



Building Trust in an Artificial Intelligence-Based Educational Support System: A Narrative Review

Membangun Kepercayaan terhadap Sistem Pendukung Pendidikan berbasis Kecerdasan Buatan: Sebuah Review Naratif

Anisa Herdiani¹, Dimitri Mahayana², Yusep Rosmansyah³

School of Computing, Telkom University¹

School of Electrical Engineering and Informatics, Bandung Institute of Technology^{2,3}

anisaherdiani@telkomuniversity.ac.id

ARTICLE INFO

Keywords:

artificial intelligence, education,
narrative review, support system,
trust

ABSTRACT

A primary challenge associated with the implementation of educational support systems is the establishment of student trust in the systems themselves. Trust is a critical factor in the acceptance and use of AI-enabled systems, as it reduces uncertainty and the perception of risk associated with new technology adoption. A literature review of existing studies on trust in AI-based systems is needed to provide a solid foundation for future studies. This research aims to identify gaps in the literature regarding the establishment of user trust in AI-based educational systems by exploring the criteria of trust and the challenges of building trust in AI systems. A narrative review of the literature is conducted to synthesize the findings of selected articles, covering (1) fundamental principles of trust and the process of establishing trust in non-human entities; (2) technical issues relating to explainable AI; (3) the utilization of explainable AI to facilitate decision-making; and (4) the use of AI systems in facilitating educational activities and its influence. This article summarizes trust criteria, including reliance, transparency, affectiveness, integrity, consistency, fairness, accountability, security, and usability. Building trust in AI systems involves addressing technical, ethical, and societal challenges to ensure the responsible and beneficial use of AI for individuals and society.

INFO ARTIKEL

Kata kunci:

kecerdasan buatan, kepercayaan,
pendidikan, review naratif, sistem
pendukung

ABSTRAK

Tantangan utama penerapan sistem pendukung pendidikan adalah membangun kepercayaan siswa terhadap sistem tersebut. Kepercayaan merupakan faktor penting dalam penerimaan dan penggunaan sistem yang mendukung AI, karena mengurangi ketidakpastian dan persepsi risiko yang terkait dengan adopsi teknologi baru. Sebuah tinjauan literatur terhadap penelitian-penelitian yang sudah ada mengenai kepercayaan terhadap sistem berbasis AI diperlukan untuk memberikan dasar yang kuat untuk penelitian-penelitian selanjutnya. Penelitian ini bertujuan untuk mengidentifikasi kesenjangan dalam literatur mengenai pembentukan kepercayaan pengguna pada sistem pendidikan berbasis AI, dengan mengeksplorasi kriteria kepercayaan dan tantangan dalam membangun kepercayaan pada sistem AI. Tinjauan naratif literatur dilakukan untuk mensintesis temuan dari artikel-artikel yang dipilih yang meliputi (1) prinsip-prinsip dasar kepercayaan dan proses membangun kepercayaan pada entitas

non-manusia; (2) isu-isu teknis yang berkaitan dengan AI yang dapat dijelaskan; (3) pemanfaatan AI yang dapat dijelaskan untuk memfasilitasi pengambilan keputusan; (4) Penggunaan sistem AI dalam memfasilitasi kegiatan pendidikan dan pengaruhnya. Peneliti merangkum kriteria kepercayaan yang meliputi kebergantungan, transparansi, afektif, integritas, konsistensi, keadilan, akuntabilitas, keamanan, dan kegunaan. Proses membangun kepercayaan dalam sistem AI melibatkan penanganan tantangan teknis, etika, dan sosial untuk memastikan penggunaan AI yang bertanggung jawab dan bermanfaat bagi individu dan masyarakat.

<https://doi.org/10.5614/sostek.itbj.2023.23.1.6>

Submitted: December 19, 2023 Accepted: March 28, 2024 Published: March 30, 2024

Introduction

Since the Intelligent Tutoring System (ITS) was introduced in the 1980s, artificial intelligence (AI) has continued to evolve to assist learning (Baker et al., 2017). ITS serves to identify student learning progress so that it can provide help that is tailored to student needs (Baker et al., 2017). The use of ITS was initially intended to complement traditional classroom learning activities. However, as more learning processes are carried out outside the system, the effectiveness of the recommendations given becomes questionable (Kärner et al., 2021). AI-based educational supporting systems then evolved with the increasingly widespread use of digital learning or e-learning using learning management systems (LMS), especially in open and distance education (ODE) (Al-Shaikhli, 2023; Chen & Cui, 2020; Hamid et al., 2022; Juhaňák et al., 2019; Karapiperis et al., 2023; Müller & Wulf, 2020; Qi et al., 2023; Sáiz-Manzanares et al., 2021; Xin & Singh, 2021). With all learning activities based on digital, a wealth of data on student behaviors and learning progress becomes an invaluable asset. The more complete the data managed, the more effective an AI-based educational support system can be. The system can optimize the learning process, making tailored recommendations for learning strategies based on the circumstances and needs of the student.

The implementation of an AI-based educational support system faces numerous obstacles. The students may face difficulties in implementing the given recommendations due to various reasons, such as time constraints caused by other activities like work or household chores, especially in open and distance education (Bağriacık Yılmaz & Karataş, 2022). Other obstacles may arise from limited access to necessary devices, such as the inability to access learning materials due to a lack of software to run a program or technical issues like the internet or power outages (Bağriacık Yılmaz & Karataş, 2022). This is beyond the capacity of an educational support system.

Another problem may come from the AI system itself. Unclear or incomplete information regarding why learners receive certain feedback or recommendations can cause learner distrust (Jin et al., 2023). Trust is a critical factor in the acceptance and use of AI-enabled systems by users (Nordheim et al., 2019). This is because trust can contribute to the reduction of uncertainty and the perception of risk associated with the adoption of new technology (Yang & Wibowo, 2022). Trust may also have an impact on the behavioral intentions of users in the use of AI-enabled systems (Wei et al., 2018). It can be concluded that establishing trust in AI systems poses a significant challenge.

