Journal of Management Practices, Humanities and Social Sciences

Vol 8 Issue 3 pp. 62-74

https://doi.org/10.33152/jmphss-8.3.6

ORIGINAL CONTRIBUTION

The Future of Learning: Building Trust and Transparency in AI Education

Saqib Raza[1 *], Iqra Fatima[2], Shumaila Arif[3], Muhammad Sharif[4], Muhammad Sarim Jalal[5], Zia Muhammad[6]

[1]University of Punjab, Lahore, Pakistan

[2]Department of Information Technology, Government College University, Faisalabad, Pakistan

[3]Faculty of Economics and Business Administration,Vilnius University, Vilnius, Lithuania

[4]Institute of Automation, Chinese Academy of Science, Beijing, China

[5]Higher School of Economics, National Research University, Moscow, Russia

[6]Department of Computer Science, North Dakota State University, Fargo, US

Abstract— Artificial intelligence has become the mark of the day; no one of the aspects of life is beyond it. Education is not only a field but a way to train and make the generation civilized, AI made this scenario easy and advanced; building confidence in AI-enabled education systems becomes a fundamental requirement for its successful application. This research article aims to investigate the connections among trust, privacy, and surveillance related to AI-facilitated educational contexts, drawing on insights from Christina Castelfranchi and Rino Falcone's work "Trust Theory: A Socio-cognitive and Computational Model." This study examines the multiple aspects of trust building, including computational dimensions and socio-cognitive dimensions, which can be helpful but difficult to investigate the attitudes and actions in an AI-mediated educational environment. This article delivers an elaborate exploration of the drivers affecting trust in AI systems deployed for educational purposes by carefully examining trustworthiness, perceived dangers, and institutional trust. Additionally, thorough examination and inherent privacy would be the foundational aspects of privacy and surveillance issues associated with AI detection, protection, and surveillance risk mitigation measures. This scholarly discourse, supported by empirical evidence, illustrative case studies, and prescriptive recommendations, articulates actionable strategies for building trust and addressing privacy and surveillance concerns in AI-enabled educational landscapes, heralding the arrival of ethically sound and trustworthy educational technologies.

Index Terms— Trust, AI-enabled education, Privacy, Surveillance, Trust theory

Received: 21 February 2024 ; **Accepted:** 17 April 2024; **Published:** 28 May 2024

Introduction

AI-enabled education has evolved into an integrative pervasive system of education, fostering trust in all aspects of its operation. This study article investigates the complex interplay of AI-enabled educational aspects such as trust, privacy, and surveillance, as well as monitoring in AI-facilitated educational situations. Christiano Castel Franchi and Rino Falcone represented a theoretical lens titled Trust Theory in the

* Email: socsaqibraza@gmail.com

book "Trust Theory: A Socio-cognitive and Computational Model," which signifies the value of the trustworthiness of AI in the education system. The purpose of this study is to investigate the complex interplay, considering trustworthy attitudes and their implications for privacy policies and surveillance issues in the education system, which elevate the best results from AI technologies.

Artificial intelligence is the productive machine having human intelligence with technologies, particularly computers. It is a powerful catalyst for enhancing and improving the education system. It is worth solving the problems, tailoring new practices, and maximizing the outcomes with advanced recommendations. Artificial intelligence is like a machine, assisting in learning and processing languages with their data analytics by creating a variety of modern ways. Intelligence tutoring system practices like AI algorithms transform the learning experience by augmenting the demands and requirements of a student with cognitive preferences. Concurrently, adaptive learning platforms use AI capabilities to dynamically adjust instructional content and tempo in response to students' changing performance metrics and academic progress.

AI, as a powerful educational tool, not only assists and provides guidance to the learning environment but also supports and instructs the students for the next. Furthermore, it also enlightened me with insightful thoughts on the performance and opportunities for improvement. However, its use can be legitimized because of privacy and surveillance in the education system. Though the AI system can handle and analyze massive or complicated data, it has a concern about security gaps, data privacy, and chances of exploitation, especially personal information. Moreover, if the purpose of usage of AI is surveillance, such as monitoring student behavior and performance, AI can be useful but with careful consideration of safeguard policies and ethical implications to protect the students' rights and autonomy.

Opposite to this environment, fostering trust in AI-enabled education is crucial to guaranteeing its ethical and responsible application. With the help of trust theory, this article recommends new insights for educators, and policymakers on how to manage the complex connection with trust, privacy, and surveillance challenges in the education system. The challenges elevate educational environments, promoting an ecosystem of trustworthy and ethically acceptable technology in the education system.

This study investigates privacy, trust, and surveillance challenges in AI integration in education, aiming to provide actionable insights to foster trust and mitigate privacy risks. Thereby, the hypothesis is to conduct research to analyze an ethically sound AI-enabled educational environment and examine its factors. Much research has been conducted on the topic as this is also the age of technology. However, the investigation on the privacy concerns issues of AI-enabled technology, particularly in the education system, is missing. This paper provides and keeps the focus of policymakers, AI developers, and educators to have ethical considerations as well as the importance of trust in privacy and surveillance concerns.

Objectives

- To explore the multifaceted interplay between trust, privacy, and surveillance concerns within AI-integrated educational environments.
- To examine the determinants influencing trust formation in AI systems deployed for educational purposes, drawing upon insights from trust theory.
- To identify and analyze the inherent privacy and surveillance challenges associated with AI safeguards in educational settings.
- Propose effective solutions for enhancing trust and addressing privacy and surveillance concerns in AI-enabled educational environments.

Research question

- How do trust, privacy, and surveillance intersect in AI-facilitated educational settings, and what factors influence confidence in AI systems used for educational purposes?

Importance of trust in AI systems

The importance of AI machines in building AI-enabled education is integrated to build confidence and ensure the effective and responsible integration of AI technology in the education system. trust is the foundation for creating an environment that overcomes privacy concerns and surveillance issues related to education. It is also essential that to promote a trustworthy environment, educators and policymakers must use a transparent and reliable AI system. Trust is the most important in the use of technical components of AI; it is also an ethical consideration while using AI between developers and consumers (Datarobot, N.d).

