



The Role of Data in Education: Benefits, Challenges, and Implications for Teaching and Learning

Georgios ALEXANDROPOULOS

Universitat Autònoma De Barcelona, Barcelona, Spain

*Corresponding author: ling.george82@yahoo.gr

Abstract

In the current digital era, data science has emerged as a transformative discipline with profound potential for reshaping the educational landscape. This paper explores the multifaceted role of data science—specifically through educational data mining (EDM) and learning analytics—in enhancing teaching and learning processes across various platforms. Through a critical literature review, the study examines the dual nature of data utilization. On one hand, it highlights significant benefits such as personalized learning, early detection of student behavioral patterns, and evidence-based decision-making. On the other hand, it addresses critical risks, including privacy concerns, ethical violations, social stereotyping (labeling), and the potential commodification of education by corporate interests. The analysis further demonstrates that an overreliance on quantitative metrics risks neglecting the psychological dimensions and sociocultural contexts inherent in human learning. To mitigate these imbalances, the paper proposes the application of the DELICATE framework (determination, explain, legitimate, involve, consent, anonymize, technical aspects, and external partners) to ensure transparency and data protection. The study concludes by emphasizing a necessary shift from a purely technocratic perspective to a human-centered design approach. The authors argue that data science should serve as a pedagogical support tool rather than a substitute for teacher intuition. By integrating quantitative methods with qualitative-ethnographic approaches, a more just, innovative, and humane educational environment can be achieved.

Keywords: Data in education, educational data analytics, digital learning, educational inequality, data ethics, platform-based learning

Acknowledgments: The author expresses sincere gratitude to the university and department for providing the necessary facilities and a supportive academic environment. The author also acknowledges the students and academic staff at the participating institutions, particularly for their cooperation during the data collection process. Finally, appreciation is given to all parties involved in supporting the integration of digital educational technologies and pedagogical innovation.

For citation:

Alexandropoulos, G. (2026). The Role of Data in Education: Benefits, Challenges, and Implications for Teaching and Learning. *Journal of Digital Learning and Distance Education*, 4(11), 1992-1997. <https://doi.org/10.56778/jdlde.v4i11.681>

Introduction

In the era of rapid digital transformation, data science has emerged as a pivotal discipline with immense potential to reshape the educational landscape. Fundamentally involving the systematic collection, analysis, and visualization of complex datasets, data science is increasingly integrated into various educational levels, ranging from primary schools to higher education institutions (Liu, 2024). The proliferation of learning management systems (LMS), such as Moodle and Canvas, has further accelerated this integration. Notably, the post-COVID-19 pandemic landscape has established distance learning and continuous data harvesting as integral elements of the modern pedagogical process (Castellanos-Reyes & Leppard, 2025).

However, the dominance of digitalization in education presents a complex paradox that warrants critical examination. On one hand, data science promises to enhance instructional quality through the implementation of personalized learning pathways and the precise prediction of student outcomes (Hsu, 2024). Conversely, this phenomenon functions as a "two-sided coin," harboring significant risks and unintended negative consequences that remain underaddressed (Bednarowska-Michael & Uprichard, 2026). There is a notable research gap between the rapid adoption of these technocratic tools and a deep understanding of their impact on the psychological and sociocultural dimensions of human learning.

The intersection of technology and education raises several crucial questions regarding institutional and ethical responsibility. It is imperative to determine how data science can substantively assist teachers and students in improving performance without undermining autonomy. Furthermore, the potential for negative impacts—such as privacy breaches, social labeling, and algorithmic bias—must be rigorously identified. Most importantly, educational stakeholders face the challenge of balancing the efficiency of advanced technological tools with the unwavering protection of core human values and ethical standards.

This report seeks to address these challenges by critically examining the technological duality of data science in education. The study offers a distinct novelty by proposing the application of the DELICATE ethical framework as a concrete solution to ensure data transparency and protection. The primary objective is to equip educators with a comprehensive perspective to adapt to a data-centric era without compromising ethical integrity. Ultimately, this analysis emphasizes a necessary shift toward a human-centered design, ensuring that data science functions strictly as a supportive instrument for a humanizing pedagogy.

Methodology

This research uses a descriptive qualitative approach with a literature review method. Information was collected and analyzed based on theoretical frameworks from relevant disciplines, namely educational data mining (EDM) and learning and teaching analytics.