While a growing body of empirical research seeks to understand the adoption of AI and its relationship to trust and organizational practice, a literature review of existing studies on the topic from different perspectives is still lacking. A comprehensive review of the relevant literature is needed to summarize previous research and provide a solid foundation for future studies, as AI applications are becoming more prevalent. This research aims to identify gaps in the literature regarding the establishment of user trust in AI-based systems, particularly in the education sector, by formulating the following research questions:

1. What are the criteria for trust in AI systems?
2. What are the current challenges of building trust in AI-based educational support systems?

Method

Trust in AI-based educational support systems is a complex topic that requires transparency and clarity of data or information. However, there are still many aspects related to trust that are widely studied. To determine the state-of-the-art in the development of trust and its impact, the study was conducted using the narrative review method. This method involves conducting various studies to conclude different interpretations and criticisms. It is suitable for obtaining state-of-the-art research with a wider scope and from various points of view (Sukhera, 2022). The narrative review was conducted in four stages: (1) defining the scope, boundaries, and definitions; (2) justifying the inclusion and exclusion criteria; (3) explaining the position of the review; and (4) analyzing and interpreting the results (Sukhera, 2022).

Coverage: Scope, Boundaries, and Definitions

Scope. This review focuses primarily on philosophical, technical, and educational aspects to ensure a comprehensive and relevant perspective. Philosophical viewpoints are discussed to determine the position of AI system development in the field of education based on the development of science. This view allows us to anticipate likely scientific developments soon. The technical point of view is discussed to find out the extent to which AI systems are developed to support learning activities. Efforts to enhance the reliability of their use are also considered. Meanwhile, the ethical point of view is discussed to find out the extent of the ethical impact of using AI systems for learners.

Boundaries. This study is limited to the utilization of AI systems to support the learning process in open and distance education environments at the higher education level. This is determined because educational activities that allow all activities to be carried out online, with a mature target age, are in higher education.

Definitions. The key terms related to the topic and research question are defined as follows:

Table I Key Terms

Key Terms	Definitions
Artificial Intelligence	the ability of a computer or computer-controlled robot to perform tasks that are commonly associated with the intellectual processes characteristic of humans, such as the ability to reason (Encyclopaedia Britannica).
Criterion (plural criteria)	a standard or principle by which something is judged or with the help of which a decision is made (Oxford Learner's Dictionaries)
Explainable Artificial Intelligence	a set of processes and methods that allow human users to comprehend and trust the results and output created by machine learning algorithms (IBM)
Learning Analytics	the measurement, collection, analysis, and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs (Lee et al., 2020)
Machine Learning	machine learning is a specialized branch of artificial intelligence that focuses on enhancing performance through the utilization of experience. Certain AI systems employ machine learning techniques to attain proficiency, whereas others do not (Russell & Norvig, 2021)
Trust	the belief that something is true or correct or that you can rely on it (Oxford Learner's Dictionaries)

Inclusion and Exclusion Criteria

To provide clear guidelines regarding the selection of articles, the inclusion and exclusion criteria were formulated, as can be seen in Table II.

Table II Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
Related to trust issues of AI systems and the effort of building trust in AI systems	Not related to trust issues of AI systems and the effort of building trust in AI systems
Systematic review or multi-case-study article	Non-systematic-review or single-case article
Publish date: 2018-2023	Publish date: 2018-2023
Describes deep explanation, analysis, limitation, and future research on the research field	Did not describe deep explanation, analysis, limitation, and future research on the research field
Published in Q1 or Q2 Journal	Published in non-Q1 or Q2 Journal

To obtain a comprehensive understanding of each topic of research, we integrate systematic reviews and multi-case study articles. Concerning the publication year, we limited our research to the past five years, focusing solely on the most recent advancements in current research. The strategy of selecting Q1 and Q2 journals is essential to ensuring that the articles demonstrate a high level of quality.

Review Position

This study will focus on literature reviews or multiple case study articles to provide a comprehensive understanding of how learners develop trust in AI-supported learning support systems. A minimum of three supporting articles will be provided for each area, with the expectation that the articles will provide different perspectives or be complementary to other articles. Selected papers are shown in Table III.

Table III Selected Papers on Narrative Review

Code	Paper	Year	Journal	SJR (2022)	Research Field	Keywords
[R1]	Supporting students' self-regulated learning in online learning using artificial intelligence applications (Jin et al., 2023b)	2023	International Journal of Educational Technology in Higher Education	2.05 – Q1	Educational Technology	Self-regulated learning, Artificial intelligence, Online learning, Student perception
[R2]	Making Trust Safe for AI? Non-agential Trust as a Conceptual Engineering Problem (Viehoff, 2023)	2023	Philosophy & Technology	1.3 – Q1	Philosophy	Trust, conceptual engineering, reliance, distrust, technology, AI
[R3]	The perils and pitfalls of explainable AI strategies for explaining algorithmic decision-making (de Bruijn et al., 2022)	2022	Government Information Quarterly	2.32 – Q1	Artificial Intelligence	Artificial intelligence, XAI Algorithms, Computational intelligence, Data-driven decision, Socio-tech, Transparency, Accountability, Trust, E-government
[R4]	Explainability of artificial intelligence methods, applications and challenges: A comprehensive survey (Ding et al., 2022)	2022	Information Sciences	2.29 – Q1	Artificial Intelligence	Black-box White-box, Explainable AI, Responsible AI, Machine learning, Deep learning

