https://doi.org/10.3311/celisr.38042

From data literacy to artificial intelligence literacy: background and approaches

Tibor Koltay
koltay.tibor@uni-eszterhazy.hu
Eszterházy Károly University, Institute of Learning Technologies
retired college professor

Received: 2024.07.28.
Accepted: 2024.09.04.
Published: 2024.09.13.

The objective of this study is to examine the characteristics of data literacy, which – on the long run – promises to become a fundamental component of artificial intelligence literacy (AI literacy). In addition to conducting a scoping review on the interrelated topics of data literacy and artificial intelligence literacy, we also draw upon our expertise in the field of data literacy and mention among others digital literacy, media literacy and their critical approaches.

In light of the considerable diversity of approaches and opinions, a significant portion of the extensive body of literature was subjected to careful examination, with a view to elucidating the nature and role of data literacy and AI literacy.

The issue of AI literacy is gaining increasing attention. It is therefore important to review its history and characteristics by examining the relationship between it and other forms of digital literacy, and in particular data literacy.

data literacy, AI literacy, critical approaches

1. Introduction

The concept of data literacy (DL) is emerging as a key means of facilitating communication between individuals, because it provides a common language for accessing and utilising existing data in a variety of ways. It is challenging to ascertain the extent to which this has been adopted beyond the confines of select professional communities. Conversely, it is our contention that – at this particular point in time – there is a greater need than ever before to place data at the heart of our activities (Corrall, 2019).

We should not forget that the domain of DL is associated with other forms of literacy, including the relatively recent addition of artificial intelligence literacy (AI literacy), the nature of which shall be described below.

2. Methods

We conducted a scoping review with the aim of summarising key concepts and definitions in order to further knowledge on the subject matter. As explained by Bowler and Shaw (2024), the aim of such reviews is to facilitate the identification of research on a given topic by presenting an initial overview on a topic by clarifying key concepts and definitions. Its diversity, and the multidisciplinary nature

Cite as: Koltay, T. (2025) From data literacy to artificial intelligence literacy: background and approaches, Central European Library and Information Science Review (CELISR), 2(1), p. 41–48. https://doi.org/10.3311/celisr.38042

of DL, as well related literacies were identified because they have the potential to identify areas that are currently lacking some kind of knowledge, but can lead to further and more comprehensive results.

3. The variety of "new" literacies

Before we start discussing our main objective, we need to briefly mention earlier literacies. One of the best known "new literacies" is information literacy (IL) that growingly adapts critical evaluation of information and its sources (American Library Association, 2000). We can also count with digital literacy (Gilster, 1997), which is often regarded to be synonymous with data literacy, given that it has become a cross-sectoral activity, leading to interdisciplinary activities that are supporting work in digital environments (Raffaghelli et al., 2020). Media literacy is also a foundational concept, but it should be acknowledged that its broad conceptualisation in this context is not necessarily media specific (Pfaff-Rüdiger and Riesmeyer, 2016). There is also media and information literacy that goes beyond knowing how to use digital technology, but focuses on knowing how to learn, how knowledge is organised, and how to find, select, evaluate, and use information. Accordingly, it includes the skills and knowledge needed to navigate the knowledge society, not only in education but also in working or everyday occupation (Varga and Egervári, 2022). We can also draw on the DigComp 2.2 Digital Competence Framework for Citizens, which also emphasises the importance of conscious data evaluation (Vuorikari, Kluzer, Punie et al., 2022).

4. Data literacy for researchers and information professionals

A significant part of the abovementioned types of data-related literacies were created almost exclusively to support research, i.e. initially, the concept of DL was predominantly interpreted through the lens of academic inquiry. This is reflected in the definition, provided by Calzada Prado and Marzal (2013) as it underlines the importance of understanding how to select and synthesise data and combine it with other sources of information and prior knowledge.

It should be noted that the focus has subsequently been extended to other professional domains and work environments. As a consequence of this evolution, there is often no clear distinction between DL for research purposes and the provision of DL services to citizens or other target audiences.

Nevertheless, effective and responsible use of data can facilitate the extraction of insights from datasets, their interpretation and translation into appropriate decisions and actions (Arreguit O'Neill, 2019). This requires us to look far beyond standard approaches by recognising the cognitive, metacognitive, affective and socio-emotional competencies associated with the use of data, which are anchored in universal moral values (Schüller, 2021).

