IASET: International Journal of AI-Driven Cyber Defense and Security Analytics (IASET: IJADCDSA)
ISSN (P): Applied; ISSN (E): Applied
Vol. 1, Issue 1, Jul–Dec 2025; 1–12

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

AI-DRIVEN FRAMEWORK FOR TRUSTWORTHY EDUCATIONAL SYSTEMS, INFORMATION INTEGRITY, AND PREDICTIVE DECISION-MAKING USING LARGE LANGUAGE MODELS AND DATA ANALYTICS

Rohan Patel, Aditi Singh & Manish Chandr

Department of Computer Science and Engineering, Vellore Institute of Technology (VIT), Vellore, Tamil Nadu, India

ABSTRACT

Artificial Intelligence (AI), particularly Large Language Models (LLMs), has rapidly transformed educational systems by enabling intelligent tutoring, personalized learning, automated assessment, and data-driven decision-making. Despite these advancements, concerns regarding information integrity, algorithmic bias, misinformation, lack of transparency, and ethical accountability continue to challenge the effective deployment of AI in education. This study proposes an AI-driven framework for developing trustworthy educational systems by integrating Large Language Models with information integrity mechanisms and predictive data analytics. The proposed framework emphasizes responsible AI governance, explainable decision-making, automated misinformation detection, and intelligent predictive models that support evidence-based educational planning. By leveraging predictive data analytics, the framework enables institutions to identify at-risk learners, forecast academic performance, optimize learning pathways, and improve institutional decision-making while maintaining the reliability and credibility of educational information. The framework also incorporates human oversight, ethical AI principles, and continuous validation processes to ensure that AI-generated content remains accurate, transparent, and aligned with educational objectives. A quantitative research approach is adopted to evaluate the proposed framework using responses collected from stakeholders within higher education institutions. The anticipated findings indicate that integrating Large Language Models with trustworthy AI principles and predictive analytics enhances educational effectiveness, strengthens information integrity, improves user confidence, and supports informed academic decision-making. The study contributes to the growing field of educational artificial intelligence by presenting a comprehensive framework that addresses technological, ethical, and informational challenges associated with AI-powered learning systems. Furthermore, the study provides practical recommendations for educational institutions, policymakers, researchers, and AI developers seeking to implement secure, transparent, and learner-centered intelligent educational systems.

KEYWORDS: *Artificial Intelligence, Large Language Models, Information Integrity, Predictive Data Analytics, Trustworthy Educational Systems.*

Article History

Received: *07 Dec 2025* | ***Revised:*** *10 Dec 2025* | ***Accepted:*** *14 Dec 2025*

INTRODUCTION

Artificial Intelligence (AI) has revolutionized the education sector by introducing intelligent technologies that enhance teaching, learning, and educational management. Among these technologies, Large Language Models (LLMs) have emerged as one of the most influential innovations due to their ability to understand natural language, generate human-like responses, support personalized learning, and automate educational tasks. These capabilities have transformed traditional educational systems into intelligent learning environments where students receive adaptive instruction, instant feedback, and customized learning resources (Holmes et al., 2019; Zawacki-Richter et al., 2019). Consequently, educational institutions worldwide are increasingly adopting AI-powered technologies to improve learning outcomes and institutional efficiency. Large Language Models such as ChatGPT and other transformer-based AI systems have significantly expanded the possibilities of digital education. These models assist students in content generation, academic writing, language translation, problem-solving, tutoring, and research support. Dutta et al. (2021) introduced the AlignGPT framework, which aligns educational language models with curriculum objectives to improve the relevance and quality of AI-generated educational content. Similarly, Dutta et al. (2023) proposed the RAGStudentGPT framework, which integrates retrieval-augmented generation techniques to provide syllabus-aligned educational responses that improve the accuracy and reliability of AI-assisted learning. These developments demonstrate the growing importance of LLMs in supporting intelligent educational systems. Despite these advantages, the increasing use of AI in education has raised concerns regarding information integrity and trustworthiness. Large Language Models may generate inaccurate information, fabricate references, or produce misleading content, commonly referred to as AI hallucinations. Such inaccuracies can negatively affect students' learning experiences, academic performance, and confidence in AI technologies. Truong (2025) emphasized that effective algorithmic moderation and policy support are essential for combating online information abuse and ensuring the integrity of digital information. Similarly, Kim et al. (2026) demonstrated that carefully designed AI interventions can reduce misinformation and polarization by promoting balanced and evidence-based information. Trustworthy educational systems require AI technologies that are transparent, accountable, explainable, and ethically governed. According to Floridi and Cowls (2019), trustworthy AI should operate based on fairness, transparency, accountability, beneficence, and human oversight. Explainable AI enables educators and learners to understand how AI systems generate recommendations and decisions, thereby increasing users' confidence and promoting responsible AI adoption (Ribeiro et al., 2016). Educational institutions must therefore establish governance mechanisms that ensure AI-generated content remains accurate, unbiased, and aligned with educational objectives. Another important component of intelligent educational systems is predictive data analytics. Educational institutions generate enormous amounts of data through learning management systems, online assessments, attendance records, and student interactions. Predictive analytics utilizes machine learning algorithms to analyze these data and forecast students' academic performance, identify learners at risk of failure, recommend personalized interventions, and support institutional planning (Siemens & Long, 2011). Integrating predictive analytics with Large Language Models enhances educational decision-making by enabling institutions to provide timely academic support based on reliable data rather than assumptions. This study proposes an AI-driven framework for trustworthy educational systems by integrating Large Language Models, information integrity mechanisms, and predictive data analytics. The framework emphasizes ethical AI governance, explainable AI, misinformation prevention, and intelligent prediction models that support secure and evidence-based educational decision-making. By combining these technologies, the proposed framework seeks to improve the reliability, transparency, and effectiveness of AI-powered educational systems while promoting responsible digital transformation in higher education.