This article highlights the trust theory ideas propounded by Franchi and Falcone, which are used in the AI system of education to address confidentiality and surveillance issues. Trust is not child's play to explain and adjust easily to the complex concept of AI-enabled education with a thorough approach from conception to execution, including assuring accountability, regulatory frameworks, and mapping stakeholders' motivations. In an AI-enabled system, trust has three dimensions: operations, performance, and ethics; however, each of them has different parameters that define the concerns (Wasson, 2023).

Furthermore, this article elaborates on the importance of interdisciplinary collaboration in enhancing a trustworthy framework in the AI-enabled education system. On the other hand, education is also an intricate web of interaction between the artificial intelligence web and the users of the regulatory environment. By addressing the stakeholders, taking into account accountability, motivation, future recommendations, and regulatory connections, this article is the bridge to building trust in the use of AI in education. Trustworthiness in the education system is beyond explainability through having fairness with human rights and supremacy of law, with careful considerations throughout the development and disposition phases (Wasson, 2023).

Trust theory: A socio-cognitive and computational model

Franchi and Falcone are the philosophers who represented trust theory as having expressive phenomena of complete paradigm for analyzing trust in AI technologies within the context of "Building Trust in AI-enabled education." This theory integrates socio-cognitive and computational perspectives to describe the dynamics of trust in human-AI interactions, highlighting the significance of transparency, accountability, and dependability in building trust among stakeholders in educational settings (Lukyanenko, Maass, & Storey, 2022). The part consists of socio-cognitive and examines the social and psychological dimensions of trust building in the education system. This concept focuses on how people think, interpret, and give feedback to AI technologies based on their cultural points of view, personal beliefs, expectations, and practices. However, it is also essential to understand the cognitive process, which enables some of the factors to discover new dimensions to influence innovative formation and keep the AI-enabled educational context updated. However, the other part of the computational component addresses the technical features of an AI system, such as bias reduction, explainability, and fairness, which are critical for establishing user confidence (Lukyanenko et al., 2022).

The theory provides a structural framework by applying the features of the theory; this article burns light on insights for educators and policymakers in AI-enabled education for obtaining new designs, implementing the task, and regulating the AI system to improve trust while effectively addressing privacy and surveillance issues. The article aims to develop ethical AI practices, data privacy procedures, and governance mechanisms that assist in building trust among students, educators, and parents. Moreover, trust theory is the formula used to analyze the connection of privacy and surveillance to AI-enabled education systems so that researchers and developers can collaborate on models and future surety. This collaboration is important to consider the ethical usage of AI in the educational environment (Lukyanenko et al., 2022).

Trust theory is significant not only at the socio-cognitive level but also for the computational framework. Learning environment through AI-enabled education is significant and beneficial while understanding and improving the trustworthy concept of AI. The ultimate model of socio-cognitive and computational values has a connection of trust, which emphasizes honesty, responsibility, accountability, and ethics in developing trust among stakeholders in an educational environment (Castelfranchi & Falcone, 2010). Trust needs an understanding of how to be used and what the requirements of stakeholders' motives, protective policy, and regulatory frameworks could be. Trust theory provides a complete methodological framework for creating an educational environment and building trust and reliability by adopting accountable policy, legitimizing processes, and ethical considerations (Castelfranchi & Falcone, 2010; Waheed & Jam, 2010).

Overall, the trust theory focuses on the significance of interdisciplinary collaborations and the creation of communication processes, regulations, and tools for addressing trust in AI-enabled education among stakeholders. The theory contributes to initiatives on legal advice in Europe in the use of AI, providing insights on competency requirements, educational offerings, and regulatory methods required to promote reliable and ethical usage of AI in educational setups. The trust theory provides a solid foundation to comprehend confidence in AI-enabled education while also efficiently revolving up privacy and surveillance issues (Park & Jones-Jang, 2023).

Literature Review

Artificial intelligence has a complex profile through explainable machines by aiming AI systems to be more reliable, understandable, and transparent to define human interpretable resistance, explanations for predictions, and worthy conduct. It reduces the opacity and unaccountability inherent in traditional black-box AI models, boosting trustworthiness and dependability in AI deployment (Raj, Köse, Sakthivel, Nagarajan, & Asirvadam, 2023; Zou & Ratana-Olarn, 2023). AI transparency ensures an explanation of the interplay of AI technology so that people can understand the system and make quick judgments. There are three levels explained by AI transparency: explainability of technical features, governance of components, and stakeholders' communication. Transparency in AI systems fosters trust and allows humans to be educated about decisions. Apart from the personalized decision, it also helps to have commercialized benefits by guaranteeing fast deployment and obvious and clear explanations of the decisions (Hilliard, 2023).

In educational settings, the criteria of trustworthiness is essential for emerging educational technology, such as AI. To ensure accountability, effectiveness, and transparency, educators must have information about the involvement in the development, testing, and application of AI-enabled instructional skills. However, adopting the AI system, it is necessary to get educational research paradigms, and

develop adopting context, increase trust and safety, implement legislation and standards specific to education to address advanced AI capabilities and hazards (Cardona, 2023)

In the educational setting, there is a call to define trustworthiness criteria for emerging educational technology, such as AI. To ensure transparency, accountability, and effectiveness, educators should be informed and involved in the development, testing, and deployment of AI-enabled instructional skills. AI for education research and development should focus on adapting to context, increasing trust and safety, and implementing legislation and standards specific to education to address new AI capabilities and hazards (Ouyang & Jiao, 2021).

Transparency, ethical consideration, and accountability are the crucial components for mitigating the issues and enhancing confidence among stakeholders in educational ecosystems (Wasson, 2023). If we talk about trust and surveillance in the case of AI-enabled education; it has a two-edged sword: who acts as monitors that can provide benefits and opportunities such as cyberbullying; technical injury, and resource allocation, while on the other hand, the personal identity can be dissolved. This sort of monitoring is significant for having a legislative framework for careful handling or control due to the biases; lack of information or consent of the individual; data privacy risk, or moral dilemmas (Pea et al., 2023).