[R5]	Trust in artificial intelligence: From a Foundational Trust Framework to emerging research opportunities (Lukyanenko et al., 2022)	2022	Electronic Markets	1.55 – Q1	Philosophy	Artificial intelligence (AI) , Trust, Foundational Trust Framework, Trust in AI, Explainable AI, Transparency, Systems
[R6]	Talking AI into Being: The Narratives and Imaginaries of National AI Strategies and Their Performative Politics (Bareis & Katzenbach, 2022)	2022	Science, Technology, & Human Values	1.17 – Q1	Philosophy	artificial intelligence, sociotechnical imaginaries, governance, discourse analysis, international comparison
[R7]	Utilising learning analytics to support study success in higher education: a systematic review (Ifenthaler & Yau, 2020)	2020	Educational Technology Research and Development	1.52 – Q1	Educational Technology	Study success, Dropout, Retention, Attrition, Higher education, Learning Analytics
[R8]	Explainable AI: A Review of Machine Learning Interpretability Methods (Linardatos et al., 2020)	2020	Entropy	0.54 – Q2	Artificial Intelligence	xAI, machine learning, explainability, interpretability, fairness, sensitivity, black-box
[R9]	Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead (Rudin, 2019)	2019	Nature Machine Intelligence	6.21 – Q1	Artificial Intelligence	Explain, interpretable
[R10]	A Survey of Methods for Explaining Black Box Models (Guidotti et al., 2018)	2018	ACM Computing Surveys	4.46 – Q1	Artificial Intelligence	Open the black box, explanations, interpretability, transparent models

Results and Discussion

Studies associated with scientific development revolve around the theoretical frameworks proposed by philosophers Thomas Kuhn and Imre Lakatos. In his book 'The Structure of Scientific Revolutions,' Thomas Kuhn introduced the concept of normal science as the regular work of scientists within a settled paradigm or explanatory framework (Politi, 2019). He described normal science as 'puzzle-solving,' in which scientists develop theories, make observations, and systematically conduct experiments to gradually gather specific information in line with established overarching theories.

Preliminary Analysis: The Philosophical View of Evolution of Educational Support Systems

The concept of normal science is characterized by scientists who are involved in three primary activities: elucidating the underlying paradigm, precisely assessing fundamental paradigmatic facts, and conducting empirical evaluations to evaluate novel aspects that challenge the theoretical paradigm (Kuhn & Hacking, 2012). This idea is an essential component of Kuhn's comprehensive theory of scientific progress and shifts in paradigms.

Kuhn articulates the iterative process of scientific advancement, encompassing the subsequent stages: (1) The pre-science phase encompasses the preliminary investigation of a specific problem domain and precedes the Kuhn cycle. During this time, scientists are unable to achieve substantial advancements. (2) Normal Science: In the realm of scientific inquiry, researchers enhance their comprehension of a certain domain by employing well-established theories and methodologies. Normal science is characterized by the accumulation of knowledge and the development of current theories. (3) Anomaly: An anomaly arises when scientists start to scrutinize the prevailing paradigm and its inherent constraints. Scientists may see inconsistencies or irregularities that cannot be accounted for by the existing model, which could result in a change in their professional obligations. (4) A 'model crisis' refers to a situation where the current paradigm is unable to adequately explain the observed data. The existing paradigm may experience a decline in confidence as a result of this catastrophe. Additionally, it presents a potential avenue for the emergence of a novel paradigm. (5) Model Revolution: The last phase of the Kuhn cycle involves the implementation of a novel paradigm that effectively tackles the constraints inherent in the preceding model. The findings possess the capacity to propel the field forward and usher in a novel era of conventional scientific practice, predicated upon the emerging paradigm. Kuhn's theory explains the cyclical nature of scientific development, which is characterized by alternating phases of normal science and revolution phases (Kuhn & Hacking, 2012).

