly toward others when they use data or AI tools. But the two literacies diverge in scope and emphasis.

GenAI literacy appears more uniformly across recent frameworks and is explicitly shaped by post November 2022 developments; its competences are tied to concrete AI capabilities (e.g., working with model outputs, giving feedback to systems, explaining AI use) and to human–AI interaction and co-creation in learning tasks.

By contrast, data literacy remains more fragmented and leans toward established “input-side” practices such as understanding data sources, quality, stewardship, and documentation. Data literacy also prioritises definitions of collecting, storing, managing, and analysing data. GenAI literacy is characterised as “output- and interaction-oriented”: it asks learners to judge AI outputs, iterate through prompts and feedback, and co-create artefacts with AI. These skills are typically newer and more situational than traditional data handling. Across the policy and practice-oriented documents reviewed, ethical safeguards and human agency (fairness, privacy, transparency, accountability) are positioned as baseline requirements for AI use. These documents also provide more actionable guidance than earlier frameworks, including practical recommendations for institutions and classrooms, which helps connect high-level principles to everyday teaching decisions.”

Data literacy frameworks (e.g., DigComp, or UNESCO) offer partial reference points for proficiency and progression, but the literature does not yet provide a single, formalised definition of “data literacy for generative AI." Our findings align with this situation and clarify the main coverage patterns: AI-Lit spans all six domains; UNESCO places stronger emphasis on creative, ethical, and pedagogical dimensions; and DigComp concentrates on stewardship and critical evaluation. Taken together, these patterns support the interpretation that the relationship between Data Literacy and GenAI Literacy is best understood as integration rather than as two separate literacies.

In the examined sources, data literacy remains foundational because it governs the conditions for trustworthy and explainable AI use (e.g., input quality, provenance, rights, and documentation). GenAI literacy typically builds on this foundation by extending competence into interactive, creative, and human-centred practice with generative systems (e.g., co-creation, communication, and responsible decision-making in context).

Proposal for a working definition

Data literacy for generative AI *is the capacity to acquire, document, govern, and critically reason about the data that GenAI systems consume or produce (including prompts, retrieval sources, training/tuning/feedback data where applicable, and generated outputs) in order to judge provenance, quality and representativeness, rights and consent/licensing, privacy and security, and downstream impacts - and to act on those judgements in the design, deployment, and use of GenAI-supported processes.*

Annex 7B contains additional analysis.

Co-funded by the European Union

4. Review of European projects

4.1 Task Overview

This task was led by the CNR, and the lead author was Manuel Gentile with support from Salvatore Perna, Giuseppe Città, Mario Allegra, and H2 Learning. In addition to reviewing EU projects this element of the desk review also considered grey literature in the field of data literacy and GenAI. When referring to grey literature we are referring to information that is not formally published through traditional publishing channels like academic journals or books.

4.2 Methodology

The CNR initially conducted an analysis of European projects on data literacy by retrieving the list of initiatives granted under the Erasmus+ program. Their analysis reveals a dynamic and evolving landscape of educational innovation and collaboration, with a total of 100 projects reviewed, the dataset provides a rich ground for understanding trends and priorities in this crucial area.

Figure 4 below illustrates the temporal distribution of European projects focused on data literacy, categorized by their funding year. It shows that interest and investment in data literacy projects have grown steadily over time, with a modest number of initiatives funded between 2016 and 2019, and a notable increase in project numbers since then**.** This suggests a rising prioritisation of data literacy within the European Union's educational agenda. This growth may be tied to broader digital transformation strategies, the emphasis on 21st-century skills, and the European Commission's commitment to digital education as articulated in action plans, such as the Digital Education Action Plan.

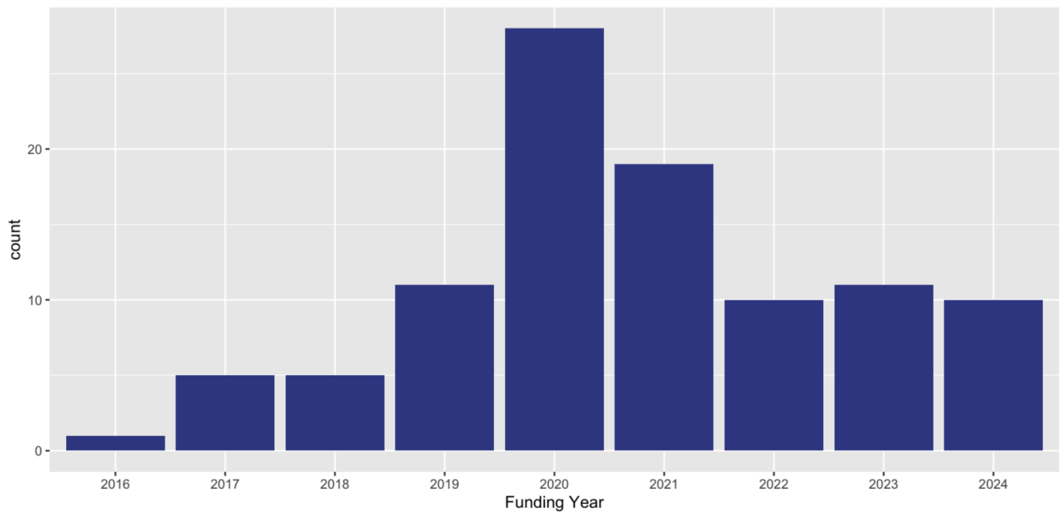


Figure 4 Data Literacy European projects by year

Co-funded by the European Union

Figure 5 organises projects according to their Key Action (KA) type as outlined in the structure of the Erasmus+ Programme[3]. It reveals a concentration of projects under Key Action 2 (KA2) Cooperation among organisations and institutions, which indicates that most data literacy projects are collaborative in nature, bringing together institutions from across the EU to co-develop and share innovative educational practices, resources, and tools. In contrast, Key Action 1 (KA1), Mobility of Individuals, has less representation, implying that while mobility and personal development are present, there is greater emphasis on institutional partnerships in contrast to individual exchanges in the context of data literacy.

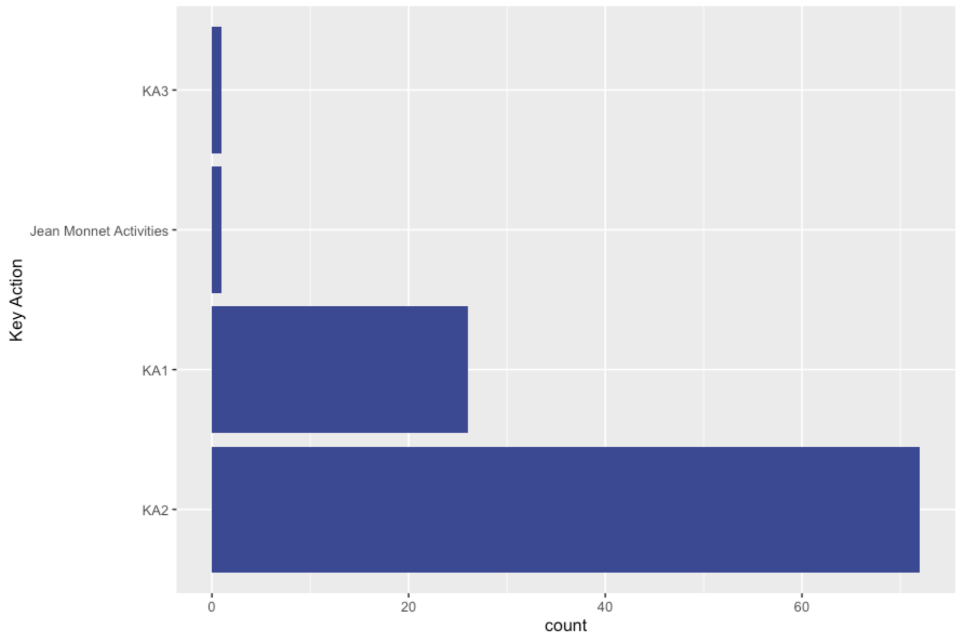


Figure 5 Data Literacy projects by Key Action

Figure 6 illustrates the details of the types of actions funded within KA2. In particular, the most common action types are:

- Strategic Partnerships for school education
- Strategic Partnerships for adult education

---

[3] The categories of action types reported in the following analyses follow the different organisations of the Erasmus+ programmes that have succeeded one another over time. In order not to alter the analyses, we have chosen to retain the original nomenclatures, not least because there is often no exact correspondence between the categories of the different structures. The last structure could be find at https://erasmus-plus.ec.europa.eu/programme-guide/part-a/priorities-of-the-erasmus-programme/structure

Co-funded by the European Union

This suggests that the integration of data literacy spans both general and vocational education systems. Projects targeting school education are likely aimed at embedding data literacy from an early age, fostering critical thinking and digital competence among young learners. Meanwhile, the focus on adults points to the importance of equipping the current and future workforce with data-related skills, aligning with labour market demands.

There are also instances of KA1 projects (Figure 7), particularly in the form of staff mobility, which likely support the professional development of educators in the field of data literacy. These initiatives help teachers and trainers gain exposure to new methods and tools, promoting a broader dissemination of data literacy practices across Europe.

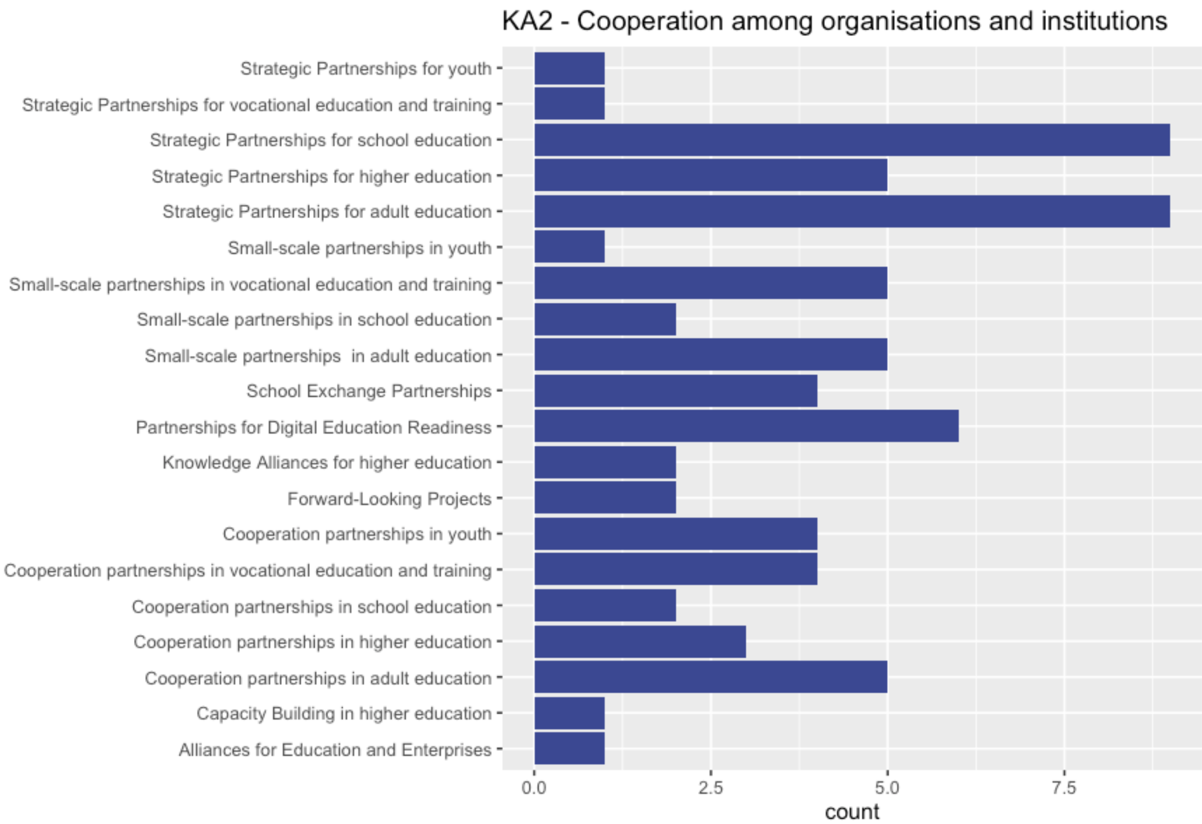


Figure 6 KA2 Data Literacy projects

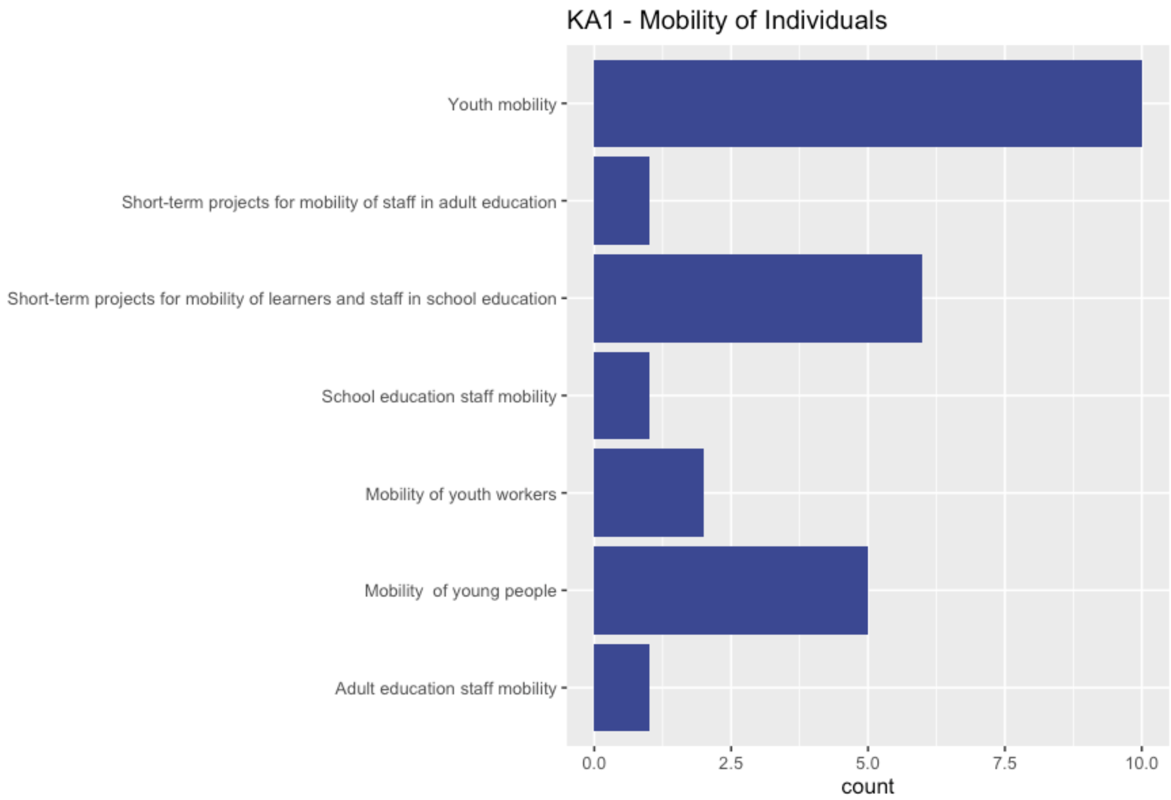


Figure 7 KA1 Data Literacy projects

Finally, the following word cloud offers a qualitative glimpse into the terms that appear most frequently across the identified 100 European projects on data literacy. Visually, the size and prominence of each word reflect its frequency and relevance within the project descriptions, objectives, and themes.

At the centre of the cloud, the most prominent terms include:

- “Data”, unsurprisingly, as the core concept of all projects analysed.
- “Digital” and “Literacy”, which reaffirms the overarching goal of equipping learners with the ability to understand, interpret, and use data effectively in digital contexts.
- “Education”, “Students”, and “Teachers”, which indicate the primary target groups and sectors involved.
- “Skills” and “Competences”, reflecting the emphasis on building practical, applicable abilities in working with data.
- Terms like “ICT”, “Tools”, “Platform”, and “Training” suggest a focus on technology-enhanced learning environments and the development of resources to support instruction.

Other terms that appear quite frequently, such as, “Critical”, “Thinking”, and “Media”, highlight the intersection of data literacy with critical media literacy and informed citizenship - key goals in combating misinformation and promoting digital resilience. The word cloud reinforces the themes

Co-funded by the European Union

captured in figures 5 and 6: themes such as education-driven, skill-oriented, and future-focused initiatives aimed at fostering data-literate individuals and institutions. It also underscores the multidisciplinary nature of these projects, which cut across education, technology, communication, and social inclusion.

