

Machine Learning-Based Tertiary Institutions' Data Harmonization with a Large Language Model Decision Support System: A Literature Review

Ekireghwo, Pius Omosehwofa¹, Blamah, Nachamada Vachaku², Ajayi, Binyamin Adeniyi³, Anyachebelu, Tochukwu Kene⁴, Salihu, Abdullahi Audu⁵

^{1,2,3,4,5}Department of Computer Science, Faculty of Natural and Applied Sciences,
Nasarawa State University, Keffi, Nigeria

Corresponding Author's Email: ekireghwo@yahoo.com

Received on: April, 2025

Revised and Accepted on: May, 2025

Published on: July, 2025

ABSTRACT

The increasing complexity of data ecosystems in tertiary institutions necessitates advanced approaches to data harmonisation. This literature appraisal investigates the purposes of Machine Learning (ML) and Large Language Models (LLMs) as groundbreaking assets for enabling data-oriented decision-making and integration. It synthesizes current research on their applications in data cleaning, feature selection, predictive analytics, and the processing of unstructured data. The findings indicate that ML and LLMs improve data quality, interoperability, and institutional responsiveness. Still, issues of morality like the bias present in algorithms, apprehensions regarding data privacy, and the imperative for clarity persist as major concerns. The review concludes with recommendations for future research to guide the responsible and effective deployment of ML and LLMs in higher education data harmonization frameworks.

Keywords: Machine learning, Data harmonisation, Tertiary institutions, large language models, Decision support systems

1.0 INTRODUCTION

1.1 Background

Data harmonisation is increasingly vital in tertiary institutions, where large volumes of data, ranging from student records to financial and academic information, are often stored in fragmented systems with inconsistent formats. This fragmentation hinders data quality, integration, and effective decision-making (García et al., 2021).

ML offers promising solutions by automating data integration, cleaning, and validation. Through pattern recognition and predictive modeling, ML enhances data consistency and supports data-driven decisions (Zhang et al., 2020).

In parallel, LLMs, such as GPT-based systems, enable the processing of unstructured data like student feedback and institutional reports. Their capacity to comprehend and produce human language renders them essential instruments for deriving insights from text-based information (Cohen & Shen, 2021).

The integration of ML and LLMs presents a compelling frontier for improving data harmonisation in higher education, although challenges such as data privacy, model transparency, and bias remain important considerations.

1.2 Purpose of the Review

This review examines the integration of ML and LLMs in data harmonisation efforts within tertiary institutions. The objective is to consolidate existing studies regarding their applications, emphasize significant challenges in implementation, and pinpoint

new opportunities for improving data-driven decision-making within higher education.

1.3 Research Questions

To guide the review, the following research questions are posed:

- i. What are the current applications of ML and LLMs in data harmonisation within tertiary institutions?
- ii. What challenges and opportunities emerge in incorporating these technologies within educational settings?

Addressing these questions will provide insights into how ML and LLMs can support more efficient, accurate, and scalable data harmonisation practices across the higher education sector.

2.0 METHODOLOGY

2.1 Literature Review Framework

This analysis embraces a systematic literature review (SLR) approach to delve into the deployment of ML and LLMs in data harmonisation within post-secondary education. The methodology is guided by the principles of Kitchenham and Charters (2007), which advocate for a transparent and replicable process in identifying, selecting, and synthesising scholarly work.

Key stages of the review included:

- a) Defining research questions and inclusion/exclusion criteria
- b) Conducting structured searches across major academic databases (e.g., Scopus, IEEE Xplore, Web of Science)



- c) Screening and quality appraisal of selected studies
- d) Thematic synthesis and critical analysis

This structured approach ensures comprehensive coverage of relevant literature, highlights prevailing trends, and uncovers research gaps. The review also integrates discussion elements to critically assess the implications, limitations, and practical relevance of ML and LLMs in institutional data harmonization.

2.2 Data Sources and Selection Criteria

The literature review is centred on published work between 2010 and 2023 from peer-reviewed journal articles, conference proceedings, and scholarly book chapters. The following inclusion criteria were applied:

- a) **Relevance:** Studies must examine ML, LLMs, data harmonisation, or decision support systems within the context of higher education.
- b) **Quality:** Only publications from reputable, peer-reviewed journals or established academic conferences were selected to ensure methodological rigor and reliability.
- c) **Diversity:** The review incorporates studies from different geographical regions and educational systems to provide a comprehensive and globally informed perspective.

