

Article

Not peer-reviewed version

Digital Twin-Based Learning Analytics with Fog Computing and LLAMA

[Abdelkawy A. Abdelaziz](#)*, [Sameh Abd EL-Ghany](#), [A. S. Tolba](#)

Posted Date: 5 August 2025

doi: 10.20944/preprints202508.0255.v1

Keywords: digital twin; learning analytics; fog computing; large language model; generative ai; personalized learning



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Digital Twin-Based Learning Analytics with Fog Computing and LLAMA

Abdelkawy A. Abdelaziz ^{1,3}, Sameh Abd EL-Ghany ^{2,3} and A. S. Tolba ⁴

- ¹ Modern College of Business and Science (MCBS), Muscat, Oman
- ² Jouf University, College of Computer and Information Sciences, Department of Information Systems, Al-Jouf, Saudi Arabia
- ³ Mansoura University, Faculty of Computers and Information, Information Systems Department, Mansoura 35516, Egypt
- ⁴ Mansoura University, Faculty of Computers and Information, Computer Sciences Department, Mansoura 35516, Egypt
- * Correspondence: abdelkawy@mcbs.edu.om

Abstract

Real-time learning analytics in higher education are often constrained by the latency, bandwidth, and privacy limitations of cloud-only architectures, which hinder the de-livery of timely, actionable feedback; this study addresses that gap. We introduce Learner’s Digital Twin, a framework that integrates fog computing at the network edge with Meta-LLAMA to interpret multimodal student data and provide instant, personalized feedback and educator insights. The architecture performs local processing on fog nodes to reduce delay and limit data movement, while LLAMA generates context-aware text analyses; predictive components include linear regression to forecast final-exam scores from attendance, assignment averages, and participation, and K-means clustering to profile learning patterns. We evaluated the framework in a real educational setting over three months, using Postman-based latency tests and user surveys. The system reduced average response latency by ~300ms. The feedback generated was personalized, and survey responses indicated positive user perceptions: for students, 80% reported overall satisfaction, with >90% perceiving the feedback as personalized and >75% finding it relevant; teachers similarly reported ~80% satisfaction. These findings indicate that combining a digital-twin paradigm with fog computing and LLM can support timely, personalized feedback and actionable insights in high-er-education contexts; future work should examine scalability and generalizability across diverse settings.

Keywords: digital twin; learning analytics; fog computing; large language model; generative ai; personalized learning

1. Introduction

In the rapidly changing face of today's digital world, educational technology has become an essential disruptor of how students learn and how instructors teach. A fitting example of this is the emergent field of real-time learning analytics, which seeks to provide immediate data collection, analysis, and interpretation for timely feedback and support directly to students (Hernández-de-Menéndez et al., 2022; L. Lim et al., 2023). However, prolonged periods of time have been required in traditional cloud-based learning analytics systems to aggregate and evaluate the requisite data; this extends latency (the delay before a transfer begins), bandwidth constraints due to excessive

transmission, and privacy issues associated with collecting raw data that can inhibit beneficial real-time feedback (Pardo et al., 2019).

These are the challenges that demand out-of-the-box and often unconventional solutions to work around cloud-based infrastructure (Ometov et al., 2022). We present a new framework known as "Learner's Digital Twin", which integrates digital twin technology with leading-edge fog computing and LLAMA to advance real-time learning analytics. The digital twin technology, which was pioneered for industrial use, is a modern take on the same concept in an educational context: giving us a live mirror of how and what our students are doing (Fuller et al., 2020).

The incorporation of fog computing in this framework resolves the inherent latency and bandwidth problems associated with conventional cloud computing; it enables data processing on the network edge (closer to the learner) (Karatza, 2020; Zarzycka et al., 2021). Therefore, to reduce the amount of time needed to process and analyze the data, the system should be closer so that feedback can be delivered with minimum latency (Wise et al., 2014). In addition, LLAMA enhances the system by providing answers in personalized forms with greater context sensitivity, which are produced after analyzing more complex student processed data (Devlin et al., 2018).

A linear regression model is built to predict student performance — leveraging historical statistics for the system to foresee academic results. Next, students are clustered using K-means clustering by means of learning behavior and engagement patterns to provide detailed interventions (Hudli et al., 2012; Moubayed et al., 2020).

In this research, the problem has been formulated as that of exploring whether fog computing can contribute to processing big data generated within Learner's digital twin framework by using LLAMA and then performing linear regression and K-means clustering on processed results in real time (Roumeliotis et al., 2024). We predict that such a framework will enhance the relevance and timeliness of feedback provided to students, with clear gains (that the above hypotheses suggest) in learning outcomes and the ability to evolve teaching strategies. This paper is structured as follows: Section 2 presents related work on digital twin technology, real-time learning analytics and fog computing before introducing the proposed LLAMA model, which uses regression and clustering techniques. In Section 3, we will describe our proposed framework, "Learner's Digital Twin", including the building blocks, underlying models utilized, and procedure for generating feedback personally suitable for a learner from his usage data. Section 4 presents the results obtained when applying our proposed framework, and in Section 5, we discuss the outcomes, highlighting some of the typical challenges faced by similar measures and pointing out future directions. Next, Section 6 concludes the paper.

2. Theoretical and conceptual Framework

Digital Twin-based learning analytics, empowered by Fog Computing and LLAMA, represents a sea change in personalized education. Such a framework leverages real-time data processing and advanced AI models for dynamically responding to the learning environment. Integration of such technologies enables immediate feedback and tailored learning experiences, improving educational outcomes significantly (Abdelaziz et al., 2024, Qin et al., 2024).

Digital Twin Framework The "Learner's Digital Twin" enables real-time analytics by interpreting multimodal data, providing personalized alerts and insights for educators. It employs predictive analytics through Linear Regression and clustering techniques like K-means to identify learning patterns (Abdelaziz et al., 2024). **Role of Fog Computing** Fog Computing processes data at the network edge, reducing latency and enhancing the responsiveness of learning systems (Pushpa & Kalyani, 2020). This architecture supports efficient handling of data generated from different educational activities for timely feedback to students and educators Pushpa & Kalyani, 2020. **Integrating LLAMA** The LLAMA model extracts context from texts. With this, the nuances in the interaction between different students and their learning behaviors come into light Abdelaziz et al., 2024. This enables a great degree of personalization of the learning experience, adapting to individual needs in real time Abdelaziz et al., 2024. While the integrated approach has a lot of benefits, challenges like data privacy and the need for robust infrastructure need to be addressed if it must be put to optimal use in educational settings.

Learning analytics theories encompasses a range of frameworks that inform the use of data to enhance educational practices. These theories have evolved significantly, integrating insights from various disciplines such as educational psychology, sociology, and cognitive science. The interplay between learning analytics and established educational theories is crucial for developing effective pedagogical strategies and improving student outcomes (Drugova et al., 2024).

2.1. Learning Analytics Theories

1. **Learning Analytics Theory (LAT):** Focuses on using data to inform teaching practices and improve learning outcomes by analyzing student interactions and engagement (Alam, 2023).
2. **Cognitive Load Theory:** Addresses the mental effort required for learning, guiding the design of instructional materials to optimize learning experiences (Giannakos & Cukurova, 2023).
3. **Control-Value Theory of Achievement Emotions:** Explores how students' emotions influence their learning processes and outcomes, providing insights for emotional support in educational settings (Giannakos & Cukurova, 2023).

2.2. Learning Analytics Theory

Provides a series of frameworks and methodologies that are based on data for improving education. It captures insights from several disciplines, including education research and sociology, that inform pedagogical strategy and the improvement of learning outcomes. Understanding student engagement and learning processes has become more nuanced with the development of learning analytics; this nuanced understanding shows that theoretical considerations are central to interpreting data effectively.

Key Theoretical Foundations Cognitive and Affective Theories, Many MMLA studies use theoretical frameworks such as cognitive load theory and control-value theory for analyzing learning behaviors and emotional responses of learners. Examples include Giannakos and Cukurova (2023). Self-Regulated Learning, Many studies use theories of self-regulated learning to interpret data and improve student engagement through analytics, such as Wang et al. (2022). Application in Learning Design Data-Driven Decisions, Learning analytics also offer insights into students' activities, thus enabling educators to understand where to improve for raising course completion rates (Kesylé & Melnikova, 2024). Peer Review Engagement, Theory-based approaches have demonstrated their potential for learning analytics in analyzing engagement patterns by students in peer review activities to find unexpected behaviors and refine exiting theories (Er et al., 2021).

A gap still exists in the integration of theory within learning analytics, with many studies remaining firmly data-driven (Wang et al., 2022). This contributes to further theoretical development if the full impact of the field is to be realized in improving educational practices. In contrast, others say that over-reliance on theory can hold back the innovative potential of learning analytics; thus, a balance between empirical data and theoretical frameworks in future research and practice is necessary.

The field of learning analytics has transitioned from a focus on data and systems to a broader integration of educational theories, enhancing the understanding of learning processes (Lodge et al., 2023). The establishment of organizations like the Society of Learning Analytics Research (SoLAR) has facilitated the growth of this field, promoting collaboration and research dissemination ("Learning Analytics in Higher Education", 2023). Learning analytics can identify student learning styles and predict academic performance, allowing for personalized learning experiences (Alam, 2023). Ethical considerations, such as student privacy and data security, remain paramount as the field expands (Čalopa, et al., 2023). While the integration of learning analytics with educational theories shows promise for enhancing learning outcomes, challenges remain in ensuring that these theories are effectively applied in practice. The ongoing dialogue about the role of theory in learning analytics is essential for addressing these challenges and advancing the field.

2.3. Cognitive Load Theory (CLT)

Provides a framework for appreciating the mental effort required by learners in education. It classifies cognitive load into three categories: intrinsic, extraneous, and germane, which differently affect learning outcomes. CLT has been used in a variety of educational areas, including medical training, nursing education, and machine learning, showing its usefulness in improving teaching methods. Major Components of Cognitive Load, Theory Intrinsic Load, this refers to the complexity of the material one is learning. For example, students in anesthesia training reported a high intrinsic load because the tasks were complex (Spijkerman et al., 2024). Extraneous Load, this consists of distractions and unhelpful teaching methods. In nursing education, extraneous load was found to largely affect decision-making skills, implying that handling this load is very important for good learning (Tabatabaee et al., 2024). Germane Load, This refers to the mental effort put into processing and making sense of information. Increased germane load might lead to improved learning outcomes, as demonstrated in specialized training for clinical reasoning among medical students (Si, 2024). Although CLT gives quite a strong base for the understanding of learning processes, some would criticize that it does not sufficiently account for individual differences in cognitive processing, which hints toward the necessity of more personalized strategies in education.

2.4. The Control-Value Theory (CVT) of Achievement Emotions

Posits that individuals' emotional experiences in academic contexts are influenced by their perceptions of control and value regarding their tasks. Emphasized in this theory is the fact that emotions such as enjoyment, anxiety, and boredom can have strong effects on motivation and academic performance. The following sections elucidate important aspects of CVT. Key Components of Control-Value Theory, first, Control Appraisals, refers to students' beliefs about their ability to influence outcomes. Higher perceptions of control are related to positive emotions such as enjoyment and hope (Armstrong, 2023). Second, Value Appraisals, Relates to the perceived meaningfulness or utility of a task. Intrinsic value in particular exerts strong influence as it can overpower negative feelings like anxiety and boredom (Abuzant et al., 2023). Affective Outcomes in Learning Settings Positive Emotions: Positive leadership and clear instructional communication enhance teachers' and students' positive emotions, fostering a supportive learning environment (Goetz et al., 2024) (Armstrong, 2023). Negative Emotions: Factors like test anxiety and boredom, therefore, discourage students from continuing with higher education, since it indicates the negative outcomes of achievement emotions (Yim et al., 2023). Implications for Teachers Educators can utilize CVT by increasing students' sense of control and value through clear communication and relevant content to promote positive emotional experiences, which in turn fosters academic success (Armstrong, 2023). On the other hand, while CVT focuses on the centrality of emotions in learning, others argue that cognitive aspects like prior knowledge and learning strategies may equally be strong determinants of academic achievement, suggesting a more holistic approach toward understanding student outcomes.