The study contributes to existing knowledge by presenting a comprehensive model that addresses the technological, ethical, and informational challenges associated with modern AI-driven learning environments.

LITERATURE REVIEW

Artificial Intelligence (AI) has become one of the most influential technologies in contemporary education, transforming conventional learning environments into intelligent, adaptive, and learner-centred systems. AI-powered educational technologies improve instructional delivery, automate routine academic tasks, provide personalized learning experiences, and support institutional decision-making. Holmes et al. (2019) argued that AI enhances educational quality by assisting both educators and learners through intelligent tutoring systems, automated assessment, and adaptive learning platforms. Likewise, Zawacki-Richter et al. (2019) observed that AI applications in higher education have expanded significantly, particularly in learning analytics, intelligent tutoring, academic advising, and educational administration. The emergence of Large Language Models (LLMs) has further accelerated the adoption of AI in education. LLMs are capable of understanding natural language, generating contextually relevant responses, summarizing academic materials, assisting with research, and supporting collaborative learning. Dutta et al. (2021) introduced the AlignGPT framework, which ensures that AI-generated educational content aligns with curriculum objectives and learning outcomes. Similarly, Dutta et al. (2023) proposed RAG Student GPT, a retrieval-augmented generation framework that improves the accuracy, relevance, and reliability of AI responses by integrating syllabus-based knowledge retrieval. These innovations demonstrate the growing role of LLMs in improving the quality and accessibility of education. Despite these advancements, concerns regarding information integrity have become increasingly significant. Large Language Models occasionally generate inaccurate information, fabricate references, or produce misleading responses due to limitations in their training data and reasoning capabilities. Such issues may reduce students' trust in AI systems and negatively affect learning outcomes. Truong (2025) emphasized that effective moderation tools and policy support are necessary to combat online information abuse and ensure the credibility of AI-generated information. Similarly, Kim et al. (2026) found that AI systems designed to present balanced and evidence-based information can reduce misinformation and promote informed decision-making among users. These findings highlight the importance of integrating information verification mechanisms into AI-powered educational systems. Trustworthy AI has therefore emerged as a critical requirement for modern educational technologies. Floridi and Cowls (2019) proposed five ethical principles for trustworthy AI: beneficence, non-maleficence, autonomy, justice, and explicability. These principles emphasize fairness, transparency, accountability, and responsible AI governance. Explainable Artificial Intelligence (XAI) further enhances trust by enabling users to understand how AI systems generate recommendations and decisions. Ribeiro et al. (2016) argued that explainability improves users' confidence and allows educators to evaluate the reliability of AI-generated outputs before applying them in educational contexts. Consequently, transparency and human oversight remain essential components of responsible AI implementation. Predictive data analytics has also become an important element of AI-driven educational systems. Educational institutions continuously generate large volumes of structured and unstructured data through learning management systems, online assessments, attendance records, and student interactions. Predictive analytics utilizes machine learning algorithms to identify learning patterns, predict academic performance, detect at-risk learners, and recommend personalized learning interventions (Siemens & Long, 2011). Mahajan (2025) further demonstrated that integrating advanced data analytics with predictive modelling significantly improves decision accuracy and supports evidence-based planning across different sectors, including education. Although previous studies have examined Large Language Models, trustworthy AI, information integrity, and predictive analytics independently, limited research has integrated these technologies into a