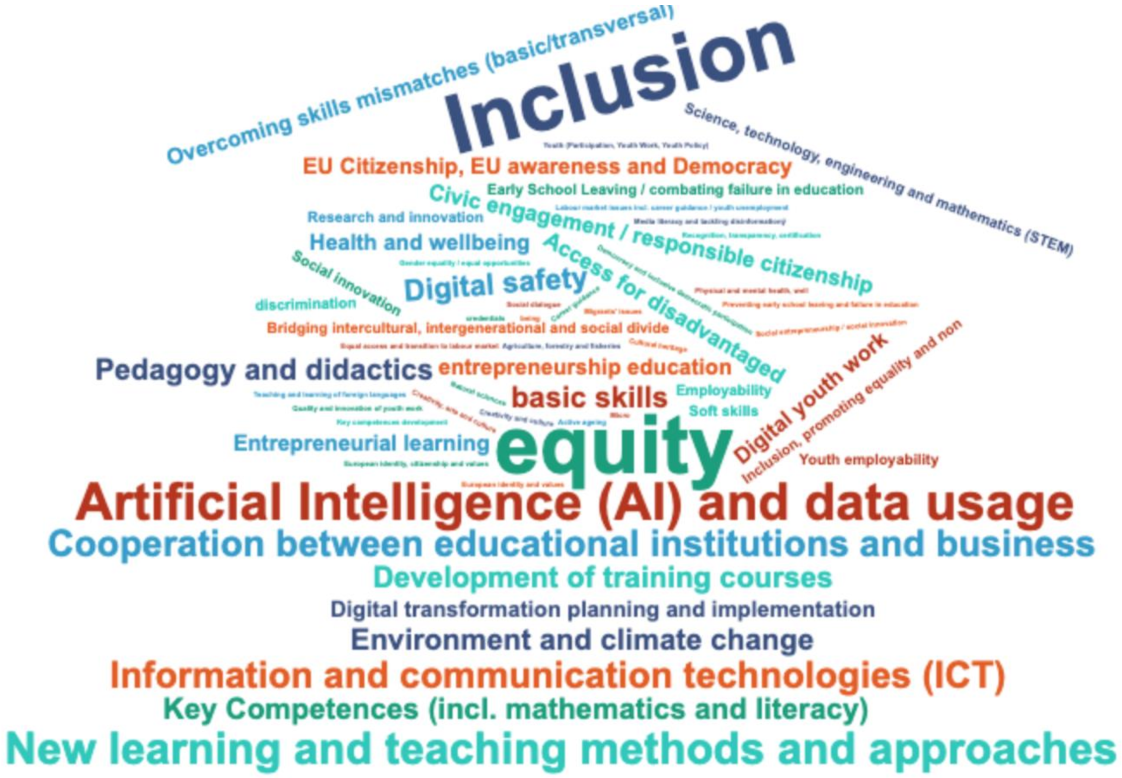


Figure 8 Words cloud of projects keywords

Overall, the analysis suggests that data literacy is gaining traction as a key educational goal across Europe, with a variety of actors involved and a noticeable increase in dedicated initiatives over the years. This growing commitment lays the foundation for more informed, empowered, and critically thinking European citizens.

The full CNR Analysis report (see Annex 7C) reviewed a subset of the data literacy projects collected. To select the projects, the CNR teams analysed the project cards reported on the Erasmus+ platform and the projects' websites with the aim of identifying the main results emerging from the projects' activities. From the analysis of the projects results', projects have been selected for a detailed analysis, including the Data Literacy for Citizenship (DALI) project, the ongoing Data Literacy for Upper Primary Schools project (DALI4US), the Teacher Training for Data Literacy and Computer Science Competences (TRAIN-DL), the MILES Project – MIL and Pre-bunking approaches for Critical thinking in the education sector and finally EDUCABILITY: Building the Capacity of Educators and Librarians in Information Literacy.

Co-funded by the European Union

In the following section we will present the key findings of the projects that most influenced the definition of the AI-DL Framework; these are the DALI project, DALI4US, TRAIN-DL and the MILES Project.

4.3 Key findings

The DALI Data Literacy for Citizenship Project

The DALI – Data Literacy for Citizenship project (2020-1-NO01-KA204-076492) was launched in 2020 and ends in 2023.

- The project aims to strengthen citizens' capacity to navigate an increasingly data-driven society. Its core objectives are to:
- Define the competences that characterise a data-literate citizen, based on a comprehensive and research-informed perspective.
- Support individuals in acquiring and further developing essential data-related skills, enabling them to understand, interpret, manage, and use data confidently.
- Increase participation in lifelong learning, using effective strategies for outreach, guidance, and learner motivation.

The project's principal achievement is the creation of the DALI Data Literacy Framework, a structured model describing the knowledge, skills, attitudes, and critical capacities required to engage meaningfully with data in everyday life.

The framework is organised into three main components:

- **Understanding Data** – encompassing knowledge, awareness, and critical thinking.
- **Acting on Data** – covering the practical skills needed to collect, manage, analyse, and communicate data.
- **Engaging Through Data** – highlighting the ability to use data for informed decision-making, civic participation, advocacy, and collective action.

A fourth transversal dimension, **Ethics & Privacy**, runs across all components, emphasising responsible and ethical data practices.

The full framework, including progressive levels of expertise (from basic to advanced) for each competence area, is available for download in English, Spanish, German, and Norwegian.

The DALI Data Literacy Framework is a comprehensive model describing what it means to be a data-literate citizen. It provides a structured set of competences, organised into three main components and three levels of progression, supported by a transversal ethical dimension. The framework is designed to guide the creation of learning activities, courses, and tools aimed at empowering citizens to understand, use, question, and act on data in everyday life.

Co-funded by the European Union

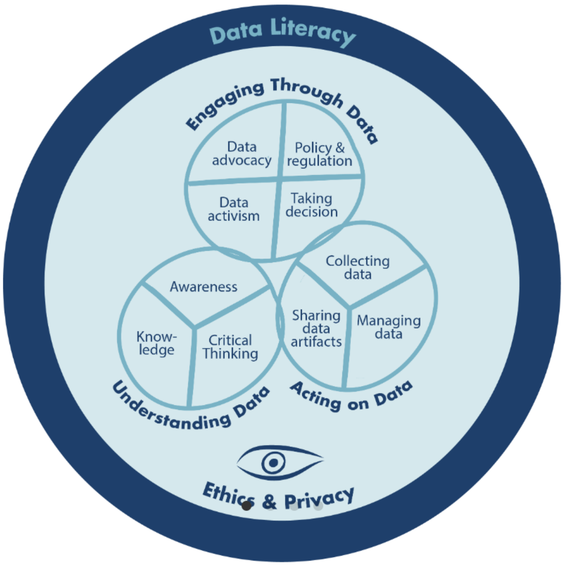


Figure 9 The DALI Framework

DALI4US – Data Literacy for Upper Primary Schools Project

The Erasmus+ project DALI4US (Data Literacy for Upper Primary Schools) is a three-year project, which is currently in progress that aims to equip primary school teachers with the necessary skills to effectively teach data literacy. The initiative is developing an evidence-based data literacy framework, including tools and resources, to help students interact critically with data. Among the main resources is OrangeEDU, an adapted version of Orange Data Mining software for primary education, which incorporates gamification elements. The project involves partners from Slovenia, Luxembourg, and Ireland, and focuses on professional training for teachers. School pilot projects are planned to evaluate and refine the effectiveness of the proposed solutions with the aim of preparing students for a data-driven society. The goal is to create a practical, developmentally appropriate data literacy framework for upper primary students.

The DALI4US data literacy framework: what it's built on

The framework was developed through a structured process: a needs analysis in partner countries, a review of existing definitions/models of data literacy, and iterative refinement through testing and feedback. It draws on well-known statistical education approaches such as the PPDAC cycle (Problem–Plan–Data–Analysis–Conclusion) and GAISE II and expands them with exploratory data analysis (EDA) ideas so that pupils can discover patterns creatively rather than only follow a linear "confirm-then-conclude" path.

Competences in the framework (expressed as an inquiry cycle)

For a newcomer, the framework can be understood as the competencies pupils (and teachers) build by moving through an eight-part cycle:

Co-funded by the European Union

- **Trigger**: notice something interesting and form meaningful questions.
- **Collect**: gather real-world data (or use authentic datasets).
- **Organize**: structure and prepare data so it can be explored.
- **Explore**: inspect data using visuals and simple tools to find initial insights.
- **Patterns**: identify trends, relationships, or differences in the data.
- **Predict**: use what's been found to make reasoned predictions or simple models.
- **Reflect**: evaluate findings, limitations, and what the data does (and doesn't) show.
- **Share**: communicate results clearly to others.

Figure 10 captures the process visually in the framework visual below.

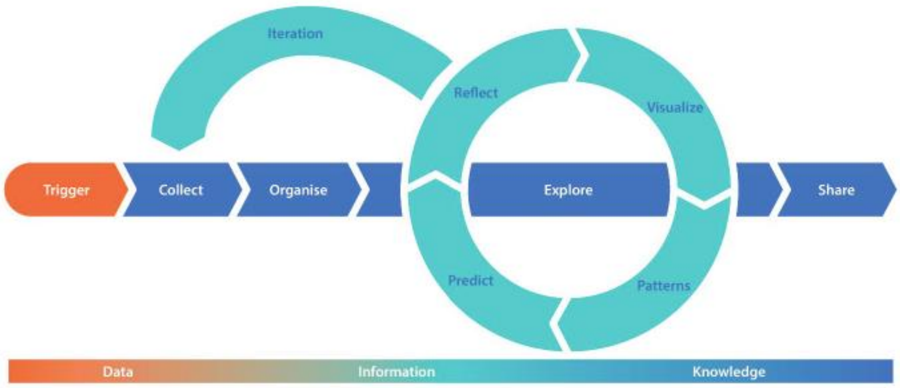


Figure 10 The DALI4US Framework

TRAIN-DL (Teacher training for Data Literacy & Computer Science competences) Project

TRAIN-DL (Teacher training for Data Literacy & Computer Science competences) was a European project (ended February 2024) focused on helping education systems embed AI and data literacy in schools at scale. Its core aim was to support the "structural integration" of Artificial Intelligence (AI) and Data Literacy (DL) skills by producing teacher-training guidelines, teaching-methodology evidence, and policy recommendations that can be adopted across different school contexts and subjects. Rather than focusing on GenAI, TRAIN-DL looks broadly at AI in education, and the data competences teachers need to teach it responsibly and effectively.

Although AI technologies are increasingly present in everyday life, many people - including teachers - still lack a clear understanding of how these systems function. In recent years, several educational frameworks have been developed to promote AI literacy, but:

- Data literacy is only minimally addressed.
- Data literacy frameworks rarely make explicit reference to AI.

Co-funded by the European Union

- No existing framework provides a holistic integration of AI and DL.

This separation is problematic because machine learning relies on data, and working with data requires at least a basic understanding of AI processes.

Key Findings from TRAIN-DL

The review shows that existing AI literacy frameworks include only fragmented and isolated elements of data literacy, and none of them cover the entire data lifecycle.

The study concludes that:

- Existing AI literacy frameworks provide insufficient and fragmented coverage of data literacy.
- A combined and holistic approach is urgently needed.
- The data lifecycle offers a promising structure for integrating AI and DL.
- Future work should evaluate integrated educational interventions for teachers based on this model.

The competence framework: integrating AI literacy + data literacy

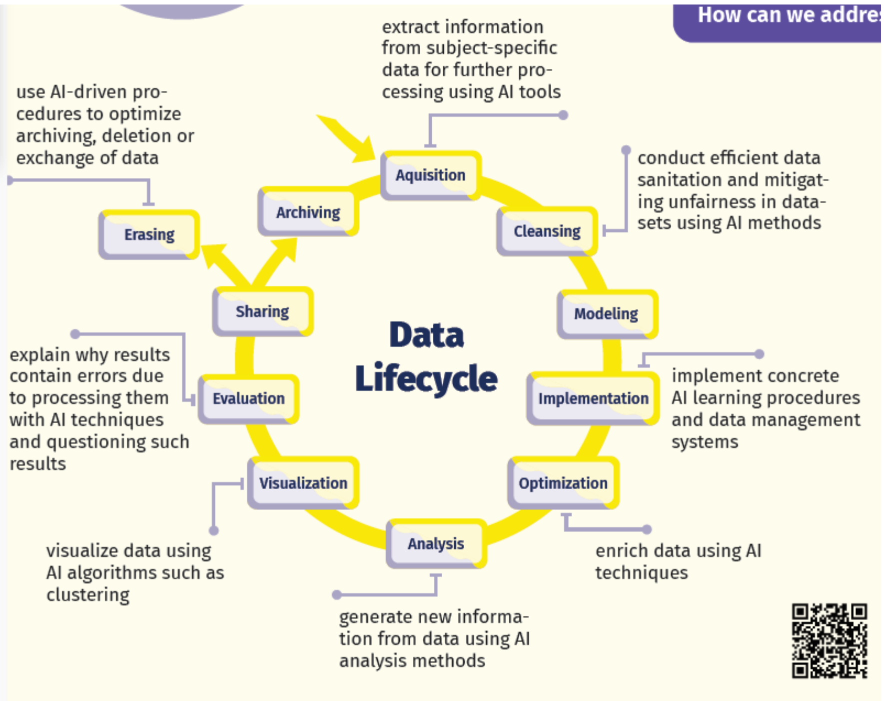


Figure 11 The competence framework: integrating AI literacy + data literacy

The framework's competence areas are as follows:

- **Acquisition**: gather/choose relevant datasets (including subject-specific data) for AI-supported work.
- **Cleansing**: clean data and address issues like bias/unfairness in datasets (including with AI support).

Co-funded by the European Union

- **Modelling**: understand that models are built from data and relate model behaviour to data choices.
- **Implementation**: put AI learning procedures and data-management processes into practice.
- **Optimization**: improve or enrich data using AI techniques.
- **Analysis**: generate insights from data using AI-assisted analysis methods.
- **Visualization**: use AI methods (e.g., clustering) to represent patterns in data.
- **Evaluation**: interpret results critically, explain errors/limits, and question outcomes (e.g., bias, reliability).
- **Sharing / Archiving / Erasing**: manage responsible data governance - sharing appropriately, storing safely, and deleting/exchanging data using well-founded procedures (including AI-driven ones).

MILES (MIL and PRE-BUNKING approaches for Critical thinking in the education sector)

MILES is an Erasmus+ cooperation project (11 partners across 10 EU countries) focused on strengthening resilience to disinformation and digital manipulation in education. Its purpose was to help schools develop informed, critical citizens by equipping teachers (and future teachers) to recognise, prevent, and address misleading information - especially in online environments where information spreads quickly and source-checking is often skipped.

At the time of analysis, only deliverable 2.1 of the project was public. However, this result was considered particularly useful because it provides a detailed analysis of good practices related to training in the field of media literacy. These good practices will certainly be useful in defining the training programme that will be established as part of WP3 of the project.

Specifically, the MILES project identifies 44 practices, with 8 specific examples detailed:

- **Generazioni Connesse (Italy)**: A multi-layered platform supported by the Ministry of Education, developing educational materials, guides, and training programs for teachers, parents, and students to promote digital, media, and information literacy.
- **Infuzarea educatiei media (Romania)**: A methodology guidebook for Romanian language and literature teachers, offering practical resources and 25 lesson plans to infuse media literacy into formal education.
- **Podcast – the Truth in Times of Corona (Germany)**: An audio podcast created by the Bundeszentrale für politische Bildung, dedicated to conspiracy theories around the Corona virus, offering deep understanding through expert discussions.
- **Mediterranean Digital Media Observatory (MedDMO) (Cyprus)**: A regional hub of the European Digital Media Observatory (EDMO) covering Greece, Cyprus, and Malta, focusing on collaborative fact-checking and promoting media literacy and news verification methods.
- **#FakeHunter-Edu (Poland)**: An educational project for secondary school students and teachers, providing video training, activity kits, and lesson plans to raise awareness about disinformation, verify online content, and practice safe internet behavior.

- **The Fake News Commissioner is on the loose (Austria)**: An online game combined with psychological inoculation theory to enhance media literacy and resilience against fake news among students, featuring an interactive game and educational materials for teachers.
- **Stampmedia (Belgium)**: The first youth media agency in Flanders, offering training programs and practical opportunities for young people to experiment with media production, foster critical media and data literacy, and engage in active citizenship.
- **Be Internet Awesome & "Interland" Game (Greece)**: A Google-developed campaign with a serious game (Interland) designed to educate children, teachers, and families on digital risks, online safety, and digital citizenship in an interactive and fun manner.

4.4 Conclusion

The analysis of the projects highlighted two main approaches to defining data literacy. The first approach starts from the definition of the data analysis process and is geared towards guiding training practices based on the stages that characterise this process (e.g., DALI4US). The second approach is more closely linked to the analysis of the competencies related to the concept of data literacy. It is precisely from this second approach, used in particular in the DALI project, that we began to build the AI-DL Framework that will be described in deliverable D2.3, asking ourselves specifically whether and how the introduction of GenAI changed the very definition of data literacy and the set of competencies required.

5. Review of Data Literacy Policy, Strategy and Curricula in Partner countries

5.1 Task Overview and Approach

The proposal stated that this task would review existing data literacy curricula in partner countries to identify best practices and areas for improvement. This task was led by H2 Learning, and Harvey Mellar was the lead author, with support from other partners, specifically the ministries of education of the project participating countries. H2 Learning developed an online survey (See Annex 7D) to gather information on existing data literacy policies, strategies, and curricula in secondary education (i.e. post-primary education) across the seven partner countries.

The survey was shared with the ministries of education in July 2025, and it sought to capture current data literacy and GenAI practices within each of the education systems. The information collected during this sub-task will ultimately inform a broader analysis of data literacy policies and curricula and will inform the development of recommendations and resources for all participating countries, specifically the design of teaching training programmes and approaches (WP3 – Training design and contents).

5.2 Key findings

Overview

Table 9 below provides an overview of the findings from the survey in relation to existing data literacy policies, strategies, and curricula in secondary education (i.e. post-primary education) across the seven partner countries.