3. Digital Twin Technology

Refers to the process of making virtual models of physical ones to enhance real-time monitoring, simulation, and optimization in various industries. Technology enhances operational efficiency, predictive maintenance, and decision-making, therefore being an important tool in manufacturing, healthcare, and renewable energy. The following sections discuss the major aspects of DTT. Manufacturing Applications Smart Manufacturing: DTT allows real-time data gathering and analysis, thus bringing improvements in operational effectiveness and product quality. Reference or citations: Sethi et al., 2024 and Singh & Gameti, 2024. Predictive Maintenance: Under simulated real-world conditions, the DTT helps in forecasting failures in equipment, thus lessening downtime. Reference: Singh & Gameti, 2024 Impact on Renewable Energy Microgrid Management: DTT helps in optimizing the performance of renewable energy microgrids through demand-supply balancing for improving efficiency and ensuring stability. Resilience and Contingency Planning: It models the impact of extreme weather, which helps develop appropriate contingency plans (Bassey et al., 2024). Challenges and Future Directions Implementation Challenges are Technical, ethical, and privacy issues hamper the acceptance of DTT (Zhou et al., 2024). Future Trends are Advancements in AI, edge

computing, and 5G are expected to further increase the capability of DTT. While DTT comes with several advantages, it also has its set of challenges, such as management and security of data. Further developments on this technology show promise in the future across several sectors (Delerm & Pilottin, 2024).

Digital Twin Technology faces significant challenges and future directions in various fields, such as interoperability, legal frameworks, data integrity, and security issues. With the advancement of digital twins, these challenges must be addressed to ensure their effective deployment and integration into existing systems. The following sections outline the main challenges and future directions for the development of digital twin technology.

Interoperability Issues Inadequate Interoperability, Nowadays, interoperability problems are common in existing digital twin systems, which undermines their effectiveness in integrated environments (Dávid et al., 2024). Setting common standards is vital in enabling efficient communication between different digital twins and hence enhancing their general functionality (Dávid et al., 2024).

Legal and Ethical Considerations Data Privacy Issues: The application of digital twins in the health sector faces significant legal challenges regarding data privacy and possible surveillance, which requires stronger regulatory frameworks (Delerm & Pilottin, 2024).

Ethical Consideration: Integration of artificial and machine learning into digital twins raises several ethical issues within the digital health sector regarding data usage and patients' rights (Abayadeera & Ganegoda, 2024).

Data Quality and Security Data Quality Assurance: High-quality data is imperative for the accurate functioning of digital twins, but many industries face challenges in terms of data integration and the control of quality (Zhou et al., 2024).

Security Measures is very important to protect sensitive data from breaches, especially in applications involving personal health information (Zhou et al., 2024).

Future Directions Research on Standardization: Future research should focus on developing robust standards and methodologies to enhance interoperability and data integration across domains (Abayadeera & Ganegoda, 2024).

Regulatory Improvements: The legal frameworks will need to be updated to deal with complexities brought about by digital twins for safe and efficient use in public health and other sectors (Delerm & Pilottin, 2024). While digital twin technology holds transformative potential, its successful implementation hinges on overcoming these challenges. Conversely, the rapid evolution of digital twins may outpace regulatory and ethical considerations, leading to potential misuse or unintended consequences if not carefully managed (Yang et al., 2024).

- **Simulation and replication theories**

Are at the core of any advance in empirical research and theoretical development in most fields. They stress the need for the replication of simulation experiments in establishing results that strengthen the resilience of theoretical conclusions. The main elements of these theories are discussed in the following sections.

The Role of Replication in Theory Development Replicated simulations have the potential to strengthen existing theoretical models by exploring new situations, such as organizational mergers, and testing the impact of variables like memory on performance (Hauke et al., 2020). Initiatives in standardization, such as the ODD protocol and DOE principles, now make replication easier, increasing transparency and reproducibility of simulations (Hauke et al., 2020).

Challenges in Replicability The replicability of statistical simulation studies varies highly; some provide enough detail that a reader can accurately replicate them, while others omit crucial information (Luijken et al., 2023) ("Replicability of Simulation Studies for the Investigation of Statistical Methods: The RepliSims Project", 2023). Factors that increase reproducibility include open access to coding data and detailed descriptions of methods, which are usually not included in original publications (Luijken et al., 2023).

Applications in Education In teacher education, simulations offer a safe space for active engagement and successfully model real-life situations for improved educational outcomes (Orland-Barak & Maskit, 2017). The theoretical frameworks underpinning such simulations highlight their efficacy in delivering experiential activities for teacher learning (Orland-Barak & Maskit, 2017). While replication is essential for validating simulation studies, it is also crucial to recognize that not all simulations yield universally applicable results. Contextual

factors can significantly influence the outcomes, necessitating careful interpretation of findings (Edmonds & Hales, 2005).

- **Continuous feedback mechanisms**

Are integral, be it in neuroscience, quantum physics, or even organizational management. These systems enable the making of real-time adjustments and improvements with ongoing input, which improves performance and adaptability. The subsequent sections develop the applications and implications of continuous feedback across the various domains. **Neuroscience and Reward Processing** In neuroscience, continuous feedback refers to the processing of the brain in terms of ongoing rewards. Evidence has shown that midbrain dopaminergic activity can track moment-to-moment changes in reward, suggesting a nuanced understanding of reward prediction errors, or RPEs (Hassall et al., 2023). EEG studies have demonstrated that continuous feedback can be measured through scalp potential coupled with reward anticipation, reflecting the adaptability of the brain based on the expectation of rewards versus no rewards (Hassall et al., 2023). **Quantum Systems** In quantum physics, continuous feedback is used to stabilize the state of a quantum gas in real time. A micro-processor-controlled feedback architecture maintains a constant intra-cavity photon number, enabling precise control near critical phase transitions (Kroeger et al., 2020). This technique illustrates the possibility to drive complex many-body phases in quantum systems with tailored feedback mechanisms (Kroeger et al., 2020). **Material Science** Continuous feedback mechanisms are also applied in the tuning of vanadium dioxide films for precise phase transitions. A robust feedback control approach suppresses hysteresis and thus enables continuous phase tuning without detailed modeling (Dai et al., 2019). **Organizational Management** In business, continuous performance management (CPM) replaces traditional annual reviews with regular feedback sessions. This agile approach fosters ongoing communication between managers and employees, adapting goals to meet evolving business needs (Traynor et al., 2021). While continuous feedback mechanisms have a very high degree of advantage when it comes to adaptability and performance in many fields, they also have their pitfalls, such as the need for quite sophisticated systems to process and respond with feedback. Balancing these facets is important to maximize these advantages of continuous feedback.

Fog Computing Paradigm

Represents an emerging paradigm that extends the capability of cloud computing toward the network edge, enabling real-time processing and storage of data closer to users and devices. It helps solve the shortfalls that traditional cloud computing presents when faced with scenarios that involve many devices generating huge amounts of data, such as IoT applications. Fog computing improves performance, decreasing latency while serving various applications for health, education, and marketing. Further sections discuss its architecture, application, and challenges. **Architecture of Fog Computing** The decentralized structure, fog computing has an architecture based on a decentralized system that spreads computing resources throughout the network rather than being confined to some cloud servers in the center (Paul, 2024). **Integration with IoT:** It acts as an intermediary between IoT devices and cloud services, enabling local data processing and reducing bandwidth usage (Swarnakar, 2024). **Applications of Fog Computing** **Healthcare:** Fog computing provides real-time monitoring and diagnostic services, addressing challenges like doctor shortages and high treatment costs (Datta & Datta, 2024). **Marketing:** It supports data management for marketing research, allowing for more agile and responsive strategies in fast-paced markets (Hornik & Rachamim, 2024). **Education:** It enhances online learning management and operational efficiency in educational institutions by 2024. **Challenges and Future Directions** **Data Security:** Data integrity and security are paramount concerns in the implementation of fog computing. **Standardization:** Standardized protocols and models are a must if this technology must find wider acceptance. Fog computing offers a bunch of advantages, but there is always a challenge that requires more research and development to make it completely implementable for every industry.

- **Distributed Computing Theory:**

Distributed computing theory refers to concepts and methodologies that, when put into practice, help to process and handle data across several computing resources. It has become prominent since there has been an increasing need for better handling of data, which big data and machine learning bring forth. The following sections outline key aspects of distributed computing theory. Overview of Distributed Computing Distributed computing is a technique of performing complex computations with multiple interconnected computers sharing resources and data. It plays a crucial role in several applications, including cloud computing, where resources are managed across different service providers (Thakur et al., 2023). Synchronization Methods Synchronization plays a critical role in distributed systems to ensure consistency and coordination among processes. There exist four major synchronization methods or barriers that are commonly used, each with different performance trade-offs (Wang & Zhao, 2022). Big Data and Machine Learning Applications Distributed computing has helped deal with handling vast sets of data, thus solving some problems involved with processing and storing data (Aggarwal, 2020). In Machine Learning applications, several methodologies involving Queuing theory maximize the correct usage of resources that ensures processes are well-streamlined Azarnova & Polukhin 2021 Ethics The rapid evolution of distributed computing brings along different ethical concerns related to privacy, security, and potential biases in handling data (Aggarwal, 2020). Although distributed computing ensures huge efficiency and scalability advantages, at the same time, it also raises various challenges that require due care to be shown, particularly in ethical consequences and potential abuse of technology.

• Edge Processing Models:

signify a radical re-technology in computing that analyzes and makes decisions on the real-time data at the edge of the network. Such an evolution addresses the challenges-ensuring low latency, band limitations-associated with cloud-centric AI by embedding artificial intelligence right at the edge devices themselves. The following sections discuss important aspects of edge processing models. Overview of Edge Processing Model Real-time Processing: Edge AI enables the processing of data in real time, which is critical in applications such as autonomous vehicles and smart cities. - Chandrasekaran et al. (2024) Resource Optimization: Complex models such as SSD and YOLO are optimized using techniques such as model quantization and pruning for resource-constrained environments. Paul & Patel (2024); Babaei (2024) Improved Privacy and Security: In that the data is processed locally, edge models reduce the risk of data breaches associated with cloud storage. - (Jain et al., 2023). Applications and Use Cases Smart Surveillance: Real-time object detection systems enhance security measures in urban areas (Paul & Patel, 2024). Healthcare Monitoring: Edge AI enables timely health data analysis, improving patient outcomes (Jain et al., 2023). IoT Integration: Edge processing supports efficient resource management in IoT applications, optimizing power usage and task scheduling (Nandhakumar et al., 2023). As powerful as these edge processing models might sound, challenges are abound-for one, the need for rigorous security measures, coupled with edge-oriented algorithm development. Meeting such critical issues will form the cornerstone for wide-scale acceptance of Edge AI technologies.

4. Large Language Models (LLMs) with LLAMA

Large Language Models have really opened the door to many interesting applications in NLP, everything from chatbots to more complex data analysis. Among these, LLaMA stands out because of its open-source nature and advanced capabilities developed by Meta, formerly Facebook. Key Features of LLaMA Model Architecture LLaMA uses transformer architecture, meaning a model having layers of attention and feed-forward networks. The model can take in text as input to predict the next word or generate coherent responses. The latest version, LLaMA , is better pre-trained and post-trained, increasing performance in multiple tasks such as reasoning and code generation (Xie et al., 2024).

Parameter Efficiency The model is optimized for efficiency in its parameters. In the case of LLaMA, for example, parameters are stored using a 16-bit floating-point format, allowing compact representation. This allows for versions like LLaMA 3 to go up to 70 billion parameters, enabling its

deployment on consumer hardware while requiring substantial storage. Training Data and Processes LLaMA was trained on 10 terabytes of text, all scraped from various internet content (Wu et al., 2024).