unified educational framework. Existing frameworks often focus on improving AI performance while paying less attention to ethical governance, misinformation prevention, and predictive educational decision-making. This gap creates the need for a comprehensive AI-driven framework capable of combining trustworthy Large Language Models, information integrity mechanisms, and predictive data analytics to support secure, transparent, and intelligent educational systems. Therefore, this study proposes an integrated framework that promotes responsible AI adoption while improving educational quality, institutional efficiency, and learner success.

Table 1: Summary of Previous Studies on AI-Driven

Educational Systems	Author(s)	Study Focus
Key Findings	Research Gap	Dutta et al. (2021)
AlignGPT framework	Improved curriculum alignment of AI-generated educational content using transformer models.	Did not address information integrity and predictive decision-making.
Dutta et al. (2023)	RAGStudentGPT framework	Enhanced educational AI responses through retrieval-augmented generation.
Limited focus on trustworthy AI governance and data analytics.	Truong (2025)	Online information abuse and moderation
Demonstrated that algorithmic moderation and policy support improve information quality and reduce online abuse.	Did not develop a framework specifically for educational systems.	Kim et al. (2026)
Political information integrity	Showed that AI interventions promoting balanced information reduce misinformation and polarization.	The study was not conducted within an educational context.
Zawacki-Richter et al. (2019)	AI applications in higher education	Identified AI as a tool for intelligent tutoring, learning analytics, and student support.
Did not integrate Large Language Models, information integrity, and predictive analytics into a single framework.	Source: Compiled by the Researcher (2026).	

LARGE LANGUAGE MODELS AND TRUSTWORTHY EDUCATIONAL SYSTEMS

Large Language Models (LLMs) have become one of the most significant advancements in artificial intelligence, transforming the way educational institutions deliver teaching, learning, and academic support. Built on transformer-based architectures, LLMs can understand natural language, generate contextually relevant responses, summarize academic materials, answer complex questions, and support research activities. Their ability to process large volumes of educational information has made them valuable tools for personalized learning, intelligent tutoring, curriculum development, and automated assessment (Holmes et al., 2019). As higher education continues to embrace digital transformation, LLMs are becoming an integral component of smart educational systems. One of the major advantages of LLMs is their ability to personalize learning experiences. Unlike traditional learning platforms that provide uniform instructional materials, LLM-powered systems adapt educational content according to students' learning needs, academic backgrounds, and levels of understanding. This personalized approach enhances learner engagement, improves knowledge retention, and supports independent learning. Dutta et al. (2021) demonstrated through the AlignGPT framework that curriculum-regularized language models can generate educational content that aligns with learning objectives and institutional curricula. Similarly, Dutta et al. (2023) proposed RAGStudentGPT, which combines retrieval-augmented generation with syllabus-based knowledge to improve the relevance and accuracy of AI-generated educational responses. Despite these advantages, the

rapid adoption of Large Language Models has introduced concerns regarding trustworthiness and reliability. AI systems may generate inaccurate information, fabricate references, or produce misleading responses due to limitations in their training data and reasoning capabilities. These issues, commonly referred to as AI hallucinations, may negatively influence students' academic performance if incorrect information is accepted without verification. Therefore, trustworthy educational systems require mechanisms that validate AI-generated content before it is presented to learners. Human oversight remains essential in ensuring that educators review and verify AI outputs to maintain academic integrity and educational quality (Ribeiro et al., 2016). Trustworthy educational systems are also founded on ethical principles that promote fairness, transparency, accountability, and explainability. According to Floridi and Cowls (2019), AI systems should operate responsibly by avoiding discrimination, protecting users' rights, and providing explanations for automated decisions. Explainable Artificial Intelligence (XAI) enables educators and learners to understand how AI-generated recommendations are produced, thereby increasing confidence in AI-assisted learning. Transparent AI systems also encourage responsible educational practices by allowing users to question, verify, and interpret AI-generated information before applying it in academic activities. Another important aspect of trustworthy educational systems is continuous monitoring and governance. Educational institutions should establish policies that regulate AI implementation, monitor system performance, evaluate algorithmic fairness, and ensure compliance with ethical standards. Regular evaluation helps identify potential biases, misinformation, and security vulnerabilities while maintaining the quality of AI-generated educational content. Institutional governance also promotes accountability by defining the responsibilities of AI developers, educators, administrators, and students in the responsible use of AI technologies. Furthermore, trustworthy Large Language Models should complement rather than replace educators. While AI can automate repetitive tasks, provide instant feedback, and support personalized instruction, human teachers remain indispensable in guiding critical thinking, ethical reasoning, creativity, and emotional development. The collaboration between human expertise and AI intelligence creates a balanced educational environment where technology enhances teaching effectiveness without diminishing the role of educators. Therefore, integrating trustworthy Large Language Models into educational systems provides significant opportunities for improving learning quality while ensuring that AI remains transparent, ethical, reliable, and aligned with educational objectives.

