

A Bibliometric Analysis of Datafication in Education: Trends, Impact and Future Implications

Nining Anggeraini

Universitas Bumigora, Mataram, Indonesia

Email Korespondensi: nining@universitasbumigora.ac.id

Genesis Artikel: Diterima: 10 Januari 2025 **Diterbitkan:** 25 Januari 2024

ABSTRACT: The massive use of digital learning platforms and AI in education has generated large quantities of data on teacher and student information, online interactions, and teaching and learning practices. This data can be collected, analyzed, and interpreted to improve educational outcomes. Following the interest and attention in this topic, numerous research studies have been conducted to gain a better understanding in this area. Thus, the purpose of this study is to summarize the literature on datafication in education from 2020 to 2024 and to explore the key terms related to its influence on educational practices, and its potential future implications. The method used in this study is a bibliometric analysis. A total of 200 articles were found in Google Scholar through a search using the keywords "datafication in education." The study found that research on datafication in education has grown significantly in recent years. Initially focused on technological aspects, the research has shifted towards practical applications and critical perspectives. Key themes identified include data literacy, AI, and the ethical implications of data use in education. As an implication of this study, this overview aims to assist future research by enriching and expanding the scope of research and encouraging progress in related disciplines. This will contribute to a fuller understanding of the potential benefits and risks of datafication in transforming education.

Keyword: AI in education, Datafication, Data Literacy, Educational Data, Ethical Considerations

Ini adalah artikel akses terbuka dibawah lisensi [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/).



Cara Sitasi:

Anggeraini, N. (2025). A bibliometric analysis of datafication in education: Trends, impact and future implications. *UPGRADE: Jurnal Pendidikan Teknologi Informasi*, 2(2), 21-30. <https://doi.org/10.30812/upgrade.v2i2.4803>

INTRODUCTION

Datafication, the large-scale collection, analysis, and interpretation of data, has become a defining feature of the modern world, impacting numerous sectors, including education. This trend, particularly relevant in the context of "Education 5.0," presents both opportunities and challenges for teaching and learning (Williamson et al., 2020). In educational settings, datafication encompasses a vast array of information, from student performance and learning behaviors to social interactions within online platforms and digital learning environments (Perrotta and Selwyn, 2020; Hooper et al., 2022; Pangrazio et al., 2023). Proponents believe this data offers valuable insights into student engagement, learning patterns, and academic performance, potentially even enabling predictions of future success (Eynon, 2022). It is important to highlight the crucial role of educators' understanding of data use, as it shapes how results are interpreted, informs teaching practices, and ultimately influences student progress (Mills et al., 2021; Howard et al., 2022). Datafication has the potential to fuel various educational advancements (Crompton et al., 2024; Vezzoli et al., 2020), including automated interventions (Hansen and Komljenovic, 2023), personalized learning experiences (Witzenberger and Gulson, 2021), and predictive analytics to guide student progress (Selwyn et al., 2023).

However, alongside these promising applications, datafication in education raises critical questions about privacy, equity, and potential unintended consequences. Responsible and ethical data use is paramount to ensure that all students benefit from this trend while safeguarding their privacy and well-being (Stoilova et al., 2021; Nottingham et al., 2022). Previous studies (Kizilcec and Lee, 2022; Baker and Hawn, 2022) raise concerns regarding the fairness of algorithmic predictions, urging scrutiny of measurement processes, model learning, and actions taken, to identify and mitigate potential bias and discrimination (Helsper, 2022). The potential impact of algorithmic bias on different student demographics (Robinson, 2020), and the entities involved in creating and implementing these algorithms in educational settings require further investigation (Kordzadeh and Ghasemaghaei, 2022).

The previous researches reveal key areas of datafication in education such as educational practices, legal frameworks, and data governance (Erstad et al., 2023), the effectiveness of data mining algorithms, and the role of student data in predicting academic success (Admiraal et al., 2020; Camacho et al., 2020) data-driven learning potential for personalized learning and improved outcomes (Yunita et al., 2021), data privacy, and the ethical implications of AI in education (Williamson et al., 2023). Furthermore, studies highlight the limitations of current data governance frameworks and emphasize the need for more equitable approaches to data use in education (Hillman, 2023). These findings underscore the importance of ongoing research to address the challenges and opportunities presented by datafication in education, ensuring that data is used responsibly and ethically to improve learning outcomes for all students (Sefton-Green and Pangrazio, 2022).