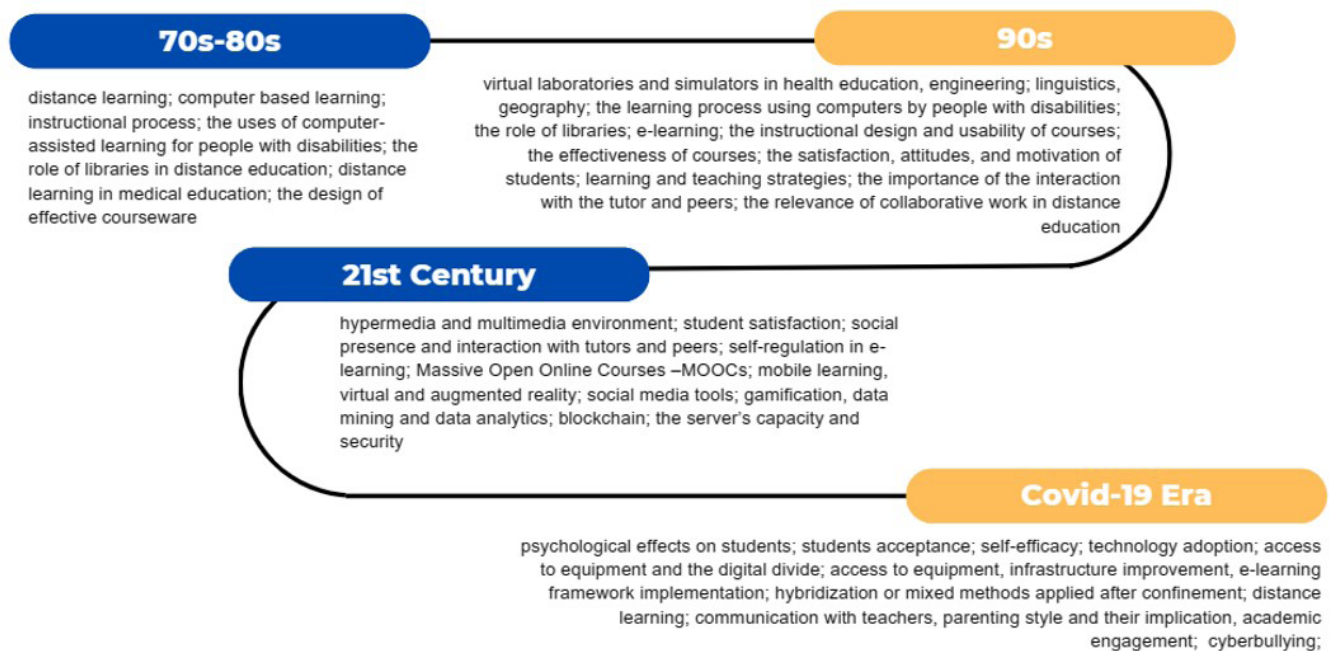


Figure 1 The Evolution of Research in E-Learning based on (Martinez-Garcia et al., 2023)

To explain the position of scientific progress in the development of an AI-based educational support system, we will first present the evolution of e-learning over time, as shown in Figure 1. The bibliometric study conducted on e-learning research spanning the years 1970 to 2022 reveals a number of significant themes and patterns that exemplify the progression of this subject. At first, the research primarily concentrated on 'distance learning', instructional methods, assisted learning for individuals with disabilities, the involvement of libraries in distance education, and courseware development (Martinez-Garcia et al., 2023). However, by the 1990s, the scope of research broadened to encompass the efficacy of courses, the contentment and drive of students and teachers, learning techniques, and interactions with tutors and peers (Martinez-Garcia et al., 2023). The emergence of the 21st century and the subsequent development of the 'world wide web' have given rise to key areas of research, namely multimedia, hypermedia, and online environments. During this period, there was a notable focus on the pedagogical approach to e-learning (Renshaw & Taylor, 2000).

Over the past ten years, there have been notable advancements in the use of portable devices, social media, augmented/virtual reality, and particularly, the adoption of e-learning during the COVID-19 epidemic, which have been the focus of extensive research (Martinez-Garcia et al., 2023). This shift signifies an increasing inclination towards exploring how technology might augment educational experiences, as well as the imperative to adjust to abrupt transformations in the educational domain, such as those instigated by the pandemic (Boonroungrut et al., 2022; Rodrigues et al., 2019). These trends underscore the expansion of the discipline beyond conventional distance learning to encompass technical improvements and the incorporation of digital tools into the learning process. This reflects the ever-evolving character of e-learning research and its adaptability to technological and societal shifts.

Within Martinez's compilation of articles, there is a notable absence of research on artificial intelligence. Nevertheless, artificial intelligence (AI) has been extensively employed to facilitate the process of learning, as shown by intelligent tutoring systems. Hence, we searched the Scopus database to identify relevant works that were published by the end of the year 2023. The search was conducted solely by inputting keywords, without any limitations on other variables. The keywords are (a) e-learning (125,353 documents), (b) open and distance education (6,074 documents), (c) intelligent tutoring systems (7,432 documents), and (d) learning analytics (41,264 documents). Figure 2 displays the graphs representing the search results.