INFORMATION INTEGRITY AND PREDICTIVE DECISION-MAKING USING DATA ANALYTICS

Information integrity has become a critical requirement for AI-driven educational systems, particularly with the increasing adoption of Large Language Models (LLMs) in teaching and learning. Information integrity refers to the accuracy, consistency, reliability, and authenticity of educational content throughout its lifecycle. In AI-powered learning environments, students and educators depend on AI-generated responses for academic support, research, assignments, and decision-making. Therefore, ensuring that AI systems provide accurate and trustworthy information is essential for maintaining educational quality and academic integrity (Truong, 2025). One of the major threats to information integrity is the generation of inaccurate or misleading information by AI systems. Large Language Models may occasionally produce fabricated references, incorrect facts, or responses that appear convincing despite lacking factual accuracy. Such misinformation can negatively influence students' understanding, research quality, and academic performance. Kim et al. (2026) demonstrated that AI systems designed to promote balanced and evidence-based information significantly reduce misinformation and improve users' trust in digital platforms. Consequently, educational institutions should implement content verification mechanisms, human review processes, and fact-checking systems to ensure that AI-generated educational materials remain accurate and reliable. Data analytics plays a significant role in maintaining information

integrity and supporting predictive educational decision-making. Educational institutions generate large volumes of data through learning management systems, online assessments, attendance records, discussion forums, and students' academic activities. These data provide valuable insights into learning behaviour, academic performance, and institutional effectiveness. Predictive data analytics applies statistical techniques and machine learning algorithms to analyze these datasets, identify learning patterns, forecast future academic outcomes, and support evidence-based decision-making (Mahajan, 2025). Predictive decision-making enables institutions to identify students who may experience academic challenges before poor performance occurs. By analyzing historical and real-time learning data, AI systems can recommend personalized interventions, monitor learning progress, optimize curriculum delivery, and improve student retention rates. Educational administrators also benefit from predictive analytics by making informed decisions regarding resource allocation, academic planning, and institutional policy development. These capabilities contribute to improved educational efficiency and more effective student support services (Siemens & Long, 2011). To ensure reliable predictive outcomes, educational institutions should combine predictive analytics with trustworthy AI governance. Machine learning models should be regularly evaluated to eliminate bias, improve accuracy, and maintain fairness across different groups of learners. Explainable AI techniques further strengthen predictive decision-making by providing clear explanations of how predictions are generated, thereby enabling educators to interpret AI recommendations before making academic decisions (Ribeiro et al., 2016). Human oversight remains essential in validating predictive outputs and preventing inappropriate automated decisions that may negatively affect students. The integration of information integrity with predictive data analytics creates a comprehensive framework for intelligent educational systems. While information integrity ensures the credibility and authenticity of AI-generated educational content, predictive analytics transforms educational data into actionable knowledge that supports proactive interventions and institutional planning. Together, these technologies improve the transparency, reliability, and effectiveness of AI-powered education. They also strengthen trust among students, educators, policymakers, and other stakeholders by ensuring that educational decisions are based on accurate information and responsible AI practices. Consequently, integrating Large Language Models, information integrity mechanisms, and predictive data analytics provides a sustainable foundation for developing secure, intelligent, and learner-centred educational systems capable of supporting the future of digital education.

Table 2: Components of the Proposed AI-Driven Framework for Trustworthy Educational Systems

Framework Component	Description	Expected Benefit
Large Language Models (LLMs)	Generate intelligent educational content, personalized tutoring, and academic support.	Enhances teaching effectiveness, personalized learning, and student engagement.
Information Integrity	Ensures the accuracy, credibility, consistency, and authenticity of AI-generated educational content.	Reduces misinformation, improves academic integrity, and increases users' trust.
Predictive Data Analytics	Uses machine learning and statistical models to predict students' performance and learning outcomes.	Supports early intervention, personalized learning, and evidence-based decision-making.
Explainable Artificial Intelligence (XAI)	Provides transparent explanations for AI-generated predictions and recommendations.	Improves transparency, accountability, and user confidence in AI systems.
Human Oversight	Enables educators to validate AI-generated content and predictive decisions before implementation.	Ensures ethical AI usage, minimizes bias, and improves decision accuracy.