The legislative framework and examination of policy implication should be carefully measured, although it is the most significant for maintaining the integration of surveillance and trust in an AI-enabled education system. The highest priority should be transparency, data privacy, and trust; however, ethical challenges represent the need for all the features integration (Ahmad, Khan, Shah, & Khan, 2022; Pea et al., 2023). Chen (2021), in one of their articles, reported after reviewing the impact of AI-enabled education, they conducted the quantifying data and analyzed AI's application in instructional fields and learning environments.

Thus, by addressing the connection of AI technology to the educational environment, this research work seeks to develop trustworthy settings by maintaining data privacy and surveillance issues, though the promotion of interdisciplinary collaboration, and following the ethical guidelines must also be prioritized for the well-being of the learner.

Methodology

This research study employs a qualitative approach to examine the complex interaction of trust, privacy, and surveillance issues in AI-enabled education, ultimately providing a full knowledge of the phenomenon. The research methodology begins with a thorough examination of the existing literature on trust theory, AI in education, privacy concerns, and surveillance issues. Data gathered for this research work come from textual mining of research articles, documented texts, academic papers, reviews, and policy documents on the topics of trust, privacy, and surveillance in AI-enabled education. Though the topic of this study is unique as not previously exist, however available documents would assist in improving the hypothesis that trust, privacy, and surveillance interact in AI-facilitated educational environments, and what factors influence them.

The documents in textual sources are from sampling approaches using purposeful assessing, with emphasis on scientific publications, and the scholarly sources from academic journals, conference proceedings, and institutional records. The sampling criteria are fixed by ethical considerations to prioritize relevancy, trustworthy concerns, and diversity of perspectives to ensure a thorough examination of the research topic.

Textual data analysis entails a detailed process of coding, categorization, and thematic analysis, however, the textual data is rigorously analyzed to discover themes, patterns, and the theoretical framework of trust theory by Castelfranchi and Falcone (2010), for trust, privacy, and surveillance in AI-enabled education. This research utilizes textual evidence to provide a comprehensive interpretation of AI-enabled education systems, highlighting the complex interplay between trust, privacy, and surveillance concerns, ensuring the validity and reliability of the hypothesis through qualitative paradigm measures.

To strengthen the reliability and validity of the study's conclusions, triangulation processes are utilized, such as cross-referencing many textual sources and involving multiple researchers in data processing. Ethical considerations of accurate and factual citations and sources used are the attributions that are followed strictly through the research work. To maintain ethical standards and safeguard intellectual property rights, plagiarism would be avoided, and the work would have its uniqueness.

Tools, procedures, and data sources used

The tools used in this paper are an extensive and detailed study of literature associated with existing literature and an analysis of the canon related to trust in AI, data privacy concerns, and ethical oversight in educational contexts. The primary data source includes monographs, case studies, scholarly articles, and books, particularly focusing on Christina Castelfranchi and Rino Falcone's "Trust Theory: A Socio-cognitive and Computational Model." Additionally, the data would be qualitative study; however, the sources also used case studies that illustrate the empirical conclusion for the practical application of AI in education. The qualitative analysis of the study emphasized the sources to identify key themes and insights associated with building trust in AI-enabled education.

Discussion

Trust is a fundamental aspect of human interaction, enabling cooperation, collaboration, and building meaningful relationships. Trust theory, particularly the socio-cognitive and computational models, provides insights into the mechanisms behind trust formation and its impact on decisions, especially when we make ourselves vulnerable. The socio-cognitive model focuses on the psychological processes within the person placing trust (the trustor) as they evaluate the person being trusted (the trustee). Trust is seen as a complex puzzle with five pieces: the trustor, the trustee, the context, the desired outcome, and any potential downsides.

Trustworthiness is a multi-layered profile based on the trustor's evaluation of the trustee's skills, intentions, and past behavior. Attribution matters, as the trustor assigns positive attributes to the trustee based on their perceived behavior. The computational model translates the psychological processes explored in the socio-cognitive model into mathematical formulas, allowing for the creation of computer simulations or models that represent trust dynamics in complex systems. However, the combined approach has limitations, such as neglecting emotions and social influences, the complexity of the computational model, and cultural and social contexts.

In one of the research articles, Alissa fosters the trust theory and its applications of socio-cognition with a computation model for considering ethical concerns in building trust and transparency (2010). She explains that Trust theory can guide AI-enabled education by examining student trust in AI tutors. The socio-cognitive model helps design AI systems that address student needs, focusing on perceived capabilities, intentions, and past performance. Quantifying trust through performance metrics, algorithmic transparency, and ethical AI design can unlock the full potential of AI in education.

In the context of AI-enabled education, trust theory is an intricate lens that provides valuable insights into how trust is the foundation for establishing and maintaining confidence between educators and AI technologies. An AI system can change the learning environment by providing transparent, reliable, and ethical considerations. Reliability ensures transparency, data protection, clear operations, accurate information, and avoidance of biases. AI is a software tool that modifies the content, pacing, and assessment capacity to meet the requirements of the educator. It is the confidence of AI to change the environment, and the tool is responsible for designing the priority and updated version of the hypothesis or the question in mind by learning educated environments and better progressive outcomes (Shachar, Gerke, & Adashi, 2020).

The trust theory also defines the same concept as in AI-enabled education; it emphasizes the importance of information and data protection with confidentiality between individuals or institutions for the use of effective use. The theory delves into the account values of fairness, consistency, security, confidentiality, inclusivity, transparency, and responsibility by using AI systems in educational settings. These values foster trust among educators, stakeholders, learners, and even institutions, enhancing the learning environment and allowing it to be a more practical experience. By integrating the given principles, the AI system can be established and implemented with clear responsibilities, policies, and trustworthy concerns (Park & Jones-Jang, 2023).