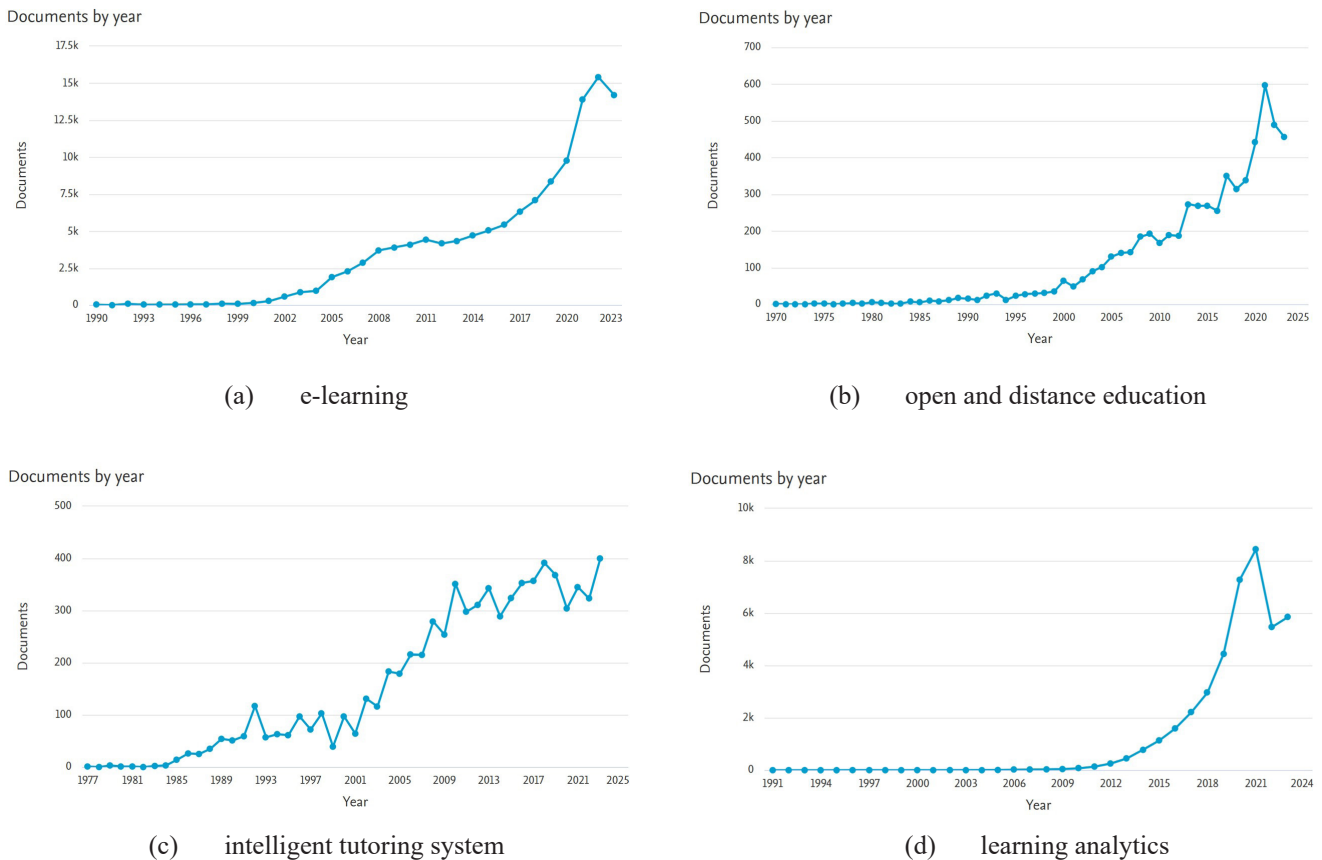


Figure 2 Graph of the number of published documents based on search results in the Scopus database with different keywords.

Source: Scopus Data Analytics

The data depicted in Figure 2 indicates that the domain of intelligent tutoring systems and learning analytics is now seeing continuous progress, aligning with the developments observed in e-learning and open and distance education research. The study reached its highest point during the COVID-19 pandemic in 2020 and 2021, and there is still potential for more advancement in the future with the escalating advancements and utilization of artificial intelligence in education.

Intelligent tutoring systems and learning analytics are frequently employed technologies to enhance the educational experience through the integration of e-learning. Research on Intelligent Tutoring Systems (ITS) has been ongoing since the 1970s to enhance student learning, before the advent of the Internet and e-learning. The number of studies conducted in this field has progressively grown over time, resulting in several advancements aimed at enhancing the efficacy of the learning process to achieve the desired learning outcomes and skills (Ifenthaler & Yau, 2020; Jin et al., 2023b; Martinez-Garcia et al., 2023). The integration of learning analytics with Intelligent Tutoring Systems (ITS) facilitates the collection of student learning behavior data, enabling the system to offer tailored learning strategies that align with the individual student's specific circumstances and requirements.

Rapid developments in artificial intelligence have led to the emergence of highly intelligent assistive devices that can answer any question a student might ask. The issue that subsequently emerges is the presence of uncertainties regarding the responses provided by an AI-based system (Lukyanenko et al., 2022; Yang & Wibowo, 2022). Is the offered response valid? The viability of AI development is called into doubt considering these factors. The research transitioned from focusing on the advancement of AI in education to addressing the issue of ensuring education is protected from any adverse effects caused by AI. This period can be characterized as the shift from a normal science to an anomaly.

Building trust in AI-based educational support systems is then increasingly scrutinized, including by building explainable and interpretable AI (Bareis & Katzenbach, 2022; de Bruijn et al., 2022; Ding et al., 2022; Guidotti et al., 2018; Khosravi et al., 2022; Linardatos et al., 2020; Rudin, 2019). This signifies a transition to the subsequent stage of scientific advancement, commonly referred to as the revolution phase.

Main Findings

The articles present varying viewpoints on the AI system. Table IV presents the main findings of each reviewed article, providing a clear understanding of the background.

Table IV Main Findings

Code	Authors	Main Findings
[R1]	(Jin et al., 2023b)	AI applications were found to be beneficial in facilitating metacognitive, cognitive, and behavioural control in self-regulated learning, but were not viewed as effective in regulating motivation by learners. The research provides practical insights for developing AI applications in online education to facilitate students' self-regulated learning. Moreover, learners adeptly employed artificial intelligence (AI) tools specifically developed to facilitate self-regulated learning processes.
[R2]	(Viehoff, 2023)	The study examines the controversy surrounding the expansion of the trust concept to include non-human agents and artefacts. It utilizes insights from conceptual engineering to investigate how both agential and non-agential explanations might satisfy the functional requirements established for trust. It also explores the influence of unfavourable situations, when there is a difference between the way a notion is expressed and how it is used, on the decision of which interpretation of trust should be preferred.
[R3]	(de Bruijn et al., 2022)	The article examines the difficulties associated with incorporating explainable artificial intelligence (XAI) into government decision-making procedures. It suggests targeted approaches to surmount these obstacles and attain societal approval of AI-driven decisions.
[R4]	(Ding et al., 2022)	The paper presents a comprehensive classification system for classifying XAI studies and offers valuable perspectives on unresolved research inquiries and potential avenues for furthering research in XAI.

[R5]	(Lukyanenko et al., 2022)	The primary outcomes encompass the creation of the Foundational Trust Framework, which serves as a basis for trust investigation, specifically in the field of AI. Additionally, a research agenda has been established to promote empirical, theoretical, and design studies on trust in AI.
[R6]	(Bareis & Katzenbach, 2022)	The primary conclusions of the study indicate a consistent narrative framework throughout national AI policies that presents AI as an unavoidable and transformative technological advancement. Additionally, the study emphasizes the influential role of political rhetoric in molding public conversations and facilitating political engagement.
[R7]	(Ifenthaler & Yau, 2020)	The success of studying can be influenced by a variety of factors, including individual dispositions and the qualities of the educational environment. There are a multitude of learning analytics methodologies that effectively facilitate academic achievement and identify students who are in danger of attrition. There is a lack of rigorous empirical evidence regarding the effective utilization of learning analytics in facilitating and enhancing students' learning and achievement in higher education.
[R8]	(Linardatos et al., 2020)	The research examines the influence of current advancements in artificial intelligence on the adoption of AI in industries and its exceptional performance. The paper offers a comprehensive literature analysis and taxonomy of interpretability methods in the field of machine learning, serving as a valuable reference. The text explores the concept of interpretability and explainability being used interchangeably.
[R9]	(Rudin, 2019)	The utilization of black box machine learning models is resulting in challenges when making critical decisions in diverse industries. The proposal of developing models that possess intrinsic interpretability is put forward as a potential resolution to the challenges arising from black box models. This research highlights the significance of developing interpretable models from the beginning to prevent the recurrence of undesirable practices.
[R10]	(Guidotti et al., 2018)	The primary outcomes of the study encompass the significance of description in the field of data science, the necessity for categorization of issues about explanations in academic literature, and the diverse viewpoints that different scientific communities adopt when addressing the matter of explaining machine-learning decision models.