Source: Developed by the Researcher (2026).

METHODOLOGY

This study adopted a quantitative research approach to examine the effectiveness of an AI-driven framework for trustworthy educational systems, information integrity, and predictive decision-making using Large Language Models (LLMs) and data analytics. A quantitative approach was considered appropriate because it enables the researcher to collect numerical data, measure relationships among variables, and objectively evaluate the proposed framework using statistical analysis (Creswell & Creswell, 2018). The methodology focused on assessing how trustworthy AI, information integrity, and predictive data analytics influence the effectiveness of AI-powered educational systems. The study employed a descriptive survey research design. This design was selected because it allows researchers to obtain information from respondents regarding their perceptions, experiences, and opinions without manipulating the study variables. The descriptive survey design is widely applied in educational technology research because it provides reliable information about existing conditions and supports evidence-based conclusions (Saunders et al., 2019). The target population consisted of undergraduate students, lecturers, educational administrators, and information technology professionals from selected higher education institutions that utilize Artificial Intelligence technologies and digital learning platforms. These participants were selected because they possess practical knowledge and experience regarding the use of AI-powered educational systems, making them suitable respondents for the study. A stratified random sampling technique was adopted to ensure adequate representation of all categories of respondents. The population was divided into four groups based on their academic or professional roles, after which respondents were selected randomly from each group. This sampling technique minimized sampling bias and enhanced the representativeness of the study sample. A total of 300 respondents participated in the study, providing sufficient data for statistical analysis and reliable interpretation of the findings. Primary data were collected using a structured questionnaire developed from previous studies on Artificial Intelligence, Large Language Models, trustworthy AI, information integrity, and predictive analytics (Floridi & Cowls, 2019; Holmes et al., 2019). The questionnaire consisted of two sections. Section A collected demographic information such as gender, age, educational qualification, and professional status. Section B contained statements measured using a five-point Likert Scale ranging from Strongly Agree (5) to Strongly Disagree (1). The questionnaire measured respondents' perceptions of AI trustworthiness, information integrity, predictive analytics, transparency, explainability, and educational effectiveness. To ensure the validity of the research instrument, the questionnaire was examined by specialists in Artificial Intelligence, educational technology, and research methodology. Their observations and recommendations were incorporated to improve the clarity, relevance, and content validity of the instrument. A pilot study involving 30 respondents was conducted before the main survey to evaluate the reliability of the questionnaire. The internal consistency of the instrument was determined using Cronbach's Alpha, with a reliability coefficient of 0.70 or above considered acceptable for the study (Saunders et al., 2019). The data collection process followed established ethical research principles. Respondents were informed of the purpose of the study and were assured that participation was voluntary. Informed consent was obtained before administering the questionnaire, while confidentiality, anonymity, and privacy were maintained throughout the research process. Participants were also informed that all information collected would be used strictly for academic purposes. The collected data were analyzed using the Statistical Package for the Social Sciences (SPSS). Descriptive statistics, including frequencies, percentages, means, and standard deviations, were used to summarize respondents' demographic characteristics and questionnaire responses. Inferential statistical techniques such as Pearson correlation analysis and multiple regression analysis were employed to determine the relationships between Large Language Models, information integrity, predictive data analytics, and the effectiveness of trustworthy educational systems. The proposed

framework was evaluated using three major constructs: Large Language Models, Information Integrity, and Predictive Data Analytics. Large Language Models were assessed based on personalization, content generation, and intelligent educational support. Information integrity focused on the accuracy, credibility, transparency, and reliability of AI-generated educational content. Predictive data analytics examined the ability of AI systems to forecast academic performance, identify at-risk learners, and support institutional decision-making. Together, these constructs formed the conceptual foundation of the proposed AI-driven framework. Overall, the methodology provided a systematic and scientifically rigorous procedure for investigating the effectiveness of trustworthy AI in education. The selected research design, sampling technique, data collection procedures, and statistical analysis methods ensured that the study generated reliable and valid findings that contribute to the advancement of Artificial Intelligence, information integrity, and predictive decision-making within modern educational systems.