Searches were conducted across major academic databases, including Google Scholar, JSTOR, IEEE Xplore, and Scopus, using combinations of the following keywords: “machine learning,” “large language models,” “data harmonisation,” and “higher education.” Boolean operators and filters were applied to refine results and exclude duplicates or non-relevant works.

2.3 Data Extraction and Analysis

Data from the selected studies were extracted and organized using predefined categories:

- a) **Author(s) and Year:** To trace the progression of research over time
- b) **Methodological Approach:** Classification of studies as qualitative, quantitative, or mixed methods
- c) **Key Findings:** Summary of insights related to the application of ML and LLMs in data harmonisation

To ensure analytical rigor, a thematic analysis was conducted following the guidelines of Braun and Clarke (2006). This involved systematically coding the data, identifying initial patterns, and refining emergent themes to develop a coherent understanding of prevailing trends, recurring challenges, and research gaps in the literature.

2.4 Limitations of the Methodology

Although the systematic literature review approach enhances rigor and transparency, certain limitations

remain. First, restricting the review to English-Language publications may exclude valuable studies published in other languages, potentially skewing the global perspective (Aldossari et al., 2021). Second, the strict inclusion criteria, while ensuring quality, may inadvertently omit relevant but less formal or interdisciplinary research.

Future reviews could mitigate these limitations by incorporating multilingual sources and adopting more inclusive selection parameters to capture a broader spectrum of insights.

2.5 Validation of Findings

To enhance the validity of the findings, triangulation was employed by synthesizing insights across multiple sources and perspectives (Denzin, 1978). Themes and conclusions were cross-checked against a diverse set of studies to ensure consistency. Additionally, expert review was conducted with scholars in educational technology and data science to strengthen the reliability of interpretations and the practical relevance of the recommendations.

3.0 OVERVIEW OF DATA HARMONIZATION IN TERTIARY INSTITUTIONS

3.1 Definition and Importance

Data harmonisation is the systematic integration and standardisation of data from diverse sources to produce a coherent, interoperable dataset. In the realm of tertiary education, this mechanism is important for fostering decisions based on solid evidence, optimising management activities, and raising academic performance (García et al., 2021).

As universities adopt data-driven approaches to areas such as curriculum planning, enrollment forecasting, and resource distribution, the value of harmonised data becomes increasingly apparent. By aligning varied data types, student demographics, academic performance, and financial records, institutions can generate comprehensive insights into operational effectiveness and student success (Hassan et al., 2021). Moreover, harmonised datasets play a crucial role in regulatory compliance and institutional accreditation by ensuring accurate, transparent, and timely reporting (Pritchard, 2020). Thus, data harmonisation serves as a foundational pillar in the advancement of institutional efficiency and accountability in higher education.

3.2 Current Practices and Challenges

Although data harmonisation is critical to institutional efficiency, many tertiary institutions continue to face significant implementation challenges. A significant obstacle is the ongoing existence of data silos, where information is segregated among departments, leading to discrepancies, redundancies, and restricted interoperability (Hassan et al., 2021). For example, student records may be fragmented across admissions,



financial aid, and academic departments, complicating efforts to construct a unified student profile.

Another key issue is the lack of standardised data formats and governance structures. Legacy systems, commonly used in many institutions, often lack compatibility with modern platforms, hindering seamless data integration (García et al., 2021). Without uniform governance policies, data quality and integrity suffer, leading to inaccurate or incomplete datasets (Kahn et al., 2019).

Moreover, institutions must now contend with the growing volume and diversity of data. From structured metrics like exam scores to unstructured sources such as textual feedback, different data types require tailored harmonisation strategies (Romero & Ventura, 2020). The swift progression of data technologies amplifies the complexity of this setting, as organisations find it increasingly difficult to realize and uphold established best practices in data management.

3.3 The Role of Machine Learning and LLMs in Data Harmonization

Emerging technologies such as ML and LLMs offer promising solutions to the challenges of data harmonisation in tertiary education. ML algorithms can automate critical processes, including data cleaning, integration, and validation, reducing manual effort, minimizing human error, and improving the reliability of harmonised datasets (Zhang et al., 2020). By uncovering patterns and correlations across distinct data sources, ML enhances consistency and supports predictive analytics for strategic planning.

Complementing ML, LLMs excel at processing and interpreting unstructured data such as student feedback, course evaluations, and institutional reports. These models can extract actionable insights from textual information, enriching the harmonisation process with qualitative dimensions often overlooked in traditional data systems (Cohen & Shen, 2021). The integration of ML and LLMs not only streamlines data management but also enables a more nuanced and comprehensive view of institutional operations, facilitating evidence-based decisions and improved academic and administrative outcomes.