Referring to these findings, the novelty of the current research is not only to confirm and expand upon the insights from previous studies but also to provide a valuable roadmap for future research on datafication in education using a bibliometric approach. This study aims to summarize the literature on datafication in education from 2020 to 2024 utilizing a bibliometric analysis and to explore the key terms related to its influence on educational practices, and its potential future implications. The results of this study are expected to be useful as a comprehensive overview in assisting new research to gain a deeper understanding of the current state, challenges, and opportunities on datafication in education.

METHODS

This research was carried out using a bibliometric analysis method. Bibliometric analysis is a quantitative method that applies mathematical and statistical tools to evaluate the inter-relationships and impacts of publications, authors, institutions, and countries in a specific research area. Precisely, this type of analysis enables the researcher to identify and display previous research findings (Muhammad et al., 2022) figure out the latest trend related to a certain research topic (Muhammad et al., 2022), and measures the relationship between the keywords, the authors, the countries, and the number of citations in a related article that has been published. The data were collected from the Google Scholars database using Perish or Publish software using "datafication in education" keywords. There were 200 articles found in the last 5 years (2020-2024). Then, the available data is then stored in CSV and RIS formats, which are later analyzed using Vosviewer. Moreover, to get more solid information related to the correlation of the data, network analysis was carried out through data clustering and visualization (Susilowati et al., 2024; Huang et al., 2022; Oyewola and Dada, 2022).

FINDINGS AND DISCUSSION

Datafication in Education Trends in the last 5 years

The results of database exploration using the “datafication in education” keywords at intervals of five years (2020-2024) from the Google Scholar database are presented in Table 1. The data has been screened based on the article title, abstract, and keywords in the literature to represent datafication in education research trends. The data revealed that in the last five years, 200 papers related to datafication in education published with over 8,000 citations, indicating a relatively productive field. The average paper received 40.55 citations, and the h-index of 43 suggests a significant portion of papers have been highly cited. The g-index of 87 provides a more nuanced view of research productivity, indicating a larger cumulative impact. The half-life of citations is 31 years, meaning half of a paper’s citations will occur within that timeframe. The hA-index of 22 suggests that individual authors have also made significant contributions to the field of datafication in education research. Following these findings, the data have been sorted from the most to the less cited paper. The top 10 most cited articles are identified in Table 2.

Table 1. Data Matrix Citation

Matrix Data	Results
<i>Keywords</i>	Datafication in Education
<i>Publication years:</i>	2020 -2024
<i>Citation years:</i>	5 (2020-2025)
<i>Papers:</i>	200
<i>Citations:</i>	8110
<i>Cites/years:</i>	1622.00
<i>Cites/papers:</i>	40.55
<i>Cites/authors:</i>	3787.15
<i>Papers/author:</i>	121.51
<i>Authors/paper:</i>	2.17
<i>h-Index:</i>	43
<i>g-index:</i>	87
<i>hI-normal:</i>	31
<i>hI annual:</i>	6.20
<i>hA-index:</i>	22

Table 2. Top 10 Articles

Year	Author	Title	Publication	Cite
	B Williamson, R Eynon, J Potter	Pandemic politics, pedagogies and practices: digital technologies and distance education during the coronavirus emergency	Learning, media and technology	1472
	M Teräs, J Suoranta, H Teräs, M Curcher	Post-Covid-19 education and education technology 'solutionism': A seller's market	Science and Education	825
	N Selwyn, T Hillman, R Eynon, G Ferreira...	What's next for Ed-Tech? Critical hopes and concerns for the 2020s	Learning, Media and Technology	330
2020	B Williamson, S Bayne, S Shay	The datafication of teaching in Higher Education: critical issues and perspectives	Teaching in Higher Education	322
	J Manolev, A Sullivan, R Slea	The datafication of discipline: ClassDojo, surveillance and a performative classroom culture	The Datafication of Education	292
	J Knox, B Williamson, S Bayne	Machine behaviourism: Future visions of 'learnification' and 'datafication' across humans and digital technologies	Learning, Media and Technology	263
	A Bradbury	Datafied at four: The role of data in the 'schoolification' of early childhood education in England	The Datafication of Education	183
	B Williamson, N Piattoeva	Objectivity as standardization in data-scientific education policy, technology and governance	The Datafication of Education	130
	JE Raffaghelli, B Stewart	Centering complexity in 'educators' data literacy' to support future practices in faculty development: A systematic review of the literature	Teaching in Higher Education	125
2021	AM Cox	Exploring the impact of Artificial Intelligence and robots on higher education through literature-based design fictions	Journal of Educational Technology in Higher Education	208