Fundamental principles of trust and the process of establishing trust in non-human entities are explored by Lukyanenko et al. (2022) and Viehoff (2023). The process of building an AI system to be trusted by humans is highly intricate due to the multitude of real-world circumstances that the AI system must be able to recognize. This study is necessary to ensure that the system developer considers not just the system's performance but also the users' perceptions of the system's suggestions or feedback and its overall impact. Conceptual engineering provides a framework for critically evaluating and potentially reshaping our understanding of trust in AI systems, encouraging a nuanced and purposeful approach to conceptual choices in this evolving domain.

The Foundational Trust Framework, based on systems theory, provides a comprehensive foundation for trust research in artificial intelligence (AI). It posits trust as a process within and between systems, utilizing systems thinking and general systems theory to understand the dynamics of trust in human-AI systems. Key insights from the framework are: (1) trust development varies by AI system type; (2) trust is purpose dependent; (3) trust is influenced by dispositional, cultural, and psychological factors; (4) personality traits and predispositions can influence trust in technology; (5) the presence of structural assurances, such as legal recourse, guarantees, and certifications, positively influences trust in e-vendors and, by extension, can be applied to AI systems. Incorporating human oversight into AI systems, particularly in sensitive domains like healthcare, can enhance trust by validating decisions and ensuring the AI's confidence.

In another way, technical issues are raised by Ding et al. (2022); Guidotti et al. (2018); Linardatos et al. (2020); and Rudin (2019). They are considering the approach to achieving explainable AI, one of the fundamental concepts that enables users to have trust in AI systems. The system user needs to comprehend what is behind each of the feedback rationally provided by the AI system. Further research is necessary to comprehend the mechanisms that establish trust in AI systems. Explainable AI (XAI) addresses the ethical challenges posed by the lack of explanation in AI decisions by promoting transparency, interpretability, and trust. XAI techniques aim to make AI models transparent and understandable, enhance the interpretability of decisions, generate human-comprehensible explanations, build trust by providing justifiable outcomes, and facilitate collaboration between humans and AI systems. By promoting transparency, interpretability, trust, and collaboration, XAI plays a crucial role in addressing ethical concerns related to AI decision-making. However, the relationship between transparency, trust, and acceptance of AI outcomes is influenced by a number of circumstances, and the necessary level of trust is not always achieved.

Another crucial aspect is the utilization of explainable AI to facilitate decision-making, which includes the support of governments as policymakers, as pointed out by (Bareis & Katzenbach, 2022; de Bruijn et al., 2022). The government must acknowledge the reality that the advancement of artificial intelligence (AI) systems is no longer disrupted. Governments must consider the varying levels of comprehension and intellectual capacity among citizens. Consequently, it is unwise to delegate the responsibility of addressing the consequences of AI system utilization solely to the citizen. The primary function of governments is to establish regulations that facilitate the advancement of AI systems, ensuring their positive impact on the public while safeguarding them from potential negative consequences.

Understanding emergent behavior in socio-technical systems is crucial for designing adaptive interventions to enhance the performance of AI systems in decision-making processes. This is due to (1) complex interactions: socio-technical systems involve complex interactions between humans, technology, and structures, leading to unpredictable behaviors; and (2) unintended consequences: Emerging behaviors can undermine AI systems' effectiveness in decision-making. Designers can anticipate negative outcomes and design adaptive interventions to mitigate risks, ensuring positive socio-technical system contributions. (3) Dynamic Environments: AI systems can adapt to changing socio-technical systems by understanding emergent behavior and enhancing their effectiveness and relevance in decision-making processes over time. (4) Human-AI collaboration: Designing AI systems that effectively collaborate with humans requires understanding emergent behavior, recognizing how technology and human behavior interact, and creating systems that complement human decision-making.

The use of AI systems in facilitating educational activities and their influence are demonstrated by Ifenthaler & Yau (2020) and Jin et al. (2023b). The authors highlighted several applications of artificial intelligence (AI) systems in facilitating learning processes, particularly in the domains of metacognitive, cognitive, and self-regulated learning capabilities. AI systems typically offer learning analytics to help students achieve academic goals and identify students facing difficulties in the learning process. However, the available empirical information regarding the efficacy of employing an artificial intelligence system remains insufficient.

The lack of statistical information regarding the long-term effects of AI systems can be attributed to the limited duration of experimentation. Undoubtedly, we all envision students receiving beneficial results. Nevertheless, it is plausible that adverse consequences may arise, such as students depending on assistance facilities. The designer of AI systems can adapt the concept of scaffolding to gradually reduce the aid given so that students can finally independently regulate their learning process to achieve their goals. In another case, students do not take advantage of the assistance offered by the system due to insufficient trust in the system compared to their faith in their teachers. Students prefer human support for motivational regulation due to trust in relationships with instructors. Understanding mechanisms to build trust in AI systems and developing strategies for motivation in online learning environments is crucial to addressing AI's challenge.