4.0 MACHINE LEARNING APPLICATIONS IN DATA HARMONIZATION

4.1 Data Integration and Cleaning

A key application of ML in data harmonisation is automated data integration and cleaning, both essential for ensuring data accuracy, consistency, and usability in tertiary institutions. ML algorithms can efficiently detect and merge duplicate records, reducing redundancy and improving data reliability. Approaches like sorting and labeling assist in the structuring of alike records, whereas natural language processing

(NLP) contributes to the normalisation of written information (Wang et al., 2021).

For example, K-means clustering has been applied to classify student information based on aspects like demographics and academic results, consequently facilitating a smoother integration process (Chen et al., 2020). Additionally, ML plays a vital role in detecting anomalies and outliers, critical for maintaining data quality. Supervised learning techniques can also predict missing values by leveraging patterns in existing data, significantly enhancing dataset completeness and integrity (Zhang et al., 2020).

4.2 Feature Engineering and Selection

Feature engineering and selection are critical components in machine learning-driven data harmonisation. These processes enhance model performance by isolating the most relevant variables that influence data integrity and analytical outcomes. By analysing historical and contextual data, ML algorithms can identify features that most significantly impact student outcomes and institutional operations (Kumar et al., 2022).

Approaches including recursive feature elimination (RFE) and regularisation strategies help in discarding unnecessary or low-importance variables, thereby boosting model accuracy and lowering computational load (Huang et al., 2020). This is particularly valuable in educational contexts, where datasets often contain numerous attributes, but only a subset may meaningfully contribute to decision-making or predictive modeling.

4.3 Predictive Analytics

Machine learning-driven predictive analytics has become an instrumental asset for tertiary institutions in advancing data harmonisation. By applying ML algorithms to historical and real-time datasets, institutions can generate forecasts on key metrics such as enrollment trends, student retention, and academic performance (Papamitsiou & Economides, 2014). These predictive insights support proactive decision-making and help address systemic challenges, such as fluctuating enrollment or increasing dropout rates. For example, regression models are used to forecast future enrollment based on historical trends, enabling more strategic resource allocation and infrastructure planning (Siemens, 2013). Similarly, classification algorithms can detect students at risk of underperforming or dropping out by analysing multiple variables, thereby informing early intervention strategies and improving institutional retention outcomes.

4.4 Enhancing Data Interoperability

ML also plays a pivotal role in enhancing data interoperability across diverse systems within tertiary institutions. By learning data mappings across varying formats and schemas, ML algorithms enable seamless



integration of data from heterogeneous sources (Baker & Inventado, 2014). For instance, transfer learning facilitates the modification of models trained on a specific institutional dataset for application on another, thereby improving how efficiently data from various platforms can be integrated (Huang et al., 2020).

Furthermore, ML addresses semantic inconsistencies that often arise between datasets, which can hinder harmonisation efforts (D'Amato et al., 2020). Through automated semantic mapping, ML techniques reconcile variations in data definitions and ensure uniform interpretation of data elements across systems, thereby improving both data quality and usability. Collectively, these applications demonstrate that ML not only automates and refines the harmonisation process but also significantly improves the interoperability and consistency of institutional data, foundational for strategic planning and evidence-based decision-making in higher education.

5.0 LARGE LANGUAGE MODELS IN DECISION SUPPORT SYSTEMS

5.1 Definition and Functionality of Large Language Models

LLMs are refined artificial intelligence systems that utilize deep learning, particularly transformer architectures, to understand, generate, and interact with human language. Leading examples feature OpenAI's GPT series alongside Google's BERT, both developed by harnessing vast repositories of text to identify language patterns, meanings, and contextual relationships (Devlin et al., 2019).

LLMs can create logical, contextually fitting replies and simultaneously uncover valuable information from unstructured text. Their ability to comprehend nuanced language makes them especially valuable in decision support systems (DSS) within higher education, where data often includes diverse forms of textual input such as student feedback, policy documents, and academic reports. Through natural language understanding and generation, LLMs enhance data-driven decision-making by enabling more intuitive access to institutional knowledge and more dynamic analysis of qualitative data.