Data Literacy (DL) Policy

- Five of the seven partner countries report having a data literacy policy with Lithuania and Slovenia being the exceptions.

Data Literacy (DL) Curriculum

- All partner countries have a data literacy curriculum

Data Literacy (DL) Support

- Five countries report providing data literacy support with Ireland and Slovenia being the exceptions.

GenAI Policy

- Four countries have a GenAI policy, while Ireland, Luxembourg and Slovenia don’t.

GenAI Use

- In France GenAI use is officially supported or endorsed by Ministry or regional/local education body, while in Spain it is primarily used informally or endorsed at individual school level only. In the other five countries its use is mainly pilot or experimental use.

MS	DL POLICY	DL CURRICULUM	DL SUPPORT	GEN AI POLICY	GEN AI USE
France	Yes	Yes	Yes	Yes	Officially supported or endorsed by Ministry or regional/local education body
Ireland	Yes	Yes	No	No	Pilot or experimental use
Italy	Yes	Yes	Yes	Yes	Pilot or experimental use
Lithuania	No	Yes	Yes	Yes	Pilot or experimental use
Luxembourg	Yes	Yes	Yes	No	Pilot or experimental use
Slovenia	No	Yes	No	No	Pilot or experimental use
Spain	Yes	Yes	Yes	Yes	Used informally or endorsed at individual school level only

Table 7 Overview of existing data literacy policies, strategies, and curricula in secondary education across the seven partner countries.

Approaches to implementing data literacy in the curriculum

The survey found that there are a range of approaches currently in place to implement data literacy in the curriculum across the participating countries and we outline these below.

- **Compulsory, named provision**
  - **Luxembourg** provides a mandatory Digital Sciences course that integrates data literacy, privacy/security, ethics and AI/algorithms and includes a data-science module for upper-secondary level.
- **Multi-subject**
  - **Lithuania** specifies that data literacy is to be presented in Informatics ("Data exploration and Information"), Mathematics ("Data & Probability"), Citizenship, Natural Sciences, Ethics and Geography courses - progressing to inclusion of data analysis, AI/ML and big data by upper secondary.

- **France** develops the acquisition of skills related to data management, analysis and protection for all students, from primary school to secondary school, in various subjects (mathematics, life and earth sciences, economics and management) as part of projects or through the use of digital tools in the classroom.

- **Whole-curriculum digital competence + subject exemplars**
  - **Spain** positions data literacy within Digital Competence, with some specific subject level descriptors (e.g., for Technology & Digitalisation; Mathematics; Biology & Geology)

- **Cross-curricular models**
  - **Ireland** explicitly includes data literacy in Junior Cycle Key Skills: Managing Information & Thinking, and in Senior Cycle Key Competencies: Thinking and solving problems, Communication and Numeracy, but does not indicate specific subject areas.
  - **Italy** integrates activities, methodologies, and content aimed at developing STEM, coding, artificial intelligence, digital and data literacy and innovation skills following a fully interdisciplinary approach across all school levels.
  - **Slovenia** does not address data literacy as a standalone subject, rather it is integrated primarily through subjects like mathematics, statistics, and informatics.

Data literacy projects

- **Slovenia**
  - PUMICE – in this project educational activities that can be used to enrich various school subjects were developed. In doing so, data related to the learning material and explore it using artificial intelligence and machine learning approaches were used.
  - NAPOJ 2 (see: Activation of Computer Science Teachers in Slovenia - Archive ouverte HAL). The project (2018 – 2024) created examples of good practice of intertwining the subjects of mathematics, informatics (computer science), STEM, arts, and technology (described as 'MINUT' subjects) in upper-secondary and lower-secondary (upper primary) schools; the main goal of the project was to create an environment that would enable active cooperation between teachers who want to familiarise students with the intertwining of MINUT fields, including Data Science and Data Literacy.

- **Luxembourg, Slovenia and Ireland**
  - DALI4US (Data Literacy for Upper Primary Schools) was a forward-looking, transnational initiative funded under the ERASMUS+ programme that addressed the growing demand for digital and data literacy in European education. It focused on equipping teachers and students with the skills to understand, interpret and critically engage with data, thereby laying the foundation for responsible AI use

- **Lithuania**

Co-funded by the European Union

- Data Literate - an Erasmus+ project for capacitating educators of secondary schools in Digital Data Literacy to identify disinformation and manage the overload of information (with Portugal, Spain, Italy)
- TrainDL - Teacher Training for Data Literacy & Computer Science Competences – an Erasmus+ project aiming to develop educational concepts for data literacy and artificial intelligence competencies and embed them in teacher and school education (with Germany and Austria)

Examples where GenAI and DL are taught together

- **France.** In secondary schools the subject Digital Sciences and Technology integrates DL concepts
- **Spain.** The Technology option for 15–16-year-olds incorporates artificial intelligence modules and data engineering techniques
- **Luxembourg.** The compulsory Digital Sciences course in lower secondary schools teaches DL, ethics/privacy and AI/algorithms within one strand
- **Lithuania.** The compulsory Informatics course includes AI/ML and big data by Grade 10 (age 16), and big data, artificial intelligence and neural networks in the optional Informatics course offered in Grades 11-12 (age 17-18)
- **Slovenia.** The Innovative Pedagogy 5.0 project includes computer science and data science/literacy including data analytics with Orange Data Mining

Examples of activities linking data literacy to the social impacts of technology/AI

- PUMICE – Gairdín - Lessons with a dash of artificial intelligence, including:
    - Cartoon Recommendations lesson - This lesson uses students' favorite cartoons (or TV series/music) to demonstrate how recommendation systems work without analysing content. (Slovenia, Luxembourg, Ireland)
- Opération Cactus
    - A fake phishing campaign, to raise awareness among students of digital risks, in particular phishing. Cactus aims to raise awareness among students about digital risks through a national educational phishing campaign. (France)
- data.education.gouv.fr platform
    - Sharing of public data on school education, in conjunction with the data already available on higher education and research. The aim is to enable everyone – citizens, partners, public and private stakeholders – to benefit from transparency and encourage innovation by: -learning about, identifying, understanding and enriching data relating to school education, -proposing useful services in the field of education. The ministry is thus

Co-funded by the European Union

publishing new reference data sets that are long-lasting, high-quality and easy to explore and reuse. The availability of educational datasets promotes educational uses (school data journalism projects, statistical analysis in mathematics or life and earth sciences, etc.) for example [Challenge Wikidata en classe — Éducation Nationale - Accueil]

Approaches to implementing data literacy in the curriculum

The survey found that there are a range of approaches currently in place to implement data literacy in the curriculum across the participating countries and we outline these below.

- **Compulsory, named provision**
  - **Luxembourg** provides a mandatory Digital Sciences course that integrates data literacy, privacy/security, ethics and AI/algorithms and includes a data-science module for upper-secondary level
- **Multi-subject**
  - **Lithuania** specifies that data literacy is to be presented in Informatics ("Data exploration and Information"), Mathematics ("Data & Probability"), Citizenship, Natural Sciences, Ethics and Geography courses - progressing to inclusion of data analysis, AI/ML and big data by upper secondary
  - **France** develops the acquisition of skills related to data management, analysis and protection for all students, from primary school to secondary school, in various subjects (mathematics, life and earth sciences, economics and management) as part of projects or through the use of digital tools in the classroom
- **Whole-curriculum digital competence + subject exemplars**
  - **Spain** positions data literacy within Digital Competence, with some specific subject level descriptors (e.g., for Technology & Digitalisation; Mathematics; Biology & Geology)
- **Cross-curricular models**
  - **Ireland** explicitly includes data literacy in Junior Cycle Key Skills: Managing Information & Thinking, and in Senior Cycle Key Competencies: Thinking and solving problems, Communication and Numeracy, but does not indicate specific subject areas
  - **Italy** integrates activities, methodologies, and content aimed at developing STEM, coding, artificial intelligence, digital and data literacy and innovation skills following a fully interdisciplinary approach across all school levels
  - **Slovenia** does not address data literacy as a standalone subject, rather it is integrated primarily through subjects like mathematics, statistics, and informatics

5.3 Conclusions

The survey showed that the implementation of GenAI in schools is at a similar early experimental stage in all the partner countries, and that no country has developed the use of GenAI to support data literacy as yet. The situation for data literacy shows quite distinct approaches being adopted in different countries. The contribution that the AI-DL project can make to developing GenAI approaches to supporting data literacy is therefore clear, but the way in which this can be integrated into national approaches will need to vary according to existing data literacy approaches and this will be a challenge for the AI-DL project. The training program needs to be written in such as way that it can be adapted to fit with national approaches to data literacy.

Co-funded by the European Union

6. General conclusions and implications for the AI-DL project

The project proposal described the desk review (Task 2.1) as conducting a "comprehensive desk research to understand the current state of data literacies and the adoption of GenAI technologies in education in the context of enhancing democratic competences in our young people" Project partners the following decisions on how best to achieve this deliverable:

- The desk research will primarily focus on data literacy in the context of GenAI use.
- GenAI usage in schools, by teachers and students, is the context within which data literacy competences would be developed.
- The project is not, primarily, focused on developing teacher and student AI or GenAI literacy.
- The focus is on developing teacher and student data literacy in the context of using GenAI to enhance their democratic competences.

The desk review was carried out in four parallel work streams: a systematic literature review of academic research; an analysis of a selection of existing competency frameworks; a review of selected Erasmus+ projects and a survey of national policies and curricula in data literacy across the partner countries.

The key findings of these workstreams have been described above in Sections 2.3, 3.2, 4.3 and 5.2. In this section we will revisit these key findings and attempt to draw some over-arching conclusions from this research to inform a range of work packages and activities in the AI-DL project.

6.1 Key Findings

The four parallel work streams addressed the objectives in complementary ways in order to obtain the widest possible range of perspectives.

The **Literature Review** addressed the objective of the desk review through an examination of recent academic publications in order to establish current thinking about data literacy, which would inform all aspects of the AI-DL project. This showed that there is no consensus on the definition of data literacy, and that data literacy is a rather fluid, multidimensional concept. Minimal connection was found between existing data literacy frameworks and GenAI usage, indicating.

The review identified three categories of data literacy components:

- **Disciplinary data literacy components,** include understanding data basics, designing investigations, evaluating data quality and sources, creating datasets, sense-making that considers bias and social factors, processing and analysing data, making data-based claims, and publishing analyses ethically.
- **Personal data literacy components**, drawing on Pangrazio and Selwyn's (2023) framework "Critical Data Literacies: Rethinking Data and Everyday Life," which encompasses five key dimensions: *Data identification*: *Data understandings*: *Data reflexivity*: *Data strategies*: *Data*

Co-funded by the European Union

tactics. This critical data literacies framework shifts the emphasis from purely technical skills toward understanding data as a social, political, and cultural phenomenon that shapes power relationships.

- **Teacher-specific components,** address handling student data ethically, using data for iterative decision making, and understanding recommended pedagogical approaches for teaching data literacy.

Other key findings from the literature review were:

- Data literacy definitions are highly contextual and evolve with technological change
- Traditional frameworks focus on "input-side" data practices (collection, management, evaluation) rather than GenAI-specific competencies
- Critical thinking, metacognition, and collaboration are essential aspects of data literacy
- Teachers require pedagogical approaches that encourage curiosity and creativity with emerging technologies

The **Framework Analysis** reviewed existing major frameworks (DigComp 2.2/3.0, UNESCO AI frameworks, CDC, AILit) for Data Literacy and GenAI in order to provide a basis from which the AI-DL framework could be constructed. The Framework Analysis demonstrated uneven coverage across six competence domains. For Data Literacy it found that coverage was uneven, reflecting each framework's original purpose and development context. While GenAI coverage, across the six domains, was comparatively broad and more uniform, reflecting the fact that contemporary GenAI-oriented frameworks were developed with an explicit intention to integrate technical understanding with responsible use and classroom-relevant practices.

The analysis revealed the following:

- Core knowledge of AI and data is well-covered in modern frameworks but absent from pre-2022 documents, specifically in the CDC framework
- Data Stewardship" and "Model Documentation" remain comparatively less visible. The analysed frameworks prioritise user-facing competences, such as critical evaluation of AI outputs and creative or pedagogical use, over less visible, but consequential, competences related to data preparation, provenance, documentation, and transparency
- While Data Literacy stays a foundational competence, GenAI Literacy extends it into interaction, creativity, and co-creation with AI systems.

Data literacy frameworks (e.g., DigComp, or UNESCO) offer partial reference points for proficiency and progression, but the literature does not yet provide a single, formalised definition of "data literacy for generative AI.

The analysis of **European Projects** sought to establish what other projects had been carried out in related areas that the AI-DL projects might learn from. Analysis of 100 Erasmus+ projects showed growing interest in data literacy education, and that most projects focused on collaborative institutional partnerships (KA2). Projects of relevance to AI-DL's aims included DALI (adult data literacy), DALI4US (primary education), TRAIN-DL (teacher training integrating AI and data literacy), and MILES (media literacy and disinformation). These projects provide valuable frameworks and methodologies that are relevant to the AI-DL project, but they did not address GenAI specifically.

The analysis of National Policies and Curricula was undertaken in order to provide the national contexts for the AI-DL Framework development and the training programmes. The survey results showed that five of seven partner countries have data literacy policies, and all have some form of data literacy curricula. However, implementation approaches vary widely – ranging from compulsory courses under the title of data science (Luxembourg) to cross-curricular integration (Ireland, Italy, Slovenia). This variability in ways of delivering data literacy presents both challenges and opportunities for the AI-DL project. The implementation of GenAI was somewhat similar across countries with most usage still experimental, and most usage is still experimental, and with four of the seven countries having GenAI policies in place.

6.2 Implications for AI-DL

The research confirms the project's core hypothesis: there is an urgent need for frameworks and training that specifically address data literacy in the context of GenAI usage. Key implications include:

1. **Framework Development:** AI-DL should not attempt to create rigid definitions but rather develop flexible, co-constructed approaches that adapt to local contexts and evolving technologies.
2. **Teacher Training:** Training should balance technical understanding with critical, ethical, and pedagogical competencies. Teachers need support in developing students' ability to evaluate AI outputs, understand data provenance, and engage in informed decision-making.
3. **Pedagogical Approaches:** The project should emphasize active learning, collaboration, real-world problem-solving, and creative engagement with both data and GenAI tools.
4. **Cross-Curricular Integration:** Given the diverse implementation approaches across countries, AI-DL resources should be flexible enough to support both subject-specific and cross-curricular applications.
5. **Ethical and Democratic Competences:** Data literacy in the GenAI context must explicitly address citizenship, privacy, bias, and the social impacts of AI-driven decision-making.

The desk research provides a foundation for WP3 – Training design and content and the creation of an emerging AI-DL Framework that will be refined throughout the project lifecycle through co-construction with teachers, students, and school leaders.

Co-funded by the European Union

7. Annexes

Annex 7A Publications and references

Publications analysed in the literature report

Ambarwati, A., Sarwanto, & Sukarmin. (2020). The profile of students' data literacy at SMA Negeri 1 Karanganyar. *Journal of Physics: Conference Series*, *1511*(1), 012028. https://doi.org/10.1088/1742-6596/1511/1/012028

Amels, J., Krüger, M., Suhre, C., & Van Veen, K. (2019). Impact of inquiry-based working on the capacity to change in primary education. *Journal of Educational Change*, *20*(3), 351–374. https://doi.org/10.1007/s10833-019-09337-3

Ansyari, M. F., Groot, W., & De Witte, K. (2020). Tracking the process of data use professional development interventions for instructional improvement: A systematic literature review. *Educational Research Review*, *31*, 100362. https://doi.org/10.1016/j.edurev.2020.100362

Banowati, E., Juhadi, & Sanjoto, T. (2019). The Utilization of Smartphone Communication Technology the as Digital Literacy Learning School Instruments in 4.0 Era. *Journal of Physics: Conference Series*, *1387*(1), 012111. https://doi.org/10.1088/1742-6596/1387/1/012111

Bolhuis, E., Voogt, J., & Schildkamp, K. (2019). The development of data use, data skills, and positive attitude towards data use in a data team intervention for teacher educators. *Studies in Educational Evaluation*, *60*, 99–108. https://doi.org/10.1016/j.stueduc.2018.12.002

Børte, K., Lillejord, S., Chan, J., Wasson, B., & Greiff, S. (2023). Prerequisites for teachers' technology use in formative assessment practices: A systematic review. *Educational Research Review*, *41*, 100568. https://doi.org/10.1016/j.edurev.2023.100568

Burnett, C. (2016). *The digital age and its implications for learning and teaching in the primary school: a report for the Cambridge Primary Review Trust*. Cambridge Primary Review Trust.