It required thousands of GPUs, considering the huge computation that needed to be handled for this process. It is mended with language patterns by a next-word prediction task, the major training objective of the model. Not only does it improve the understanding of languages, but it also has aided in compressing information into its parameters. Enhancements in LLaMA 3 Where the LLaMA 3 differs significantly from its earlier models are the critical features listed below: Better Tokenization: New tokenizer featuring a 128K-token vocabulary and therefore giving better encoding efficiency. GQA, or Grouped Query Attention, enhances the inference efficiency of both the 8B and 70B parameter models (Li et al., 2025).

Quality Data Filtering: There were long filtering pipelines that guaranteed high-quality training data featuring heuristic filters and semantic deduplication methods. Applications and Implications Besides generating text, LLaMA can be fine-tuned for a wide variety of applications, making it quite versatile for use in many different NLP tasks. Its open-source nature invites and encourages research and development by the AI community, thus promoting innovation and collaboration. In conclusion, LLaMA represents an important advance in the area of large language models, combining efficiency with powerful performance across diverse applications in natural language processing Minaee et al., (2024).

5. Clustering and Predictive Analytics

Clustering and predictive analytics form the backbone of data mining and machine learning, especially in educational data. These methodologies allow pattern identification and forecasting of outcomes from historical data. Clustering Techniques K-Means Clustering K-means clustering is an unsupervised learning algorithm that groups data points into clusters based on their similarities. The algorithm works in a greedy manner: it iteratively assigns each data point to the closest cluster centroid, recomputes centroids, and refines cluster assignments in an iterative fashion until convergence (Srividhya et al., 2024).

This method is very effective in educational contexts to group learners based on metrics about performance and behavioral patterns. Applications in Education: K-means help educators identify distinct student profiles, allowing targeted interventions that can enhance learning outcomes. It allows institutions to adapt educational strategies to diverse needs by clustering students according to similar performance features. Predictive Analytics Predictive analytics involves the use of historical data to predict future outcomes. Techniques such as linear regression analyze past behaviors in order to predict future performance. In conjunction with clustering, predictive models can be trained separately for each identified cluster, thereby improving accuracy and relevance of forecasts (Valli, 2024).

Integration with Clustering, the integration of clustering with predictive analytics will help organizations bring out hidden insights from their data. Segmentation of learners into clusters can be followed by building predictive models for each cluster to predict learners' academic performance or to identify at-risk students³⁵. Segmentation Theory Segmentation theory plays a crucial role in understanding learner profiles. By creating distinct segments based on various characteristics—such as demographics, behaviors, and performance metrics, educators can implement targeted interventions (Bhaskaran, 2024, Goriparthi 2024).

This approach fosters inclusive and effective learning environments, ensuring that diverse learner needs are addressed appropriately. Conclusion The integration of K-means clustering techniques with predictive analytics provides powerful tools for informed decision-making in education. The application of these methodologies can help educational organizations increase their knowledge about learners' behaviors and improve academic performance for better interventions (Li et al., 2024).

5.1. Practical Work

There is an evolution of learning analytics that continues to drive beyond simple accumulation and signaling work being performed toward a greater focus on understanding student behavior,

engagement, and performance in real time (Pardo et al., 2019; Wise et al., 2014). Empowered through data, we have seen amazing innovation over the last ten years in the tools and frameworks available to support new ways of understanding better user experiences for learning. This section discusses the related literature on digital twins, real-time learning analytics, and fog computing and discusses how technologies such as the large language model (LLM), regression models and clustering techniques are being applied in educational settings to create the groundwork for our proposed framework.

5.2. Digital Twins in Education

Originating in the industrial sector, digital twins have been extrapolated to a multitude of other fields, such as healthcare, urban planning and, recently, education (Fuller et al., 2020). A digital twin is a simulation model of a physical entity that behaves and interacts with the real world in the same way. As described, the "digital twin" of the student acts as an up-to-date simulation reflecting their learning activities and interactions with educational content or devices (Eriksson et al., 2022).

There is less research on the use of digital twins in education, and many studies are either conceptual or concerned with pilot applications (Zhang et al., 2022). Indeed, as discussed, the possibility of digital twins providing a highly personalized learning paradigm in which customized content and assessment approaches are used for each student in real time is performative data. Digital twins were found to be fundamental in the provision of adaptive learning pathways because they allow real-time feedback and interventions aimed at addressing a learner's challenge when those challenges arise.

For instance, in 2021, scholars studied the possibility of using digital twins to monitor remote learners to reduce student academic dropout (Kinsner, 2021). They also warn that the real-time nature of responses will be essential, particularly in online learning, as physical presence is not suggestive of learner engagement or understanding.

5.3. Real-Time Learning Analytics

Learning analytics represents an emergent scientific field of research that explores potential analysis techniques for the data stored in educational databases and is based on a process of capacitance-based collection information for learners (Hernández-de-Menéndez et al., 2022). The rise in the availability of digital learning platforms that capture rich data on student interactions, behaviors and performance has made this possible. The aim of real-time learning analytics is to operationalize these data into practical insights that can help improve students' learning experience (L.-A. Lim et al., 2021), diagnose those who are at risk and enable decision-making in education.

However, many analytics systems currently in use are based on architectures that reside in the cloud and can add latency and interfere with delivering feedback quite promptly. Fog computing can be seen as a potential solution for these issues, which were presented in 2012 (Gerla et al., n.d.). Fog computing is an extension of the Cloud where data processing and analytics are performed closer to the source of the data, typically via the EDGE network. It has controls to reduce latency, minimize bandwidth usage and address confidentiality concerns, which are the absolute necessity for real-time learning analytics.

The findings showed that on-time and real-time feedback to students is possible in scenarios where instant interventions are mandatory for achieving certain learning outcomes through fog computing. This is one of the studies that showed how fog computing can lead to real-time learning analytics, enabling local data processing and no longer depending on cloud infrastructure to reduce other similar delays due to latency.

5.4. Fog Computing and Edge Analytics

Fog computing is an extension of cloud computing to network edges that places computations, storage, and networking close to data sources, where it provides advantages over traditional Cloud-based approaches (Ometov et al., 2022). They are located closer to the data and process it for applications that require real-time responses, such as learning analytics (Alshammari et al., 2020; Karatza, 2020).

Education-related: This is one of the use cases where fog computing has substantial benefits in comparison to conventional cloud-based systems. Fog computing reduces latency (time lag) and allows quicker delivery of real-time feedback, which translates into faster responsive learning experiences (Karatzas, 2020). It also decreases the bandwidth needed to move data into or out of the cloud, which is instructive in some areas where both network access and excellent quality are continuous limitations.

Moreover, fog computing increases data privacy and safety by allowing personal information to be stored at its place of origin, which reduces online transfer risks (Alwakeel, 2021). Learning literacy is especially critical in the context of learning environments, as educational institutions are obligated to protect student data. Fog computing for real-time learning analytics and improving system performance and scalability without compromising data privacy and security.

5.5. Meta-LLAMA and Large Language Models

The meta-LLAMA breaks new ground in large language models (LLMs) and provides remarkable power to generate personalized education feedback. We leverage pretrained models that are specifically trained for understanding and generating human language in context; these models are particularly well suited for analyzing student-generated content, e.g., essays or discussion posts (Radford et al., 2021). Through LLAMA, educational systems could generate feedback tailored more to each student's individual learning styles.

LLAMA Standard Setting Official Release, such as BERT and GPT-3, are built on top of the original models by using massive amounts of data as a seed question," writes the Twitter user Ben Levine in response to some examples given (Michelet & Breiteringer, 2024). Models such as these are trained by learning repeated patterns in different languages, which enables them to understand how language works and then use their understanding across many tasks. LLAMA can then assess a student's work by identifying where they require improvements and recommending resources or strategies that will enable these areas to be addressed in the context of personalized feedback, which, as suggested, leads to improved learning outcomes (Devlin et al., 2018; Touvron et al., 2023).

The integration of LLAMA into education platforms is an example of how AI-powered tools are designed to evolve feedback systems that have been ingrained in tradition. Unlike traditional feedback systems, tailored responses are formulated during production that are not only relevant to the content itself but also respectful of previous exchanges and progress. This enables educators to provide timely, specific, and actionable feedback, which will create a more engaging and supportive learning experience.

5.6. Regression Models and Clustering Techniques

Regression models and clustering techniques are the most common tools used in educational data mining to forecast student performance and determine clusters of students with comparable learning behavior. A core regression is used to understand the relationship between a dependent variable such as student performance and one or more independent variables (such as engagement metrics and attendance). Linear regression can also predict future outcomes by means of historical data — which is how the model helps to identify students who might need more individual attention.

Clustering techniques, such as K-means clustering, are applied to cluster students according to their learning behavior and engagement pattern as well as their performance metrics (Moubayed et al., 2020). K-means is a well-known unsupervised learning algorithm based on clustering; this method helps to separate the dataset into K groups of clusters where each cluster has similar students. By identifying these clusters, educators can develop better positions to develop interventions and support targeted toward each group's unique need for educational programs to have a greater impact.

7. Proposed Learner's Digital Twin-LDT Framework

In the following section, we describe a framework called "Learner's Digital Twin", which leverages Fog Computing and LLAMA using linear regression and K-means clustering to generate actionable information for educators as well as real-time personalized feedback for students. The

framework is, in fact, devised to circumvent the limitations of conventional cloud-based systems by handling data at on-network edges that could diminish latencies and support responsive times for time-sensitive responses.

7.1. System Architecture and Components

The architecture used in our learner's digital twin framework to process the data and provide feedback is based on combining fog computing with LLAMA. The system consists of several components that fulfill a particular purpose to make the framework capable and efficient. **Fig 1.** shows the proposed system architecture, the components and layers of interactions and the data flow.

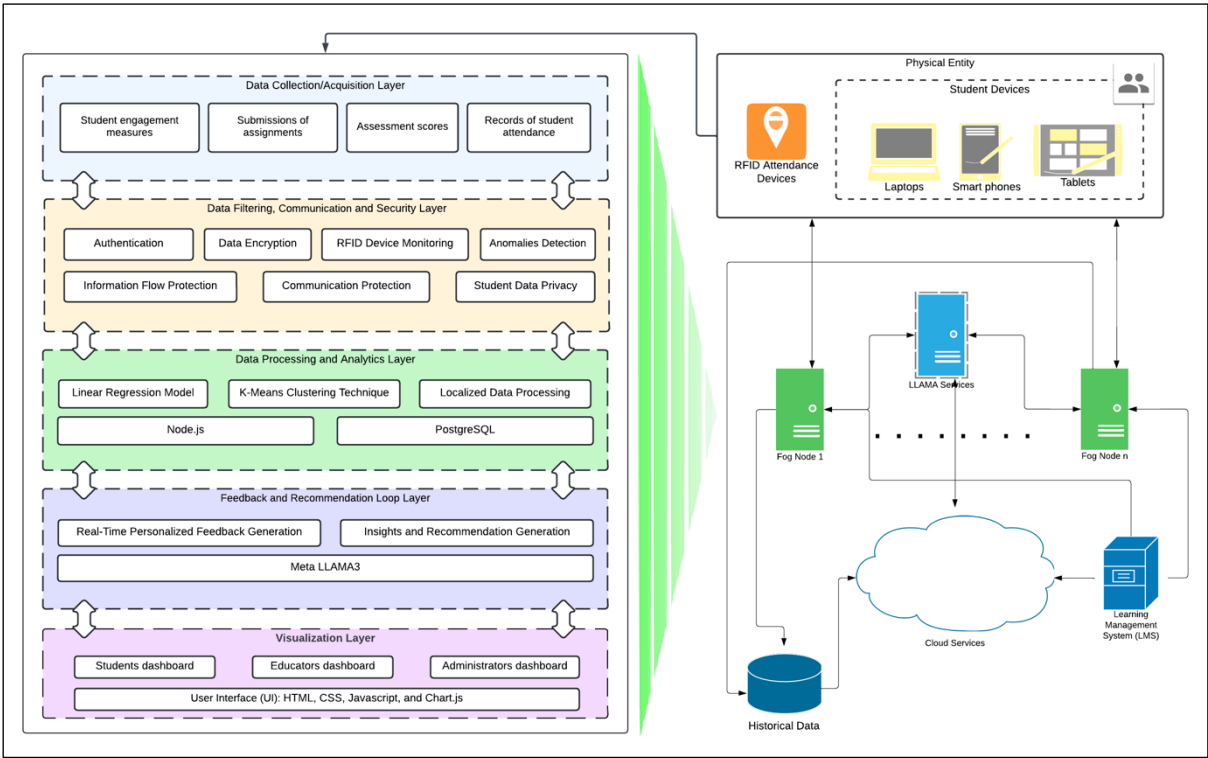


Figure 1. System Architecture and Components.