5.2 Enhancing Data Interpretation and Analysis

Within decision support systems, LLMs significantly enhance the interpretation and analysis of unstructured textual data. By applying advanced natural language processing, LLMs can analyse content such as open-ended survey responses, course evaluations, and institutional reports to identify key themes, sentiments, and actionable insights (Cohen & Shen, 2021). This talent nurtures a stronger comprehension of stakeholder perspectives, allowing organisations to arrive at decisions that are both highly informed and exceptionally responsive.

Furthermore, LLMs streamline the synthesis of complex documents by extracting and summarising essential information. For instance, they can generate concise overviews of lengthy policy documents or academic reports, allowing decision-makers to focus on critical findings without being overwhelmed by detail (Bhanot et al., 2021). This function is particularly valuable in resource-constrained settings, where rapid and accurate analysis of qualitative data is essential for timely institutional responses.

5.3 Supporting Predictive Decision-Making

LLMs play an important role in enhancing predictive decision-making within tertiary institutions. By analysing historical data alongside external contextual factors, such as economic trends and social dynamics, LLMs can generate informed forecasts related to student enrollment, retention, and academic performance (Kumar et al., 2022). For example, integrating unstructured data like student feedback with quantitative metrics allows LLMs to provide deeper insights into factors shaping institutional outcomes.

This fusion of diverse data types supports the development of advanced predictive analytics tools, which aid strategic planning by offering a holistic view of institutional performance drivers (Siemens, 2013). Consequently, decision-makers can leverage these insights to make proactive, data-driven choices aligned with institutional goals and priorities.

5.4 Natural Language Processing for Improved Communication

The integration of LLMs in decision support systems significantly enhances communication among stakeholders within tertiary institutions. LLMs facilitate the automatic generation of reports, memos, and policy documents, ensuring that information is articulated and aligned with institutional goals (Zhou et al., 2022). This capability reduces the burden on administrators and improves the consistency and professionalism of institutional communications.

Additionally, LLMs enable the customisation of content for diverse audiences, including students, faculty, and administrators, thereby promoting transparency and stakeholder engagement (Khan et al., 2023). By tailoring language and presentation style, LLMs help foster a data-informed culture where all members of the institution can effectively participate in and contribute to decision-making processes.

5.5 Ethical Considerations in the Use of LLMs

Despite their potential, the use of Large Language Models in decision support systems raises critical ethical concerns. Key issues include data privacy, algorithmic bias, and the need for transparency in AI-generated recommendations (Binns, 2018). Tertiary



institutions must ensure that LLM deployment adheres to ethical standards, with thorough oversight to maintain fairness and accuracy in decision-making processes.

Ongoing evaluation of model performance is necessary to identify and mitigate biases or misinterpretations that could impact institutional outcomes. By addressing these concerns proactively, institutions can responsibly harness the power of LLMs to enhance data harmonization and support informed, equitable strategic decisions.

6.0 DISCUSSION AND SYNTHESIS OF FINDINGS

6.1 Integration of Machine Learning and Large Language Models

The reviewed literature demonstrates a growing recognition of the complementary strengths of ML and LLMs in advancing data harmonisation within tertiary institutions. ML shows remarkable skill in enhancing efficiency for actions like data purification, combining datasets, and exploring arranged data collections, while large-scale language models equip users with advanced resources for acquisitive and creating human language, which facilitates the extraction of valuable insights from unorganised text information (Zhang et al., 2020; Devlin et al., 2019). The combination of these technologies creates a collaborative decision support system that motivates organisations to tap into a wider array of data sources, ultimately delivering deeper and more insights into institutional effectiveness and student achievement.

6.2 Addressing Data Quality Challenges

A consistent finding across the literature is the persistent challenge of maintaining high data quality within higher education institutions. Key concerns encompass the disjointedness of data scattered among departmental silos, variations in data entries, and the lack of uniform formats and governance guidelines (Hassan et al., 2021; García et al., 2021). These challenges hinder comprehensive data harmonisation and limit the utility of institutional data in strategic planning and decision-making.

ML approaches, especially in the contexts of anomaly identification and forecasting models, yield effective answers by refining the mechanisms of spotting and fixing mistakes, decreasing overlap, and boosting the overall thoroughness of datasets (Wang et al., 2021). Concurrently, LLMs strengthen these efforts by extracting meaningful, contextually grounded insights from unstructured sources such as open-ended surveys and institutional documents. This dual approach fosters a more integrated and accurate data environment, paving the way for reliable, data-driven decision-making in tertiary education.