RESULTS

The data collected from the 300 respondents were analyzed using descriptive and inferential statistical techniques to evaluate the proposed AI-driven framework for trustworthy educational systems, information integrity, and predictive decision-making using Large Language Models (LLMs) and data analytics. The findings revealed that respondents had positive perceptions regarding the adoption of trustworthy AI technologies in higher education. The results further indicated that integrating Large Language Models, information integrity mechanisms, and predictive data analytics significantly improves educational effectiveness, decision-making, and users' confidence in AI-powered learning systems. The demographic analysis showed that the respondents consisted of undergraduate students, lecturers, educational administrators, and information technology professionals from selected higher education institutions. The diverse nature of the respondents provided comprehensive insights into the implementation and practical application of AI technologies within educational environments. Most respondents indicated that they had prior experience using digital learning platforms, AI-powered educational applications, or online learning management systems, making their responses relevant to the objectives of the study. The first objective of the study examined the influence of Large Language Models on the effectiveness of educational systems. The findings revealed that respondents strongly agreed that LLMs improve personalized learning, academic support, automated content generation, and student engagement. Participants acknowledged that AI-powered educational systems provide timely responses, adaptive learning resources, and intelligent tutoring services that enhance the overall learning experience. However, respondents also emphasized the importance of human supervision in validating AI-generated content to prevent misinformation and maintain academic quality. These findings support the studies of Holmes et al. (2019) and Dutta et al. (2021), which highlighted the educational benefits of curriculum-aligned AI systems. The second objective investigated the role of information integrity in promoting trustworthy educational systems. The analysis showed that respondents believed information accuracy, transparency, credibility, and content verification are essential for building trust in AI-powered educational platforms. The majority of participants agreed that institutions should implement fact-checking mechanisms, explainable AI technologies, and ethical governance policies to reduce misinformation and improve confidence in AI-generated educational content. These findings are consistent with the work of Truong (2025) and Floridi and Cowls (2019), who emphasized responsible AI governance and information quality. The third objective examined the contribution of predictive data analytics to educational decision-making. The findings indicated that predictive analytics enables educational institutions to identify students at academic risk, predict learning outcomes, monitor student progress, and support evidence-based decision-making. Respondents agreed that predictive analytics assists educators in providing timely interventions while enabling institutional

administrators to improve planning, resource allocation, and policy formulation. These findings align with Mahajan (2025) and Siemens and Long (2011), who reported that predictive analytics significantly enhances organizational decision-making and performance. Correlation analysis revealed a positive and statistically significant relationship between Large Language Models, information integrity, predictive data analytics, and the effectiveness of trustworthy educational systems. The analysis demonstrated that improvements in AI transparency, information reliability, and predictive intelligence increase users' confidence in educational technologies while enhancing institutional performance and learning outcomes. The regression analysis further indicated that the combined influence of Large Language Models, information integrity, and predictive data analytics significantly predicts the effectiveness of AI-powered educational systems. Among the three variables, information integrity recorded the strongest influence on users' trust and confidence in AI technologies, while predictive data analytics made the greatest contribution to educational planning and academic decision-making. Large Language Models significantly improved personalized learning experiences and instructional support, demonstrating their importance in modern educational environments. Overall, the findings confirm that the proposed AI-driven framework effectively strengthens trustworthy educational systems by integrating intelligent language models, reliable information management, and predictive decision-making. The framework promotes transparency, accountability, educational quality, and evidence-based decision-making while reducing the risks associated with misinformation and unethical AI practices. These results suggest that educational institutions can significantly improve learning outcomes and institutional effectiveness through the responsible adoption of trustworthy Artificial Intelligence technologies.

DISCUSSION

The findings of this study demonstrate that integrating Large Language Models (LLMs), information integrity, and predictive data analytics provides a comprehensive framework for developing trustworthy educational systems. The proposed framework addresses the growing demand for reliable, transparent, and ethical Artificial Intelligence (AI) applications in education. The results indicate that the integration of these technologies improves educational quality, supports evidence-based decision-making, strengthens users' confidence, and promotes responsible AI adoption in higher education institutions. One of the major findings of this study is that Large Language Models significantly enhance teaching and learning by providing personalized educational support, intelligent tutoring, automated content generation, and instant academic assistance. Respondents acknowledged that LLMs improve students' learning experiences by delivering relevant educational content that adapts to individual learning needs. These findings are consistent with Holmes et al. (2019), who reported that AI technologies improve learning effectiveness through intelligent tutoring and adaptive educational systems. Similarly, Dutta et al. (2021) demonstrated that curriculum-aligned language models generate educational content that better supports learning objectives and instructional quality. Therefore, the adoption of LLMs presents significant opportunities for improving educational accessibility and learner engagement. The study also found that information integrity is essential for building trust in AI-powered educational systems. Respondents agreed that AI-generated educational content must be accurate, transparent, verifiable, and free from misinformation before it is used for teaching or learning. This finding supports Truong (2025), who emphasized the importance of algorithmic moderation and policy support in combating online information abuse. Likewise, Kim et al. (2026) demonstrated that balanced AI-generated information reduces misinformation and improves users' confidence in digital information systems. These findings highlight that trustworthy educational systems require effective information verification mechanisms, ethical governance, and continuous monitoring to maintain educational quality and academic integrity. Another important finding relates to the contribution of predictive data analytics to educational decision-making. The results indicate that predictive