Table 3. Keywords That Represent Each Cluster

No.	Cluster	Elements
1.	Cluster 1 (11 items)	Datafication process, disruption, educational data, educational process, future research, information, insight, platformization, teacher accountability, theory, trust
2.	Cluster 2 (11 items)	Benefit, data doppelganger, datafied society, digital tool, educational response, exploration, opportunity, profile, researcher, social justice, student life
3.	Cluster 3 (10 items)	Assessment, consequence, critique, datafied school, dataveillance, digital data, education technology, management,school data,schooling
4.	Cluster 4 (8 items)	Complexity,data activism,data literacy,early childhood education, educational space, gap, narrative, systematic review
5.	Cluster 5 (7 items)	Algorithms, artificial intelligence shape policy, competitiveness, datafication solidify market, early years education, education sector, university
6.	Cluster 6 (6 items)	Algorithms, artificial intelligence,automation, calculation,expansion, potential
7.	Cluster 7 (5 items)	Evaluation, evaluation policy, paper investigates datafication, performativity, quality assurance
8.	Cluster 8 (5 items)	Big data, critical perspective, digital transformation, digitization, media education
9.	Cluster 9 (4 items)	Inequality, negative effect, perception, privacy
10.	Cluster 10 (2 items)	Algorithmic automation, transformation

The datafication process (cluster 1) centers around the collection and analysis of educational data (Roslan and Chen, 2022; Admiraal et al., 2020; Camacho et al., 2020). The ‘benefits and opportunities’ cluster (cluster 2) highlights the potential positive outcomes of datafication (Yunita et al., 2021). The ‘critical and societal perspectives’ cluster (cluster 3, 8, dan 9) underscores concerns about potential downsides and societal implications (Hillman, 2023). The ‘data literacy and early childhood ducation’ cluster (cluster 4 dan 5) emphasizes the importance of data literacy skills from an early age. The ‘algorithms and artificial intelligence’ cluster (5, 6, dan 10) highlights the increasing role of AI in education, raising concerns about equity and fairness Williamson (2023).

The ‘evaluation and policy’ cluster (cluster 7) suggests that datafication is being used to inform and improve educational policies (Erstad et al., 2023). These clusters are interconnected, with ‘data literacy’ acting as a bridge between technology and education, and ‘digitization’ enabling data collection and analysis (Mills et al., 2021; Hansen and Komljenovic, 2023). ‘Algorithms and Artificial Intelligence’ are linked to various clusters, highlighting their significant influence. ‘Critique’ connects to multiple clusters, reflecting the growing awareness of the potential downsides of datafication in education (Kordzadeh and Ghasemaghahi, 2022). This interconnectedness underscores the complex interplay between technological advancements, educational applications, and critical perspectives in shaping the framework of datafication in the education sector.