Operational Framework

This research method is divided into three main phases:

Identification of the Theoretical Foundation: Integrating data science theories according to (Leung et al., 2023), which incorporate machine learning, data visualization, and statistical algorithms for organizational decision-making; Educational Data Categorization: The data for the study came from online assessment activities, virtual learning environments, and academic information systems. The data was classified into subsets based on common characteristics, such as text patterns, comments, and student participation levels; Critical Comparative Analysis: Reviewing expert perspectives on benefits (such as personalized pedagogical strategies) and barriers (such as privacy concerns, imbalanced collaboration between actors, and a lack of uniform legal framework).

Analysis Instrument

The analysis was conducted by comparing literature focused on information automation (EDM) with literature focused on improving instructional design (Teaching Analytics). Furthermore, this study employed an ethical analysis approach to evaluate the balance between data collection and privacy protection, as proposed by (Ray et al., 2018).

Methodological Objectives

Through a synthesis of various theories and case studies in the literature, this method aims to formulate practical solutions to achieve a balance between technological advancement and student protection in the digital education ecosystem.

Results and Discussion

At this point, we will shed light on the positive and negative aspects of data in education for the teacher and the student, and then present ways in which we can achieve balance. First, the positive aspects of using data in education will be mentioned. In general, the benefits of data in education are many. Firstly, we can focus on detecting behavioral patterns, for example, to prevent absenteeism. In addition, teaching can be personalized with accurate reports and progress predictions. Moreover, teaching and its learning outcomes can be improved. Furthermore, students can be protected if they are at risk, to comply with inclusion and equity rules. Lastly, economic and social information can be linked to performance, and appropriate measures can be taken to provide assistance. The use of data in education offers significant benefits, particularly in improving the quality of decision-making related to educational issues. The use of data science enables the development of more personalized and adaptive learning systems, which in turn can improve student motivation and retention (Lin et al., 2025). Furthermore, data integration facilitates continuous monitoring of student progress to ensure their academic success (Agus, 2022). At the organizational level, extracting information from various administrative documents helps educational institutions and governments design more efficient policies, create effective curricula, and build more flexible education systems (Ogunniran & Makinde, 2024). Other benefits include: Problem Identification: Accelerating the detection and resolution of educational obstacles; Stakeholder Participation: Involving parents in decision-making and helping teachers choose the most appropriate teaching methods; Collaboration & Support: Increasing interaction between students through interactive learning environments and enabling teachers to provide more targeted support; Resource Optimization: Allowing qualitative resource analysis and more effective redistribution of teaching staff (Alsayer, 2023).

The use of data in education raises serious concerns about objectivity and reliability. Often, data is perceived as a neutral instrument, when in fact, it is a construct that can lead policymakers to erroneous decisions (Li, 2025). Furthermore, there are fundamental issues regarding privacy and intellectual property rights violations. Students often lose control over their personal data and lack the "right to be forgotten" or the right to opt out of data collection. Transparency in data management is crucial for building trust, yet it is often overlooked (Yun & Kim, 2025). Data use also risks: Labeling and Stereotyping (Ciabatti, 2023): Students are often treated as mere numbers and categorized into specific groups, which exacerbates social inequalities and perpetuates prejudice; False Predictions: Predictive analytics can create inaccurate profiles of students regarding their future performance, unfairly predetermining their success or failure. There is a worrying trend toward the commodification of education (Armila et al., 2024). Data infrastructure is often controlled by global corporations that exploit student activity to enhance profit-driven systems (Semião et al., 2023). This transforms education into a system based solely on measurable outcomes, promotes strict surveillance, and even commodifies the emotional aspects of learning. Constant monitoring creates psychological pressure in the form of: Anxiety and Fear: Transforming learning into mere numbers creates a mental burden for students; Loss of Autonomy: Creativity and personal initiative of both

teachers and students are limited because decisions are always dictated by available data. Finally, (Hardy et al., 2025) discuss aspects of how datafication impacts policies. Policies are used in school to guide teachers' actions with relation to students, interventions, and actions that should be taken based on data collected on students. We should not treat the student and the teacher coldly, detached from the social and cultural conditions, because these also affect teaching (Torres Soto et al., 2022).