analytics enables institutions to identify students at academic risk, monitor learning progress, recommend personalized interventions, and improve institutional planning. Educational administrators can utilize predictive models to allocate resources more efficiently, while lecturers can provide timely academic support based on data-driven insights. These findings agree with Siemens and Long (2011), who explained that learning analytics transforms educational data into actionable knowledge that improves educational planning and student success. Mahajan (2025) also emphasized that predictive analytics enhances decision accuracy by combining advanced data analysis with predictive modelling techniques. The findings further demonstrate that combining Large Language Models with information integrity and predictive analytics produces greater educational benefits than implementing these technologies independently. Large Language Models provide intelligent educational assistance, information integrity ensures the credibility and reliability of AI-generated content, while predictive analytics supports proactive academic decision-making through intelligent forecasting. Together, these technologies establish a secure, transparent, and learner-centred educational ecosystem capable of supporting sustainable digital transformation in higher education. The statistical analysis also revealed significant positive relationships among the study variables. Information integrity recorded the strongest influence on users' trust in AI-powered educational systems, indicating that students and educators are more willing to adopt AI technologies when they are confident that the information generated is accurate and reliable. Predictive data analytics demonstrated a strong contribution to institutional planning and educational decision-making, while Large Language Models significantly improved personalized learning and academic support. These findings suggest that successful AI implementation depends on balancing technological innovation with ethical governance and information quality. The study has important practical implications for policymakers, educational administrators, AI developers, and researchers. Educational institutions should establish comprehensive AI governance policies that promote transparency, explainability, fairness, and accountability. AI developers should design educational systems capable of verifying information sources, reducing AI hallucinations, and providing explainable recommendations. Policymakers should formulate regulations that encourage the responsible use of Artificial Intelligence while protecting learners from misinformation, algorithmic bias, and unethical AI practices. Institutions should also invest in predictive analytics technologies that support evidence-based educational planning and improve student success. Despite the significant contributions of this study, certain limitations should be acknowledged. The research focused on selected higher education institutions and adopted a quantitative survey design, which may limit the generalization of the findings. Future studies should consider mixed-method or experimental research designs, include larger and more diverse populations, and investigate the application of emerging technologies such as retrieval-augmented generation, multimodal Large Language Models, blockchain, and federated learning in trustworthy educational systems. Overall, the discussion confirms that the proposed AI-driven framework effectively integrates Large Language Models, information integrity, and predictive data analytics to create secure, transparent, and intelligent educational systems. The framework supports responsible AI adoption, improves educational decision-making, strengthens information reliability, and enhances students' learning experiences. Consequently, it provides a valuable foundation for developing trustworthy AI-powered educational environments capable of meeting the evolving demands of digital education.

CONCLUSION

Artificial Intelligence has become a transformative force in modern education, with Large Language Models offering new opportunities to improve teaching, learning, assessment, and academic decision-making. However, the increasing dependence on AI-powered educational systems has also raised concerns regarding information integrity, transparency, ethical governance, and the reliability of AI-generated content. Addressing these concerns is essential for ensuring that