Effective and responsible use of data can facilitate the extraction of insights from datasets, their interpretation and translation into appropriate decisions and actions (Arreguit O'Neill, 2019).

5. The need for critical approaches

At the same time, literates we have begun to pay attention to critical attitudes, as we have recognised the need to present a complete picture that reveals a broader picture neglected in some cases. In light of this we can understand that the treatment of datasets should begin with a process of deconstructing their unquestioned authority by identifying the context in which the data were

"As a consequence of this evolution, there is often no clear distinction between DL for research purposes and the provision of DL services to citizens or other target audiences."

 https://doi.org/10.3311/celisr.38042

produced and subsequently reused (Calzada Prado, Marzal, 2013). In addition, data literacy should be dedicated to promoting cross-disciplinary thinking, for example, with the aim of breaking down academic silos and encouraging and applying creative and critical thinking (Ridsdale et al., 2015). The ability to interpret and use data is of paramount importance to researchers, citizens and information professionals (librarians) alike, and it has become clear that the general public also needs data (Carlson and Johnston, 2015). With this comes the interest of dataset owners, as they should also be held accountable for the ethical use of data, because they are in a position to answer questions and thus become responsible for the ethical use of data by taking into account the following issues. In the light of the potential limitations of data-driven analysis, all players need to discern the discrepancies between data and reality. It is therefore the final responsibility of researchers, librarians and the general public to learn that data only provides a partial view of the subject matter (Fontichiaro and Oehrli, 2016). We also know that examining the subjectivities of individuals can help to construct appropriate messages, which – not incidentally – go hand in hand with the biases, associated with the sharing of data and information (Funk et al., 2016).

To achieve this requires

- Proper data literacy education, which can be integrated into a core skillset that encompasses the capacity to exercise discernment. It can be seen as a development of information discernment that is evaluating the information component of digital literacy (Walton, Pickard, Dodd, 2018).
- The development of the skills needed to transform data into non-raw, non-numerical representations necessitates the verification of the origin and reliability of data, with particular attention to factors that may affect the patterns, errors, and omissions of data (Bhargava, 2019).

6. Data literacies for citizens

The majority of the aforementioned concepts are beneficial for researchers and information professionals, but they can also prove invaluable to the general public. Namely, the concept of DL holds great promise for empowering individuals to understand and solve their own problems through the use of existing data or the generation of new data, as well as through the use and monetisation of diverse datasets (Bhargava et al., 2015). The optimal use of approaching data literacy is essential to facilitate decision making and to generate new knowledge beyond the boundaries of scientific research. Such knowledge can benefit not only scientific research, but also the promotion of community engagement and citizen empowerment.

"The optimal use of approaching data literacy is essential to facilitate decision making and to generate new knowledge beyond the boundaries of scientific research."

Data literacy is an essential skill to face a multitude of everyday challenges. In order to meet the associated responsibilities, it is necessary to have the necessary knowledge and skills to protect, secure and maintain the privacy of data on social networks (Corrall, 2019).

Managing data related tasks fall upon any individuals, who utilise data. They must be aware that varied data types can be used to serve a number of functions. However, it is recommended also for citizens to use data in a professional manner, for example by distinguishing between a simple opinion, a dataset or piece of information.

It is thus incumbent upon any data user to be aware that a variety of data types can serve for varied reasons. Searching for information and data to complete diverse narratives also can assist in the validation of claims, derived from

datasets. The development of democracy in several countries has led to the realisation that informed citizens are better equipped to make informed decisions as voters.

The first step towards more expansive definitions, approaches and practices of data literacy (Leu and Kinzer, 2000), which are necessary given the potential for data to be misinterpreted or employed in an inappropriate manner. Secondly, it is of utmost importance to recognise the limitations of reductionist perspectives on the world. This suggests that the preliminary step in working with data should be to question the apparent authority of a text, examining how it was constructed, by whom and under what circumstances (Špiranec, Kos, George, 2019). Data-driven thinking also can be employed in order to facilitate the translation of data into abstract concepts (Davies, Fidler, Gorbis, 2020). Cultivating a creative data literacy that appeals to a more diverse population and aims to alleviate the symptoms of data inequality (D'Ignazio, 2017). One of its manifestation is data infrastructure literacy, which aims to foster a critical approach that enables responding to the impact of datafication. To do this, it is paramount to adopt a broader view of literacy, which requires embracing personal data, generated by our digital activities, while recognising the importance of data identification, understanding, reflexivity, applications and strategies (Selwyn, 2019, p. 215).