Carmichael, P. (2020). Post digital Possibilities: Operaismo, Co-research, and Educational Inquiry. *Postdigital Science and Education*, *2*(2), 380–396. https://doi.org/10.1007/s42438-019-00089-0

Chaka, C. (2019). Re-imagining literacies and literacies pedagogy in the context of semio-technologies. *Nordic Journal of Digital Literacy*, *14*(1–2), 54–69. https://doi.org/10.18261/issn.1891-943x-2019-01-02-05

Cowie, B., Edwards, F., & Trask, S. (2021). Explicating the Value of Standardized Educational Achievement Data and a Protocol for Collaborative Analysis of This Data. *Frontiers in Education*, *6*, 619319. https://doi.org/10.3389/feduc.2021.619319

Dander, V., & Macgilchrist, F. (2022). School of Data and Shifting Forms of Political Subjectivity. In P. Bettinger (Ed.), *Educational Perspectives on Mediality and Subjectivation* (pp. 45–67). Springer International Publishing. https://doi.org/10.1007/978-3-030-84343-4_3

Co-funded by the European Union

Dangol, A., & Dasgupta, S. (2023). Constructionist approaches to critical data literacy: A review. *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference*, 112–123. https://doi.org/10.1145/3585088.3589367

Department of Curriculum and Instructions Studies, School of Teacher Education, University of South Africa, Pretoria, South Africa, & Plessis, E. D. (2025). Embracing Project-Based Assessments in the Age of AI in Open Distance e-Learning. *International Journal of Information and Education Technology*, *15*(2), 372–381. https://doi.org/10.18178/ijiet.2025.15.2.2249

Dunlap, K., & Piro, J. S. (2016). Diving into data: Developing the capacity for data literacy in teacher education. *Cogent Education*, *3*(1), 1132526. https://doi.org/10.1080/2331186X.2015.1132526

Fagerlund, J., Palsa, L., & Mertala, P. (2025). Exploration of domains of educational purpose in K-12 data literacy education research. *Educational Research Review*, *46*, 100663. https://doi.org/10.1016/j.edurev.2024.100663

Gal, I. (2022). Critical Understanding of Civic Statistics: Engaging with Important Contexts, Texts, and Opinion Questions. In J. Ridgway (Ed.), *Statistics for Empowerment and Social Engagement* (pp. 323–343). Springer International Publishing. https://doi.org/10.1007/978-3-031-20748-8_13

Hansen, C. J., & Wasson, B. (2016). Teacher Inquiry into Student Learning: - The TISL Heart Model and Method for use in Teachers' Professional Development. *Nordic Journal of Digital Literacy*, *11*(1), 24–49. https://doi.org/10.18261/issn.1891-943x-2016-01-02

Henderson, J., & Corry, M. (2021). Data literacy training and use for educational professionals. *Journal of Research in Innovative Teaching & Learning*, *14*(2), 232–244. https://doi.org/10.1108/JRIT-11-2019-0074

Hoogland, I., Schildkamp, K., Van Der Kleij, F., Heitink, M., Kippers, W., Veldkamp, B., & Dijkstra, A. M. (2016). Prerequisites for data-based decision making in the classroom: Research evidence and practical illustrations. *Teaching and Teacher Education*, *60*, 377–386. https://doi.org/10.1016/j.tate.2016.07.012

Ilomäki, L., Lakkala, M., Kallunki, V., Mundy, D., Romero, M., Romeu, T., & Gouseti, A. (2023). Critical digital literacies at school level: A systematic review. *Review of Education*, *11*(3), e3425. https://doi.org/10.1002/rev3.3425

Irish, T., Berkowitz, A., & Harris, C. (2019). Data Explorations: Secondary Students' Knowledge, Skills and Attitudes Toward Working with Data. *EURASIA Journal of Mathematics, Science and Technology Education*, *15*(6). https://doi.org/10.29333/ejmste/103063

Israel-Fishelson, R., Moon, P. F., Pauw, D., & Weintrop, D. (2024). Using Participatory Design to Gain Insight into How Students Make Sense of Data in Their Lives. *Proceedings of the Symposium on Learning, Design and Technology*, 85–94. https://doi.org/10.1145/3663433.3663461

Jackson, C. J. (2022). The utility of NAPLAN data: issues of access, use and expertise for teaching and learning. *The Australian Journal of Language and Literacy*, *45*(2), 141–157. https://doi.org/10.1007/s44020-022-00009-z

Juergens, C. (2020). Digital Data Literacy in an Economic World: Geo-Spatial Data Literacy Aspects. *ISPRS International Journal of Geo-Information*, *9*(6), 373. https://doi.org/10.3390/ijgi9060373

Kahila, J., Vartiainen, H., Tedre, M., Arkko, E., Lin, A., Pope, N., Jormanainen, I., & Valtonen, T. (2024). Pedagogical framework for cultivating children's data agency and creative abilities in the age of AI. *Informatics in Education*, 323–360. https://doi.org/10.15388/infedu.2024.15

Kim, J., Lee, H., & Cho, Y. H. (2022). Learning design to support student-AI collaboration: perspectives of leading teachers for AI in education. *Education and Information Technologies*, *27*(5), 6069–6104. https://doi.org/10.1007/s10639-021-10831-6

Kim, M. S., & Yu, F. (2023). 'Teacher data literacies practice' meets 'pedagogical documentation': A scoping review. *Review of Education*, *11*(2), e3414. https://doi.org/10.1002/rev3.3414

Kippers, W. B., Poortman, C. L., Schildkamp, K., & Visscher, A. J. (2018). Data literacy: What do educators learn and struggle with during a data use intervention? *Studies in Educational Evaluation*, *56*, 21–31. https://doi.org/10.1016/j.stueduc.2017.11.001

Lee, V. R., Pimentel, D. R., Bhargava, R., & D'Ignazio, C. (2022). Taking data feminism to school: A synthesis and review of pre-collegiate data science education projects. *British Journal of Educational Technology*, *53*(5), 1096–1113. https://doi.org/10.1111/bjet.13251

Lee, V. R., Pope, D., Miles, S., & Zárate, R. C. (2024). Cheating in the age of generative AI: A high school survey study of cheating behaviours before and after the release of ChatGPT. *Computers and Education: Artificial Intelligence*, *7*, 100253. https://doi.org/10.1016/j.caeai.2024.100253

Lee, V. R., Wilkerson, M. H., & Lanouette, K. (2021). A Call for a Humanistic Stance Toward K–12 Data Science Education. *Educational Researcher*, *50*(9), 664–672. https://doi.org/10.3102/0013189X211048810

Lestari, W. Y., & Rosana, D. (2020). Analysis of Junior High School students' data literacy in Ciamis with local potential kampung adat kuta. *Journal of Physics: Conference Series*, *1440*(1), 012097. https://doi.org/10.1088/1742-6596/1440/1/012097

Lichti, D., Mosley, P., & Callis-Duehl, K. (2021). Learning from the Trees: Using Project Budburst to Enhance Data Literacy and Scientific Writing Skills in an Introductory Biology Laboratory During Remote Learning. *Citizen Science: Theory and Practice*, *6*(1), 32. https://doi.org/10.5334/cstp.432

Mandinach, E. B., & Jimerson, J. B. (2022). Data ethics in education: a theoretical, practical, and policy issue. *Studia Paedagogica*, *26*(4), 9–26. https://doi.org/10.5817/SP2021-4-1

Mandinach, E. B., & Schildkamp, K. (2021). The complexity of data-based decision making: An introduction to the special issue. *Studies in Educational Evaluation*, *69*, 100906. https://doi.org/10.1016/j.stueduc.2020.100906

Matuk, C., DesPortes, K., Amato, A., Vacca, R., Silander, M., Woods, P. J., & Tes, M. (2022). Tensions and synergies in arts-integrated data literacy instruction: Reflections on four classroom implementations. *British Journal of Educational Technology*, *53*(5), 1159–1178. https://doi.org/10.1111/bjet.13257

Melnikova, J., Batuchina, A., Ahrens, A., & Zascerinska, J. (2023). TEACHERS' DATA LITERACY SKILLS FOR PEDAGOGICAL DECISION MAKING: NEEDS ANALYSIS IN LITHUANIA AND GERMANY. *ENVIRONMENT. TECHNOLOGIES. RESOURCES. Proceedings of the International Scientific and Practical Conference*, *2*, 182–188. https://doi.org/10.17770/etr2023vol2.7287

Mendez-Carbajo, D., Illinois Wesleyan University, Jefferson, C., California State University-Northridge, Stierholz, K., & Federal Reserve Bank of St. Louis. (2019). Keeping It Real: Information Literacy, Numeracy, and Economic Data. *Numeracy*, *11*(1). https://doi.org/10.5038/1936-4660.12.2.5

Mertala, P., (2020). Data (il)literacy education as a hidden curriculum of the datafication of education. *Journal of Media Literacy Education*, 12(3), 30-42. https://doi.org/10.23860/JMLE-2020-12-3-4

Michos, K., Schmitz, M.-L., & Petko, D. (2023). Teachers' data literacy for learning analytics: a central predictor for digital data use in upper secondary schools. *Education and Information Technologies*, *28*(11), 14453–14471. https://doi.org/10.1007/s10639-023-11772-y

Moon, P. F., Israel-Fishelson, R., Tabak, R., & Weintrop, D. (2023). The Tools Being Used to Introduce Youth to Data Science. *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference*, 150–159. https://doi.org/10.1145/3585088.3589363

Mouta, A., Pinto-Llorente, A. M., & Torrecilla-Sánchez, E. M. (2024). Uncovering Blind Spots in Education Ethics: Insights from a Systematic Literature Review on Artificial Intelligence in Education. *International Journal of Artificial Intelligence in Education*, *34*(3), 1166–1205. https://doi.org/10.1007/s40593-023-00384-9

Ndukwe, I. G., & Daniel, B. K. (2020). Teaching analytics, value and tools for teacher data literacy: a systematic and tripartite approach. *International Journal of Educational Technology in Higher Education*, *17*(1), 22. https://doi.org/10.1186/s41239-020-00201-6

Nwagwu, W. E. (2025). Citation analysis of global research on data literacy. *Education and Information Technologies*. https://doi.org/10.1007/s10639-025-13397-9

Pellegrino, M., & Antelmi, A. (2023). At School of Open Data: A Literature Review: *Proceedings of the 15th International Conference on Computer Supported Education*, 172–183. https://doi.org/10.5220/0011747500003470

Physics Education Department, Faculty of Mathematics and Natural Sciences, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia, & Dewi, C. T. (2024). The Effectiveness of Virtual Laboratory

Integrated Blacksmith Ethnophysics in Enhancing Data Literacy and Graphic Representation Skills. *International Journal of Information and Education Technology*, *14*(8), 1155–1163. https://doi.org/10.18178/ijiet.2024.14.8.2144

Rahmita, & Rosana, D. (2020). Profile analysis of data literacy capability based on NGSS junior high school students in Takalar, South Sulawesi. *Journal of Physics: Conference Series*, *1440*(1), 012082. https://doi.org/10.1088/1742-6596/1440/1/012082

Robertson, J., & Tisdall, E. K. M. (2020). The importance of consulting children and young people about data literacy. *Journal of Media Literacy Education*, *12*(3), 58–74. https://doi.org/10.23860/JMLE-2020-12-3-6

Saddiqa, M., Rasmussen, L., Magnussen, R., Larsen, B., & Pedersen, J. M. (2019). Bringing open data into Danish schools and its potential impact on school pupils. *Proceedings of the 15th International Symposium on Open Collaboration*, 1–10. https://doi.org/10.1145/3306446.3340821

Schildkamp, K. (2019). Data-based decision-making for school improvement: Research insights and gaps. *Educational Research*, *61*(3), 257–273. https://doi.org/10.1080/00131881.2019.1625716

Schildkamp, K., Van Der Kleij, F. M., Heitink, M. C., Kippers, W. B., & Veldkamp, B. P. (2020). Formative assessment: A systematic review of critical teacher prerequisites for classroom practice. *International Journal of Educational Research*, *103*, 101602. https://doi.org/10.1016/j.ijer.2020.101602

Suryadi, Mahardika, I. K., Supeno, & Sudarti. (2021). Data literacy of high school students on physics learning. *Journal of Physics: Conference Series*, *1839*(1), 012025. https://doi.org/10.1088/1742-6596/1839/1/012025

Suryadi, S., Mahardika, I. K., Sudarti, S., Supeno, S., & Wicaksono, I. (2024). Development of object-oriented physics learning model to promote data literacy in physics instructional. *Journal of Infrastructure, Policy and Development*, *8*(9), 7402. https://doi.org/10.24294/jipd.v8i9.7402

Vacca, R., DesPortes, K., Tes, M., Silander, M., Matuk, C., Amato, A., & Woods, P. J. (2022). ”I happen to be one of 47.8%”: Social-Emotional and Data Reasoning in Middle School Students’ Comics about Friendship. *CHI Conference on Human Factors in Computing Systems*, 1–18. https://doi.org/10.1145/3491102.3502086

Vartiainen, H., Tedre, M., Kahila, J., & Valtonen, T. (2020). Tensions and trade-offs of participatory learning in the age of machine learning. *Educational Media International*, *57*(4), 285–298. https://doi.org/10.1080/09523987.2020.1848512

Wang, N., & Lester, J. (2023). K-12 Education in the Age of AI: A Call to Action for K-12 AI Literacy. *International Journal of Artificial Intelligence in Education*, *33*(2), 228–232. https://doi.org/10.1007/s40593-023-00358-x

Watson, J., & Smith, C. (2022). Statistics education at a time of global disruption and crises: a growing challenge for the curriculum, classroom and beyond. *Curriculum Perspectives*, *42*(2), 171–179. https://doi.org/10.1007/s41297-022-00167-7

Wilkerson, M. H., Kim, J., Lee, H. S., Stokes, D. J., & Ferrell, M. (2025). How Teachers Envision Using Data Visualization Discussion Tasks in Classroom Instruction. *International Journal of Science and Mathematics Education*. https://doi.org/10.1007/s10763-024-10521-y

Woods, P. J., Matuk, C., DesPortes, K., Vacca, R., Tes, M., Vasudevan, V., & Amato, A. (2024). Reclaiming the right to look: making the case for critical visual literacy and data science education. *Critical Studies in Education*, *65*(5), 441–459. https://doi.org/10.1080/17508487.2023.2298198

Zhang, H., Lee, I., & Moore, K. (2024). An Effectiveness Study of Teacher-Led AI Literacy Curriculum in K-12 Classrooms. *Proceedings of the AAAI Conference on Artificial Intelligence*, *38*(21), 23318–23325. https://doi.org/10.1609/aaai.v38i21.30380

Other references in the literature review

Banegas, D., & Villacañas, L. (2016). Criticality. ELT Journal, 70(4), 455–457. https://doi.org/10.1093/elt/ccw048

Bashir, N., Donti, P., Cuff, J., Sroka, S., Ilic, M., Sze, V., Delimitrou, C., Olivetti, E. (2024) The Climate and Sustainability Implications of Generative AI, An MIT Exploration of Generative AI. https://doi.org/10.21428/e4baedd9.9070dfe7

Brunton, F., & Nissenbaum, H. F. (2015). Obfuscation. A user's guide for privacy and protest. MIT Press.

Costa, C., & Murphy, M. (2025). Generative artificial intelligence in education: (what) are we thinking? Learning, Media and Technology, 1–12. https://doi.org/10.1080/17439884.2025.2518258

Chen, B. (2025). Beyond Tools: Generative AI as Epistemic Infrastructure in Education. ArXiv, abs/2504.06928.

D'Ignazio, C., & Klein, L. F. (2020). Data feminism. MIT press.

Frankenstein, M., (2013). Reading the world with math: Goals for a critical mathematical literacy curriculum. *Rethinking mathematics: Teaching social justice by the numbers*. 2nd ed. 30–39.

Lehrer, R., & Schauble, L. (2000). Inventing data structures for representational purposes: Elementary grade students' classification models. *Mathematical Thinking and Learning*, 2(1&2), 51–74. https://doi.org/10.1207/S15327833MTL0202_3

Kitchin, R. (2014). The data revolution. Sage.

Kitto, K., & Knight, S. (2019). Practical ethics for building learning analytics. *British Journal of Educational Technology*, 50(6), 2855–2870. https://doi.org/10.1111/bjet.12868

Mandinach, E., & Gummer, E. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher*, 42, 30–37. http://dx.doi.org/10.3102/0013189X12459803

Mandinach, E. and Gummer, E. (2016), "What does it mean for teachers to be data literate: laying out the skills, knowledge, and dispositions", *Teaching and Teacher Education*, Vol. 60 No. C, pp. 366-376, https://doi.org/10.1016/j.tate.2016.07.011Nichols, T. P., & Stornaiuolo, A. (2019). Assembling "digital literacies": Contingent pasts, possible futures. *Media and Communication*, 7(2), 14–24. https://doi.org/10.17645/mac.v7i2.1946

Pangrazio, L., & Selwyn, N. (2020). Towards a school based 'critical data education'. *Pedagogy, Culture & Society*. Advance online publication. https://doi.org/10.1080/14681366.2020.1747527

Pangrazio, L., & Selwyn, N. (2023). Critical Data Literacies: Rethinking Data and Everyday Life. MIT press. https://doi.org/10.7551/mitpress/14155.001.0001

Selwyn, N. (2019). What's the problem with learning analytics? *Journal of Learning Analytics*, 6(3), 1119. http://dx.doi.org/10.18608/jla.2019.63.3

Wolff, A., Gooch, D., Cavero Montaner, J., Rashid, U., & Kortuem, G. (2017). Creating an understanding of data literacy for a data-driven society. *Journal of Community Informatics*, 12(3), 9-26. www.cijournal.net/index.php/ciej/article/view/1286

Wolff, A., Wermelinger, M., and Petre, M. (2019). Exploring design principles for data literacy activities to support children's inquiries from complex data. International *Journal of Human-Computer Studies*, 129:4154.