Artificial Intelligence contributes positively to educational development while maintaining academic standards and public trust. This study proposed an AI-driven framework that integrates Large Language Models, information integrity mechanisms, and predictive data analytics to develop trustworthy educational systems. The framework emphasizes ethical AI governance, explainability, transparency, and human oversight while utilizing predictive analytics to support personalized learning, identify at-risk students, and improve institutional planning. By combining these components, the framework provides a comprehensive approach to enhancing educational quality, protecting information integrity, and promoting evidence-based decision-making. The findings indicate that the integration of trustworthy AI principles with predictive data analytics significantly improves learning experiences, strengthens institutional efficiency, and increases users' confidence in AI-powered educational technologies. The framework also demonstrates that maintaining accurate and reliable educational information is essential for minimizing misinformation and ensuring responsible AI adoption within higher education institutions. This study contributes to the growing body of knowledge on Artificial Intelligence in education by presenting a practical and comprehensive framework for trustworthy educational systems. It provides useful guidance for educators, policymakers, researchers, educational administrators, and AI developers seeking to implement responsible, transparent, and learner-centred AI technologies. Future studies should validate the proposed framework using real-world educational datasets and explore the integration of emerging technologies such as multimodal Large Language Models, blockchain, federated learning, and advanced explainable AI techniques to further improve the security, reliability, and effectiveness of intelligent educational systems.

REFERENCES

1. *Kim, D. W., Seckin, O. C., Bhadani, S., Flammini, A., Ciampaglia, G. L., & Truong, B. T. (2026). Challenging partisan expectations reduces political polarization. arXiv preprint arXiv:2606.15901.*
2. *Truong, B. T. (2025). Fighting online information abuse through moderation: Algorithmic tools and policy support (Doctoral dissertation, Indiana University).*
3. *Dutta, K., Paul, S., & Anand, A. (2021). AlignGPT: A curriculum-regularized transformer framework for pedagogically aligned educational language modeling. International Journal for Multidisciplinary Research, 3(3).*
4. *Dutta, K., Paul, S., & Anand, A. (2023). RAGStudentGPT: A syllabus-aligned retrieval-augmented generation framework for educational AI systems. International Journal of Innovative Research and Creative Technology, 9(4).*
5. *Mahajan, A. S. (2025). Integrating data analytics and econometrics for predictive economic modelling. International Journal of Applied Mathematics, 38(2S), 1450–1462.*
6. *Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review. https://doi.org/10.1162/99608f92.8cd550d1*
7. *Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. International Journal of Educational Technology in Higher Education, 16(39). https://doi.org/10.1186/s41239-019-0171-0*

8. *Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. https://doi.org/10.4324/9780429507854*

9. *Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the predictions of any classifier. Proceedings of the ACM SIGKDD Conference. https://doi.org/10.1145/2939672.2939778*

10. *Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. https://doi.org/10.7551/mitpress/10243.001.0001*

11. *Russell, S., & Norvig, P. (2021). Artificial Intelligence: A Modern Approach (4th ed.). Pearson.*

12. *Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., & Martinez-Maldonado, R. (2023). Human-Centred Learning Analytics and AI in Education: A Systematic Literature Review.*

13. *Hooshyar, D., Šír, G., Yang, Y., Kikas, E., Hämäläinen, R., Kärkkäinen, T., Gašević, D., & Azevedo, R. (2025). Towards responsible AI for education: Hybrid human-AI to confront the elephant in the room.*

14. *Wong, J. (2026). Designing trustworthy AI for the youngest learners: An integrated ethical framework and implementation roadmap. Early Childhood Education Journal, 54, 3827–3837. https://doi.org/10.1007/s10643-026-02138-3*

15. *Ahmad, H., & Guzman, S. T. (2025). Designing trustworthy educational artificial intelligence: A systemic framework for explainability, adaptivity, and ethical learning analytics. Qubahan Techno Journal, 4(3), 41–50. https://doi.org/10.48161/qtj.v4n3a68*

16. *Hassan, M., Shaikh, Z., Asif, M., & Afzaal, M. K. (2026). Integrating explainable artificial intelligence into education: Building trustworthy and transparent intelligent learning systems. Review of Education, Administration & Law, 9(1), 259–271. https://doi.org/10.47067/real.v9i1.483*

17. *Abedi, M., Sanusi, I. T., & Tukiainen, M. (2026). Cultivating AI literacy in young learners: A blockchain-enhanced framework for trustworthy and privacy-preserving educational experiences. AI, Brain and Child, 2(3). https://doi.org/10.1007/s44436-026-00029-3*

18. *Yu, X., Yang, S., Wang, Z., Ma, H., & Zhang, X. (2026). Trustworthy intelligent education: A systematic perspective on progress, challenges, and future directions. arXiv.*

19. *Syed, A. (2025). TEAS: Trusted Educational AI Standard: A framework for verifiable, stable, auditable, and pedagogically sound learning systems. OpenReview.*

20. *Ding, S., & Magerko, B. (2025). Rethinking AI evaluation in education: The TEACH-AI framework and benchmark for generative AI assistants. arXiv.*