"One of its manifestation is data infrastructure literacy, which aims to foster a critical approach that enables responding to the impact of datafication."

It is also essential to critically examine the technological infrastructure and political economy of digital platforms. This must be accompanied by an analysis of strategies and tactics for managing and protecting privacy and resisting profiling and tracking (Pangrazio and Sefton-Green, 2020). In such environments, we need to place even more emphasis on the approaches outlined above. One of the related key challenges is to understand the processes by which data is generated, extracted, managed, analysed and aggregated. It is also crucial to draw conclusions from the data, provide a summary and offer an explanation. It is also crucial to resist the invasion of privacy, the influence of algorithms and the commodification of data. Promoting the responsible use of data is also of paramount importance because it promotes open science, which is the responsible use of data. On a broader horizon, we should understand the ethical implications of data-driven processes (Raffaghelli et al., 2020). These and similar goals can be achieved by improving the data literacy skills of a wide range of current and potential users by encouraging a more sophisticated, analytical approach to data-driven processes (Gebre and Morales, 2020).

The aforementioned efforts stipulate that any framework for data citizenship should be designed to assist laypersons in their individual and collective critical inquiry, enabling them to engage in their communities in a meaningful and proactive manner (Carmi et al., 2020). Yates et al. (2021) reminds us that data thinking can foster citizens' critical understanding of data and data doing, which motivates citizens' everyday engagements with data, including deleting and using data in an ethical way. Data participation, which entails enhancing citizens' proactive engagement with data and their networks of literacy, is meant to with the objective of safeguarding individual and collective privacy and wellbeing, and providing assistance to others in developing their data literacy.

The willingness and ability to apply it transcends disciplinary boundaries, encompassing knowledge and willingness. This type of data literacy encourages critical questioning, imagination and intervention around the broader socio-technical infrastructures for creating, using and sharing data (Gray et al., 2018). Interrogating the socio-technical systems of big data practices requires

 https://doi.org/10.3311/celisr.38042

forming opinions about current debates around data analytics in order to make informed decisions about our personal choices (Sander, 2020).

7. Data literacy and AI literacy

AI (artificial intelligence) literacy can be defined as "the ability of a system to correctly interpret external data, learn from that data, and use that learning to achieve specific goals and tasks through flexible adaptation" (JISC, 2022). It is also understood as "a set of competencies that enable individuals to critically evaluate AI technologies, communicate and collaborate effectively with AI, and use it as a tool online, at home and at work" (Long and Magerko, 2020, p. 2).

However, what we see today is that the forms of this literacy do not yet seem to be fully clear, and - at the very least - AI applications themselves in general, and in libraries in particular, seem to be in their infancy.

On the other hand, we can already see that it includes the ability to understand, use, monitor and critically reflect on AI applications, and that it can be used without a background in computer science (Laupichler et al., 2022).

All above considerations require a return to the original conceptualisations of DL, as AI literacy relies on the ability to understand the intrinsic nature of data and to use this understanding to inform decision-making processes. It must also include the ability to identify, collect, organise, analyse and synthesise prioritised data (Mandinach and Gummer, 2013), recognising that the knowledge associated with data varies across disciplines (Carlson and Johnston, 2015).

It is therefore important to emphasise that in order to be considered data literate, it is necessary to identify the circumstances in which DL may be useful (Bhargava, 2019). This leads to the recognition that data often interacts with information and therefore its value depends on its intended use (Makani, 2015).

Given the potential of AI, it seems prudent to consider the principles of digital humanities (DH). In this and similar contexts, it may be helpful to recognise that information literacy is typically created directly by people, whereas the production of data is. This means that IL typically involves forms of knowledge, whereas the production of data may not require direct input, assuming that texts can be interpreted as data and partially as machine-generated abstractions (Schöch, 2013). Despite the fact that digital humanities literacy is less frequently mentioned, it has been largely overlooked in related discussions, as it is often seen as an ancillary outcome of the iterative process of inquiry.