Pati, D., & Lorusso, L. N. (2017). How to write a systematic review of the literature. *HERD: Health Environments Research & Design Journal, 11*(1), 15–30. https://doi.org/10.1177/1937586717747384

References from the Frameworks Review

Cosgrove, J., & Cachia, R. (2025). DigComp 3.0: European Digital Competence Framework. *JRC Publications Repository*. https://doi.org/10.2760/0001149

Vuorikari, R., Kluzer, S., & Punie, Y. (2022, March 17). *DigComp 2.2: The Digital Competence Framework for Citizens - With new examples of knowledge, skills and attitudes*. JRC Publications Repository; Publications Office of the European Union. https://publications.jrc.ec.europa.eu/repository/handle/JRC128415

UNESCO. (2024). *AI competency framework for teachers*. Unesco.org. https://www.unesco.org/en/articles/ai-competency-framework-teachers

UNESCO. (2024). *AI competency framework for students*. Unesco.org. https://www.unesco.org/en/articles/ai-competency-framework-students

Council of Europe. (2016). *Reference Framework of Competences for Democratic Culture*. Council of Europe. https://www.coe.int/en/web/reference-framework-of-competences-for-democratic-culture/rfcdc

OECD. (2025). Empowering learners for the age of AI: An AI literacy framework for primary and secondary education [Review draft]. https://ailiteracyframework.org

Annex 7B List of evidence found for single Skills/ Competencies ordered by domain

Data Literacy

A. Core Knowledge of AI & Data

AILit Framework

- K2.1 (human labour, global conditions)
- "Data are the raw material of every AI system…"
- "Machine-learning models learn from data, not by grasping meaning…"
- "Humans shape every dataset…"
- "Datasets come from many sources…"
- "Interactive systems keep gathering new data…"
- "Bias travels through data…"
- "Large-scale data use has externalities…"
- "Training data may include copyrighted or sensitive material…"

UNESCO Teacher Framework

- AI foundations and applications:
    - Basic AI techniques and applications
    - Application skills
    - Creating with AI
    - Competent operation of AI tools

UNESCO Students Framework

- AI techniques and Applications – Conceptual knowledge of AI
- AI techniques and Applications – AI foundations: "Build basic knowledge … regarding data and algorithms"
- AI techniques and Applications – Application skills: "Critically evaluate and leverage free/open-source AI tools, libraries, datasets"

B. Data Stewardship & Model Documentation

AILit Framework

- Collect and curate data … (p. 37)
- Describe an AI model's purpose, intended users and limitations (p. 38)
- Computational thinking – break problems into attributes that can be captured, labelled and structured as data

Co-funded by the European Union

- Communication – explain, in plain language, what data went in, how it was prepared, and where the limitations lie
- Collaboration – work with peers to gather, label and validate datasets efficiently

DigComp 2.2

- 1.1 Browsing, searching and filtering data, information and digital content
- 1.3 Managing data, information and digital content

DigComp 3.0

- 1.1 Browsing, searching and filtering data, information and digital content
- 1.3 Managing data, information and digital content

C. Cognitive & Analytical Domain

CDC Framework

- Analytical and critical thinking skills are the skills required to analyse, evaluate and make judgments about materials of any kind (e.g. texts, arguments, interpretations, issues, events, experiences, etc.) in a systematic and logical manner. (p.13)

AILit Framework

- Critical Thinking (evaluate AI output for accuracy, fairness, bias)
- Computational Thinking (decompose problems, give structured instructions to AI)
- Problem Solving (decide when and how to use AI; assess capabilities, risks, ethics)
- Critical thinking – judge data quality, spot gaps or biases, cross-check AI outputs against other evidence

DigComp 2.2

- 1.2 Evaluating data, information and digital content

DigComp 3.0

- 1.2 Evaluating data, information and digital content

UNESCO Teacher Framework

- AI pedagogy – Valutare criticamente l'impatto dell'AI su insegnamento, apprendimento e valutazione

Co-funded by the European Union

D. Creative & Innovative Domain

AILit Framework

- Creativity (collaborate with AI to create and refine original ideas while considering ownership, attribution, responsible use)

UNESCO Teacher Framework

- AI pedagogy – AI-enhanced pedagogical innovation
- AI for professional development – AI to support professional development (personalizzare e modificare strumenti AI)

UNESCO Students Framework

- AI techniques and Applications – Creating AI tools: "Customize existing AI toolkits and create AI tools... develop social & emotional skills..."
- AI system design – General description: "Systemic design thinking and comprehensive engineering skills..."
- AI system design – Problem scoping: "Understand when to use AI... devise and plan implementation"
- AI system design – Architecture design: "Acquire skills to design an AI system architecture, build prototypes"
- AI system design – Iteration and feedback: "Evaluate and improve AI models... develop identity as co-creators"

E. Ethical, Self & Social Domain

AILit Framework

- Self and Social Awareness (recognise AI influence on choices, relationships, communities; reflect on societal & environmental impact)
- Self & social awareness – recognise how data practices affect different people and communities

UNESCO Teacher Framework

- Human-centred mindset – Human agency Ethics of AI – Ethical principles
- Human-centred mindset – Human accountability Ethics of AI – Safe & responsible use
- Human-centred mindset – Social responsibility Ethics of AI – Co-creating ethical rules

UNESCO Students Framework

- Human-centred mindset – Human agency: "Recognize that AI is the result of human decisions"

Co-funded by the European Union

- Human-centred mindset – Human accountability: "Recognize that human responsibility includes legal and social obligations"
- Human-centred mindset – Citizenship in the AI era: "Critically understand the social impact of AI, promote its responsible and inclusive use"
- Human-centred mindset – Students' critical thinking, sense of social and civic responsibility
- Ethics of AI – Ethics of AI: "Know how to make the ethical value judgments..."
- Ethics of AI – Embodied ethics: "Understand key ethical issues... (human rights, social justice, inclusion, equity, climate change)"
- Ethics of AI – Safe and responsible use: "Use AI in a safe and ethical manner, protecting data privacy and security"
- Ethics of AI – Ethics-by-design: "Integrate ethical principles from design through the AI life-cycle"

F. Interaction, Collaboration & Communication

AILit Framework

- Collaboration (work effectively with AI and humans, communicate clearly, give feedback, manage shared tasks)
- Communication (explain how AI is used, promote transparency, avoid anthropomorphism, encourage responsible use)

UNESCO Teacher Framework

- AI pedagogy – AI-assisted teaching
- AI for professional development – Enabling lifelong professional learning
- AI pedagogy – Pedagogy integration
- AI for professional development – AI to enhance organisational learning

GenAI Literacy

A. Core Knowledge of AI & Data

AILit Framework

- K1.3 – The Nature of AI: Generative AI uses probabilities to create human-like outputs but lacks true understanding
- K2.1 – AI Reflects Human Choices and Perspectives (knowledge aspect): Systems mirror the data, design decisions and labour behind them
- K4.3 – AI's Capabilities & Limitations: LLMs can blur fact and fiction, enabling misinformation and deepfakes

Co-funded by the European Union

DigComp 2.2

- Awareness that search engines, social media and content platforms often use AI algorithms for personalisation
- Awareness that AI algorithms often act as black boxes whose internal logic is not visible to users
- Awareness that the data on which AI depends may include biases that can be automated and amplified
- (46) Understanding how AI systems (e.g., recommender systems, image generators, voice assistants) work and when it is appropriate to use them

DigComp 3.0

- Know & understand AI – recognising AI, distinguishing what AI is / is not, grasping core concepts and techniques

UNESCO Teacher Framework

- Basic knowledge of AI concepts

UNESCO Students Framework

- Digital awareness

B. Data Stewardship & Model Documentation

DigComp 3.0

- 1.1 Browsing, searching, filtering

UNESCO Teacher Framework

- Validation of effective AI use strategies
- Use of data and feedback to innovate pedagogy

UNESCO Students Framework

- Data literacy

C. Cognitive & Analytical Domain

AILit Framework

- Critical Thinking: Judge AI-generated content for accuracy, fairness and bias

- Computational Thinking: Break problems down and craft instructions that AI can follow
- Problem Solving: Decide when and how to use AI by weighing its strengths, risks and ethics

DigComp 2.2

- Weighs the benefits and disadvantages of using AI-driven search engines (e.g., privacy risks vs convenience, commercial bias, etc.)

DigComp 3.0

- Use & apply AI – integrating AI outputs to accomplish tasks
- Evaluate & create AI – critical analysis & interpretation of outputs
- 1.2 Evaluating information
- 3.4 Programming & computational thinking (problem decomposition, algorithmic logic)
- Problem-solving aspect of 5.3 Creatively using

UNESCO Teacher Framework

- Ability to evaluate and use appropriate AI tools in educational contexts
- Use of AI for reflective practice, assessment of learning needs, and personalized professional learning path

UNESCO Students Framework

- Critical thinking
- Decision-making skills
- Systemic thinking

D. Creative & Innovative Domain

AILit Framework

- Creativity: Co-create with AI while managing attribution, ownership and responsible use

DigComp 2.2

- (33) To know how generative-AI tools (text, image, music) can be used to create content

DigComp 3.0

- Evaluate & create AI – designing/coding AI applications (creative build side)
- 3.1 Developing digital content
- 3.2 Integrating & re-elaborating content
- 5.3 Creatively using (generative remix, artistic experimentation)

Co-funded by the European Union

- Creating aspect of generative-AI tools from earlier lists

UNESCO Teacher Framework

- Competent operation of AI tools
- Skills to customize and modify AI tools for inclusive learning environments and professional development
- Use of AI tools to improve lesson planning, teaching, and assessment
- Integration of AI in student-centered teaching practices
- Design and facilitation of AI-supported learning scenarios
- Customization of AI tools to support personal and community professional growth
- Experimentation with innovative strategies

UNESCO Students Framework

- Design skills
- Creativity
- Co-creator identity

E. Ethical, Self & Social Domain

AILit Framework

- K2.1 – AI Reflects Human Choices and Perspectives (ethical dimension): Highlights unequal labour conditions and assumptions embedded in AI
- K3.3 – AI Reshapes Work & Roles: Humans remain accountable for judgement-laden decisions
- K5.3 – AI's Role in Society: GenAI raises questions of authenticity, authorship and copyright
- Self & Social Awareness: Reflect on how AI shapes personal choices and community

DigComp 2.2

- (17) To understand how AI can be used to produce and distribute disinformation
- (47) Understanding the environmental impact of training large AI models

DigComp 3.0

- Ethics & societal implications of AI – fairness, accountability, transparency, sustainability
- 2.3 Engaging in citizenship (digital civic participation)
- 4.2 Protecting personal data & privacy
- 4.3 Protecting wellbeing

UNESCO Teacher Framework

Co-funded by the European Union

- Critical understanding of AI's impact on human rights, autonomy, and agency
- Application of fundamental ethical principles for human-AI interaction
- Assuming human responsibility in AI use and critical evaluation of AI's role in decision-making
- Integration of ethical rules for safe and responsible AI use
- Advocacy for AI ethics
- Contribution to building inclusive societies

UNESCO Students Framework

- Ethical skills
- Civic responsibility
- Autonomy
- Adaptability
- Environmental awareness
- Cross-cultural

F. Interaction, Collaboration & Communication

AILit Framework

- Collaboration: Work productively with both humans and AI, giving and receiving feedback
- Communication: Explain AI use transparently, avoid anthropomorphism and promote responsible practice

DigComp 2.2

- (21) Providing feedback to AI systems (voice assistants, chatbots) to improve interaction

DigComp 3.0

- Providing feedback to AI systems (from earlier list – voice assistants, chatbots)

UNESCO Teacher Framework

- Active participation in AI-supported professional communities

UNESCO Students Framework

- Social-emotional skills
- Teamwork skills
- Complex communication

Co-funded by the European Union

Annex 7C Analysis of the existing projects

1. Introduction

We conduct an analysis of European projects on data literacy by retrieving the list of initiatives granted under the Erasmus+ program.

The analysis reveals a dynamic and evolving landscape of educational innovation and collaboration. With a total of **100 projects** examined, the dataset provides a rich ground for understanding trends and priorities in this crucial area.

The Fig. 1 illustrates the temporal distribution of European projects focused on data literacy, categorized by their funding year. From the visual trend, it's clear that **interest and investment in data literacy projects have grown steadily over time**. While earlier years show a more modest number of funded initiatives, the more recent years demonstrate a notable **increase in project volume**, suggesting a rising prioritization of data literacy within the European Union's educational agenda. This growth may be tied to broader digital transformation strategies, the emphasis on 21st-century skills, and the European Commission's commitment to digital education as articulated in frameworks like the Digital Education Action Plan.

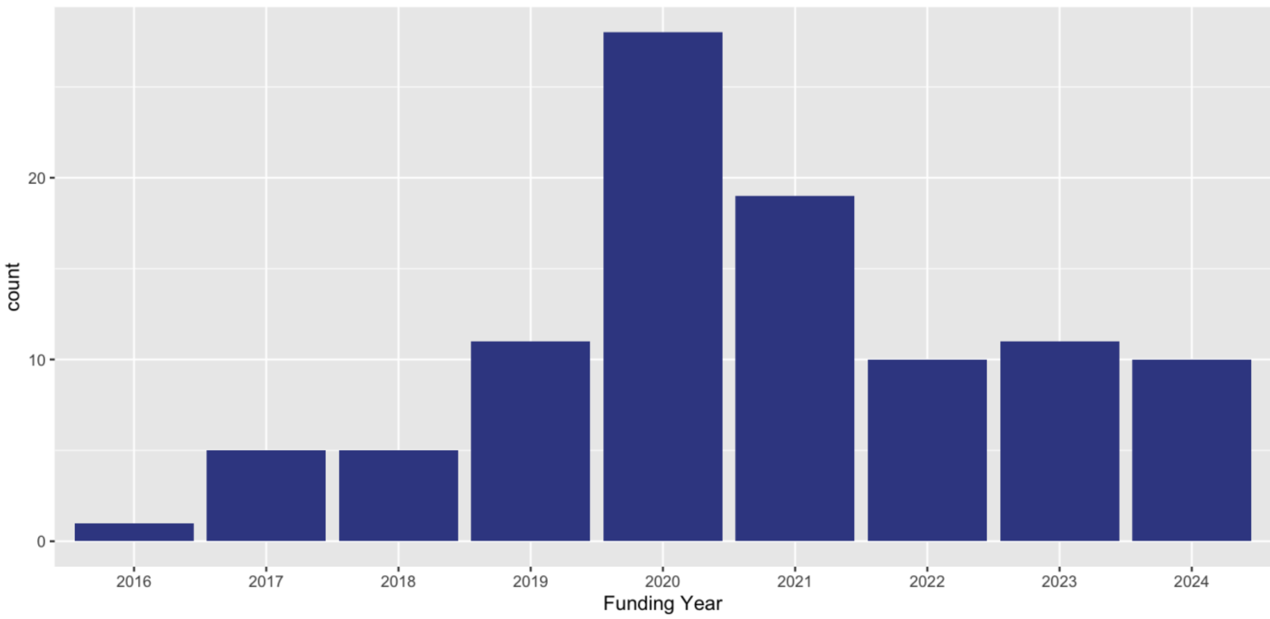


Figure 1: "Data Literacy" European projects by year

The second graph organizes projects according to their **Key Action (KA)** under the Erasmus+ framework[4]. It reveals a concentration of projects under **Key Action 2 (KA2)** – "Cooperation for

[4] The categories of action types reported in the following analyses follow the different organisations of the Erasmus+ programmes that have succeeded one another over time. In order not to alter the analyses, we have chosen to retain the original nomenclatures, not least because there is often no

innovation and the exchange of good practices." This indicates that **most data literacy projects are collaborative in nature**, bringing together institutions across borders to co-develop and share innovative educational practices, resources, and tools. In contrast, **Key Action 1 (KA1)** – "Learning mobility of individuals" – is less represented, implying that while mobility and personal development are present, the emphasis is more on institutional partnerships than on individual exchanges in the context of data literacy.

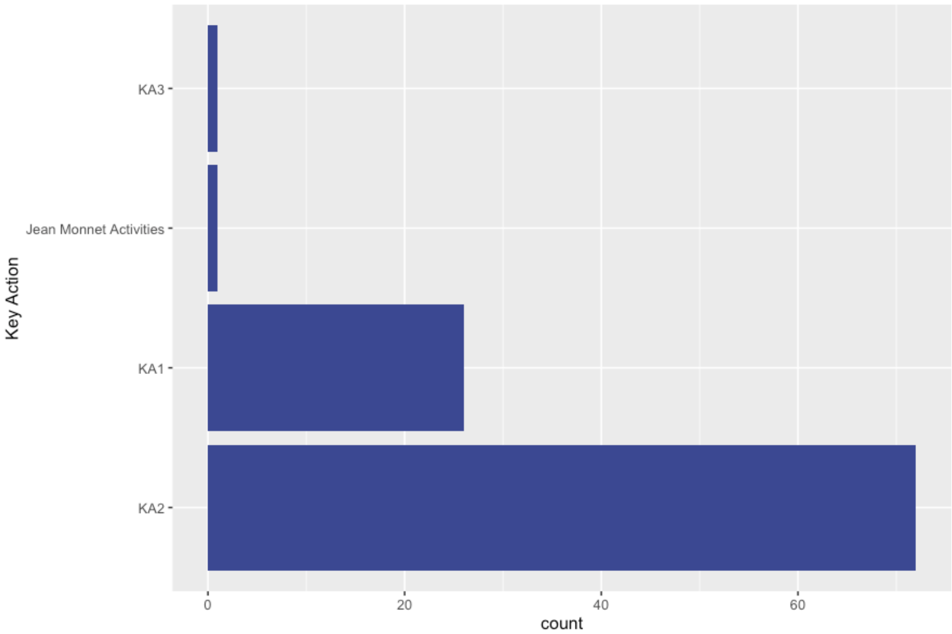


Figure2: "Data Literacy" projects by Key Action

The third graph illustrates the details of the types of actions funded within KA2. In particular, the most common action types are:

- **Strategic Partnerships for school education**
- **Strategic Partnerships for adult education**

This suggests that the integration of data literacy spans both general and vocational education systems. Projects targeting **school education** are likely aimed at embedding data literacy from an early age, fostering critical thinking and digital competence among young learners. Meanwhile, the focus on **adult** points to the importance of equipping the current and future workforce with data-related skills, aligning with labor market demands.