7.1.1. Edge Devices

Student Devices — Laptops, Tablets and Smartphones the students use to interact with LMS. The devices create data, including student engagement measures, submission of assignments and exam scores.

RFID Attendance Devices: Installed in classrooms that keep the record of student attendance. In turn, it feeds real-time data that allows students to understand when they are engaged, which is vital to their learning.

7.1.2. Fog Nodes

Data Aggregation: Fog nodes collect the data from edge devices and RFID attendance devices and aggregate it for local processing. In other words, this approach helps to reduce the amount of raw data that needs to be sent down to the cloud, hence reducing latency and bandwidth consumption.

Localized processing- Fog nodes perform initial data processing operations, such as computing engagement scores, identifying outliers and elevating real-time feedback to students. Because fog nodes conduct analyses on site, the response time is reduced to zero compared with the number of cloud processing steps. Fog nodes will work with Meta-LLAMA3 to analyze students’ data and provide personalized feedback.

Data Filtering: This method acts as a filter between edge devices to the cloud and filters the raw data into relevant high-quality data before they are transmitted to the Cloud. This approach also

helps take some load off the cloud servers and increases data privacy, as many features can work without sending sensitive information to the network.

7.1.3. Cloud Services

Additionally, cloud services perform heavy-duty tasks such as complex data analyses, predictive modeling, analyzing trends, and aggregating large volumes of data. Furthermore, they will store the processed data and analytics results in a central place accessible by educators, students, and other college staff for further analysis and reporting; finally, the cloud services will work with Meta-LLAMA3 to analyze big data and provide insights and recommendations.

7.1.4. LMS (Learn Management System)

Content Management: The LMS acts as the central hub from which educational content can be delivered, assignments and assessments can be planned and graded, etc. It is also the face where students access their learning publications.

Data Collection: The LMS captures and logs student interactions (e.g., quiz results, assignment submissions, time spent on tasks), and the fog nodes then process these data.

7.1.5. Dashboards

Student Dashboard–Represents personal feedback and recommendations and visualizes performance metrics for the student.

Teacher Dashboard–represents insights into the informed class, at-risk students' identification and teaching strategy suggestions and recommendations.

Admin Dashboard: Provides a performance dashboard showing trends, common issues across the system and proposed intelligent policies. The dashboard gives administrative staff all the information they need to successfully run and develop their institute.

7.2. Applying the Digital Twin Model

As such, a critical part of the learner's digital twin framework is its implementation: the digital twin model (DT), which captures learner interactions, behaviors, and progress in real time. As new data are generated, the digital twin constantly updates to deliver immediate feedback and adapt learning strategies based on what could be happening "now" in that system for a given learner.

7.2.1. Data Collection

The digital twin model is based on continuous data from various sources, such as student interactions with learning management systems (LMSs), online tests, assignment submissions and other digital tools. Sources from which data are collected:

- **Behavioral Data:** Information about the level of engagement by a student, including time spent on the task and interactions with the content.
- **Performance Data:** Assessments of scores, exam grades and other performance metrics along with feedback from course instructors.
- **Context Data:** included the learning environment, such as the type of device used while conducting the session and the time.

These data are fed back into the digital twin, creating a full picture of the student's journey through learning. These data are used by the system to recognize patterns, predict future performance and recommend individualized interventions.

7.2.2. Real-Time Updates

With every new data generated, the same data generated in real time are updated and stored in a digital twin. In this way, the system provides immediate feedback, which adapts automatically according to the learner's current needs. If a student has trouble with a specific concept, for instance,

the digital twin can identify this challenge and recommend resources or interventions that will assist him/her in tackling that obstacle.

7.3. Predicting Student Outcomes using Linear Regression

In the context of our framework, linear regression is used to predict a desired result (e.g., final exam grade) based on an independent dataset (features). It has features such as attendance, scores in assignment and participation rates that are directly proportionate to a successful academic result.

The objective is to determine the score on the final exam given the student's attendance, assignment grade and participation.

Assumptions:

Independent Variables (Features):

- x_1 : Attendance rate (as a percentage);
- x_2 : Average assignment score (as a percentage);
- x_3 : Participation rate (as a percentage);

Dependent Variable (Target):

- y : Final Exam Score (as a percentage).

Initial Coefficients:

- ($\beta_0 = 10$) (initial intercept, assuming all features have minimal effect).
- ($\beta_1 = 0.5$) (initial coefficient for the attendance rate).
- ($\beta_2 = 0.3$) (the initial coefficient for assignment score).
- ($\beta_3 = 0.2$) (initial coefficient for the participation rate).

Linear Regression Equation

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \quad (1)$$

where

\hat{y} is the predicted final exam score.

x_1, x_2, x_3 are the independent variables (features).

$\beta_0, \beta_1, \beta_2, \beta_3$ are the coefficients.

The Linear Regression Algorithm used in LDT

Input:

- A dataset X with n students, where each student has features x_1, x_2, x_3 .
- The corresponding target variable y (final exam score) for each student.

Output:

- Final exam score prediction via linear regression.

Steps:

1- Initialize the parameters: Set initial values for the coefficients $\beta_0, \beta_1, \beta_2, \beta_3$.

2- Compute Predictions: For each student i in the dataset, calculate the predicted final examination score \hat{y}_i using the current coefficients:

$$\hat{y}_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} \quad (2)$$

3- Compute the cost function (mean squared error):

Do for each student i Calculate the error:

$$e_i = y_i - \hat{y}_i \quad (3)$$

Calculate the mean squared error (MSE):

$$MSE = \frac{1}{n} \sum_{i=1}^n e_i^2 \quad (4)$$

4- Optimization of Parameters (Gradient Descent): The coefficients are adjusted to minimize the MSE:

$$\beta_j := \beta_j - \alpha \frac{\partial}{\partial \beta_j} MSE \quad (5)$$

where α is the learning rate.

5- Iterate:

Replay steps 2 to 4 until the changes in the MSE at convergence become extremely compact or the maximum ranges of iterations are reached.

Final Model:

Once the algorithm converges, the coefficients $\beta_0, \beta_1, \beta_2, \beta_3$ represent the final model.

Final equation:

$$\hat{y} = 10 + 0.3x_1 + 0.4x_2 + 0.2x_3 \quad (6)$$

Example of Prediction:

For a student with 85% attendance, 78% average assignment score, and 90% participation, the predicted final exam score would be

$$\hat{y} = 10 + 0.3(85) + 0.4(78) + 0.2(90) = 10 + 25.5 + 31.2 + 18 = \mathbf{84.7\%}$$

7.4. Clustering Students into Groups with K-means Clustering

The system clusters students using K-means clustering for personalized feedback and recommendations based on their (attendance rate, assignment scores, test scores, participation rate, submission timeliness, interaction with learning materials). The student data are automatically partitioned into K clusters, which can act as clusters of similar kinds of students. $K=3$ to $K=5$ —the 1st bucket—we can put students in high-level groupings but not get too granular, which is a good place to start. We can then return, tuning that K to represent more of the subtleties in student behavior and performance for better targeted insights/interventions. For example:

$K = 3$ might produce clusters of high-achieving students, moderate-achieving students, and low-performing students

$K = 5$ will result in more detailed groupings, such as high performers, consistent performers, improving students, struggling students, and at-risk students.

K-Means Clustering Algorithm (Starting with $K = 3$)

Input:

- $K = 3$: The initial number of clusters.
- $X = x_1, x_2, \dots, x_n$: The dataset of n students, where each x_i is a vector of features representing student behavior and performance metrics (e.g., attendance rate, assignment scores, test scores, participation rate, etc.).

Output:

- A set of 3 clusters, each containing a group of students, was constructed.
- The centroid of each cluster.

Steps:

Initialize Centroids: Three students are randomly selected from X as the initial centroids μ_1, μ_2, μ_3 .

Assign Students to Clusters: For each student x_i in the dataset:

- Compute the Euclidean distance between x_i and each centroid μ_j for $j = 1$ to 3 .

$$d(x_i, \mu_j) = \sqrt{\sum_{k=1}^m (x_{ik} - \mu_{jk})^2} \quad (7)$$

$d(x_i, \mu_j)$ represents the Euclidean distance between the student vector x_i and the centroid μ_j .

m is the number of features (attributes) in the student vector.

x_{ik} is the k -th feature of the student vector x_i .

μ_{jk} is the k -th feature of the centroid μ_j .

- Assign x_i to the cluster C_j with the closest centroid μ_j .

Update centroids:

- For each cluster C_j , recalculate the centroid μ_j as the mean of all student vectors assigned to that cluster:

$$\mu_j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i \tag{8}$$

- Here, $|C_j|$ is the number of students in cluster C_j .

Check for convergence:

- If the centroids μ_1, μ_2, μ_3 do not change (or change truly little) after updating, the algorithm has converged, and you can stop.
- Otherwise, go back to Step 2, and repeat the process.

Interpret and refine K :

- The resulting clusters were analyzed to determine whether they provided meaningful groupings of students.
- If the clusters are too broad or too narrow, adjust K (e.g., increase K to 4 or 5) and rerun the algorithm.
- Methods such as the elbow method or silhouette analysis were used to determine the optimal number of clusters.

Output the final clusters:

- Once satisfied with the clustering, the final clusters C_1, C_2, C_3 and their corresponding centroids are output.
- Each cluster should represent a distinct group of students with similar behaviors and performance metrics.

7.5. Generating Personalized Feedback and Recommendation using Meta-LLAMA-3

We integrate LLAMA into the Learner's Digital Twin framework so that it can analyze student-processed data. LLAMA will return different outputs for each learner because it can understand and process text personally using advanced natural language processing (NLP).

We take advantage of advanced features provided by Meta-LLAMA3, which provides personalized feedback and recommendations matched to the specific performance measures of each student. It starts with capturing and making sense of a holistic dataset for each student — attendance, assignment scores, test results, levels of participation in class & historical data. These features are formatted into a structured input that can then be passed through Meta-LLAMA3.

After preparing the data, these actions are processed through different deep learning layers to generate an internal representation that captures complex interactions between a set of indicators and other performance metrics. The meta-LLAMA3 applies its pretrained knowledge and inference abilities to the data, generating insights that correspond with student performance and highlighting both strengths and weak areas where improvement is needed.

The feedback and recommendations-stage two Meta-LLAMA3 were generated, after which the inferred insights were mapped to personalized feedback for the students, after which the students were outfitted with coordinate bearings on how they could improve or maintain their performance.

At the same time, recommendations for teachers are generated by the model to suggest actionable strategies for how they can continue supporting a student throughout their learning. These outputs are saved into a PostgreSQL database to be available for real-time visualization on student and teacher dashboards. If applied to every student in the institution, it could certainly make both learning outcomes more personalized and teachers empowered with insights that allow for a better outcome.

Algorithm for Data Collection, Processing and Generation of Personalized Feedback and Recommendations using Meta-LLAMA3

This algorithm describes in a logical step-by-step how the system gathers and processes student data, creates a contextual prompt, and produces custom feedback involving the LLAMA 3 model.