However, both DL and AI literacy encompass a range of literacies, when critically examining the sources of data and their origins. Abner (2020, p. 20) adds that a data literate person "will ask questions about how value is determined, who assigns value, and what societal factors might influence value." Answers to these questions are not always readily available (if at all). However, "when learners understand the process of data creation as entirely human, they think critically about that process and how it affects any data they examine."

"However, both DL and AI literacy encompass a range of literacies, when critically examining the sources of data and their origins."

8. Applying data literacy

One of the potential applications of DL is in the field of education. For example, there are comprehensive approaches to teaching data literacy, based on a collaboration between several educational institutions in the United States. In order to achieve this objective, a number of potential stakeholders and vital elements were identified through a process of analysis and synthesis.

1. The articulation of the specific information or data that is required to be gathered, along with the rationale behind it, is of paramount importance.
2. It is to define the specific questions that require answers based on data.
3. The management of data entails a focus on the acquisition and storage of appropriate data.

The utilisation of data requires the analysis, interpretation, and reporting of data, with particular attention to the following:

- The selection of data sources or the definition of new data collection needs,
- The clarification of data submission procedures and requirements,
- The monitoring of data collection and entry fidelity,
- The collection and storage of datasets to ensure reliability, usability, and accessibility,
- The provision of related education.

This approach permits reflection and improvement, reducing thereby the unwarranted burden placed on educators and enhances their capacity to utilise data for a range of purposes. This may be accomplished by identifying deficiencies in the processes and defining strategies for improvement, as well as strengthening sustainability through professional development opportunities (Ruedel, Kuchle, Bailey, 2021).

In contrast to the aforementioned example, we may look at a relatively straightforward illustration of the involvement of libraries and librarians. In 2020, Georgetown University in Qatar conducted a pilot data literacy course. The specific learning objectives include the ability to locate and analyse both quantitative and qualitative data. The course addressed the fundamental competencies associated with data literacy, including the capacity to access data and critically assess its quality. The course was delivered in two formats: a lecture-based format and a hands-on format (Usova and Laws, 2021).

"In 2020, Georgetown University in Qatar conducted a pilot data literacy course."

9. Conclusion

This paper has shown that literacies can be identified in a variety of contexts, but has also highlighted the importance of recognising the similarities between them. It has highlighted the need for further research into data literacy in all its varieties, particularly in view of the fact that many questions remain unanswered in this area, so that existing knowledge on these issues needs to be both extended and consolidated.

Sources/Literature

Abner, K. (2020) *Data literacy as digital humanities literacy: Exploration of threshold concept*, In: M. Brooks, M. Hubbard, J. Perkins, J. Russel (eds.), Literacies in a Digital Humanities Context, dh+ lib, p. 19–22, Available at: https://acrl.ala.org/dh/2020/06/22/data-literacy-as-digital-humanities-literacy-exploration-of-threshold-concepts/ (Accessed: 2024.07.31.)

American Library Association (2000) *Information Literacy Competency Standards for Higher Education*, Available at: https://alair.ala.org/server/api/core/bitstreams/ce62c38e-971a-4a98-a424-7c0d1fe94d34/content (Accessed: 2024.07.31.)

Arreguit O'Neill, S. (2019) *Fomento de la alfabetización de datos en todos los ámbitos* (Fostering data literacy across field), Aula Abierta, 48(4), p. 419–434. (in [Spanish]) https://doi.org/10.17811/rifie.48.4.2019.419-434

https://doi.org/10.3311/celisr.38042

Bhargava, R. (2019) *Data literacy*, In: The International Encyclopedia of Media Literacy, Hobbs, R., Mihailidis, P. (eds.), John Wiley & Sons, Hoboken, NJ, USA, ISBN: 9781118978238, https://doi.org/10.1002/9781118978238.ieml0049

Bhargava, R., Deahl, E., Letouzé, E., Noonan, A., Sangokoya, D., Shoup, N. (2015) *Beyond data literacy: Reinventing community engagement and empowerment in the age of data*, Working Paper for Discussion, Data-Pop Alliance, Available at: https://datapopalliance.org/wp-content/uploads/2015/11/Beyond-Data-Literacy-2015.pdf (Accessed: 2024.07.31.)