Criteria for Trust in AI Systems

To establish students' confidence in AI-based educational support systems, it is imperative to ascertain the specific characteristics that define trust. As these conditions are progressively satisfied, the likelihood of students placing trust in the AI system increases. Table V outlines the diverse criteria for trust.

Table V Trust Criteria

Criteria	Description	References
Reliance	Trust should be defined clearly and objectively, distinguishing it from other related concepts in the field.	[R2][R5]
Rationality/ Explanatory Power/ Transparency	Any discussion of trust must satisfy an explanatory requirement, in that it must provide, or at least be capable of convincingly explaining, how trust may be rational for individuals and, consequently, how trust may be responsible for social cooperation.	[R2] [R3] [R5]
Affective/Emotional Responses	An account of trust should be able to explain the emotional response that occurs when trust is broken, and it is important to distinguish this response from mere disappointment because it involves a sense of betrayal.	[R2][R5]
Distrust	Explanations of trust should also address its relationship to mistrust; if our explanation cannot account for the tripartite conceptual structure that divides our practice into trust, mistrust, and non-trust, then it would be incomplete, covering only one aspect of a practice that requires holistic understanding.	[R2]
Integrity/honesty/ adherence to ethical norms	The system should uphold ethical principles and ensure honesty and transparency in its actions and communications.	[R5]
Ability	The system must have the necessary capabilities, skills, and knowledge to perform the functions for which it is intended effectively.	[R5]
Consistency/ Predictability/ Monotonicity	The system continues to behave and perform stably and predictably.	[R5][R10]
Fairness	The system provides fair and unbiased treatment to all users and stakeholders.	[R5]
Accountability	The system takes responsibility for its actions and decisions and stands ready to answer for the results.	[R5]
Security	The system ensures that confidential information is protected. It also provides a secure environment for interactions.	[R5]
Usability/Mutual Benefit	Models that allow interaction and querying are more trusted than those that only provide fixed explanations in text form.	[R10][R5]

Trust is widely recognized as a psychological process for minimizing uncertainty and increasing the possibility of an accomplishment (e.g., secure, pleasurable, convenient) connection with entities in the environment. We allocate fewer cognitive, physiological, and financial resources when interacting with someone we trust. Humans gained advantages from trust gradually, and it is suggested that trust is an essential requirement for all human relationships (Yamagishi, 2011). Human trust in AI is a human mental and physical process that considers the characteristics of a particular AI, a class of AIs, or other AIs that they are embedded in or interacting with to control the extent and parameters of the interaction with those AIs (Lukyanenko et al., 2022).

With the rise of artificial intelligence, the question of trust in this technology is emerging as a matter of paramount concern for society. Many ethical and existential questions are being raised, and fear and anxiety are being generated by applications such as AI-based surgery and medical diagnosis, driverless cars, prison and parole, automated job application screening, wealth investment, and AI-based

military weapons. Many avant-garde scientists (e.g., Stephen Hawking) and business leaders (e.g., Elon Musk and Bill Gates) see major threats to society from advanced AI solutions (Bostrom, 2017).

Building student trust in AI is important. First, it concerns students' willingness to use the system. If students are unable to utilize the system due to their lack of confidence in its dependability, how may they derive advantages from the learning support system? Furthermore, it pertains to the cognitive workload. Establishing a strong level of trust in the system can effectively decrease the cognitive burden associated with its use. This enables students to concentrate on their educational journey without being diverted by the inquiry of whether they ought to adhere to the suggestions provided by the system. Thirdly, it is imperative to cultivate trust through the provision of transparent information, enabling students to possess a sense of assurance regarding the veracity of the recommendations.

To build student confidence, the AI system needs to ensure that it meets the criteria for being trustworthy. Compliance with technical criteria is an essential need for an AI system. The primary requirement to possess is ability. It is mandatory to ensure that the system can deliver sensible and objective responses. This objective can be accomplished by engaging in a comprehensive research process to explore the diverse range of responses encompassed by the system. The inclusion of experts is crucial throughout all stages of system development. The level of expertise required for system development increases proportionally with the criticality of the study field. Before the system is utilized by the end user, it is imperative to evaluate the system, utilizing the diverse factors that are necessary. The government ought to establish criteria for the utilization of AI systems to facilitate their accessibility to the general populace. This is also related to the accountability of the AI systems.

Consistency is the subsequent requirement. The AI system must possess the capability to deliver a conclusive response. For a given question or condition, the system must provide an identical response. It is also associated with its closest predictable attribute. When the AI system produces disparate responses to identical conditions, it might be inferred that the system lacks reliability, hence posing challenges in establishing user confidence. Hence, it is important to do extensive screening to guarantee the uniformity of the AI system. The concept of consistency is intricately linked to the fairness criteria, which necessitates that the system furnish an impartial response to every user. Testing the fairness criterion is more intricate than testing consistency due to the need for the system to consider multiple aspects to get a response. When there is a minor disparity in characteristics between the user or input data, it is possible for the subsequent answers to exhibit variations as well.

Transparency is a crucial technical criterion. The system should possess the capability to effectively explain the process by which an answer is generated. The rationality of the explanation is imperative. However, the AI system must provide a meticulous explanation. Excessive complexity in explanations or intricate visualizations can pose difficulties for individuals in comprehending the provided solutions. To handle this situation, an AI system must possess a well-defined criterion for assessing the intricacy of the problem at hand, as not all solutions necessitate elaborate explanations. This criterion is closely related to integrity; the system should uphold ethical principles and ensure honesty in its actions and communications.

The trade-off of security is a common occurrence in AI systems. Increasing the amount of data sent to the AI system enhances the system's expertise, resulting in more precise solutions. However, the data handled by this system is susceptible to being exploited by other systems for financial gain. Therefore, the AI system must acquire authorization from individuals to utilize their data to generate solutions. Conversely, AI systems must guarantee that the data gathered is exclusively intended to deliver solutions in domains that have been mutually agreed upon by the user. Avoid venturing to outsiders.