It is a fact that between the positive and negative elements of the use of data in education, a balance should be achieved (L. Wang & Liang, 2025). Concerns are intense from various scientists. Many others wonder about the data that is collected and how it is presented. Data science makes us wonder what ethical principles apply to the analysis of learning data, who has access to the data analysis, whether a learner can opt out of such a data collection model, whether a legal framework for data analysis has been established, and how one can guarantee that the use of data science does not endanger the well-being of students and teachers. Education professionals should be aware of their ethical role in practice. (C. Wang et al., 2024) propose a model of balancing the data and the information obtained, so that there is a code of ethics that will prevent irresponsibility. In general, it contains several important points of ethical conduct. The model of is based on the DELICATE grid and is interpreted as follows: Determination: the added value of learning data analysis; Explain: Scope of data collection; Legitimate: Operation within legal frameworks; Involve: Involvement of interested parties; Consent: Consent and agreement; Anonymise: Anonymity, concealment of identification; Technical aspects: Privacy and security; External partners: Compliance with the rules by external partners. In conclusion, it is understood that a human-centered design is needed. The emphasis of data science seems to be on analytics and less on learning, according to (Du et al., 2023), and a transformation from a technocratic perspective to a more educational one is required with clear boundaries.

Conclusion

In general, there are many questions that come to our minds about the use of data in education. It is understood that in the new era of digitalization, data science has also become established in the educational field. The advantages and disadvantages are many. For this reason, a critical perspective is needed to maintain the balance between these opposing forces and to keep the human being and pedagogical science at the center. Data science as a tool of pedagogical science can offer many things and help improve education by providing students and teachers with security, autonomy, development, and improvement, because only in this way will they be able to look to the future with optimism. If this data is used correctly, then the results will be useful. The ethical use of it will make us better. In practice, this means making us all participate in educational planning and addressing the phenomenon holistically. Guidance and systematic research are needed. We must utilize all our capabilities. The research should be done in many schools, and conclusions should be drawn from the data, from students, parents, and teachers. The research should combine data from various scientific areas, examine student performance, improve the image of the school towards society in relation to the data it has collected, and not manipulate the data to ensure positive conclusions in the name of so-called innovation. But it should recognize weaknesses and address them not as numerical actions but as social processes. It requires a combination of both quantitative and qualitative analysis through interviews that will allow us to observe the behavior of students, teachers, and parents, and create a climate of freedom, respect, and trust. An ethnographic approach will help us understand a lot in order to make proper use of the data for the benefit of the educational community and society in general. Therefore, it is essential that we finally act responsibly. For years, we have been saying to focus on people, to solve problems collectively, but in the end, profit and the materialistic view prevail. Let the data companies that are now entering the field of education understand before they begin their analysis that they must move beyond an interdisciplinary approach and include in their analysis relevant social and cultural mindsets that are implicit and not explicitly expressed, and

observe how these influence the formation of their conclusions. Only in this way will they have qualitative results and not cursory considerations. In conclusion, no research can function without taking into account the nature of the beings it studies. Humans are communicative beings and therefore we need to take into account their nature itself, their psychosynthesis, and the cultural and social conditions in which they live and behave in data analyses, otherwise we will be led to incorrect conclusions.

Conflicts of Interest

The authors declare no conflict of interest

Funding

This research received no external funding.

Author contribution

Conceptualization; Methodology; Writing-Original draft preparation; Writing - Review & Editing; Conceptualization; Methodology; Supervision: **Georgios Alexandropoulos**.

Declaration of generative AI in scientific writing

During the preparation of this work, the author(s) used Google translate in order to improve English grammar and readability, or help in structuring the manuscript.

References

- Agus, A. R. R. (2022). Meta-Analysis: Correlation Study Between Science Process Skills and Learning Outcomes. *Journal of Digital Learning and Education*, 2(3), 196–202. <https://doi.org/10.52562/jdle.v2i3.427>
- Alsayer, A. A. (2023). Learners' Experiences in an Online Learning Environment: An Analysis of the Impact of International Collaboration. *Sage Open*, 13(4), 21582440231208524. <https://doi.org/10.1177/21582440231208524>
- Armila, P., Sivenius, A., Stanković, B., & Juutilainen, L. (2024). Digitalization of Education: Commodification Hidden in Terms of Empowerment? *Postdigital Science and Education*, 6(2), 556–571. <https://doi.org/10.1007/s42438-022-00347-8>
- Bednarowska-Michaiel, Z., & Uprichard, E. (2026). Bringing Interdisciplinary Data Science Education Challenges into the Classroom. *Journal of Statistics and Data Science Education*, 34(1), 72–87. <https://doi.org/10.1080/26939169.2025.2507366>
- Castellanos-Reyes, D., & Leppard, T. R. (2025). Online project-based learning to foster students' course choices in data science: A longitudinal case study using Sankey visualization. *International Journal of Educational Technology in Higher Education*, 22(1), 63. <https://doi.org/10.1186/s41239-025-00560-y>
- Ciabatti, N. (2023). Teaching about Culture or Learning with and from Others? *Societies*, 13(8), 194. <https://doi.org/10.3390/soc13080194>
- Du, R., Grigorescu, A., & Aivaz, K.-A. (2023). Higher Educational Institutions' Digital Transformation and the Roles of Digital Platform Capability and Psychology in Innovation Performance after COVID-19. *Sustainability*, 15(16), 12646. <https://doi.org/10.3390/su151612646>
- Hardy, I., Zhu, G., Hou, C., Hamid, M. O., Reyes, V., & Phillips, L. G. (2025). Data, datafication and