Factors influencing trust in AI-enabled education

Trust issues in building confidentiality of AI-enabled education are very important in having the influence of the system in the educational environment:

- Performance expectancy is a critical aspect as it has users' confidence and maintains the planning of adoption towards AI-enabled settings. Improved decision-making, increased work efficiency, accuracy, dependability, and less cognitive load are important components. By ensuring customization, automating tedious processes, and producing accurate and dependable assessments, AI systems also analyze large amounts of data, offer insights, and foster trust (Alanzi et al., 2023).
- Effort expectancy and ease of using AI-enabled systems significantly influence users' trust and adoption. It is influenced by user-friendly interfaces, minimal learning curves, robust technical support, task completion efficiency, seamless integration, and adaptive feedback mechanisms. In education, the seamless integration of AI tools can enhance their utility and acceptance (Dahri et al., 2024).
- Trust in AI-enabled education is shaped by social influence, including peer recommendations, collaborative use, instructor endorsement, institutional support, societal norms, and media coverage. Peer recommendations can mitigate apprehensions and foster reliability, while institutional support signals legitimacy. Societal norms and cultural context also affect trust. Strategies to establish trust include promoting success stories, engaging influential figures, facilitating peer learning, and transparently addressing concerns (Venkatesh, Morris, Davis, & Davis, 2003).
- Understanding AI technology is crucial for trust and acceptance in the educational sector. It involves understanding algorithms, decision-making processes, and applications. Transparency and explainability, such as Explainable AI (XAI), are essential. Educational initiatives, clear communication strategies, and stakeholder engagement can enhance comprehension (Cowls, King, Taddeo, & Floridi, 2019)
- Personal innovativeness significantly influences attitudes towards AI systems in educational settings. It involves willingness to

experiment, quick adaptation, curiosity, confidence, perceived usefulness, and role in educational innovation. Highly innovative individuals are open to new experiences, quick to adapt, and perceive AI as useful, fostering trust and facilitating its integration. Their proactive stance drives institutional changes and promotes broader acceptance and innovation (Ajlouni & Rawadieh, 2022).

- Trust is a key factor in the acceptance and adoption of AI-enabled technologies. It reduces perceived risks, increases confidence, facilitates user acceptance, fosters positive attitudes, enhances engagement, and encourages sustained usage. Building trust involves transparency, reliability, user-centric design, and ethical practices. Focusing on these aspects can lead to greater trust in AI, broader adoption, and meaningful integration across various domains (Lankton, McKnight, & Tripp, 2015).
- Building trust in AI systems requires ensuring their safety and explainability. The ASSET Centre focuses on creating methodologies that explain the thinking underlying AI system decisions, hence increasing transparency and trustworthiness (Lerner, 2022).
- Building trust in AI systems requires maintaining their reliability. Reliable systems make sound decisions, respond correctly in uncertain situations, detect errors, and deal with unexpected occurrences effectively. Confidentiality and reliability are important to have trust and surety that AI-enabled education operates protectively and accurately (Bailey, 2023)
- The competency of AI technology is significant for having trust and adoption. It is accurate in terms of exact task performance, reliable information, recommended predictions, and insightful ideas. In an educational environment, accurate guidance, obvious information, predicted assignment grades, learning outcomes, and identity are trustworthy features with the help of AI technology. Trust is further reinforced by well-informed decision-making, effective task management, and perceived value, such as tailored feedback and flexible learning materials (Yin, 2024).
- In AI technology, transparency is crucial for maintaining ethical standards, and educator's trust; however, it is involved in documentary descriptions with precedents, disclosure of processing data, responsible deployment, minimizing risks, and informed decision-making process (Mylrea & Robinson, 2023).
- Explainable artificial intelligence, or XAI, is a method that makes AI systems less opaque, enabling human understanding, validation, and confidence in results produced by AI. It makes AI systems more transparent and reliable by utilizing methods like feature significance evaluation, rules-based explanations, and model-agnostic approaches like LIME and SHAP. Through improved understanding, validation, verification, accountability, and regulatory compliance, XAI fosters confidence. It can be applied to administrative decision-making, personalized learning programs, and student performance prediction in educational contexts (Khosravi et al., 2022).

Perceived risks

Though AI technology helps enhance the personalized learning environment and the learner's experience, it also has privacy and surveillance issues. Transparency, accountability, ethical use of AI technology, stakeholder involvement, and avoidance of bias can mitigate the perceived risks. The risks might have a bad impact, as given below:

- The enormous accumulation of personal data that comes with integrating AI into education creates privacy issues even though it offers benefits like faster procedures and personalized learning. Institutions should acquire user consent, establish strong data protection policies, be open and honest about their data-gathering procedures, ethically design AI systems, and involve stakeholders in the process to foster confidence. This makes the atmosphere around AI technologies more trustworthy (Custers, Calders, Schermer, & Zarsky, 1866).
- Personalized learning is made possible by AI integration in education, yet privacy issues are raised. Data misuse, unauthorized access, profiling, and surveillance are among the risks. To protect students' rights and trust, institutions must prioritize data security and transparency. Openness, informed consent, robust data security procedures, ethical AI development, and stakeholder interaction are crucial elements. Educational institutions may ensure that AI is utilized responsibly while respecting students' rights and fostering confidence by implementing certain measures (Akgun & Greenhow, 2022).
- Although AI-driven learning platforms offer personalized instruction and insights derived from data, they also give rise to concerns about privacy, liberty, and self-governance over tracking. These are the worries while doing profiling, surveillance, possible biases, or relevant information gathering. To foster respectful and responsible AI-enabled education, it is essential to follow liberty, honesty, data minimizing, bias detection, and stakeholder ethical involvement (Zuboff, 2023).
- AI technology provides useful assistance and transformable instructions with monitoring tools as well. These systems also provide confidential data with security protection and better outcomes, although they also bring issues of security and privacy risks. Predictive analysis can alter behavior and limit autonomy. It is important to have in mind the ethical considerations while using the AI system: guaranteeing accountability, justice, and transparency. It must be included in the planning process of stakeholders to reduce the difficulties and preserve users' confidence (Pea et al., 2023).
- Indeed, AI technology has many potential advantages, while data privacy is a problem. To address this, transparency, robust data security protocols, a privacy-by-design approach, and stakeholder education are necessary. This process involves the reduction of

data collecting, fixing default settings, and integrating privacy safeguards into product creation. Possible difficulties might be found with regular accountability while having openness and communication to gain users' trust. It's also essential to make educated decisions and provide parents and kids with tools and information about the benefits and risks of data privacy (Akgun & Greenhow, 2022).