There are also instances of **KA1 projects (Figure 4)**, particularly in the form of **staff mobility**, which likely support the professional development of educators in the field of data literacy. These initiatives

exact correspondence between the categories of the different structures. The last structure could be find at https://erasmus-plus.ec.europa.eu/programme-guide/part-a/priorities-of-the-erasmus-programme/structure help teachers and trainers gain exposure to new methods and tools, promoting a broader dissemination of data literacy practices across Europe.

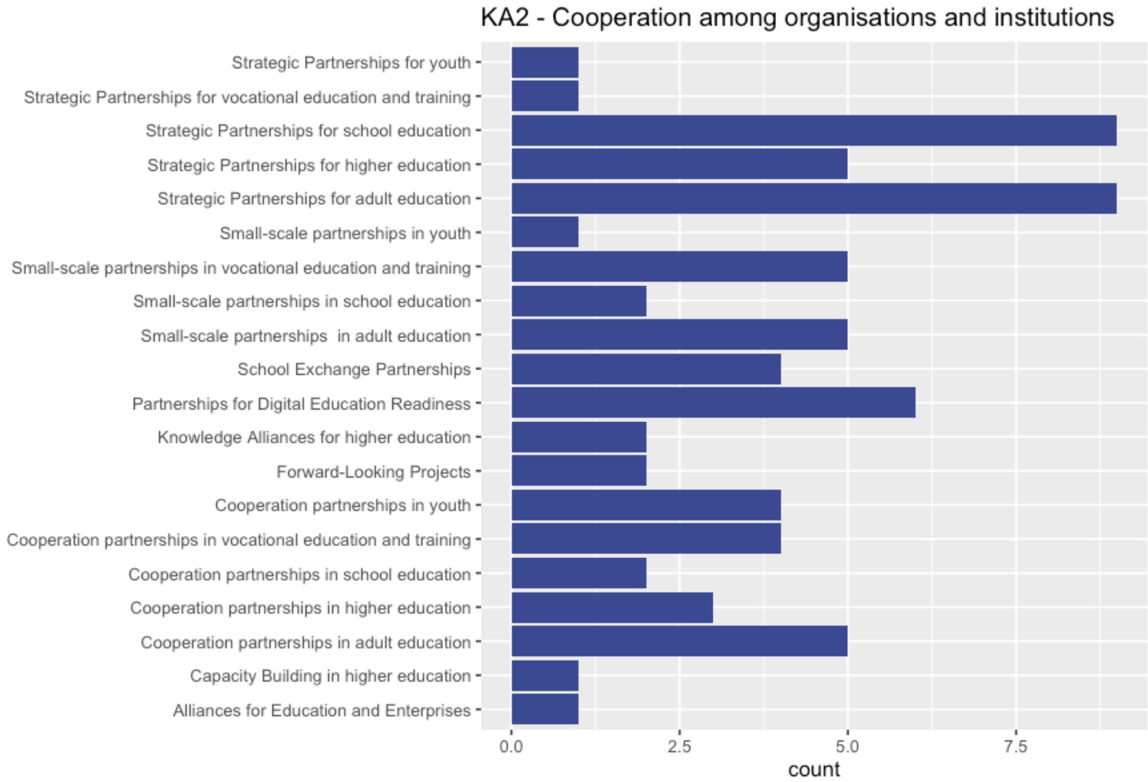


Figure 312: KA2 Data Literacy projects

Co-funded by the European Union

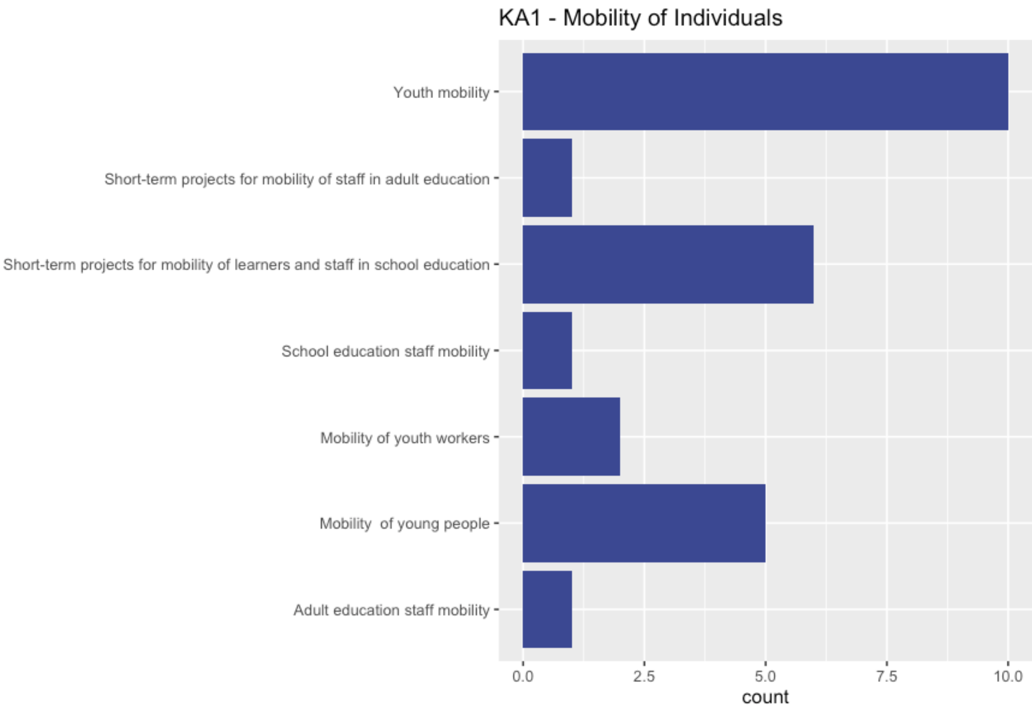


Figure 13 KA1 Data Literacy projects

Finally, the word cloud offers a qualitative glimpse into the **most frequently occurring terms** across the 100 European projects on data literacy. Visually, the size and prominence of each word reflect its frequency and relevance within the project descriptions, objectives, and themes.

At the center of the cloud, the **most prominent terms** include:

- **"Data"**, unsurprisingly, as the core concept of all projects analyzed.
- **"Digital"** and **"Literacy"**, which reaffirm the overarching goal of equipping learners with the ability to understand, interpret, and use data effectively in digital contexts.
- **"Education"**, **"Students"**, and **"Teachers"**, which indicate the primary target groups and sectors involved.
- **"Skills"** and **"Competences"**, reflecting the emphasis on building practical, applicable abilities in working with data.
- Terms like **"ICT"**, **"Tools"**, **"Platform"**, and **"Training"** suggest a focus on technology-enhanced learning environments and the development of resources to support instruction.

Other frequent terms such as **"Critical"**, **"Thinking"**, and **"Media"** highlight the intersection of data literacy with **critical media literacy** and informed citizenship - key goals in combating misinformation and promoting digital resilience.

The word cloud thus reinforces the themes seen in the graphs: **education-driven, skill-oriented, and future-focused initiatives** aimed at fostering data-literate individuals and institutions. It also underscores the multidisciplinary nature of these projects, which cut across education, technology, communication, and social inclusion.

Co-funded by the European Union

Figure 5: Words cloud of projects keywords

Overall, the analysis draws a coherent picture: data literacy is gaining traction as a key educational goal across Europe, with a variety of actors involved and a noticeable increase in dedicated initiatives over the years. This growing commitment lays the foundation for more informed, empowered, and critically thinking European citizens.

Starting from the collected list of projects, we analysed the project cards reported on the Erasmus+ platform and the projects' websites. The aim was to identify the main results emerging from the projects' activities, and particularly the results that could be useful for the definition of the AI-DL Framework. From this analysis of the projects' results, seven projects have been selected, including the Data Literacy for Citizenship (DALI) project, the ongoing Data Literacy for Upper Primary Schools project (DALI4US), the Teacher Training for Data Literacy and Computer Science Competences (TRAIN-DL), the MILES Project – MIL and Pre-bunking approaches for Critical thinking in the education sector and finally EDUCABILITY: Building the Capacity of Educators and Librarians in Information Literacy.

2. DALI Data Literacy Framework

The DALI – Data Literacy for Citizenship project (2020-1-NO01-KA204-076492) was launched in 2020 and ends in 2023.

- The project aims to strengthen citizens' capacity to navigate an increasingly data-driven society. Its core objectives are to:
- Define the competences that characterise a data-literate citizen, based on a comprehensive and research-informed perspective.
- Support individuals in acquiring and further developing essential data-related skills, enabling them to understand, interpret, manage, and use data confidently.
- Increase participation in lifelong learning, using effective strategies for outreach, guidance, and learner motivation.

The project's principal achievement is the DALI Data Literacy Framework, a structured model describing the knowledge, skills, attitudes, and critical capacities required to engage meaningfully with data in everyday life.

The framework is organised into three main components:

- Understanding Data – encompassing knowledge, awareness, and critical thinking.
- Acting on Data – covering the practical skills needed to collect, manage, analyse, and communicate data.
- Engaging Through Data – highlighting the ability to use data for informed decision-making, civic participation, advocacy, and collective action.

A fourth transversal dimension, Ethics & Privacy, runs across all components, emphasising responsible and ethical data practices.

The full framework, including progressive levels of expertise (from basic to advanced) for each competence area, is available for download in English, Spanish, German, and Norwegian.

Summary of the project's activities

The DALI Data Literacy Framework was developed through a structured, research-based, and participatory process designed to define the essential competences that characterise a data-literate citizen. Its development followed several key methodological steps:

Identification of Core Competences Through a Delphi Study

The foundation of the framework was built through a Delphi study involving experts in data literacy, education, digital competence, and related fields. This structured, multi-round consultation identified the key knowledge, skills, and attitudes citizens need to engage with data in everyday life. The Delphi process allowed consensus to emerge on the essential elements of data literacy, ensuring that the framework reflects expert understanding, societal needs, and diverse perspectives.

Co-funded by the European Union

Organisation of Competences Into Three Main Dimensions

The expert-validated competences were structured into three main components, representing the ways citizens interact with data:

- Understanding Data (knowledge, awareness, critical thinking),

- Acting on Data (collecting, managing, and sharing data),

- Engaging Through Data (policy, decision-making, data activism, data advocacy).

A transversal dimension, Ethics & Privacy, was incorporated across all components, recognising that responsible data practices underpin all data literacy skills.

Definition of Progressive Levels of Expertise

To support educational adaptation and competence development, the framework introduces three progressive levels:

Level A – basic knowledge and introductory skills,

Level B – intermediate competence and active engagement,

Level C – advanced ability to act, influence, and advocate through data.

For each sub-element, indicators were defined to help educators and designers of learning materials identify what learners should know or be able to do at different stages.

Integration of Inquiry-Based and Critical Perspectives

Data literacy was conceptualised not only as technical proficiency but also as the development of:

- inquiry skills (asking questions from data),
- critical judgement (evaluating data, claims, representations, and impacts),
- ethical awareness (privacy, bias, surveillance),
- civic competences (engaging in societal issues using data).

This ensured that the framework reflects a holistic, human-centred, and democratic view of data literacy.

Validation and Refinement Across Project Partners

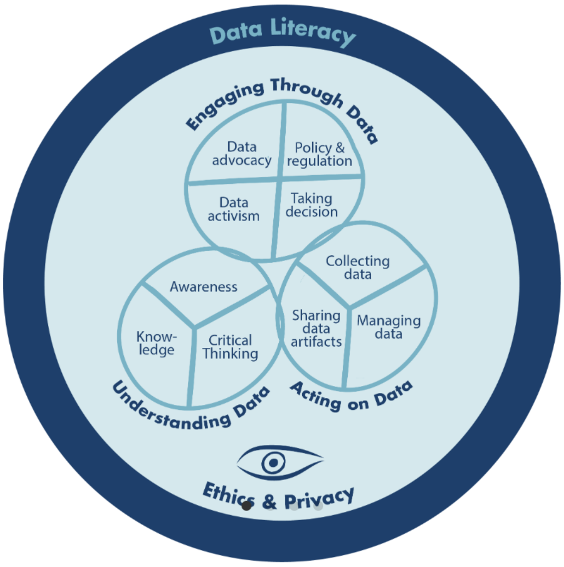
The framework was reviewed, discussed, and refined collaboratively by the DALI project partners across several European countries. This transnational validation ensured cultural relevance, applicability in multiple contexts, and alignment with broader European digital competence initiatives.

Final Result

The final DALI Data Literacy Framework is a comprehensive model describing what it means to be a data-literate citizen. It provides a structured set of competences, organised into three main

Co-funded by the European Union

components and three levels of progression, supported by a transversal ethical dimension. The framework is designed to guide the creation of learning activities, courses, and tools aimed at empowering citizens to understand, use, question, and act on data in everyday life.



DALI framework

Understanding Data

Knowedge

Citizens will know:

- what data is, what form it takes and how it can be used in society
- that data has different sources, types and formats
- how data can be collected from different environments
- there are technological pre-conditions for data creation and use
- that data is processed and manipulated by algorithms and apps
- that data is persistent and potentially stored
- the concepts of data security, data surveillance, big data and small data
- their data rights

Awareness

Citizens will be aware that:

- they generate data using apps, websites, driving their car, etc.
- there exists data about them in profiles (My Data)
- data is a representation of reality; it is not reality itself
- data is complex, and that there are variations in complexity of data

- there are potential and drawbacks of big data in different realms of society such as health, education, economics, security, etc.
- data is monetised (e.g., “data as the new oil”)
- there are trade-offs when sharing your data

Critical Thinking

Citizens will be able to think critically about:

- the relationship between humans and data: the use automated processes vs human actions
- who is making the decision (the human or the algorithm)
- how data tools work
- data being used for targeted advertisements
- the misrepresentation of data
- how data is monetised, for which purposes it is being collected

Acting on Data

Collecting Data

Citizens will have the skills to:

- configure privacy settings, revoke access, request to have your data erased
- use collected data to change your own behaviour (e.g., from a health app)
- make informed decisions when interacting with data-collecting actors (e.g. mobile apps, internet portals and employers)

Collecting Data

Citizens will have the skills to:

- organise data
- process, protect and store their own data
- move data from one application to another
- manipulate data
- evaluate the quality of data
- identify misrepresentation of data

Collecting Data

Citizens will have the skills to:

- synthesise, visualise and represent data in different formats
- translate data into everyday language (e.g., tell a story about data)
- share their data through an open repository

Engaging Through Data

Policy and regulation

Citizens will have the knowledge and skills to:

- participate in data-based policy-making processes (C)
- interact with key stakeholders (e.g., data protection agencies) as needed for the resolution
- of issues related to data use (their own or other’s data) (I & C)
- Taking Decision

Co-funded by the European Union

Citizens will have the knowledge and skills to:

- make their own decisions based on critical consideration of data (personal, professional, etc.)
- (I)
- understand the balance between individual and social benefits and the risks related to data
- use (I & C)
- be aware of their own role in acting on data from their different roles (professional, parent,
- citizen, etc.)

Data activism

Citizens will have the knowledge and skills to:

- use data as a basis or activism for data engagement (C)
- put their data rights into practice
- self-regulate their own data footprint

Data advocacy

Citizens will have the knowledge and skills to:

- communicate data meaning to stakeholders or to other peers (C)
- raise collective data awareness about the possibilities and challenges in data use (C)

3. DALI4US - Data Literacy for upper primary schools

The Erasmus+ project DALI4US (Data Literacy for Upper Primary Schools) aims to equip primary school teachers with the necessary skills to effectively teach data literacy. The initiative develops an evidence-based data literacy framework, including tools and resources to help students interact critically with data. Among the main resources is OrangeEDU, an adapted version of Orange Data Mining software for primary education, which incorporates gamification elements. The project involves partners from several European countries, such as Slovenia, Luxembourg, and Ireland, and focuses on professional training for teachers. School pilot projects are planned to evaluate and refine the effectiveness of the proposed solutions with the aim of preparing students for a data-driven society.

DALI4US is a three-year project, currently in progress. The project activity includes the following phases:

- Year 1: In this initial phase, teachers have already started to be involved. For example, teachers from Slovenia, Luxembourg, and Ireland participated in the first data literacy workshop held in Lasko, Slovenia, in March 2024. Throughout the three years of the project, a series of professional training programs will be designed to improve teachers' competence in teaching data literacy, enabling them to effectively integrate these concepts into their classrooms.
- Years 2 and 3: In these years, the DALI4US project will work with primary schools in the participating countries to conduct the piloting activities and evaluate the effectiveness of the proposed solutions. Through an iterative process of testing and refinement, the project aims to develop a sustainable and scalable model. This process will help develop an evidence-based

data literacy framework and create a digital data mining ecosystem called OrangeEDU, as well as professional training programs for teachers.