- data citizenship: Managing, moderating and ameliorating testing in Singapore. *Teaching and Teacher Education*, 159, 105014. <https://doi.org/10.1016/j.tate.2025.105014>
- Hsu, Y.-C. (2024). Mapping the Landscape of Data Science Education in Higher General Education in Taiwan: A Comprehensive Syllabi Analysis. *Education Sciences*, 14(7), 763. <https://doi.org/10.3390/educsci14070763>
- Leung, C. K., Pasi, G., & Wang, L. (2023). Theoretical and practical data science and analytics: Challenges and solutions. *International Journal of Data Science and Analytics*, 16(4), 403–406. <https://doi.org/10.1007/s41060-023-00465-x>
- Li, H. (2025). Impact of collaborative learning on student engagement in college English programs: Mediating effect of peer support and moderating role of group size. *Frontiers in Psychology*, 16, 1525192. <https://doi.org/10.3389/fpsyg.2025.1525192>
- Lin, K.-Y., Li, M.-H., Lo, F., Huang, H.-C., Matsuno, K., & Watanabe, R. (2025). Adaptive learning with human factors and Artificial Intelligence: Associations with training effectiveness in programming education. *International Journal of Industrial Ergonomics*, 110, 103834. <https://doi.org/10.1016/j.ergon.2025.103834>
- Liu, X. (2024). The educational resource management based on image data visualization and deep learning. *Heliyon*, 10(13), e32972. <https://doi.org/10.1016/j.heliyon.2024.e32972>
- Ogunniran, O. O., & Makinde, O. B. (2024). Assessment Strategies, Instructional Resources, Science Course Teaching Methods and Acquisition of Employability Skills by Potential Science Graduates: The Lecturers' Perspective. *Journal of Digital Learning and Education*, 4(2), 85–99. <https://doi.org/10.52562/jdle.v4i2.997>
- Ray, J., Johnny, O., Trovati, M., Sotiriadis, S., & Bessis, N. (2018). The Rise of Big Data Science: A Survey of Techniques, Methods and Approaches in the Field of Natural Language Processing and Network Theory. *Big Data and Cognitive Computing*, 2(3), 22. <https://doi.org/10.3390/bdcc2030022>
- Semião, D., Mogarro, M. J., Pinto, F. B., Martins, M. J. D., Santos, N., Sousa, O., Marchão, A., Freire, I. P., Lord, L., & Tinoca, L. (2023). Teachers' Perspectives on Students' Cultural Diversity: A Systematic Literature Review. *Education Sciences*, 13(12), 1215. <https://doi.org/10.3390/educsci13121215>
- Torres Soto, A., García Hernández, M. L., & Vallejo, M. (2022). Society and Its Challenges: The Teacher's Perspective on Students at Risk. *Social Sciences*, 11(11), 517. <https://doi.org/10.3390/socsci11110517>
- Wang, C., Chen, X., Yu, T., Liu, Y., & Jing, Y. (2024). Education reform and change driven by digital technology: A bibliometric study from a global perspective. *Humanities and Social Sciences Communications*, 11(1), 256. <https://doi.org/10.1057/s41599-024-02717-y>
- Wang, L., & Liang, G. (2025). Data-Driven Teaching and Learning Effect Evaluation in Online Education: *International Journal of Web-Based Learning and Teaching Technologies*, 20(1), 1–24. <https://doi.org/10.4018/IJWLTT.381307>
- Yun, G., & Kim, S. (2025). Analysis of high school students' use of digital devices: Focus on learning and instruction elements. *Cogent Education*, 12(1), 2468008. <https://doi.org/10.1080/2331186X.2025.2468008>