This approach personalizes the feedback for each individual student through statistical analysis and utilizing LLAMA 3's advanced natural language processing. By also processing statistical data and context directly into prompt generation, the feedback serves as highly relevant to every student.

Input Data Collection:

Input: Collect raw student data D where $D_s = \{A, S, G\}$

- A : Attendance records.
- S : Assignment submissions and their respective scores.
- G : Assessment scores (e.g., quizzes, exams).

Data Preprocessing:

Normalize Attendance Data- Calculate the attendance rate RA for each student s_i :

$$RA(s_i) = \frac{\text{Total Attended Classes}}{\text{Total Scheduled Classes}} \times 100 \quad (9)$$

Aggregate assignment scores: Compute the average assignment score $\bar{S}(s_i)$ for each student:

$$\bar{S}(s_i) = \frac{\sum_{j=1}^{n_s} S_{ij}}{n_s} \quad (10)$$

where S_{ij} is the score of the j -th assignment and n_s is the total number of assignments.

Normalize Assessment Scores - Calculate the overall assessment performance $G_{norm}(s_i)$ by normalizing the assessment scores against the maximum possible score:

$$G_{norm}(s_i) = \frac{G(s_i)}{G_{max}} \times 100 \quad (11)$$

where $G(s_i)$ is the score of the student s_i and G_{max} is the maximum possible score.

Statistical Data Processing:

Performance analysis - Statistical summaries, such as the mean, median, and standard deviation, were generated for RA , \bar{S} , and G_{norm} across the entire student cohort.

Identify Outliers - Identify students whose performance metrics significantly deviate from the cohort's average using Z scores:

$$Z(x) = \frac{x - \mu}{\sigma} \quad (12)$$

where x is the student's score, μ is the mean, and σ is the standard deviation.

Prompt Generation for LLAMA 3:

Data Summarization-The processed data P_s for each student s_i are summarized into a structured text prompt:

$$P_s = \text{StudentID: } s_i, \text{ Attendance: } RA(s_i), \text{ Assignment Avg: } \bar{S}(s_i), \text{ Assessment Performance: } G_{norm}(s_i) \quad (13)$$

Contextual Information - Add additional contextual information such as learning objectives, areas of difficulty, and past performance trends:

$$\text{CProm} = P_s + \text{LearningObjective: } L_i + \text{Difficulty Areas: } D_i \quad (14)$$

Construct Final Prompt - Combine the data summary with personalized instructions for LLAMA 3 to generate feedback:

$$\text{FProm} = \text{"Analyze the following student data and provide personalized feedback and recommendations:"} + \text{CProm} \quad (15)$$

Feedback Generation using LLAMA 3:

Input to LLAMA 3 - Pass the final prompt **FProm** to the LLAMA 3 model.

Text Processing - LLAMA 3 processes the prompt by leveraging its NLP capabilities to understand the context and generate feedback. This involves:

$$\text{Feedback}_s = \text{LLAMA3}(\text{FProm}) \quad (16)$$

The model outputs a structured response, potentially including praise, constructive criticism, and actionable recommendations.

Personalization- Ensure that the generated feedback is personalized based on the student's data, addressing specific strengths and weaknesses and providing guidance for improvement.

Postprocessing:

Grammar and Clarity Check - Refine the feedback $Feedback_s$ to ensure grammatical correctness, clarity, and coherence.

Final Output - Deliver the final personalized feedback $Feedback_s$ to the student through the dashboard.

Feedback loop:

Student interaction - Students were allowed to review and interact with the *feedback* and collect responses.

Model improvement - Use student responses to update and fine-tune LLAMA 3, enhancing its future performance and feedback relevance.

7.6. Process Flow in Learner’s Digital Twin

Fig 2. shows the process flow of our proposed framework. The process starts with data aggregation and preprocessing in the fog nodes, where necessary information is aggregated from the student devices and attendance RFID devices; then, the system processes the data by filtering it to remove anomalies. Instant feedback is generated and shared with students as needed using LLAMA.

The students’ processed data are subsequently saved in a PostgreSQL database and published to cloud services for further analysis. Then, on the cloud, we perform advanced analyses, such as regression models, to predict the outcomes of students and clustering based on behaviors, performances, etc., to obtain insights for each cluster. Using Meta-LLAMA3, recommendations are generated through students’ processed data stored in the database.

Additionally, it provides endpoints for obtaining detailed student performance and class analytics, which can be visualized on respective dashboards.

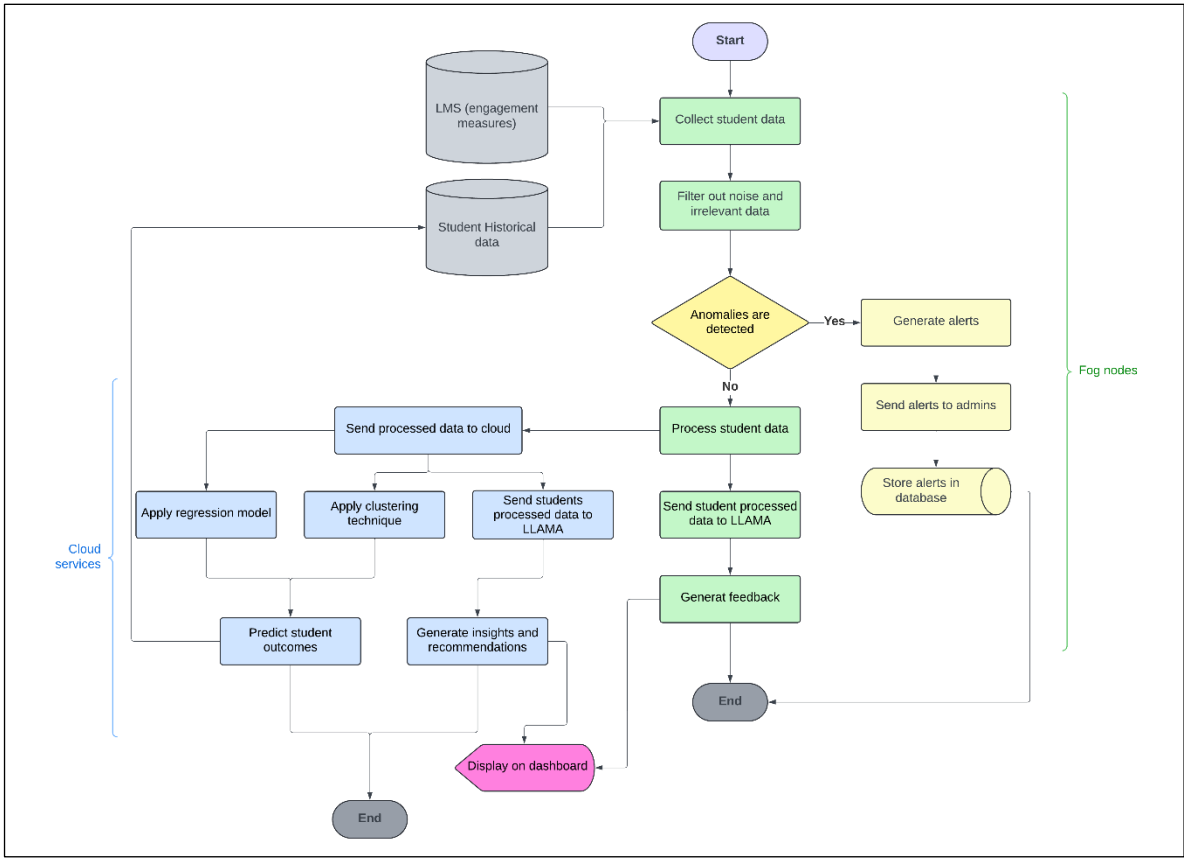


Figure 2. Learner's Digital Twin Flowchart.

8. Results

A real-world educational setup was used to evaluate the proposed Learner's Digital Twin framework for real-time personalized feedback to students and teachers’ insights and recommendations. In total, 25 students in two courses were part of this evaluation, and data were

collected for 3 months with the previous methods. The study examined metrics such as latency, feedback accuracy, student performance and user satisfaction.

8.1. Latency Reduction

First and foremost, the goal was to decrease latency in data processing as well as feedback delivery. For this reason, Fog Computing was integrated into architecture. To evaluate the performance benefits of our Learner's Digital Twin Framework using Fog Computing compared to traditional cloud-based systems, we conducted latency tests using Postman. Latency, the time taken for data to travel from the source to the destination and back, is critical in real-time learning analytics.

Lower latency ensures timely feedback, which is essential for effective learning interventions. By using Postman, a popular API development tool, we were able to simulate client requests and measure the response times for both systems under identical conditions. As shown in Fig 3., the initial test results revealed that the LDT system exhibited a significantly lower latency compared to the Cloud-Based system. On average, the LDT system was 300 MS faster in responding to requests.

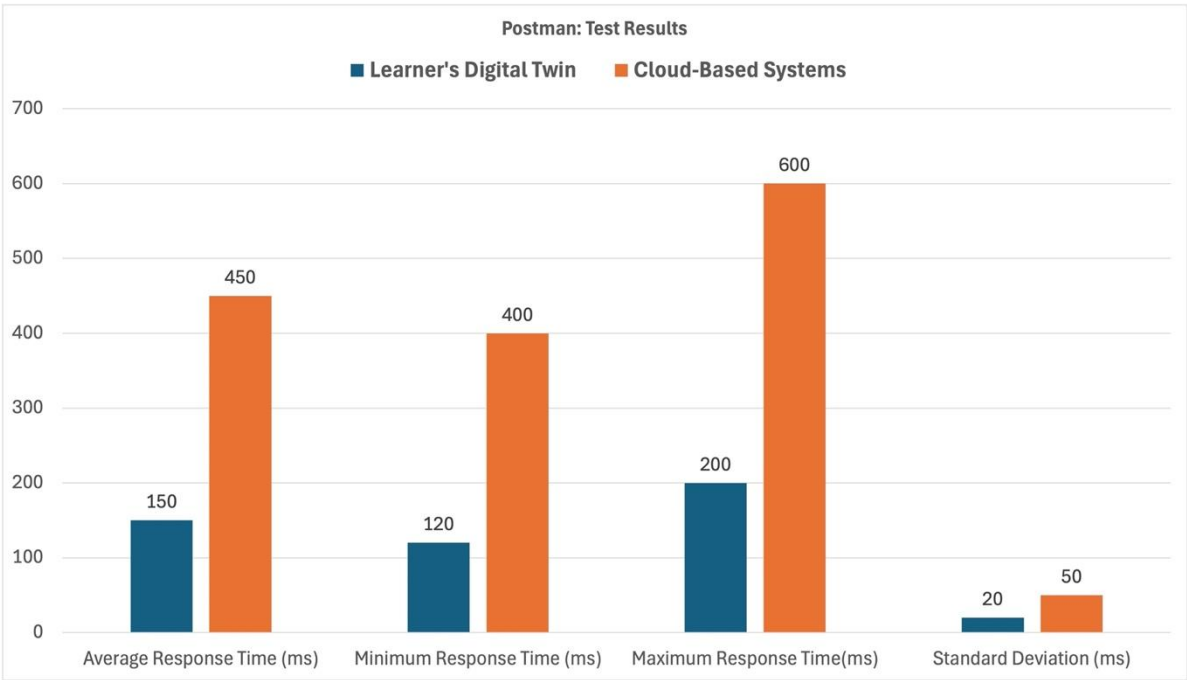


Figure 3. Postman test results.

8.2. Feedback Accuracy and Personalization

Because LLAMA can analyze student processed data sent to it in textual prompt format in real time, the feedback was highly personalized to students and therefore more likely to provide them with information that matched their current learning point.

Table 1 shows samples of LLAMA-generated feedback and recommendations for students and teachers. The sample taken from the performance data (attendance rate, assignment average, and assessment performance and difficulty areas) was measured for 4 students for two courses.