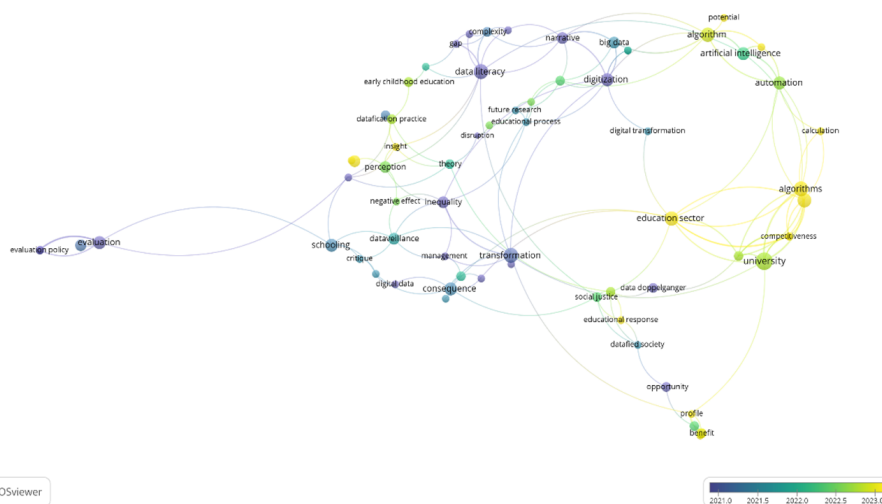


Figure 3. Overlay Visualization

CONCLUSION

This study has summarized and analyzed the development, key areas and trends of datafication in education research from 2020 to 2024 through bibliometric analysis. Datafication in education is a burgeoning field with significant impact, evidenced by substantial research output and a high citation rate. While initially focused on technological underpinnings, the field is dynamically evolving, incorporating critical perspectives and emphasizing practical applications yet there is limited discussion in particular areas like early childhood education and the societal implications of dataveillance. Given the increasing role of data in education, fostering data literacy is crucial, and careful consideration must be given to the ethical and societal implications, including concerns about privacy, equity, and fairness. Ultimately, datafication has the potential to revolutionize education, but its implementation requires a balanced approach that maximizes benefits while mitigating risks. Moreover, the implications of the findings in this study benefit future research since it presents an overview of the key area and research topic development related to datafication in education. Thus, it is recommended for future research to conduct longitudinal studies, qualitative research, interdisciplinary collaboration, and a focus on developing evidence-based policies and engaging the public in shaping the future of datafication in education.

ACKNOWLEDGMENTS

I extend my appreciation to my fellow lecturers at the Information Technology Education Department for their invaluable support and insights during the research process. I would also like to thank Bumigora University for providing the necessary resources and facilities that enabled me to conduct this study. Special thanks are due to our colleagues and peers who offered constructive feedback and encouragement throughout the writing process.

DECLARATION

Contributor Role Taxonomy

All authors contributed equally as primary contributors to this paper. All authors have read and approved the final version of the manuscript.

Funding Statement

This research received no specific grants from public, commercial, or non-profit funding agencies.

Competing Interests Statement

The authors declare that they have no financial or personal relationships that could be perceived to influence the work reported in this article.

DAFTAR PUSTAKA

- Admiraal, W., Vermeulen, J., and Bulterman-Bos, J. (2020). Teaching with learning analytics: How to connect computer-based assessment data with classroom instruction? *Technology, Pedagogy and Education*, 29(5). <https://doi.org/10.1080/1475939X.2020.1825992>.
- Al Husaeni, D. F. and Nandiyanto, A. B. D. (2022). Bibliometric using Vosviewer with Publish or Perish (using google scholar data): From step-by-step processing for users to the practical examples in the analysis of digital learning articles in pre and post Covid-19 pandemic. *ASEAN Journal of Science and Engineering*, 2(1):19–46.
- Baker, R. S. and Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, 32(4):1–41. <https://doi.org/10.1007/s40593-021-00285-9>.
- Camacho, V. L., Guia, E. D. L., Olivares, T., Julia Flores, M., and Orozco-Barbosa, L. (2020). Data capture and multimodal learning analytics focused on engagement with a new wearable IoT approach. *IEEE Transactions on Learning Technologies*, 13(4). <https://doi.org/10.1109/TLT.2020.2999787>.
- Crompton, H., Jones, M. V., and Burke, D. (2024). Affordances and challenges of artificial intelligence in K-12 education: A systematic review. *Journal of Research on Technology in Education*, 56(3). <https://doi.org/10.1080/15391523.2022.2121344>.