- AI systems in education can potentially undermine teachers' and students' autonomy, leading to biases and social discrimination. They may forecast results based solely on demographics, limiting opportunities and perpetuating inequality. Ethical concerns must be addressed in algorithm development, training, deployment, and monitoring to promote fair educational opportunities (Huang, 2023).

Institutional trust

The relationship between AI developers and educational institutions is complex, involving surveillance and privacy issues, with research on institutional trust being dynamic and influenced by various entities.

- Educational institutions play a crucial role in fostering trust in AI-enabled learning environments, ensuring transparency, data security, and moral standards through data collection, maintenance, and privacy (Nazaretsky, Ariely, Cukurova, & Alexandron, 2022).
- AI developers must keep the value of confidence for signifying educational settings, developing trustworthy environments, build confidentiality and transparency (Ma et al., 2023).
- For establishing and building surveillance and privacy, strict policies should be adopted and implemented in AI-enabled educational programs. Schools should prioritize privacy-by-design methodologies, encrypt data, use secure servers, and ensure data transparency to protect student privacy.
- To build trust in AI applications in education, transparency and accountability are crucial. AI engineers should prioritize open communication about data practices, fostering a culture of trust and ethical use among stakeholders (Nazaretsky et al., 2022).
- Ethical frameworks are crucial for fostering trust in AI education and promoting equity, equality, and privacy protection. Institutions can demonstrate commitment to moral AI application by tying AI activities to ethical principles (Culican, 2023).
- Ethics is also one of the foundational aspects to follow and maintain transparency; involved by AI developers and users. They are all bound to hold ethical considerations for appropriate applications and open development of AI technologies (Thornton, 2020).
- Regulatory structures ensure open, auditable educational information in AI-enabled systems, preventing biases and promoting ethical guidelines through government involvement and collaboration among stakeholders (Lee, 2019).
- AI educational systems may face algorithmic bias, causing unfair outcomes for certain student groups. Solutions include training models, conducting bias audits, preparing teachers, and designing ethical AI systems. (Akter et al., 2021).
- Balancing AI and human interaction in education is crucial to prevent dehumanization and potential teacher replacement. Solutions include developing AI systems with human-centered values and emphasizing social-emotional learning (Balthasar, 2023).

AI techniques in education

Explainable AI technology (XAI) has different approaches that increase trust in AI-enabled educational systems and improve data protection policy, reliability, and trustworthy information. There are some techniques for dealing with the issues raised in the education system, such as recognizing real-time data, creating the AI model for the betterment of policies, and finding flaws in AI technology. Additionally, this XAI system acts as a technique for rule-based systems, decision-making trees, and model-based policies, which are used to provide human-verifiable interpretations, hence increasing transparency and accountability. This is a situation that has been a trust issue, so it is also important to have trust concerns in mind. As said in the theory of 'Trust Theory, ' trust is a significant feature for the formation of a healthy environment between the user and the AI technology. XAI technique is based on the trust system boost for growing new knowledge for educators, as well as legislators (Nazaretsky et al., 2022).

AI technology is available for many advantages, including customized learning environments with security concerns. Apart from the benefits, it also has downsides, such as snooping, improper data use, and wrong profiling. Sometimes, AI technology is not able to detect the requirement domain of the question or has a limit to providing data for security reasons. Furthermore, a strong legislative framework is needed to safeguard students' data privacy while using AI technology in education (Gille, Jobin, & Ienca, 2020).

AI technology while having an XAI system for use, the aim is to highlight legal loopholes and the urgent need for doing legislation. The tool used for this goal is the standardization of AI-enabled Education, promoting educational settings, accountability openness, and high-quality educational pedagogy, with the protection of all the personal entities of individuals (COE, 2023).

The Council of Europe has fixed the legislation on the accountability of the use of AI in Europe, representing the significance and evaluation of AI for encouraging the ethical use of technology. It is not only easy to go with ethics but also to maintain trust in both the

education and economic sectors. It would be beneficial not only for the business but also for the individuals themselves while having trust in decision-making by using detective data, correcting errors, and avoiding biases or malicious behavior. Trust and openness enhance legal and ethical compliance, as well as innovation and collaboration in the refinement of AI-enabled education (Lee, 2019).

It is difficult to have trust and transparency while at the same time ensuring fairness and adherence while using AI in regulations. It not only entails comprehending the decision-making process but also the reasoning behind it, which is important while using data for progressive and instructional purposes. The rules set by the EU and the principles of AI usage technology are focused not only on transparency and accountability but also on ethical considerations for AI developers and the use of AI in educational settings (Wren, 2024).

The educational environment, while using AI technology, requires trust and standards to maintain its value. Educators or developers must educate the learners or the users on how to generate data and reduce management risks. Ethical research must be done a bit with multiple contexts and safety to ensure success (Woolf, 2023). AI technology in getting education raises ethical issues regarding bias, surveillance, and privacy violations. It can be continuous without legal action as would be used discrimination, threatening to the student, establishing safeguards, and ensuring transparency.