Bowler, L., Shaw, C. (2024) *Trends in data literacy, 2018-2023: a review of the literature*, Information Research, 29(2), p. 198–205. https://doi.org/10.47989/ir292822

Calzada Prado, J., Marzal, M. Á. (2013) *Incorporating data literacy into information literacy programs: Core competencies and contents*, Libri, 63(2), p. 123–134. https://doi.org/10.1515/libri-2013-0010

Carlson, J., Johnston, L. R. (2015) *Data information literacy: Librarians, data, and the education of a new generation of researchers*, West Lafayette: Purdue University Press, ISBN 9781557536969

Carmi, E., Yates, S. J., Lockley, E., Pawluczuk, A. (2020) *Data citizenship: Rethinking data literacy in the age of disinformation, misinformation, and malinformation*, Internet Policy Review, 9(2), p. 1–22. https://doi.org/10.14763/2020.2.1481

Corrall, S. (2019) *Repositioning data literacy as a mission-critical competence*, In: ACRL 2019: Recasting the Narrative, Cleveland, OH, USA, Available at: http://d-scholarship.pitt.edu/36975/1/Data%20literacy%20discussion%20handout.pdf (Accessed: 2024.07.31.)

Davies, A, Fidler, D., Gorbis, M. (2020) *Future work skills*, Institute for the Future for Apollo Research Institute, Palo Alto, California, Available at: https://uqpn.uq.edu.au/files/203/LIBBY%20MARSHALL%20future_work_skills_2020_full_research_report_final_1.pdf (Accessed: 2024.07.31.)

D'Ignazio, C. (2017) *Creative data literacy: Bridging the gap between the data-haves and data-have nots*, Information Design Journal, 23(1), p. 6–18. https://doi.org/10.1075/idj.23.1.03dig

Fontichiaro, K., Oehrli, J. A. (2016) *Why data literacy matters*, Knowledge Quest, 44(5), p. 21–27, Available at: https://files.eric.ed.gov/fulltext/EJ1099487.pdf (Accessed: 2024. 07.31.)

Funk, S., Kellner, D., Share, J. (2016) *Critical media literacy as transformative pedagogy*, In: Handbook of Research on Media Literacy in the Digital Age, p. 1–30, ISBN 9781466696679 https://doi.org/10.4018/978-1-4666-9667-9.CH001

Gebre, E. H., Morales, E. (2020) *How "accessible" is open data? Analysis of context-related information and users' comments in open datasets*, Information and Learning Sciences, 121, p. 19–36. https://doi.org/10.1108/ILS-08-2019-0086

Gilster, P. (1997) *Digital literacy*, Wiley, ISBN 978-0471249528

Gray, J., Gerlitz, C., Bounegru, L. (2018) *Data infrastructure literacy*, Big Data & Society, 5(2), https://doi.org/10.1177/2053951718786316

JISC (2022) *AI in tertiary education: A summary of the current state of play*, Available at: https://repository.jisc.ac.uk/8783/1/ai-in-tertiary-education-report-june-2022.pdf (Accessed: 2024. 07.31.)

Laupichler, M. C., Aster, A., Schirch, J., Raupach, T. (2022) *Artificial intelligence literacy in higher and adult education: A scoping literature review*, Computers and Education: Artificial Intelligence, 3, 100101, https://doi.org/10.1016/j.caeai.2022.100101

Leu, D. J., Jr., Kinzer, C. K. (2000) *The convergence of literacy instruction with networked technologies for information and communication*, Reading Research Quarterly, 35(1), p. 108–127. https://doi.org/10.1598/RRQ.35.1.8

Long, D., Magerko, B. (2020) *What is AI literacy? Competencies and design considerations*, Proceedings of the 2020 CHI conference on human factors in computing systems, p. 1–16, https://doi.org/10.1145/3313831.3376727

Mandinach, E. B., Gummer, E. S. (2013) *A systemic view of implementing data literacy in educator preparation*, Educational Researcher, 42(1), p. 30–37. https://doi.org/10.3102/0013189X12459803