- Dewantara, K. T. W. (2024). Mobile learning dalam pendidikan: Analisis bibliometrik. *UPGRADE: Jurnal Pendidikan Teknologi Informasi*, 2(1):23–32. <https://doi.org/10.30812/upgrade.v2i1.4153>.
- Erstad, O., Guðmundsdóttir, G. B., Viberg, O., Williamson, B., Gilje, Ø., Gudmundsdóttir, G. B., Rebekka, B., Wagstaffe, K., Kumpulainen, O., Viberg, B., Williamson, J. O., and Howard, S. (2023). *Datafication in and of education*. *Agile Edu*. <https://doi.org/10.13140/RG.2.2.35661.00483>.
- Eynon, R. (2022). Datafication and the role of schooling: Challenging the status quo. In *Learning to Live with Datafication: Educational Case Studies and Initiatives from Across the World*. <https://doi.org/10.4324/9781003136842-2>.
- Finandhita, A., Mega, R. U., Jumansyah, R., Rafdhi, A. A., and Oktafiani, D. (2022). VOSviewer application analysis: Computational physical chemistry case study. *Moroccan Journal of Chemistry*, 10(1):091–101. <https://doi.org/10.48317/IMIST.PRSM/morjchem-v10i1.31756>.
- Ginting, S. L. B. (2023). A computational bibliometric analysis of esports management using VOSviewer. *International Journal of Informatics, Information System and Computer Engineering (INJIISCOM)*, 4(1):31–48. <https://doi.org/10.34010/injiiscom.v4i1.9570>.
- Hansen, M. and Komljenovic, J. (2023). Automating learning situations in edtech: Techno-commercial logic of assetisation. *Postdigital Science and Education*, 5(1). <https://doi.org/10.1007/s42438-022-00359-4>.
- Helsper, E. (2022). *The digital disconnect: The social causes and consequences of digital inequalities*. <https://doi.org/10.4135/9781526492982>.
- Hillman, V. (2023). Bringing in the technological, ethical, educational and social-structural for a new education data governance. *Learning, Media and Technology*, 48(1). <https://doi.org/10.1080/17439884.2022.2052313>.
- Hooper, L., Livingstone, S., and Pothong, K. (2022). Problem with data governance in UK school: The cases of Google Classroom and Classdojo d. Technical report, London.
- Howard, S. K., Swist, T., Gasevic, D., Bartimote, K., Knight, S., Gulson, K., Apps, T., Pelocche, J., Hutchinson, N., and Selwyn, N. (2022). Educational data journeys: Where are we going, what are we taking and making for AI? *Computers and Education: Artificial Intelligence*, 3. <https://doi.org/10.1016/j.caeai.2022.100073>.
- Huang, Y. J., Cheng, S., Yang, F. Q., and Chen, C. (2022). Analysis and visualization of research on resilient cities and communities based on VOSviewer. <https://doi.org/10.3390/ijerph19127068>.
- Indriyanti, F., Fauziah, T. N., and Nuryadin, A. (2023). Analisis bibliometrik penggunaan video pembelajaran di sekolah dasar tahun 2013-2022 menggunakan aplikasi vosviewer. *Jurnal Educatio*, 9(1):23–31. <https://doi.org/10.31949/educatio.v9i1.3906>.
- Kizilcec, R. F. and Lee, H. (2022). Algorithmic fairness in education. In *The Ethics of Artificial Intelligence in Education: Practices, Challenges, and Debates*. <https://doi.org/10.4324/9780429329067-10>.
- Kordzadeh, N. and Ghasemaghaei, M. (2022). Algorithmic bias: Review, synthesis, and future research directions. <https://doi.org/10.1080/0960085X.2021.1927212>.
- Lesmana, R. and Rifaldi, M. I. (2023). A computational bibliometric analysis of e-groceries analysis using VOSviewer. *International Journal of Informatics, Information System and Computer Engineering (INJIISCOM)*, 4(1). <https://doi.org/10.34010/injiiscom.v4i1.9586>.
- Mills, M., Mockler, N., Stacey, M., and Taylor, B. (2021). Teachers' orientations to educational research and data in England and Australia: implications for teacher professionalism. *Teaching Education*, 32(1). <https://doi.org/10.1080/10476210.2020.1843617>.
- Muhammad, U. A., Fuad, M., Ariyani, F., and Suyanto, E. (2022). Bibliometric analysis of local wisdom-based learning: Direction for future history education research. *International Journal of Evaluation and Research in Education*, 11(4). <https://doi.org/10.11591/ijere.v11i4.23547>.