Table 1. Samples of personalized feedback and recommendations.

Student ID	Course	Final Prompt	Personalized Feedback	Student Recommendations	Teacher Recommendations

101	Data Science 101	Analyze the following student data and provide personalized feedback and recommendation s: Student ID: 101, Attendance: 95%, Assignment Average: 85%, Assessment Performance: 78%. Learning Objective: Mastering data visualization techniques. Difficulty Areas: Struggles with advanced Python programming concepts.	You have a strong attendance record and are performing well in assignments. However, your assessment scores suggest some challenges in advanced Python programming . You have a good grasp of data visualization but need to strengthen your Python skills to excel further.	Focus on practicing Python coding exercises, especially those related to data manipulation. Consider attending additional tutoring sessions on advanced Python topics. Regularly review Python documentation and participate in coding forums.	Consider providing more practice problems focused on Python programming, particularly on data manipulation and advanced concepts. Offering supplementary materials or tutorials on these topics might also help students who are struggling.
102	Machine Learning 201	Analyze the following student data and provide personalized feedback and recommendation s: Student ID: 102, Attendance: 88%, Assignment Average: 92%, Assessment Performance: 90%. Learning Objective: Understanding supervised learning	Your performance is excellent, with high scores in both assignments and assessments. Your understanding of supervised learning algorithms is strong, but there's room for	To deepen your understanding, focus on the mathematical foundations of algorithms, such as linear algebra and calculus. Engage in exercises that require you to derive and prove algorithmic concepts. Consider reviewing relevant math	Offer additional resources or sessions that focus on the mathematical underpinnings of machine learning algorithms. Consider organizing study groups or recommending specific texts that help reinforce these concepts.

		algorithms. Difficulty Areas: Occasionally struggles with mathematical foundations behind algorithms.	improvement in grasping the mathematical concepts.	tutorials and textbooks.	
103	Data Science 101	Analyze the following student data and provide personalized feedback and recommendation s: Student ID: 103, Attendance: 75%, Assignment Average: 68%, Assessment Performance: 65%. Learning Objective: Gaining proficiency in data cleaning and preprocessing. Difficulty Areas: Inconsistent with assignment submissions and struggles with time management.	Your attendance and assignment completion rates are lower than expected, which may be impacting your overall performance. Consistent effort in these areas is crucial for mastering data cleaning and preprocessing skills.	Improve your time management by setting aside dedicated study hours each week. Prioritize completing assignments on time to reinforce your learning. Consider using a task management tool to organize your study schedule. Seek help from peers or instructors if needed.	Monitor this student's progress more closely and consider offering regular check-ins to discuss time management strategies. Providing flexible deadlines or alternative assignments might help in addressing their consistency issues.
104	Machine Learning 201	Analyze the following student data and provide personalized feedback and recommendation s: Student ID: 104, Attendance: 80%, Assignment Average: 70%,	You have a good understanding of machine learning models, but there seems to be some difficulty in evaluating	Focus on learning and applying different model evaluation metrics such as accuracy, precision, recall, and F1-score. Participate in study groups or online courses	Consider revisiting the topic of model evaluation metrics in class, possibly with more examples or hands-on activities. Providing

Assessment Performance: 72%. Learning Objective: Applying machine learning models to real-world datasets. Difficulty Areas: Understanding model evaluation metrics.	their performance effectively. Your attendance is satisfactory, but there's room for improvement in both assignment scores and assessment performance.	that focus on these metrics. Practice with real-world datasets to apply these concepts effectively.	additional exercises or workshops that focus on these metrics could benefit students struggling in this area.
---	--	---	---

8.3. User Satisfaction

We measured user satisfaction with the system via surveys for both students and educators. Both groups reported high levels of satisfaction with the intervention in terms of ease of use and usefulness of acquiring benefits related to immediate feedback from the system.

Most users expressed the desire to continue with this system in other learning activities due to the quality of personal feedback and the option for just-in-time support from teachers. The analysis of students’ and teachers’ responses to the survey is explained below.

Student Survey Analysis (SSA)

Fig 1. SSA-Relevance of LDT-Feedback

More than 75% of students found the feedback was relevant and 24% found it relevant some how.

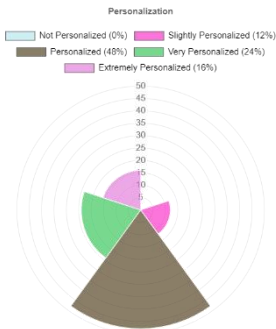


Fig 2. SSA-Personalization of LDT-Feedback

More than 90% of students found the feedback was personalized.

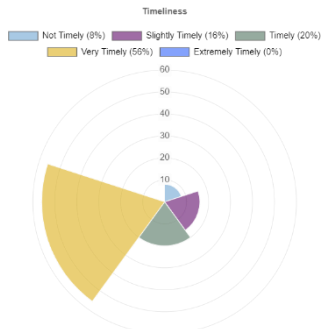


Fig 3. SSA-Timeliness of LDT-Feedback

More than 75% of students found the feedback was timely and approximately 15% found it slightly timely.

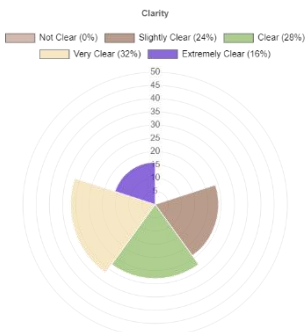


Fig 4. Clarity of LDT-Feedback

More than 85% of students found the feedback clear and approximately 24% found it slightly clear.

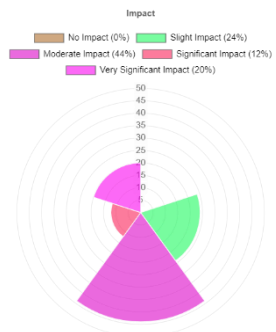


Fig 5. Impact of LDT-Feedback

More than 75% of students noticed feedback impact on their learning experience and 24% found slight impact.

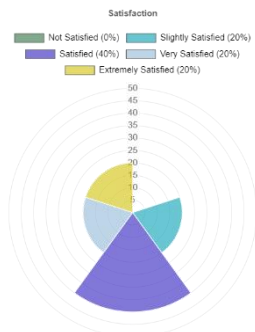


Fig 6. SSA-Students Satisfaction

80% of students are satisfied with the new system and the personalized real-time feedback and 20% are slightly satisfied.

Teacher Survey Analysis (TSA)

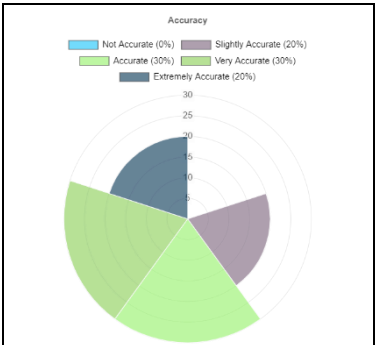


Fig 7. TSA-Accuracy of LDT-Feedback

80% of our teachers found the feedback was accurate and 20% found it slightly accurate.

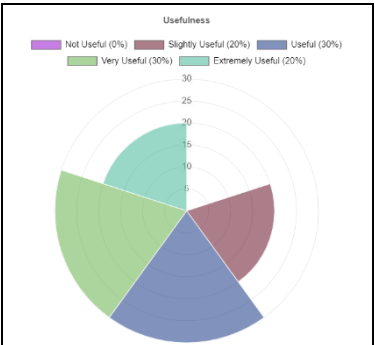


Fig 8. TSA-Usefulness of LDT-Feedback

80% of teachers found the feedback was useful and 20% of them found it slightly useful.

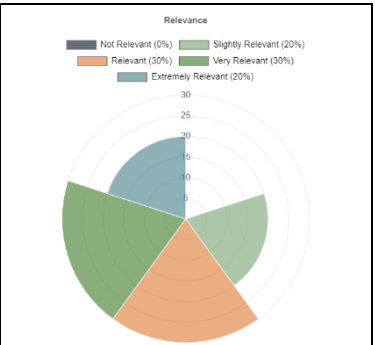


Fig 9. TSA-Relevance of LDT-Feedback

80% of teachers found the context was relevant and 20% found it slightly relevant but 10% found it not relevant.

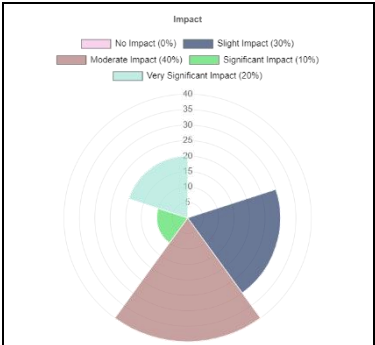


Fig 10. TSA-Impact of LDT-Feedback

70% of our teachers found impact on their teaching and 30% of them noticed slight impact.

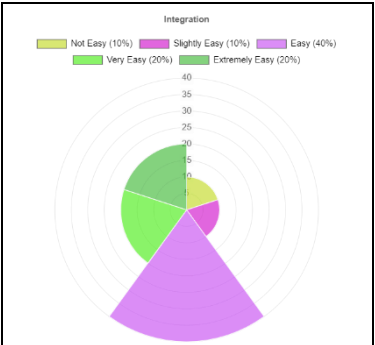


Fig 11. TSA-Integration into Curriculum

80% of teachers found it easy to integrate it in the curriculum and 10% found it slightly easy but 10% said it is not easy.

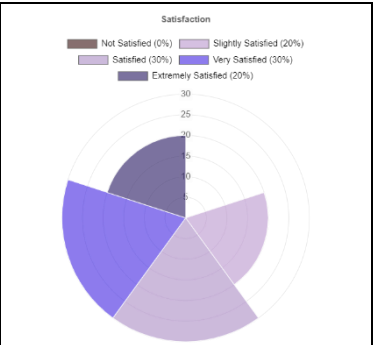


Fig 12. TSA-Teachers Satisfaction

80% of teachers are satisfied with new system and 20% of them are slightly satisfied.

8.4. Summary of Findings

Personalization/Contextual Relevancy: The LLAMA-based system was highly successful at personalizing and presenting contextually relevant feedback compared with traditional systems; there were very positive reports from both students and teachers regarding the satisfaction of almost all aspects.

Real-Time Analysis: The ability of LLAMA to analyze student data in real time is directly related to how the feedback can be useful; thus, this approach improves the overall learning experience.

Satisfaction among students and teachers: The survey showed that both students and educators found LLAMA system feedback to be more beneficial for learning than standard nonadaptive feedback was, as LLAMA feedback is very modernized, which in turn increases the motivation responsible for improved performance.

9. Discussion

The experimentation phase of this research revealed a significant opportunity to improve real-time learning models using Learner's digital twin model. The study was complemented by Fog Computing and LLAMA (LLM). Using linear regression and k-means clustering together enables them to analyze large sets of learning data in a talent-based setting. A learner's digital twin is a living embodiment of ongoing learning, an up-to-date model that incorporates history and real-time data (and possibly even future plans) into what the learner knows today.

A key advantage of using fog computing in this model is the decreased latency. The processing and return times in cloud-based systems are longer than those in traditional systems, which can slow learning interventions. In contrast, processing data from a nearby source via fog computing enables quick feedback (near real-time). Fog computing reduces latency so that educators can offer quick assistance when needed, providing a more responsive and efficient learning environment.

Incorporating LLAMA provides accurate and highly personalized feedback. The unique circumstances of each student and their responses to their needs, strengths, and weaknesses should be considered. These breakthroughs are crucial for addressing the wide range of learning styles and places present in education environments, making it easier for each student to receive guidance on how they learn best.

9.1. Implications for Educational Practice

The prospective success of the framework has important implications for educational practice. It provides students with immediate and personalized feedback, allowing them to find learning gaps more efficiently. This approach also allows for differentiated instruction, where educators can adapt teaching strategies to meet the needs of learners. In addition, this system has the capacity to provide timely insights for making better decisions by educators and improving learning experience through real-time data.