- Legal frameworks and international collaboration are being used to control artificial intelligence in education. The development of a dedicated legally binding instrument on the use of AI in education aims to ensure a human rights-based approach and address gaps in existing regulations (COE, 2023).
- Transparency in AI algorithms promotes trust, accountability, and ethical use of technology. It enables an understanding of how AI systems make judgments, detect flaws or biases, and adhere to legal and ethical requirements. Explainable AI approaches, and transparency standards are proposed ways to increase transparency in AI systems (Mylrea & Robinson, 2023).
- The White House issued an Executive Order to ensure the safe, secure, and trustworthy development and deployment of AI. It includes efforts to reduce AI-related intellectual property threats, enhance responsible AI innovation in healthcare, and improve the quality of veterans' healthcare using AI technologies (Biden Jr, 2023).
- XAI makes AI technological systems more reliable and transparent by performing human-readable elaborations for adopting behaviors and predictions. It also focuses on how to explain and express the reasons asked by the user. Its output not only increases the trustworthy information but also provides a responsible protective application (Raj et al., 2023).

Trust-building strategies in AI-enabled education

Explainable Artificial Intelligence (XAI) is an AI technology system that refers to a paradigm in AI-enabled education that aims to make the transparency in AI technology more reliable, visible, and comprehensive. It gives knowledge of graphs and explanatory details, though some AI systems deal with picture and graphic designs; AI protection and prediction promote judgments, trust, and liberty to use. It always favors considering ethical standards such as the European Union Artificial Intelligence Act and the OECD AI principles (Wren, 2024).

Though the Trust issue is a significant foundation for making AI technology more reliable, however, trust in AI-enabled education must be improved by empowering users; ensuring transparency; clearing the data protection policies, and improving the education sector not only for improving data but digital citizenship also. Developers should prioritize transparency, training material, and ethical guidelines to foster trust and better outcomes.

AI technology might be prioritized by social concerns, such as bias, privacy, and transparency. The lack of transparency of trust hinders the development of education. The researcher must keep in mind the development and definition of trust, which can create unified and interdisciplinary collaboration while establishing principles that are also driven by trust theory. The European Union and OECD have set some principles for ensuring trustworthiness and also focus on human autonomy with accountability and explainability but with prevention of harm. These factors are necessary to evaluate the factors of reliability, human interaction, and public acceptance.

AI can transform the educational environment by signifying learning adaptability and feedback. However, the AI-generated material must be adopted and delivered ethically, taking into account privacy concerns and the reliability of technology. Apart from the developers, educators must also understand the responsibility on their part as they must deliver the value of AI-enabled education with the acceptance of accountability and surveillance issues while customizing learning environments and experiences. This learning environment could create future-driven technology and empower the learners as well. There are different options for learning through AI-enabled education, such as multidisciplinary case studies and project-based learning environments, to provide benefits such as comprehension and recommended engagements. AI technology also helps to prepare curriculum guidelines for future generations not only to advocate the application development of AI technology but also to benefit them (Jowallah, 2024).

AI-enabled education is crucial for enabling students to grasp the implications of AI beyond expectations, including an active listening environment, practical exercises, designs, and technical features with ethical conversations. Educators play a vital role in changing the method and mind of the learner while delivering AI-enabled education; their collaboration with AI developers is also significant in this case. The application of AI helps to create a holistic approach that can be adapted to change, minimize biases, and maximize impactful

outcomes. By using impact full collaboration between educators and developers, maximum benefits can be achieved and prepared for the next generation to make it more successful in an AI-driven society (Saputra, Astuti, Sayuti, & Kusumastuti, 2023). Clear guidelines must be part of the protection policy focusing on the characteristics, fairness, and regularly updated policy regulations. With clear policy, proper planning, and ethical use, technology can be more useful in the future for promoting responsible and regular updates.

The rise of AI in education presents opportunities for personalized learning but also raises ethical concerns about data security and student privacy. Current legislation faces challenges like limited transparency, ambiguity, and multinational differences. Building comprehensive frameworks for AI developers, students, and data privacy would increase trust, promote ethical AI development, and create safer learning environments.

Building trust through education

The use of AI technology in the education system is beneficiary, helpful, and also assists in the improvement of ideas and experiences. Though the AI system is a tool for supporting the required information and data protection, however, it is not a remedy to remove the flaws of the education system, so it can also be said that it must be used carefully.

- Some educational programs should train trained students to have information about AI, its applications, usage, and the value of transparency, particularly ethical concerns.
- Trust in AI is essential at every level, from procurement to testing to public acceptance, highlighting the importance of educating stakeholders about the ethical principles and implications of AI.
- Successful implementation of trust-building measures in AI-enabled education, particularly in addressing privacy and surveillance issues, requires a comprehensive approach that involves educators, policymakers, and technology providers (Lukyanenko et al., 2022).
- Researchers and funders should prioritize investigations on how AI can adapt to context, address diversity among learners, and enhance trust and safety in AI-enabled systems for education (Castelfranchi & Falcone, 2010).
- Privacy breaches and surveillance issues concerns stem from the vast amounts of data AI systems require, raising issues like unintentional data exposure, inaccurate or biased AI models, lack of data minimization, legal compliance challenges, and potential loss of rust in educational institutions (Culican, 2023). To solve these issues, schools and educators should increase data literacy, implement data collection strategies, prioritize transparency, and establish education-specific norms and safeguards (Culican, 2023).
- Prioritizing privacy by design, robust data security rules, openness, education on data privacy, and a strong regulatory framework are crucial for integrating AI into education (Datarobot, N.d).

Implication of the study

The study provides various implications which are important for AI-enabled education; emphasizing the importance of trust, ethics, privacy and surveillance.

Policy development

- The study underscores the need of robust affiliation of privacy, and policies that protect students' personal information. Educational institutions and policymakers must develop clear guidance for the procedure of data collection, storage, and usage, ensuring compliance with legal standards and ethical values.
- The research suggests the transparency regulations to inquire AI systems to disclose the data sources, and decision-making processes. This transparency regulations would assist to build trust among the users by allowing them to understand and evaluate AI-enabled systems.