3.1 Summary of the Process Used to Develop the DALI4US Data Literacy Framework

The DALI4US framework was developed through a systematic, multi-step process grounded in needs analysis, established statistical education models, and iterative refinement. The goal was to create a practical, developmentally appropriate data literacy framework for upper primary students. The process can be summarised in the following key phases.

Needs Analysis Across Partner Countries

The project began with a structured needs assessment in Luxembourg, Ireland, and Slovenia to identify:

- existing curriculum expectations regarding data competencies,
- teacher challenges and training needs,
- the need for user-friendly tools (e.g., OrangeEDU),
- the importance of leadership engagement,
- and the need for iterative framework testing.

Review of Existing Definitions and Models of Data Literacy

The team conducted a literature review to identify key components of data literacy, analysing definitions and sub-competencies such as reading, working with, analysing, and communicating data.

Selection of Foundational Frameworks

DALI4US adopted two major statistical education models:

- the PPDAC cycle (Problem, Plan, Data, Analysis, Conclusion),
- and the GAISE II framework, which emphasises inquiry, authentic datasets, visualisation, and technology use.

Construction of the Initial DALI4US Framework

The first version synthesised PPDAC, GAISE II, the DIKW model, and the “three C’s” (comprehension, communication, critical thinking). It emphasised contextualisation, cross-curricular learning, and alignment with DIGCOMP.

Iterative Development Through Testing and Feedback

The framework went through multiple cycles of testing, feedback collection, and refinement to ensure practicality and curricular alignment.

Integration of Exploratory Data Analysis (EDA)

Co-funded by the European Union

To overcome limitations of a linear confirmatory approach, the framework was expanded with EDA principles inspired by Tukey. This added flexibility, pattern discovery, and creativity. The extended cycle includes:

- Trigger
- Collect
- Organize
- Explore
- Patterns
- Predict
- Reflect
- Share

Final Result

The final DALI4US framework is a hybrid, research-based model combining PPDAC, GAISE II, DIKW, critical-creative competences, technology integration, and EDA. It is designed specifically for upper primary learners and will guide training, resource development, and classroom implementation.

3.2 Analysed frameworks

PPDAC Cycle

Source: https://dataschools.education/about-data-literacy/ppdac-the-data-problem-solving-cycle/

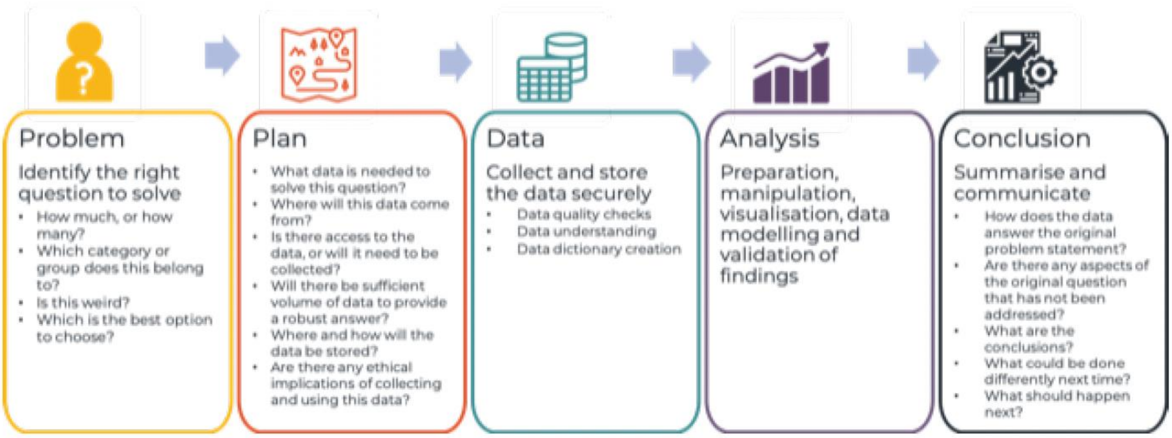


Figure 2:PPDAC cycle (Source:https://dataschools.education/about-data-literacy/ppdac-the-data-problem-solving-cycle/)

Problem

The first step is to identify and define the problem to be addressed. This includes formulating a statistical research question that will guide the entire investigation.

Plan

The next stage is the planning stage. In this phase a plan is drawn up for collecting the data needed to answer the research question. This includes selecting appropriate data sources, determining appropriate data collection methods, and selecting the tools and technologies to be used.

Data

In this phase, data is collected according to the plan established earlier. This may involve collecting new data through surveys or experiments, or using pre-existing data sets originally collected for other purposes.

Analysis

Once the data have been collected, they are analysed to extract valuable insights. This phase involves using statistical methods and tools to interpret the data, identify patterns and draw meaningful conclusions.

Conclusion

The final stage is to evaluate the results of the analysis and draw conclusions. This involves reflecting on the original research question, evaluating the results in relation to the research question, and considering the wider implications of the results.

Gaise II Framework

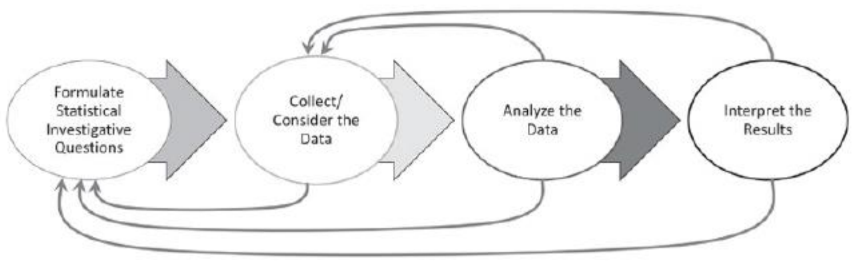


Figure 3: GAISE II Framework (Source: https://www.nctm.org/uploadedFiles/GAISEIIPreK-12_Full.pdf)

Formulate statistical investigative questions

The formulation of statistical questions is a fundamental aspect of statistical literacy. It is important to understand when and how to formulate statistical research questions that are relevant and manageable for students, often using data they have collected or have access to.

Collect/Consider data

The next step is to collect and analyse the data. The ability to collect data through surveys, experiments or examination of existing data sets is developed, as is an understanding of the different types of variables (categorical or quantitative).

Co-funded by the European Union

Analyze the data

This stage involves the analysis of the data. Appropriate graphical representations such as tables, scatterplots and bar charts are used to visualise the data. Students are introduced to basic statistical concepts such as mean, median, range and variability.

Interpret the results

The final stage of the process is to interpret the results. Students reflect on their results, evaluate them in the context of their initial question and learn about the limitations of generalising conclusions beyond the data set. The interpretation of results is an essential aspect of data analysis. At this stage, students begin to extrapolate their results beyond the immediate sample, taking into account uncertainty and variability in their conclusions.

Level B is for older students (intermediate to advanced) and builds on Level A. It involves formulating more complex questions, advanced concepts and more sophisticated analytical tools in data analysis and visualisation.

3.3 Dali4Us framework

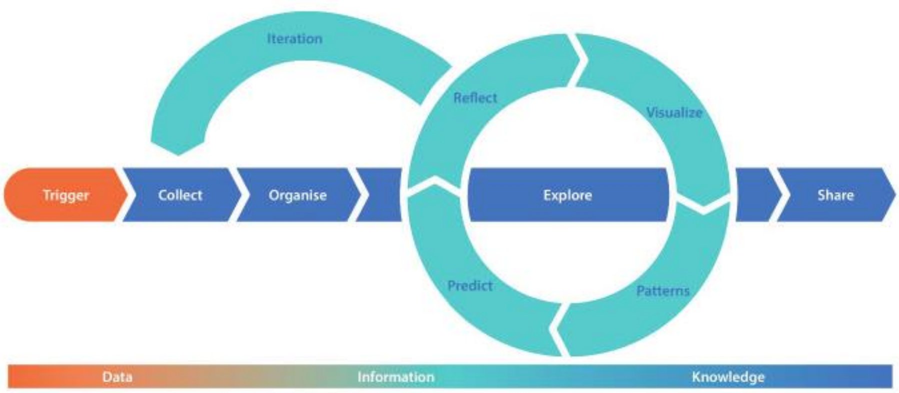


Figure 6: The extended DALI4US framework (own representation)

Trigger

This stage initiates curiosity, encouraging students to ask meaningful questions and engage with real-world data - aligning with Moore and Cobb's emphasis on using real data and fostering interpretation.

Collect

Gathering data from the world mirrors Moore's recommendation to ground learning in meaningful, authentic contexts, integrating computational tools as part of the process.

Co-funded by the European Union

Organize

Structuring data to facilitate easier analysis is consistent with the pedagogical goal of fostering skills in managing and preparing data for exploration and interpretation.

Explore

Students uncover initial insights, consistent with Tukey's emphasis on exploratory data analysis (EDA) as detective work, where patterns and trends are sought iteratively to build understanding.

Patterns

Identifying trends and relationships within the data reflects Moore's call for a focus on higher-order thinking, problem-solving, and making connections within the data.

Predict

Developing predictive models builds on this understanding, encouraging application and judgment, as highlighted in Moore's focus on combining computational activity with interpretation.

Reflect

Reviewing insights, predictions, and findings emphasizes critical thinking and self-assessment, encouraging a deeper engagement with data.

Share

Communicating and presenting findings aligns with Moore's stress on communication as a vital component of statistical practice.

4. Train-DL - Teacher training for Data Literacy & Computer Science competences

The TrainDL project ended in February 2024. It aimed to "disseminate AI and data literacy skills in all schools" by creating teacher training guidelines and recommendations for the "structural integration of AI and DL skills" (data literacy) in educational systems. The results of the project demonstrated the feasibility of teacher training in AI and data literacy. This highlights AI as a critical element that defines part of the set of data literacy skills that teachers need to acquire and integrate.

The project does not refer to GenAI but focuses more broadly on AI and data literacy in school education and teacher training. The documents produced cover competence areas, pedagogical approaches, teaching methodologies, and policy recommendations for the introduction of AI and data literacy.

Overview of the project's results

A variety of materials and results focusing on the current state of AI and data literacy in school education and teacher training have been produced. The results of the projects offer insights into teaching materials, concepts, and outcomes from policy experimentations.

The main results are organized into clusters.

Co-funded by the European Union

Education Cluster

In the education cluster, several results have been produced. The main results useful for the aim of this report are Deliverable 2.1 and Deliverable 3.1.

Deliverable 2.1 is a literature review titled "Introducing Artificial Intelligence Literacy in Schools: A Review of Competence Areas, Pedagogical Approaches, Contexts, and Formats". It examines practices and formats already evaluated with students, identifying those that are effective and those that require further exploration to facilitate AI teaching and encourage the development of new activities.

Deliverable 3.1 provides a research summary on Data Literacy (DL) and AI teaching methodologies, as well as primary teacher education formats, which encapsulates contemporary methods of teaching AI and data literacy and scrutinises the instructional resources offered in elementary schools.

Policy Cluster

This cluster focused on developing policy recommendations regarding data literacy and AI in school and teacher education, analysing existing works, and integrating these competencies.

The following deliverables fall within this area:

- Deliverable 1.1, a policy research summary aimed at developing a policy monitor on digital literacy, with a focus on assessing the state of data literacy and AI education.
- Deliverable 1.2, which reports the policy comparison among different school systems.
- Deliverable 5.1 provides a summary of the research and state-of-the-art.
- Policy and curricula recommendations for computer science teachers (Deliverable 5.3) were developed and underwent two iterations of refinement and strategy testing.
- Prototype policy and curriculum recommendations for STEAM (Science, Technology, Engineering, Arts, and Mathematics) teachers (Deliverable 5.5) were created, building on the recommendations for computer science educators.
- Prototype policy and curricula recommendations for primary teachers (Deliverable 5.7) were also developed, similarly building on the recommendations for computer science teachers.
- A set of consolidated recommendations (Deliverable 5.9) is the latest report in this series, containing a compilation of recommendations for computer science, STEAM, and primary education, following the iterative TrainDL Policy Experimentation approach.

Evaluation Cluster

Also in the evaluation cluster, several results have been produced. This cluster developed guidelines for policy experimentation and supervised research partners in their implementation and evaluation of results.

The main result useful for the aim of this report is Deliverable 4.1, a State of the Art Report providing an overview of relevant literature and results from related projects that contribute to the further design of project activities.

Co-funded by the European Union

TRAIN-DL Framework

As part of the project, the partners conducted an in-depth analysis of the main Artificial Intelligence (AI) literacy and Data Literacy (DL) frameworks, with the aim of understanding how these models address the competencies required by K-12 teachers across different subjects. This activity builds on the study by Olari and Romeike (Addressing AI and Data Literacy in Teacher Education: A Review of Existing Educational Frameworks, WiPSCE 2021), which we cite as the reference analysis and of which we provide a summary below.

The article examines how current AI and data literacy frameworks support teachers' competences and highlights a significant gap: AI literacy and data literacy are generally treated as separate domains, despite being fundamentally interdependent.

Framework	Acquisition	Cleansing	Modeling	Implementation	Optimization	Analysis	Visualization	Evaluation	Sharing	Erasing	Archiving
Touretzky et al. (2019) – *Envisioning AI for K-12*			✓					✓			
Long & Magerko (2020) – *What is AI Literacy?*	✓		✓			✓		✓			✓
Blakeley & Breazeal (2019) – *Ethics of AI Curriculum (Middle School)*	✓							✓			
Williams et al. (2021) – *How to Train Your Robot (Teacher Perspectives)*	✓							✓			
Clarke (2019) – *AI Alternate Curriculum Unit*	✓					✓		✓			
Vazhayil et al. (2019) – *Teacher Education to Introduce AI in Schools*	✓							✓			
Higuera (n.d.) – *Report on Education, Training Teachers and Learning AI*											

Co-funded by the European Union

Summary of the review

Identified Problem

Although AI technologies are increasingly present in everyday life, many people - including teachers - still lack a clear understanding of how these systems function. In recent years, several educational frameworks have been developed to promote AI literacy, but:

- Data literacy is only minimally addressed.
- Data literacy frameworks rarely make explicit reference to AI.
- No existing framework provides a holistic integration of AI and DL.

This separation is problematic because machine learning relies on data, and working with data requires at least a basic understanding of AI processes.

Methodological Approach

The authors reviewed major AI literacy frameworks - both for teachers and for K-12 students—through the lens of data literacy.

The analysis used the data lifecycle (acquisition, cleansing, modeling, analysis, visualization, evaluation, sharing, erasing, archiving) as a conceptual grid.

The guiding question was:

- Which stages of the data lifecycle are covered by current AI literacy frameworks, and in what way?

Main Findings

The review shows that existing AI literacy frameworks include only fragmented and isolated elements of data literacy, and none of them cover the entire data lifecycle.

Most frequently covered stages

- Acquisition: collecting datasets, ensuring diversity, selecting training data.
- Evaluation: training models, understanding bias and fairness.

Occasionally mentioned

- Modeling: understanding that models are built from data.
- Analysis: using clean/labeled data to train simple models.
- Archiving: privacy concerns related to stored personal data.

Not addressed at all

- Cleansing
- Implementation
- Optimization
- Visualization
- Sharing
- Erasing

Co-funded by the European Union

The results clearly show that AI literacy frameworks are incomplete from a data literacy perspective.

Contribution of the Study

Olari and Romeike propose using the data lifecycle as an organizing structure to design integrated AI + DL curricula for teachers. This would allow:

Anchoring AI-related competencies within each stage of data work.

Making explicit the connections between AI processes (training, evaluating, optimizing models) and data skills (cleaning, visualizing, managing, sharing data).