Education: Creating digital twins in education is a game changer for utilizing learning analytics to better serve students. Each virtual model adapts as individuals learn more about the individual learner in real time to create a big data-driven understanding of what learners are doing and where educators can most effectively intervene. This method leads not only to enhancing overall learning experience but also to ensuring a more engaging and personalized way of learning.

9.2. Limitations and Future Research

Despite these promising results, several limitations of this study must be considered. The system was implemented in a controlled setting with few subjects, and the results may not reflect the decreased performance of this target group for generalization. Since the design and participants can sometimes affect the outcomes of a study (e.g., whether at home or institution or older than younger people), we should take note of this. As such, additional research is needed in diverse educational environments to examine whether the system scales with rural and under resourced institutions where technology access may be limited.

Closed laboratories and other traditional teaching spaces are limited, although perhaps appropriately, given the need to limit physical interaction. The reliance on technology of some kind automatically excludes anyone without access (remembering students who may be in remote locations) or who is simply less confident in using digital tools. More efforts must be made to address how digital learning resources are distributed fairly and other possible alternatives for personalized feedback in low-tech environments.

In addition, although this particular research was concerned with the effects of the system on learner outcomes and user experience after using digital twin-based learning analytics, in an attempt to determine whether it is suitable for immediate application (e.g., can we use them now?), as future work aims at how a special type of educational technology such as digital twins that employs visual feedback facilitated by multidimensional data visualization opportunities might affect long-term student engagement, motivation and retention beyond standard statistics. The use of this technology over time—especially over multiple years—will likely yield important longitudinal studies about the effect of this technology on educational outcomes.

A multiagent system can significantly enhance the performance of a learner's digital twin. The digital twin can be even more personalized and effective when multiple independent autonomous agents are integrated, which address different facets in the development of the learner. For example, one agent might consider the learner's cognitive capacities, another emotional state and a third their social interactions. These agents can work together to determine individual learning needs, tailor educational content and provide real-time feedback, which should eventually lead to better learning outcomes.

10. Conclusions

In this research, we demonstrate a framework for implementing real-time learning analytics using fog computing, LLAMA with linear regression and K-means clustering to provide immediate feedback, recommendations and predictions. The framework reduces latency, improves the personalization of feedback, and supports personalized and adaptive learning journeys. The results support the conclusion that combining digital twin technology with recent advances in computational models presents a more alarming, nimbler response to educational challenges than before—one which can harness data at scale. Digital twins have the ability to rectify wide educational practices. This allows educators to create 'virtual models' of the learner that demonstrate their interactions and progress in real time, providing insight never before possible into learning paths that might be particularly beneficial for one student over another. This potential is further enhanced by the integration of Fog Computing and LLAMA, ensuring timely personal feedback, which is a crucial element of student success.

There is ample room for further work to adapt and refine the tool and investigate its scalability as well as issues related to widening access. Digital twins, fog computing and large language models promise to offer much potential in regard to creating more efficient educational technology applications as the industry continues to evolve.

References

1. Abayadeera, M. R., & Ganegoda, G. U. (2024). Digital twin technology: A comprehensive review. *International Journal of Scientific Research & Engineering Trends*, 10(4), 1485-1506. <https://doi.org/10.61137/ijset.vol.10.issue4.199>.
2. Abdelaziz, A. A., Abd EL-Ghany, S., & Tolba, A. S. (2024). Digital Twin-Based Learning Analytics with Fog Computing and LLAMA. <https://doi.org/10.21203/rs.3.rs-5051420/v1>.
3. Abuzant, M., Daher, W., & Mahamid, F. (2023). Control-Value Appraisals and Achievement Emotions: A Moderation Analysis. *Psych*, 5(4), 1207-1223. <https://doi.org/10.3390/psych5040080>.
4. Aggarwal, S. (2020). Distributed computing for efficient data processing and storage. *Turkish Online Journal of Qualitative Inquiry (TOJQI)*, 11(4), 2201-2210. <https://doi.org/10.52783/tojq.v11i4.10015>.
5. Alam, A. (2023, May). Improving Learning Outcomes through Predictive Analytics: Enhancing Teaching and Learning with Educational Data Mining. In *2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 249-257). IEEE. <https://doi.org/10.1109/ICICCS56967.2023.10142392>.
6. Alshammari, H., El-Ghany, S. A., & Shehab, A. (2020). Big IoT Healthcare Data Analytics Framework Based on Fog and Cloud Computing. *Journal of Information Processing Systems*, 16(6), 1238-1249. <https://doi.org/10.3745/JIPS.04.0193>.
7. Alwakeel, A. M. (2021). An overview of fog computing and edge computing security and privacy issues. *Sensors*, 21(24). <https://doi.org/10.3390/s21248226>.
8. Armstrong, K. E. (2023). A Test of the Control Value Theory of Achievement Emotions in an Instructional Communication Context. West Virginia University. <https://doi.org/10.33915/etd.12144>.
9. Armstrong, K. E. (2023). A Test of the Control Value Theory of Achievement Emotions in an Instructional Communication Context. West Virginia University. <https://doi.org/10.33915/etd.12144>.
10. Azarnova, T. V., & Polukhin, P. V. (2021, May). Distributed computing systems synchronization modeling for solving machine learning tasks. In *Journal of Physics: Conference Series* (Vol. 1902, No. 1, p. 012050). IOP Publishing. <https://doi.org/10.1088/1742-6596/1902/1/012050>.

11. Babaei, P. (2024, March). Convergence of Deep Learning and Edge Computing using Model Optimization. In *2024 13th Iranian/3rd International Machine Vision and Image Processing Conference (MVIP)* (pp. 1-6). IEEE. <https://doi.org/10.1109/MVIP62238.2024.10491145>.
12. Bhaskaran, E. (2024). Business Analytics in Steel Product Fabrication Cluster. *SEDME (Small Enterprises Development, Management & Extension Journal)*, 09708464241233023. <https://doi.org/10.1177/09708464241233023>.
13. Čalopa, M. K., Kokotec, I. Đ., & Kokot, K. (2023). THE ETHICAL AND PEDAGOGICAL ISSUES OF LEARNING ANALYTICS IN HIGHER EDUCATION INSTITUTIONS. In *EDULEARN23 Proceedings* (pp. 6150-6159). IATED. <https://doi.org/10.21125/edulearn.2023.1600>.
14. Chandrasekaran, S., Athinarayanan, S., Masthan, M., Kakkar, A., Bhatnagar, P., & Samad, A. (2024). Edge Intelligence Paradigm Shift on Optimizing the Edge Intelligence Using Artificial Intelligence State-of-the-Art Models. In *Advancing Intelligent Networks Through Distributed Optimization* (pp. 1-18). IGI Global. <https://doi.org/10.4018/979-8-3693-3739-4.ch001>.
15. Dai, J., Annasiwatta, C., Bernussi, A., Fan, Z., Berg, J. M., & Ren, B. (2019, July). Continuous phase tuning of vanadium dioxide films using robust feedback mechanism. In *2019 American Control Conference (ACC)* (pp. 5743-5748). IEEE. <https://doi.org/10.23919/ACC.2019.8815224>.
16. Damaševičius, R., & Zailskaitė-Jakštė, L. (2024). Digital Twin Technology: Necessity of the Future in Education and Beyond. *Automated Secure Computing for Next-Generation Systems*, 1-22. <https://doi.org/10.1002/9781394213948.ch1>.
17. Datta, S., & Datta, P. (2024). Fog Computing in Healthcare: Application Taxonomy, Challenges and Opportunities. *Fog Computing for Intelligent Cloud IoT Systems*, 173-201. <https://doi.org/10.1002/9781394175345.ch8>.
18. David, I., Shao, G., Gomes, C., Tilbury, D., & Zarkout, B. (2024, October). Interoperability of Digital Twins: Challenges, Success Factors, and Future Research Directions. In *International Symposium on Leveraging Applications of Formal Methods* (pp. 27-46). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-75390-9_3.
19. Delerm, F., & Pilottin, A. (2024). Double edged tech: navigating the public health and legal challenges of digital twin technology. *European Journal of Public Health*, 34(Supplement_3), ckae144-1510. <https://doi.org/10.1093/eurpub/ckae144.1510>.
20. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. <http://arxiv.org/abs/1810.04805>.
21. Drugova, E., Zhuravleva, I., Zakharova, U., & Latipov, A. (2024). Learning analytics driven improvements in learning design in higher education: A systematic literature review. *Journal of Computer Assisted Learning*, 40(2), 510-524. <https://doi.org/10.1111/jcal.12894>.
22. Edmonds, B., & Hales, D. (2005). Computational simulation as theoretical experiment. *Journal of Mathematical Sociology*, 29(3), 209-232. <https://doi.org/10.1080/00222500590921283>.
23. Eriksson, K., Alsaleh, A., Behzad Far, S., & Stjern, D. (2022). Applying Digital Twin Technology in Higher Education: An Automation Line Case Study. *Advances in Transdisciplinary Engineering*, 21, 461-472. <https://doi.org/10.3233/ATDE220165>.
24. Fuller, A., Fan, Z., Day, C., & Barlow, C. (2020). Digital Twin: Enabling Technologies, Challenges and Open Research. *IEEE Access*, 8, 108952-108971. <https://doi.org/10.1109/ACCESS.2020.2998358>.
25. Gerla, M., Huang, D., Association for Computing Machinery. Special Interest Group on Data Communications, ACM Digital Library., & ACM SIGCOMM Conference (2012: Helsinki, F. (n.d.). *MCC'12: proceedings of the 1st ACM Mobile Cloud Computing Workshop: August 17, 2012, Helsinki, Finland*.
26. Giannakos, M., & Cukurova, M. (2023). The role of learning theory in multimodal learning analytics. *British Journal of Educational Technology*, 54(5), 1246-1267. <https://doi.org/10.1111/bjet.13320>.
27. Goetz, T., Botes, E., Resch, L. M., Weiss, S., Frenzel, A. C., & Ebner, M. (2024). Teachers emotionally profit from positive school leadership: Applying the PERMA-Lead model to the control-value theory of emotions. *Teaching and Teacher Education*, 141, 104517. <https://doi.org/10.1016/j.tate.2024.104517>.
28. Goriparthi, R. G. (2024). AI-Driven Predictive Analytics for Autonomous Systems: A Machine Learning Approach. *Revista de Inteligencia Artificial en Medicina*, 15(1), 843-879.