Educational practices

- Educators and students must need to be trained on the use of AI-enabled technologies, including the limitations and the benefits. Training should be focused on developing digital literacy, critical thinking, and ethical considerations as well.
- The findings of consulted privacy and trust in AI-enabled systems advocate for integrating AI literacy into the curriculum, enabling students to engage with AI technologies critically and responsibly. This has the ethical implications of AI on understanding its societal impact.

Ethical frameworks

- The study emphasizes the implications of ethical frameworks to reduce biasness in AI systems. Educators and developers must work together to ensure the design of AI tools and must introduce ways of fairness and equity.
- One of the implications is the emphasized student autonomy; research AI-mechanism; and allow users to consent the data collection and its use. The users must have control over their personal data and the ability to opt out of AI-driven activities if desired.

Stakeholders' collaborations

The study highlights the implicated demands for interdisciplinary collaborations among AI developers, educators, policymakers, and ethicists to create AI systems that are trustworthy and beneficial. Regular monitoring would assist the emerging challenges and align AI technologies with educational purposes.

Future research assessments

- In order to comprehend the long-term implications of AI on the educational system, the study suggests that longitudinal research is necessary. These kinds of studies can shed light on the long-term effects of AI on student engagement, learning results, and ethical considerations.
- Subsequent research results should focus on assessing of AI on different student demographics, particularly those that are underrepresented. Understanding how AI affects a range of learners can help in the development on inclusive and equitable AI technologies.

Ultimately, this research work offers a comprehensive framework for integrating AI into education ethically and responsibly. By addressing transparency, promoting privacy concerns, and fostering interdisciplinary collaborations, stakeholders can build trust in SI-enabled education and ensure their positive impact on teaching as well as learning,

Conclusion

Confidence building in AI-enabled education requires privacy and surveillance issues while ensuring the ethical concerns to responsible incorporation of AI into educational settings. This research delves deeply into the intricate dynamics of trust in AI-enabled education, drawing on Christian Casterfranchi and Rino Falcone's Trust theory. The study emphasizes the critical interplay of reliability, validity, accountability, and transparency in fostering stakeholders' trust. It also highlights the importance of ethical considerations, thorough inspections of honesty, perceived risks, and institutional trust, particularly privacy and surveillance concerns inherent in AI-enabled education.

Furthermore, in order to reduce biasness, establish student autonomy, secure personal information, and ethical frameworks derived from the trust theory must be put into practice. AI-enabled education must be committed to openness, strong data protection regulations, and user involvement in order to promote ethical values. Educators, AI trust builders, and stakeholders are responsible entities for creating an atmosphere of trust by implementing educational technologies that respect individual privacy, liberty, and rights while improving learning outcomes.

We can produce a culture of confidence in AI-enabled education, assuring a productive, and positive impact on learners and educators alike, by making trust of fundamental concept and taking proactive measures to resolve privacy and surveillance issues. This is the strategy which helps to assist the moral application of AI technology in an inclusive, and reliable learning environment.

Future recommendations

After the extensive study on the title "The Future of Learning: Building Trust and Transparency in AI Education", the researchers have some outlines of recommendation for the future researchers.

- Conduct an empirical study across diverse educational settings to validate the theoretical framework applied in this research.
- Establish a monitoring mechanism to check the continuous evaluation of trust-building and ethical adaptation challenged new technical advancements.
- Provide an interdisciplinary collaboration among educators, technical developers, policymakers, and ethicists to develop comprehensive guidelines for the practice of AI in education.

Limitations of the study

There are some limitations of the study: first, the focus of the study emphasizes theoretical frameworks. However, extensive empirical data from diverse educational contexts is not included in it. This would limit the generalizability of the study. Second, there are some recommendations, but as AI technologies and ethical issues are trends nowadays, and because of rapid advancements, they might be outdated soon. Lastly, the study relies on extensive literary studies, which may not fully capture the nuances of the implementation of AI technology across different educational settings.

References

Ahmad, S., Khan, N. U., Shah, M. H., & Khan, M. N. (2022). An analysis of the dynamic linkages of Pakistani stock market with the world's leading stock markets. *Journal of Management Practices, Humanities and Social Sciences*, *6*(5), 22-33.

Ajlouni, A., & Rawadieh, S. (2022). Technophobia and technophilia among undergraduates: Cross-national research in jordan, qatar, and egypt. *Journal of Social Studies Education Research*, *13*(4), 24-55.

Akgun, S., & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics*, *2*(3), 431-440.

Akter, S., Dwivedi, Y. K., Biswas, K., Michael, K., Bandara, R. J., & Sajib, S. (2021). Addressing algorithmic bias in AI-driven customer management. *Journal of Global Information Management*, *29*(6), 1-27.

Alanzi, T., Almahdi, R., Alghanim, D., Almusmili, L., Saleh, A., Alanazi, S., … Alzahrani, H. (2023). Factors affecting the adoption of artificial intelligence-enabled virtual assistants for leukemia self-management. *Cureus*, *15*(11).

Bailey, J. (2023). Ai in education: The leap into a new era of machine intelligence carries risks and challenges, but also plenty of promise. *Education Next*, *23*(4), 28-35.

Balthasar, M. (2023). Balancing strengths and weaknesses in human-machine decision making. India: GI.

Biden Jr, J. R. (2023). *Executive order on the safe, secure, and trustworthy development and use of artificial intelligence.* Retrieved from `https://shorturl.at/nr4ts`

Cardona, R. J. R. . K. I., Miguel A. (2023). *Artificial intelligence and the future of teaching and learning* (Tech. Rep.). United States: Office of Educational Technology.

Castelfranchi, C., & Falcone, R. (2010). *Trust theory: A socio-cognitive and computational model*. Hoboken, NJ: John Wiley & Sons.

Chen, P. . Z. L., Lijia; Chen. (2021). Artificial intelligence in education: A review. *IEEE Access*, *8*, 75264-75278.