- Nottingham, E., Stockman, C., and Burke, M. (2022). Education in a datafied world: Balancing children's rights and school's responsibilities in the age of Covid 19. *Computer Law and Security Review*, 45. <https://doi.org/10.1016/j.clsr.2022.105664>.
- Oyewola, D. O. and Dada, E. G. (2022). Exploring machine learning: A scientometrics approach using bibliometrix and VOSviewer. *SN Applied Sciences*, 4(5). <https://doi.org/10.1007/s42452-022-05027-7>.
- Pangrazio, L., Selwyn, N., and Cumbo, B. (2023). A patchwork of platforms: Mapping data infrastructures in schools. *Learning, Media and Technology*, 48(1). <https://doi.org/10.1080/17439884.2022.2035395>.
- Perrotta, C. and Selwyn, N. (2020). Deep learning goes to school: Toward a relational understanding of AI in education. *Learning, Media and Technology*, 45(3). <https://doi.org/10.1080/17439884.2020.1686017>.
- Ria, R. R. P. and Susilowati, D. (2023). Local wisdom-based computational thinking diagnostic Test: A bibliometric analysis mapping atate-of-the-art and research gaps. *Nusantara: Jurnal Pendidikan Indonesia*, 3(3):371–394. <https://doi.org/10.14421/njpi.2023.v3i3-2>.
- Robinson, S. C. (2020). Trust, transparency, and openness: How inclusion of cultural values shapes Nordic national public policy strategies for artificial intelligence (AI). *Technology in Society*, 63. <https://doi.org/10.1016/j.techsoc.2020.101421>.
- Roslan, M. B. and Chen, C. (2022). Educational data mining for student performance prediction: A systematic literature review (2015-2021). *International Journal of Emerging Technologies in Learning (iJET)*, 17(5):147–179.
- Sefton-Green, J. and Pangrazio, L. (2022). The death of the educative subject? The limits of criticality under datafication. *Educational Philosophy and Theory*, 54(12). <https://doi.org/10.1080/00131857.2021.1978072>.
- Selwyn, N., Hillman, T., Bergviken-Rensfeldt, A., and Perrotta, C. (2023). Making sense of the digital automation of education. <https://doi.org/10.1007/s42438-022-00362-9>.
- Soegoto, E. S., Luckyardi, S., Rafdhi, A. A., and Oktafiani, D. (2022). Computational analysis of waste management and entrepreneur using VosViewer application. *Moroccan Journal of Chemistry*, 10(3):542–552. <https://doi.org/10.48317/IMIST.PRSM/morjchem-v10i3.33142>.
- Stoilova, M., Nandagiri, R., and Livingstone, S. (2021). Children's understanding of personal data and privacy online—a systematic evidence mapping. *Information Communication and Society*, 24(4). <https://doi.org/10.1080/1369118X.2019.1657164>.
- Susilowati, D., Saputri, D. S. C., Ria, R. R. P., and Hidjah, K. (2024). Bibliometric analysis of artificial intelligent studies in education and pedagogy. *Nusantara: Jurnal Pendidikan Indonesia*, 4(3):621–634. <https://doi.org/10.14421/njpi.2024.v4i3-1>.
- Vezzoli, Y., Mavrikis, M., and Vasalou, A. (2020). Inspiration cards workshops with primary teachers in the early co-design stages of learning analytics. In *ACM International Conference Proceeding Series*. <https://doi.org/10.1145/3375462.3375537>.
- Williamson, B. (2023). Governing through infrastructural control: Artificial intelligence and cloud computing in the data-intensive state. In *The SAGE Handbook of Digital Society*. <https://doi.org/10.4135/9781529783193.n29>.
- Williamson, B., Bayne, S., and Shay, S. (2020). The datafication of teaching in Higher Education: critical issues and perspectives. <https://doi.org/10.1080/13562517.2020.1748811>.
- Williamson, B., Macgilchrist, F., and Potter, J. (2023). Re-examining AI, automation and datafication in education. <https://doi.org/10.1080/17439884.2023.2167830>.

Witzenberger, K. and Gulson, K. N. (2021). Why edtech is always right: students, data and machines in pre-emptive configurations. *Learning, Media and Technology*, 46(4). <https://doi.org/10.1080/17439884.2021.1913181>.

Yunita, A., Santoso, H. B., and Hasibuan, Z. A. (2021). Research review on big data usage for learning analytics and educational data mining: A way forward to develop an intelligent automation system. In *Journal of Physics: Conference Series*, volume 1898. IOP Publishing Ltd. <https://doi.org/10.1088/1742-6596/1898/1/012044>.