6.3 Enhancing Decision-Making Processes

The convergence of ML and LLMs elevates the decision-making capabilities of tertiary institutions. ML-powered predictive analytics empower institutions to accurately project enrollment trends, enhance resource distribution, and foresee student performance results with remarkable precision (Kumar et al., 2022). These predictive insights facilitate proactive and strategic responses to emerging challenges.

Simultaneously, LLMs provide robust support in processing and interpreting qualitative data, such as student feedback, course evaluations, and institutional narratives, extracting key themes and sentiments that would otherwise remain underutilised (Cohen & Shen, 2021). The synergy between ML's quantitative strength and LLMs' contextual understanding fosters a more agile and data-informed decision-making environment. This integrated approach ensures that institutional policies and strategies are grounded in a holistic interpretation of both structured and unstructured data, thereby improving responsiveness and effectiveness in governance and planning.

6.4 Implications for Institutional Practices

The adoption of ML and LLMs in data harmonisation presents opportunities for transforming institutional practices. Data integration, cleansing, and predictive analytics processes are streamlined through ML, which subsequently boosts operational workflows and bolsters resource planning informed by data insights (Huang et al., 2020). LLMs, by extracting insights from unstructured data, support nuanced understanding of stakeholder feedback, informing curriculum design and institutional policy (Baker & Inventado, 2014).

Together, these technologies promote a data-informed culture, fostering cross-departmental collaboration and improving institutional responsiveness. The resulting unified datasets support transparency, compliance, and strategic decision-making, ultimately advancing academic quality and student outcomes. Institutions must, however, invest in technical capacity and ethical oversight to maximize the benefits of these innovations.

6.5 Ethical Considerations and Challenges

Despite the transformative potential of ML and LLMs in data harmonisation, their application raises critical ethical concerns. Key issues include algorithmic bias, data privacy, and the opacity of AI-driven outputs (Binns, 2018; Kahn et al., 2023). Without robust oversight, these technologies risk reinforcing systemic unfairness or compromising user trust.

To meet these challenges, entities are required to formulate governance frameworks that sustain ethical norms, advance clarity in AI functions, and facilitate



equitable choices. This includes establishing protocols for data protection, auditing ML/LLM outputs, and ensuring inclusivity in model design and deployment. Ethical stewardship is essential to sustaining trust and equity in AI-supported educational systems.

6.6 Future Research Directions

This synthesis reveals key avenues for future inquiry. Longitudinal studies are needed to assess the sustained impact of ML and LLM integration on institutional performance across varied educational settings. Additionally, empirical case studies documenting successful implementations can offer actionable insights. Research should also focus on developing standardised frameworks for the ethical deployment of AI in higher education, ensuring equity, transparency, and accountability (Siemens, 2013; Papamitsiou & Economides, 2014). Such efforts will deepen our understanding of both the transformative potential and limitations of these technologies in academic environments.

7.0 CONCLUSION

The integration of ML and LLMs offers a transformative pathway for advancing data harmonisation in tertiary institutions. This investigation has revealed the way these technologies confront ongoing challenges including data silos, variances, and the nuances related to the management of both structured and unstructured data. ML contributes by automating data cleaning, integration, and predictive analytics, while LLMs enhance the interpretation of textual data and support refined decision-making processes. Together, they enable institutions to build comprehensive decision support systems that improve operational efficiency, optimise resource allocation, and elevate the quality of educational outcomes. As higher education systems face increasing demands for accountability and alertness, adopting AI-driven solutions will be critical for cultivating a data-informed institutional culture. Continued research and ethical oversight will be essential to fully realize the benefits while safeguarding fairness, privacy, and transparency.

REFERENCES

Aldossari, M., Alkhalifah, A., & Alkhaldi, A. (2021). A systematic review of data harmonization techniques for educational data mining. *Journal of Educational Computing Research*, 59(3), 463–489. <https://doi.org/10.1177/0735633120960422>

Baker, R. S. J. D., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds.), *Learning analytics: From research to Practice* (pp. 61–75). Springer. https://doi.org/10.1007/978-1-4614-3305-7_4

Bhanot, K., Kaushik, M., & Kumar, A. (2021). Leveraging transformer-based models for automated

policy document summarization. *IEEE Access*, 9, 51345–51357. <https://doi.org/10.1109/ACCESS.2021.3068532>

Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. *Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency (FAT)*, 149–159. <https://doi.org/10.1145/3287560.3287598>,

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>.

Chen, H., & Zhang, X. (2021). Machine learning for data harmonization in higher education: Challenges and prospects. *Journal of Educational Technology*, 12(2), 15-27.