COE. (2023). Regulating artificial intelligence in education. In *26th session of the council of europe standing conference of ministers of education, strasbourg, france.*

Cowls, J., King, T., Taddeo, M., & Floridi, L. (2019). Designing ai for social good: Seven essential factors. *Available at SSRN 3388669 (Online First)*, 1-26.

Culican, J. (2023). *Ai in education: privacy concerns and data security: Navigating the complex landscape.* Retrieved from `https://shorturl.at/zju1a`

Custers, B., Calders, T., Schermer, B., & Zarsky, T. (1866). Studies in applied philosophy, epistemology and rational ethics. In (chap. Discrimination and privacy in the information society). Springer.

Dahri, N. A., Yahaya, N., Al-Rahmi, W. M., Vighio, M. S., Alblehai, F., Soomro, R. B., & Shutaleva, A. (2024). Investigating AI-based academic support acceptance and its impact on students' performance in Malaysian and Pakistani higher education institutions. *Education and Information Technologies*, 1-50. doi:doi={10.1007/s10639-024-12599-x}

Datarobot. (N.d). *Introduction to the dimensions of trust.* Retrieved from `https://shorturl.at/ZsiDT`

Gille, F., Jobin, A., & Ienca, M. (2020). What we talk about when we talk about trust: Theory of trust for ai in healthcare. *Intelligence-Based Medicine*, *1*, 100001. doi:https://doi.org/10.1016/j.ibmed.2020.100001

Hilliard, A. (2023). *What is AI transparency?* Retrieved from `https://shorturl.at/l8YYf`

Huang, L. (2023). Ethics of artificial intelligence in education: Student privacy and data protection. *Science Insights Education Frontiers*, *16*(2), 2577-2587.

Jowallah, R. J. (2024). Integrating Artificial Intelligence (AI) Into the curriculum: Empowering the next generation through proactive AI education. In *Generative AI in teaching and learning* (p. 355-368). Hershey, PA: IGI Global.

Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., … Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *3*, 100074.

Lankton, N. K., McKnight, D. H., & Tripp, J. (2015). Technology, humanness, and trust: Rethinking trust in technology. *Journal of the Association for Information Systems*, *16*(10), 1.

Lee, S. (2019). *The ethics and accountability of AI for learning.* Retrieved from `https://shorturl.at/dFyby`

Lerner, E. (2022). *The new asset center will focus on the safety, explainability, and trustworthiness of ai systems.* Retrieved from `https://shorturl.at/VNtxD`

Lukyanenko, R., Maass, W., & Storey, V. C. (2022). Trust in artificial intelligence: From a Foundational Trust Framework to emerging research opportunities. *Electronic Markets*, *32*(4), 1993-2020.

Ma, C., Li, J., Wei, K., Liu, B., Ding, M., Yuan, L., … Poor, H. V. (2023). Trusted AI in multiagent systems: An overview of privacy and security for distributed learning. *Proceedings of the IEEE*, *111*(9), 1097-1132.

Mylrea, M., & Robinson, N. (2023). Artificial Intelligence (AI) trust framework and maturity model: Applying an entropy lens to improve security, privacy, and ethical AI. *Entropy*, *25*(10), 1429.

Nazaretsky, T., Ariely, M., Cukurova, M., & Alexandron, G. (2022). Teachers' trust in AI-powered educational technology and a professional development program to improve it. *British journal of educational technology*, *53*(4), 914-931.

Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence*, *2*, 100020.

Park, Y. J., & Jones-Jang, S. M. (2023). Surveillance, security, and AI as technological acceptance. *AI & society*, *38*(6), 2667-2678.

Pea, R. D., Biernacki, P., Bigman, M., Boles, K., Coelho, R., Docherty, V., … others (2023). Learning: Designing the future. In (chap. Four surveillance technologies creating challenges for education). New York, NY: Springer.

Raj, P., Köse, U., Sakthivel, U., Nagarajan, S., & Asirvadam, V. S. (2023). *Explainable Artificial Intelligence (XAI): Concepts, enabling tools, technologies and applications*. Savoy, London: Institution of Engineering & Technology.

Saputra, I., Astuti, M. D., Sayuti, M., & Kusumastuti, D. P. (2023). Integration of artificial intelligence in education: Opportunities, challenges, threats and obstacles. A literature review. *Indonesian Journal of Computer Science*, *12*(4), 1590-1600. Retrieved from `https://api.semanticscholar.org/CorpusID:261413770`

Shachar, C., Gerke, S., & Adashi, E. Y. (2020). AI surveillance during pandemics: Ethical implementation imperatives. *Hastings Center Report*, *50*(3), 18-21.

Thornton, L. (2020). *Artificial intelligence and ethical accountability.* Retrieved from `https://shorturl.at/yETrl`

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, *27*(3), 425-478.

Waheed, M., & Jam, F. A. (2010). Teacher's intention to accept online education: Extended TAM model. *Interdisciplinary Journal of Contemporary Research in Business*, *2*(5), 330-344.

Wasson, B. (2023). *Artificial intelligence in education: Layers of trust (trust AI).* Retrieved from `https://shorturl.at/DnOLs`

Woolf, I. A. A. L. . M. G. . C. D., Beverly. (2023). *Data directed education: The future of AI in education* (Tech. Rep.). US: NSF Convergence Accelerator Ideation Workshop.

Wren, H. (2024). *What is ai transparency? a comprehensive guide.* Retrieved from `https://shorturl.at/x563w`

Yin, J., William. (2024). *Will our educational system keep pace with ai? a student's perspective on ai and learning.* Retrieved from `https://shorturl.at/Ih6qa`

Zou, X., & Ratana-Olarn, T. (2023). Empowering preservice teachers: A mentored blended learning approach with MOOCs to elevate skills and self-efficacy. *Pakistan Journal of Life & Social Sciences*, *21*(2), 1727-4915.

Zuboff, S. (2023). The age of surveillance capitalism. In *Social theory re-wired* (p. 203-213). England, UK: Routledge.