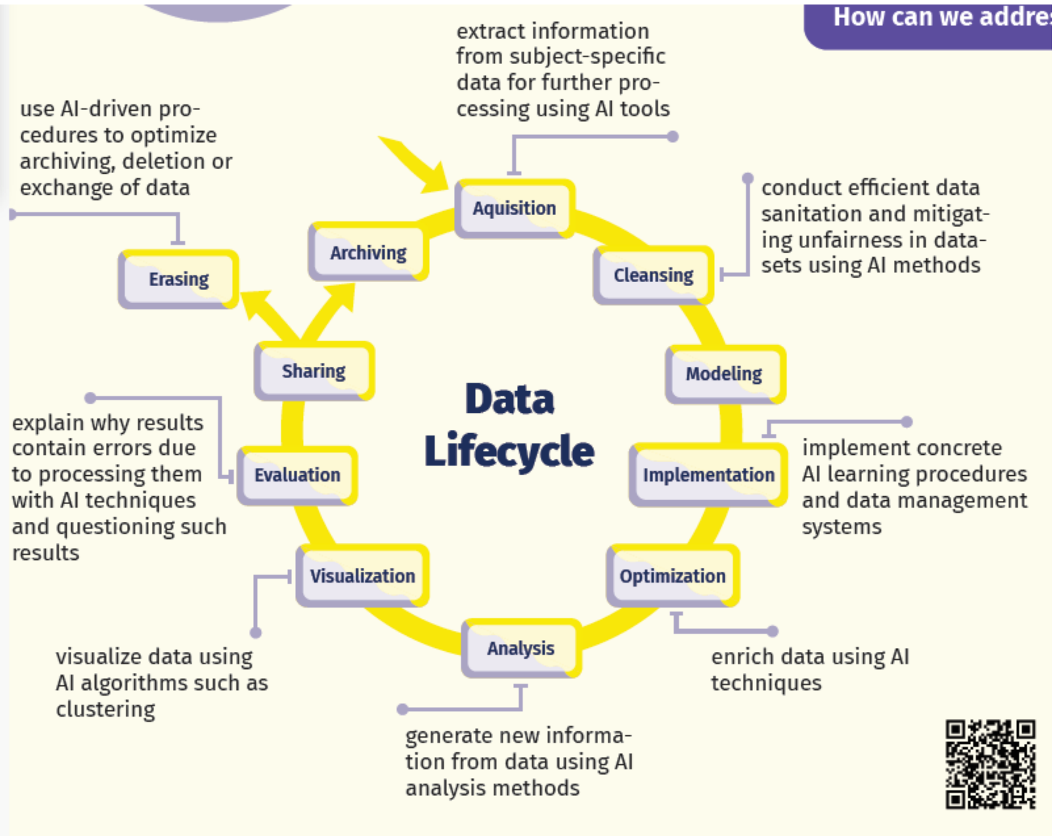
Such an approach could guide the development of holistic teacher education programs that prepare K-12 teachers from all subject areas to understand and teach both AI and data responsibly.

Conclusions

The study concludes that:

- Existing AI literacy frameworks provide insufficient and fragmented coverage of data literacy.
- A combined and holistic approach is urgently needed.
- The data lifecycle offers a promising structure for integrating AI and DL.
- Future work should evaluate integrated educational interventions for teachers based on this model.

Framework



Co-funded by the European Union

Acquisition

Extract information from subject-specific data for further processing using AI tools.

Cleansing

Conduct efficient data sanitation and mitigate unfairness in datasets using AI methods.

Modeling

Definition not available

Implementation

Implement concrete AI learning procedures and data management systems.

Optimization

Enrich data using AI techniques.

Analysis

Generate new information from data using AI analysis methods.

Visualization

Visualize data using AI algorithms such as clustering.

Evaluation

Explain why results contain errors due to processing them with AI techniques and question such results.

Sharing

Definition not available

Archiving

Definition not available

Erasing

Use AI-driven procedures to optimize archiving, deletion, or exchange of data.

5. MILES - MIL and PRE-BUNKING approaches for Critical thinking in the education sector

The MILES project (MIL and PRE-BUNKING approaches for Critical thinking in the education sector) is an ongoing cooperation initiative funded by the Erasmus+ program, involving 11 partners from 10 European Union countries.

Its main purpose is to tackle the growing challenges of disinformation and digital manipulation that directly impact the education sector and the formation of informed and critical citizens. The MILES project promotes **media literacy** (MIL), **pre-bunking**, and **data literacy** as essential skills in an age where anyone can quickly access information, often without checking sources.

The project aims to train teachers and future educators to be agents of transformation, equipped to strengthen resilience against disinformation and cultivate critical thinking in the classroom and in society.

Specific objectives of the MILES projects include:

- **Developing and implementing training programs for teachers and future educators** in media literacy, pre-bunking, and data literacy, covering two levels of training: Initial Teacher Education (ITE) and Continuous Professional Development (CPD).
- **Training teachers and educators** to become multipliers of critical thinking practices, creating a network of educators prepared to apply and disseminate innovative methodologies in their daily practices.
- **Integrating pre-bunking and data literacy methods into the school curriculum**, training teachers to develop pedagogical tools; and running workshops that involve students in identifying and critically analysing false and misleading information.
- **Engaging families and educational communities in digital literacy activities**, using resources such as student-developed board games to raise awareness about the dangers of disinformation and promote a culture of critical thinking beyond the school environment.
- Creating a lasting and sustainable impact on European digital education, **involving policymakers and educational stakeholders** to ensure the continuity and replicability of the methodologies developed.

Current results of the project

At the time of the current analysis, the only publicly available result is deliverable D2.1 (MIL & Debunking: datasheet, needs, good practices), developed within the WP2 (Mapping existing resources and training materials development) of the MILES project. The deliverable reports interviews, desk research, and online surveys in an initial research phase performed by the project's partners. The specific objective of this deliverable is to explore the current situation regarding media literacy (MIL) in their respective countries, identifying the main trends, learning needs, and inspiring practices that can serve as a reference for the next steps of the project.

The MILES project's research, conducted through desk research, interviews, and online surveys in 10 European countries (Italy, Portugal, Romania, Germany, Greece, Belgium, Cyprus, Austria, Netherlands, and Poland), revealed a diverse landscape of approaches and perceptions towards media literacy and critical thinking in addressing fake news.

Current Situation and Trends in Media and Data Literacy

Varying Digital Skills Across EU: As of 2023, 55% of individuals aged 16-74 in the EU possess at least basic digital skills, which include essential data literacy capabilities. However, significant disparities exist, with the Netherlands and Finland leading (79%), while Romania and Bulgaria are at the lower end (28% and 31% respectively). Higher education levels correlate with better digital skills, with large

Co-funded by the European Union

gaps noted in Portugal, Greece, and Malta, and smaller gaps in Estonia, Finland, and Lithuania. Younger individuals generally have higher digital skills compared to older age groups.

Media Literacy Confidence: Approximately 65% of Europeans feel confident in their media literacy skills, but significant country-level disparities exist. Northern European countries like Finland and Sweden are leaders in media literacy education, integrating it into their national curricula from early childhood (e.g., 90% of Finnish students receive formal MIL education). Countries such as Bulgaria, Romania, and Greece face challenges, including limited resources and insufficient educator training.

Data Literacy Gaps: A 2019 survey indicated that 58% of Europeans feel unprepared to handle data effectively. Western European countries (Germany, UK, France) have invested significantly, while Southern and Eastern European countries (Italy, Spain, Poland) lag behind.

Fake News Perception: 83% of EU citizens perceive fake news as a problem for democracy, with 73% concerned about its impact on elections. Social media platforms are seen as the main sources of fake news by 70% of respondents. The prevalence of disinformation varies, with higher incidences reported in Southern and Eastern Europe (e.g., 42% in Romania vs. 16% in Finland). Fake news often spreads faster due to its sensational nature and algorithmic amplification.

Main Needs and Challenges Identified by Stakeholders (including Policy Makers and School Actors)

Integration into Curriculum: There is a consensus on the urgent need to systematically incorporate comprehensive media literacy and critical thinking education into school curricula, but current initiatives are often sporadic and lack standardized guidelines or resources. The rapid advancement of technology makes it difficult for curricula to stay up to date.

Teacher Training and Readiness: A key challenge is the lack of continuous professional development for educators. Teachers often lack sufficient training, specific knowledge, and pedagogical strategies to effectively teach media literacy and combat disinformation. Many feel overwhelmed by the volume of digital tools and resources.

Resource Constraints: Schools frequently face limited access to modern technology, updated educational materials, and sufficient funding for specialized training programs.

Student Engagement: While students are proficient in using digital platforms, they often lack the critical thinking skills to evaluate the reliability and credibility of online information, accepting content at face value. Engaging them in critical thinking, especially on emotionally charged topics, is challenging.

Trust Erosion: Public trust in traditional media and institutions has eroded, making people more susceptible to fake news and driving them to less reliable alternative sources.

Educational Disparities: Unequal access to media literacy education across Europe creates a digital divide, leaving some citizens more vulnerable to disinformation.

Co-funded by the European Union

Platform Role and Anonymity: Social media algorithms often amplify sensational or misleading content, and the borderless nature of the internet, coupled with anonymity, complicates regulation and accountability.

Policy and Collaboration: The absence of comprehensive education policies and coordinated efforts across governmental, educational, and media sectors is a significant challenge. Stakeholders emphasize the need for cohesive policies and strong partnerships. Balancing regulatory measures with freedom of speech is also a concern.

Promising Practices and Resources Highlighted

The report identifies 44 practices, with 8 specific examples detailed:

Generazioni Connesse (Italy): A multi-layered platform supported by the Ministry of Education, developing educational materials, guides, and training programs for teachers, parents, and students to promote digital, media, and information literacy.

Infuzarea educatiei media (Romania): A methodology guidebook for Romanian language and literature teachers, offering practical resources and 25 lesson plans to infuse media literacy into formal education.

Podcast – the Truth in Times of Corona (Germany): An audio podcast created by the Bundeszentrale für politische Bildung, dedicated to conspiracy theories around the Corona virus, offering deep understanding through expert discussions.

Mediterranean Digital Media Observatory (MedDMO) (Cyprus): A regional hub of the European Digital Media Observatory (EDMO) covering Greece, Cyprus, and Malta, focusing on collaborative fact-checking and promoting media literacy and news verification methods.

#FakeHunter-Edu (Poland): An educational project for secondary school students and teachers, providing video training, activity kits, and lesson plans to raise awareness about disinformation, verify online content, and practice safe internet behavior.

The Fake News Commissioner is on the loose (Austria): An online game combined with psychological inoculation theory to enhance media literacy and resilience against fake news among students, featuring an interactive game and educational materials for teachers.

Stampmedia (Belgium): The first youth media agency in Flanders, offering training programs and practical opportunities for young people to experiment with media production, foster critical media and data literacy, and engage in active citizenship.

Be Internet Awesome & "Interland" Game (Greece): A Google-developed campaign with a serious game (Interland) designed to educate children, teachers, and families on digital risks, online safety, and digital citizenship in an interactive and fun manner.

Other noted resources and practices include: IREX's Learn to Discern Program (Poland), NASK's Department for Counteracting Disinformation (Poland), Digital Poland Foundation's 'Digital Youth'

Program, public websites for verifying videos/articles, cybersecurity workshops, and continuous professional development for educators.

Insights from Questionnaires (Adults and Students)

Demographics: The survey involved 1159 respondents (795 adults, 364 minors). Adults were primarily aged 25-35, with a majority identifying as female (58%). Most adults held bachelor's or master's degrees, and many were educators/teachers (197) or university students (187). Student respondents were mostly aged 16-18, followed by 12-15.

News Consumption (Adults): Digital platforms, especially social media, are the predominant news sources, followed by online newspapers/websites, and then television, radio/podcasts, and printed media.

Trust in News Sources (Adults): Social media platforms (TikTok, Twitter, Facebook) are most likely to spread fake news, according to respondents. Online newspapers/websites, television, radio/podcasts, and printed media are generally viewed as more reliable.

Perception of Fake News (Adults): The vast majority of respondents (blue segment dominating chart) have encountered what they believe to be fake news. Most believe fake news has a substantial impact on democracy, personal life, health, and politics.

Credibility Assessment (Adults): Respondents primarily assess credibility by checking sources and cross-referencing with other media outlets. The "source" and "author" are considered the most important criteria.

Information Seeking (Adults): A majority frequently or very often seek additional information to verify news accuracy, mainly using fact-checking websites and trusted news sources.

Media Literacy Education (Adults): 51% reported receiving some formal or non-formal media literacy education, but nearly half (49%) have not. Face-to-face training and watching videos were the most used means for skill development. Online courses and in-person learning are preferred formats, along with articles and videos. Topics of interest include identifying fake news, accessing suitable media, and assessing quality media.

Student Findings:

Fake News Definition: Students understand fake news as content purposefully written to be incorrect, false/fabricated, or false stories spread online.

Identifying Fake News: They primarily rely on fact-checking, mainstream media denouncement, or their own judgment.

Social Media Usage: TikTok, Instagram, and WhatsApp are the most popular platforms. Social media is the most frequently used news channel.

Co-funded by the European Union

Exposure to Fake News: A significant portion (largest segment) are "not sure" if they've encountered fake news, while some say "yes". Most students report "almost never" encountering fake news, and generally believe friends/family do not share fake news on purpose.

Flagging: Most students (larger segment) have not flagged fake news themselves or know anyone who has.

Discussions: Fake news is most commonly discussed at school, then with friends and family.

Adequate Teaching: A significant portion feel they have "not" been adequately taught to deal with fake news or are "not sure".

Impact on Behavior: Students generally perceive fake news as having a "moderate" impact on their own and their friends' behavior.

Trustworthiness of Sources: Students have a mixed perception of social media's trustworthiness (tendency towards moderate trust but also skepticism). Online newspapers/websites are generally rated higher in trustworthiness, followed by television, radio, and printed newspapers/magazines.

In conclusion, the report highlights the critical need for comprehensive and coordinated media literacy initiatives across Europe, addressing educational gaps, supporting educators, and leveraging technology to build a more resilient and critically thinking society capable of navigating the complexities of the modern information landscape.

6. Educability: Building the Capacity of Educators & Librarians in Information Literacy

The EDUCABILITY project, started in the autumn of 2020 and completed in 2023, aimed to fulfil urgent requirements of the Information and Knowledge Society:

1. Development of an electronic, unified, and freely available **Information Literacy Training Package (ILTP)** for educators and librarians.
2. Free, asynchronous, distance learning for educators of all levels and librarians on the core competencies of information literacy as a horizontal objective and on six emerging competencies as vertical objectives: digital literacy, mobile literacy, media and information literacy, critical information literacy, data literacy, and sustainable development literacy.
3. Convergence of strategy, know-how, and infrastructure for information literacy initiatives through a transnational memorandum of cooperation, sustainability, and transferability between EDUCABILITY partners

Current results of the project

The first intellectual output of the project is the **Transnational Information Literacy Ecosystem Mapping (TILEM)**, whose main purpose was to provide the basis for the construction of comprehensive curricula for the six emerging skills (https://educability.cut.ac.cy/I01).

Co-funded by the European Union

The definition of the TILEM involved mapping the state of the art in the field of information literacy and emerging literacies. The mapping was followed by the analysis of trainees' needs, the creation of a pool of experts, and the implementation of a Delphi study to test the validity of the proposed solutions.

The mapping was carried out through an extensive and systematic review of the international literature on the six information science competencies.

The most recent publications were examined, classifying them by type and relevance to the project objectives. Through text analysis, the reliability of sources, key concepts, and content was identified.

This information served as a reference for the development of six relevant curricula, summed up and coded into categories such as new definitions, key concepts and content, learning objectives and outcomes, teaching approaches, and assessment methods.

For each of the six literacies, the definitions available in the literature were analysed and classified, and from this analysis, the key concepts, learning objectives, and main skills to be developed were extracted.

The results were then validated by a group of experts through a Delphi study, which enabled the definition of these literacies to be refined.

The second intellectual output, **Six Information Literacy Learning Modules Curriculum Development (SILLMCD)**, provides a compilation of the modules developed (https://educability.cut.ac.cy/I02). The objective of the SILLMCD Report is to integrate and coherently present the content of all the literacies developed.

The Information Literacy Training Package (ILTP) consists of the following components:

- A generic information literacy curriculum template
- Critical Literacy (CL) curriculum
- Digital Literacy (DiL) curriculum
- Mobile Literacy (MoL) curriculum
- Media and Information Literacy (MIL) curriculum
- Data Literacy (DL) curriculum
- Sustainable Development Literacy (SDL) curriculum.

Each course consists of the following components:

- General information about the course
- Learning objectives and outcomes
- Learning content
- Complementary material for upskilling (e.g., suggested practices, strategies, resources)

Each curriculum operates independently; however, the six courses collectively constitute a comprehensive Information Literacy Training Package (ILTP) for educators and librarians.

Co-funded by the European Union

The curriculum design used several learning theories and pedagogical methods to provide end-users with a range of high-quality instructional strategies for developing distinct reading skills.

Each course included a variety of interactive activities, including quizzes, crosswords, interactive films, interactive books, interactive presentations, flashcards, and drag-and-drop exercises, allowing learners to obtain detailed feedback with the proper answers for each activity. Thus, educators and librarians acquire information and abilities that empower them to confidently design their own activities and interventions on many subjects.

Co-funded by the European Union

Annex 7D Survey for Ministries of Education: national policies, strategies and curricula on Data Literacy and Generative AI in post-primary education

Country:

1. Is there a national or regional policy or strategy on data literacy in post-primary education?
2. Is data literacy currently addressed in your national curriculum for post-primary education? If yes, please explain.
3. Has your Ministry taken any specific action to promote or support data literacy at the post-primary level?
4. Has your Ministry issued any guidance or policy documents related to the use of Generative AI in schools?
5. Are GenAI tools being used in post-primary schools in your country?
    - Yes – officially supported or endorsed
    - Yes – informally or at school level only
    - Pilot or experimental use
    - Not used
    - Not sure
6. Has your Ministry launched or supported any projects, pilots, or research involving data literacy and GenAI in education?
    - Yes
    - In planning stage
    - No

    If yes, please briefly describe:
7. Is there any teacher training or professional development offered related to:
    a. Data literacy
    - Yes
    - In development
    - No

    If yes, please specify:

    b. Using GenAI tools in teaching/learning
    - Yes
    - In development
    - No

    If yes, please specify:

Co-funded by the European Union

8. Are there any national or regional resources/platforms provided to schools that support:
   a. Teaching data literacy
      - Yes
      - No
      - Not sure

      If yes, please specify:

   b. Using GenAI in teaching/learning
      - Yes
      - No
      - Not sure

      If yes, please specify:

Co-funded by the European Union