Chen, X., Xie, H., Cheng, S., & Lei, W. (2020). Using K-means clustering to analyze student learning behavior in online learning. *Education and Information Technologies*, 25(6), 5321–5337. <https://doi.org/10.1007/s10639-020-10247-9>.

Cohen, J., & Shen, W. (2021). Comparative analysis of integrated approaches in educational data processing. *International Journal of Educational Management*, 35(6), 1023-1038.

Cohen, M., & Shen, Y. (2021). Large language models for education: Promises and perils. *AI & Society*, 38, 291–302. <https://doi.org/10.1007/s00146-021-01191-w>.

D’Amato, C., Esposito, F., Leone, P., & Redavid, D. (2020). Semantic interoperability through machine learning in education: An ontology-based framework. *Semantic Web*, 11(1), 47–65. <https://doi.org/10.3233/SW-190360>.

Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological methods* (2nd ed.). McGraw-Hill.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT 2019*, 4171–4186. <https://doi.org/10.48550/arXiv.1810.04805>.

García, C. E., López, J. M., & Romero, C. (2021). A survey on data integration and data harmonization for educational environments. *Computers in Human Behavior Reports*, 3, 100065. <https://doi.org/10.1016/j.chbr.2021.100065>.

García, E., Pérez, L., & Martínez, A. (2021). Data preprocessing and integration in educational contexts: Challenges and solutions. *Journal of Educational Technology*, 10(1), 45-58.

Hassan, S., Habib, M., & Rashid, T. (2021). Challenges in higher education data harmonization: A case study of Pakistani universities. *International Journal of Advanced Computer Science and Applications*, 12(2), 134–142.



- <https://doi.org/10.14569/IJACSA.2021.0120217>.
- Hassan, S., Sardar, S., & Ahmed, A. (2021). Text normalization techniques for data preparation in NLP applications. *Journal of Language Processing*, 8(2), 32-47.
- Huang, S., Wu, L., Zhang, Q., & He, D. (2020). Improving feature selection in educational data mining with regularization techniques. *Applied Intelligence*, 50(9), 2764–2780. <https://doi.org/10.1007/s10489-020-01727-4>.
- Kahn, J., Bowser, A., & May, C. (2019). Guidelines for data governance in education: Toward a balanced approach. Data & Society Institute Report.
- Khan, S., Saeed, U., & Mahmood, T. (2023). Ethical concerns in AI-driven decision-making: A higher education perspective. *Education and Information Technologies*, 28, 491–510. <https://doi.org/10.1007/s10639-022-11345-0>.
- Kahn, J., Freedman, M., & Hill, A. (2019). Data governance in higher education: Strategies for success. *Journal of Educational Data Science*, 6(4), 89-103.
- Kitchenham, B., & Charters, S. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering. Technical Report EBSE 2007-001, Keele University and Durham University.
- Kumar, A., Singh, A., & Jha, D. (2022). Machine learning applications in educational data mining: Predictive modeling and beyond. *Education and Information Technologies*, 27(1), 891–913. <https://doi.org/10.1007/s10639-021-10706-8>.
- Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Educational Technology & Society*, 17(4), 49–64.
- Pritchard, R. M. (2020). Accreditation and quality assurance in higher education: Global perspectives. *International Journal of Educational Development*, 78, 102244. <https://doi.org/10.1016/j.ijedudev.2020.102244>.
- Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3), e1355. <https://doi.org/10.1002/widm.1355>.
- Romero, C., & Ventura, S. (2020). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3), e1322.
- Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, 57(10), 1380–1400. <https://doi.org/10.1177/0002764213498851>.
- Wang, Y., Yu, S., & Fong, S. (2021). Data cleaning for education data mining: A comprehensive survey. *IEEE Access*, 9, 135513–135530.
- <https://doi.org/10.1109/ACCESS.2021.3117634>.
- Zhang, Y., Wu, Y., & Chen, Q. (2020). The role of large language models in enhancing data interoperability: Opportunities and challenges. *Journal of Information Science*, 46(4), 505-517.
- Zhang, Y., Ren, F., & Liu, Q. (2020). Machine learning for educational data: A systematic review. *Computers & Education*, 156, 103954. <https://doi.org/10.1016/j.compedu.2020.103954>.
- Zhou, M., Lin, S., & Xie, Y. (2022). Enhancing institutional communication through AI-generated content. *Journal of Educational Technology Development and Exchange*, 15(1), 55–72. <https://doi.org/10.18785/jetde.1501.05>