Makani, J. (2015) *Knowledge Management, Research Data Management and University Scholarship: Towards an Integrated Institutional Research Data Management Support-System Framework*, VINE, 45(3), p. 344–359. https://doi.org/10.1108/VINE-07-2014-0047

Pangrazio, L., Sefton-Green, J. (2020) *The social utility of 'data literacy'*, Learning, Media and Technology, 45(2), p. 208–220. https://doi.org/10.1080/17439884.2020.1707223

Pfaff-Rüdiger, S., Riesmeyer, C. (2016) *Moved into action. Media literacy as social process*, Journal of Children and Media, 10(2), p. 164–172. https://doi.org/10.1080/17482798.2015.1127838

Raffaghelli, J. E., Manca, S., Stewart, B., Prinsloo, P., Sangrà, A. (2020) *Supporting the development of critical data literacies in higher education: building blocks for fair data cultures in society*, International Journal of Educational Technology in Higher Education, 17(1), p. 1–22. https://doi.org/10.1186/s41239-020-00235-w

Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D., Kelley, D., Matwin, S., Wuetherick, B. (2015) *Strategies and best practices for data literacy education: Knowledge synthesis report*, Halifax, NS: Dalhousie University, Available at: https://dalspace.library.dal.ca/bitstream/handle/10222/64578/Strategies%20and%20Best%20Practices%20for%20Data%20Literacy%20Education.pdf?sequence=1&isAllowed=y (Accessed: 2024. 07.31.)

Ruedel, K., Kuchle, L. B., Bailey, T. (2021) *Essential Elements of Comprehensive Data Literacy*, National Center for Systemic Improvement at WestEd, Available at: https://files.eric.ed.gov/fulltext/ED620527.pdf (Accessed: 2024. 07.31.)

Sander, I. (2020) *What is critical big data literacy and how can it be implemented?*, Internet Policy Review, 9(2), p. 1–22. https://doi.org/10.14763/2020.2.1479

Schüller, K. (2021) *Towards a Framework for Data and AI Literacy*, In: Proceedings 63rd ISI World Statistics Congress, 11–16 July 2021, p. 843–848. Available at: https://isi-web.org/sites/default/files/import/pdf/225-day5-ips087-towards-a-framework-for-data-a.pdf (Accessed: 2024. 07.31.)

Selwyn, N. (2019) *What's the problem with learning analytics?*, Journal of Learning Analytics, 6(3), p. 11–19. https://doi.org/10.18608/jla.2019.63.3

Schöch, C. (2013) *Big? Smart? Clean? Messy? Data in the Humanities?*, Journal of the Digital Humanities, 2(3), p. 2–13. Available at: file:///C:/Users/user/Downloads/Schoech_2013_Data-in-the-humanities.pdf (Accessed: 2024. 07.31.)

Špiranec, S., Kos, D., George, M. (2019) *Searching for critical dimensions in data literacy*, In: Proceedings of the Tenth International Conference on Conceptions of Library and Information Science, Ljubljana, Slovenia, June 16-19, 2019, 24(4). Available at: http://InformationR.net/ir/24-4/colis/colis1922.html (Accessed: 2024. 07.31.)

Usova, T., Laws, R. (2021) *Teaching a one-credit course on data literacy and data visualisation*, Journal of Information Literacy, 15(1), p. 84–95. http://dx.doi.org/10.11645/15.1.2840

Varga, K., Egervári, D. (2022) *Learn, unlearn, relearn: The new way of learning is called media and information literacy*, Tudásmenedzsment, 23, p. 120–135. https://doi.org/10.15170/TM.2022.23.K1.9

Vuorikari, R., Kluzer, S., Punie, Y. (2022) *DigComp 2.2: The Digital Competence Framework for Citizens – With new examples of knowledge, skills and attitudes*, Publications Office of the European Union, Luxembourg, ISBN 978-92-76-48882-8 https://doi.org/10.2760/115376

Walton, G., Pickard, A. J., Dodd, L. (2018) *Information discernment, mis-information and pro-active scepticism*, Journal of Librarianship and Information Science, 50(3), p. 296–309. https://doi.org/10.1177/0961000618769980

Yates, S., Carmi, E., Lockley, E., Wessels, B., Pawluczuk, A. (2021) *Me and My Big Data: Understanding Citizens Data Literacies-Final report*, Liverpool: Nuffield Foundation.