29. Hassall, C. D., Yan, Y., & Hunt, L. T. (2023). The neural correlates of continuous feedback processing. *Psychophysiology*, 60(12), e14399. <https://doi.org/10.1111/psyp.14399>.
30. Hauke, J., Achter, S., & Meyer, M. (2020). Theory development via replicated simulations and the added value of standards. *The journal of artificial societies and social simulation*, 23(1). <https://doi.org/10.15480/882.2679>.
31. Hernández-de-Menéndez, M., Morales-Menendez, R., Escobar, C. A., & Ramírez Mendoza, R. A. (2022). Learning analytics: state of the art. *International Journal on Interactive Design and Manufacturing*, 16(3), 1209–1230. <https://doi.org/10.1007/s12008-022-00930-0>.
32. Hornik, J., & Rachamim, M. (2024). Internet of Marketing Things: A Fog Computing Paradigm for Marketing Research. <https://doi.org/10.5772/intechopen.114333>.
33. Hudli, S. A., Hudli, A. A., & Hudli, A. V. (2012). Identifying online opinion leaders using k-means clustering. *2012 12th International Conference on Intelligent Systems Design and Applications (ISDA)*, 416–419. <https://doi.org/10.1109/ISDA.2012.6416565>.
34. Jain, P., Pateria, N., Anjum, G., Tiwari, A., & Tiwari, A. (2023). Edge AI and on-device machine learning for real-time processing. *International Journal of Innovative Research in Computer and Communication Engineering*, 12(5). <https://doi.org/10.15680/IJIRCCE.2024.1205364>.
35. Karatza, H. (2020). Cloud vs Fog Computing — Scheduling Real-Time Applications. *2020 9th Mediterranean Conference on Embedded Computing (MECO)*, 1. <https://doi.org/10.1109/MECO49872.2020.9134125>.
36. Kelvin Edem Bassey, Jesse Opoku-Boateng, Bernard Owusu Antwi, Afari Ntiakoh, & Ayanwunmi Rebecca Juliet. (2024). Digital twin technology for renewable energy microgrids. *Engineering Science & Technology Journal*, 5(7), 2248–2272. <https://doi.org/10.51594/estj.v5i7.1319>.
37. Kinsner, W. (2021). Digital Twins for Personalized Education and Lifelong Learning. *Canadian Conference on Electrical and Computer Engineering*, 2021-September. <https://doi.org/10.1109/CCECE53047.2021.9569178>.
38. Kroeger, K., Dogra, N., Rosa-Medina, R., Paluch, M., Ferri, F., Donner, T., & Esslinger, T. (2020). Continuous feedback on a quantum gas coupled to an optical cavity. *New Journal of Physics*, 22(3), 033020. <https://doi.org/10.1088/1367-2630/ab73cc>.
39. Li, H., Sun, J., & Ke, X. (2024). AI-driven optimization system for large-scale Kubernetes clusters: Enhancing cloud infrastructure availability, security, and disaster recovery. *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, 2(1), 281–306. <https://doi.org/10.60087/jaigs.v2i1.244>.
40. Li, Y., Wang, C., & Jia, J. (2025). Llama-vid: An image is worth 2 tokens in large language models. In *European Conference on Computer Vision* (pp. 323–340). Springer, Cham. https://doi.org/10.1007/978-3-031-72952-2_19.
41. Lim, L., Bannert, M., van der Graaf, J., Singh, S., Fan, Y., Surendrannair, S., Rakovic, M., Molenaar, I., Moore, J., & Gašević, D. (2023). Effects of real-time analytics-based personalized scaffolds on students' self-regulated learning. *Computers in Human Behavior*, 139. <https://doi.org/10.1016/j.chb.2022.107547>.
42. Lim, L.-A., Gentili, S., Pardo, A., Kovanović, V., Whitelock-Wainwright, A., & Gašević, D. (2021). What changes and for whom? A study of the impact of learning analytics-based process feedback in a large course. *Learning and Instruction*, 72, 101203. <https://doi.org/10.1016/j.learninstruc.2019.03.001>.
43. Lodge, J. M., Knight, S., & Kitto, K. (2023). Theory and learning analytics, a historical perspective. *Center for Open Science*. <https://doi.org/10.35542/osf.io/3z9fm>.
44. Luijken, K., Lohmann, A., Alter, U., Claramunt Gonzalez, J., Clouth, F. J., Fossum, J. L., ... & Groenwold, R. H. H. (2024). Replicability of simulation studies for the investigation of statistical methods: the RepliSims project. *Royal Society Open Science*, 11(1), 231003. <https://doi.org/10.1098/rsos.231003>.
45. Michelet, G., & Breitingner, F. (2024). ChatGPT, Llama, can you write my report? An experiment on assisted digital forensics reports written using (local) large language models. *Forensic Science International: Digital Investigation*, 48. <https://doi.org/10.1016/j.fsidi.2023.301683>.
46. Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X., & Gao, J. (2024). Large language models: A survey. *arXiv preprint arXiv:2402.06196*. <https://doi.org/10.48550/arXiv.2402.06196>.
47. Moubayed, A., Injadat, M., Shami, A., & Lutfiyya, H. (2020). Student engagement level in an e-learning environment: Clustering using k-means. *American Journal of Distance Education*, 34(2), 137–156. <https://doi.org/10.1080/08923647.2020.1696140>.

48. Nandhakumar, A. R., Baranwal, A., Choudhary, P., Golec, M., & Gill, S. S. (2024). Edgeaisim: A toolkit for simulation and modelling of ai models in edge computing environments. *Measurement: Sensors*, 31, 100939. <https://doi.org/10.1016/j.measen.2023.100939>.
49. Ometov, A., Molua, O. L., Komarov, M., & Nurmi, J. (2022). A Survey of Security in Cloud, Edge, and Fog Computing. In *Sensors* (Vol. 22, Issue 3). MDPI. <https://doi.org/10.3390/s22030927>.
50. Orland-Barak, L., Maskit, D., Orland-Barak, L., & Maskit, D. (2017). Simulation as 'Replicating Experience'. *Methodologies of Mediation in Professional Learning*, 63-74. https://doi.org/10.1007/978-3-319-49906-2_6.
51. Pardo, A., Jovanovic, J., Dawson, S., Gašević, D., & Mirriahi, N. (2019). Using learning analytics to scale the provision of personalised feedback. *British Journal of Educational Technology*, 50(1), 128–138. <https://doi.org/10.1111/bjet.12592>.
52. Paul, D. R., & Patel, V. G. L. (2024). EDGE AI-based object detection system using TFLite. *International Journal of Scientific Research in Engineering and Management (IJSREM)*, 8(7), 1. <https://doi.org/10.55041/IJSREM36414>.
53. Paul, P. K. (2024). Fog Computing and Its Emergence with Reference to Applications and Potentialities in Traditional and Digital Educational Systems: A Scientific Review. *Fog Computing for Intelligent Cloud IoT Systems*, 331-353. <https://doi.org/10.1002/9781394175345.ch14>.
54. Pushpa, J., & Kalyani, S. A. (2020). Using fog computing/edge computing to leverage Digital Twin. In *Advances in computers* 117(1), 51-77, Elsevier. <https://doi.org/10.1016/bs.adcom.2019.09.003>.
55. Qin, B., Pan, H., Dai, Y., Si, X., Huang, X., Yuen, C., & Zhang, Y. (2024). Machine and deep learning for digital twin networks: A survey. *IEEE Internet of Things Journal*. <https://doi.org/10.1109/IIOT.2024.3416733>.
56. Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). *Learning Transferable Visual Models From Natural Language Supervision*. <http://arxiv.org/abs/2103.00020>.
57. Roumeliotis, K. I., Tselikas, N. D., & Nasiopoulos, D. K. (2024). LLMs in e-commerce: A comparative analysis of GPT and LLaMA models in product review evaluation. *Natural Language Processing Journal*, 6, 100056. <https://doi.org/10.1016/j.nlp.2024.100056>.
58. Sacoto-Cabrera, E. J. (2024, August). Fog Computing: A Systematic Review from the Perspectives of Architecture, Business Models, and Research Developments. In *2024 IEEE Colombian Conference on Communications and Computing (COLCOM)* (pp. 1-6). IEEE. <https://doi.org/10.1109/COLCOM62950.2024.10720269>.
59. Sethi, S., Ahmad, N., Rani, S., & Anjum, M. Digital Twin Technology Use Cases in Intelligent Manufacturing. In *Intelligent Manufacturing and Industry 4.0* (pp. 67-80). CRC Press.
60. Si, J. (2024). Using cognitive load theory to tailor clinical reasoning training for preclinical medical students. *BMC Medical Education*, 24(1), 1185. <https://doi.org/10.1186/s12909-024-06111-9>.
61. Singh, A. P. A., & Gameti, N. (2024). Digital Twins in Manufacturing: A Survey of Current Practices and Future Trends. <https://doi.org/10.30574/ijjsra.2024.13.1.1705>.
62. Spijkerman, S., Manning, D. M., & Green-Thompson, L. P. (2022). A Cognitive Load Theory Perspective of the Undergraduate Anesthesia Curricula in South Africa. *Anesthesia & Analgesia*, 10-1213. <https://doi.org/10.1213/ANE.0000000000007033>.
63. Srividhya, V., Gowriswari, S., Antony, N. V., Murugan, S., Anitha, K., & Rajmohan, M. (2024, February). Optimizing Electric Vehicle Charging Networks Using Clustering Technique. In *2024 2nd International Conference on Computer, Communication and Control (IC4)* (pp. 1-5). IEEE. <https://doi.org/10.1109/IC457434.2024.10486422>.
64. Swarnakar, S. (2024). Fog Computing: Architecture and Application. *Fog Computing for Intelligent Cloud IoT Systems*, 1-31. <https://doi.org/10.1002/9781394175345.ch1>.
65. Tabatabaee, S. S., Jambarsang, S., & Keshmiri, F. (2024). Cognitive load theory in workplace-based learning from the viewpoint of nursing students: application of a path analysis. *BMC Medical Education*, 24(1), 678. <https://doi.org/10.1186/s12909-024-05664-z>.

66. Thakur, K., Pathan, A. S. K., & Ismat, S. (2023). Distributed Cloud Computing. In *Emerging ICT Technologies and Cybersecurity: From AI and ML to Other Futuristic Technologies* (pp. 185-197). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-27765-8_7.
67. Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G. (2023). *LLaMA: Open and Efficient Foundation Language Models*. <http://arxiv.org/abs/2302.13971>.
68. Traynor, S., Wellens, M. A., Krishnamoorthy, V., Traynor, S., Wellens, M. A., & Krishnamoorthy, V. (2021). Continuous performance management. *SAP SuccessFactors Talent: Volume 1: A Complete Guide to Configuration, Administration, and Best Practices: Performance and Goals*, 613-643. https://doi.org/10.1007/978-1-4842-6600-7_13.
69. Valli, L. N. (2024). Predictive Analytics Applications for Risk Mitigation across Industries; A review. *BULLET: Jurnal Multidisiplin Ilmu*, 3(4), 542-553. <https://doi.org/10.1109/IC457434.2024.10486422>.
70. Wang, L., & Zhao, J. (2022). Distributed Computing. In *Architecture of Advanced Numerical Analysis Systems: Designing a Scientific Computing System using OCaml*, 243-279, Berkeley, CA: Apress.
71. Wise, A., Zhao, Y., & Hausknecht, S. (2014). Learning Analytics for Online Discussions: Embedded and Extracted Approaches. *Journal of Learning Analytics*, 1(2), 48-71. <https://doi.org/10.18608/jla.2014.12.4>.
72. Wu, C., Lin, W., Zhang, X., Zhang, Y., Xie, W., & Wang, Y. (2024). PMC-LLaMA: toward building open-source language models for medicine. *Journal of the American Medical Informatics Association*, ocae045. <https://doi.org/10.1093/jamia/ocae045>.
73. Xie, Q., Chen, Q., Chen, A., Peng, C., Hu, Y., Lin, F., ... & Bian, J. (2024). Me llama: Foundation large language models for medical applications. *arXiv preprint arXiv:2402.12749*. <https://doi.org/10.48550/arXiv.2402.12749>.
74. Yang, Z., Tang, C., Zhang, T., Zhang, Z., & Doan, D. T. (2024). Digital Twins in Construction: Architecture, Applications, Trends and Challenges. <https://doi.org/10.3390/buildings14092616>.
75. Yim, J. S. C., Sedhu, D. S., & Abdul-Rahim, P. R. M. (2024). Emotion Matters: Understanding the Role of Achievement Emotions in High School Students' Decision for Tertiary Education. *Journal of Psychoeducational Assessment*, 42(2), 223-240. <https://doi.org/10.1177/07342829231211044>.
76. Zarzycka, E., Krasodomska, J., Mazurczak-Maká, A., & Turek-Radwan, M. (2021). Distance learning during the covid-19 pandemic: Students' communication and collaboration and the role of social media. *Cogent Arts & Humanities*, 8(1), 1953228. <https://doi.org/10.1080/23311983.2021.1953228>.
77. Zhang, Y., Liang, R., Wei, L., Fu, X., & Zheng, Y. (2022). A Digital Twin of Instructor in Higher Education. *2022 8th Annual International Conference on Network and Information Systems for Computers (ICNISC)*, 866–870. <https://doi.org/10.1109/ICNISC57059.2022.00173>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.