The Challenges of Building Trust

Establishing trust in an AI system by satisfying the criteria is a challenging task. The obstacles to be encountered are shown in Table VI.

Table VI The Challenges of Building Trust

Challenge	Reference
Human Connection: Building trust in AI systems requires understanding mechanisms and addressing learners' preferences for human support. To overcome this challenge, strategies promoting motivation in online learning environments and fostering trust in AI systems are essential. Trust in AI systems should not only focus on performance and reliability but also on human connection.	[R1] [R4]
Building trust in AI systems requires efforts to increase both transparency and system openness . By making AI systems more understandable and their operations more accessible, these approaches can help mitigate concerns about complexity, opacity, and potential misuse, thereby fostering a more trusting relationship between AI systems and their users and stakeholders.	[R5]
Interpretability and Explainability: The "black box" nature of many advanced AI models, where the decision-making process is not transparent, poses a significant challenge to trust. There is a growing need for interpretable and explainable AI (XAI) that can provide understandable explanations for its decisions, making it easier to identify and correct flaws in models and data. The system must also consider trade-offs between explainability and efficiency, as complicated explanations will cause the user to leave the system.	[R3][R4] [R5] [R8] [R9]
Mitigating Bias and Ensuring Fairness: AI systems can inadvertently learn and perpetuate biases present in their training data, leading to unfair or discriminatory outcomes. Addressing these biases and ensuring fairness in AI decisions is crucial for building trust, especially in applications with significant moral and ethical considerations.	[R3][R5] [R8]
Domain-Specific Interpretability: Interpretability varies across domains, necessitating domain-specific definitions and solutions. For example, computer vision may not relate to pixel sparsity, necessitating tailored methods.	[R3][R4] [R5] [R9]
Computational Complexity: Interpretable models often require constrained optimization problems, which are more computationally intensive due to the need to incorporate application-specific constraints, making the problem harder to solve.	[R3][R5] [R9]
Requirement for Domain Expertise: Interpretable models require extensive domain knowledge to define interpretability and engineer suitable features, ensuring relevance and accuracy within the domain.	[R3][R5] [R9]
Intellectual Property and Profitability Concerns: Companies may resist adopting interpretable models due to concerns about losing proprietary advantages and profits, while black box models offer a competitive edge and revenue through prediction services.	[R4][R9]
Belief in the Superiority of Black Box Models: Black box models are often believed to uncover hidden patterns in data, but evidence suggests that interpretable models can effectively identify and leverage these patterns.	[R9]
Comprehending and Validating Machine-Learning Components: As artificial intelligence systems, especially those including machine learning, grow more intricate, it becomes increasingly challenging for firms and individuals to grasp and authenticate the fundamental reasoning behind these components. The absence of comprehension can have a significant effect on ethics, responsibility, security, and legal responsibility, particularly in safety-sensitive sectors such as driverless vehicles and personalized medicine.	[R4][R9]
Safety and Reliability: It is of utmost importance to guarantee the safety and dependability of AI systems, especially in situations where incorrect judgments could result in harm or loss of life. It is essential to develop transparent machine-learning technologies that enable the evaluation and comprehension of AI judgments to establish trust in these systems. The system must also consider its vulnerability to adversarial attacks.	[R4][R10]

The relationship between AI systems and trust is complex and evolving, influenced by many factors, including technological advances, societal perceptions, and ethical considerations. The following are our conclusions about the key aspects of this relationship:

Challenges in Trusting AI: Due to factors such as algorithmic bias, lack of transparency in decision-making processes, and concerns about accountability and control, trusting AI systems can be challenging. To ensure that AI systems are trustworthy and reliable, clear guidelines, regulations, and ethical frameworks are necessary.

Trustworthiness of AI: Issues related to fairness, accountability, transparency, and ethics (FATE) must be addressed to ensure the trustworthiness of AI systems. Trustworthy AI frameworks prioritize ethical considerations, minimize biases, and increase the transparency of AI algorithms and decision-making processes.

Human-AI Interaction: Trust in AI systems is affected by the quality of how humans and AI interact. The design of AI systems that effectively communicate their capabilities, limitations, and decision-making processes can help build trust among users. The establishment of clear channels for feedback, recourse, and human oversight can be a powerful driver of trust in AI technologies.

Conceptual Challenges: Because traditional notions of trust do not easily apply to non-agent entities, extending the concept of trust to AI systems presents conceptual challenges. To address this issue, conceptual engineering provides a framework for re-evaluating and refining the concept of trust in the context of AI technologies. This framework considers factors such as reliance, rationality, affective responses, and the distinction between trust and distrust.

To summarize, the relationship between trust and AI systems is multifaceted, requiring a nuanced understanding of trust dynamics, ethical considerations, and the evolving landscape of AI technologies. Building trust in AI systems involves addressing technical, ethical, and societal challenges to ensure the responsible and beneficial use of AI for individuals and society.

Direction For Future Research

In this section, we discuss directions for research on the development of artificial intelligence-based educational support systems. The discussion includes (1) the ontology, epistemology, and axiology of the research, and (2) strategies for avoiding pseudoscience.

The Ontology and Epistemology

Research regarding AI-based educational supporting systems uses and contributes to the knowledge domains of integrated systems engineering, artificial intelligence (AI), platform-based development, social issues and professional practice, user experience design, and systems analysis and design. In addition to the computer science knowledge domain, this research also utilizes and contributes to the knowledge domain of education, specifically educational technology and learning theory. In particular, the learning theory used is related to self-regulated learning, heutagogy, Bloom's taxonomy, and scaffolding. Overall, the research ontology that illustrates the relationship of this research to the fields of computing and education can be seen in Figure 3.

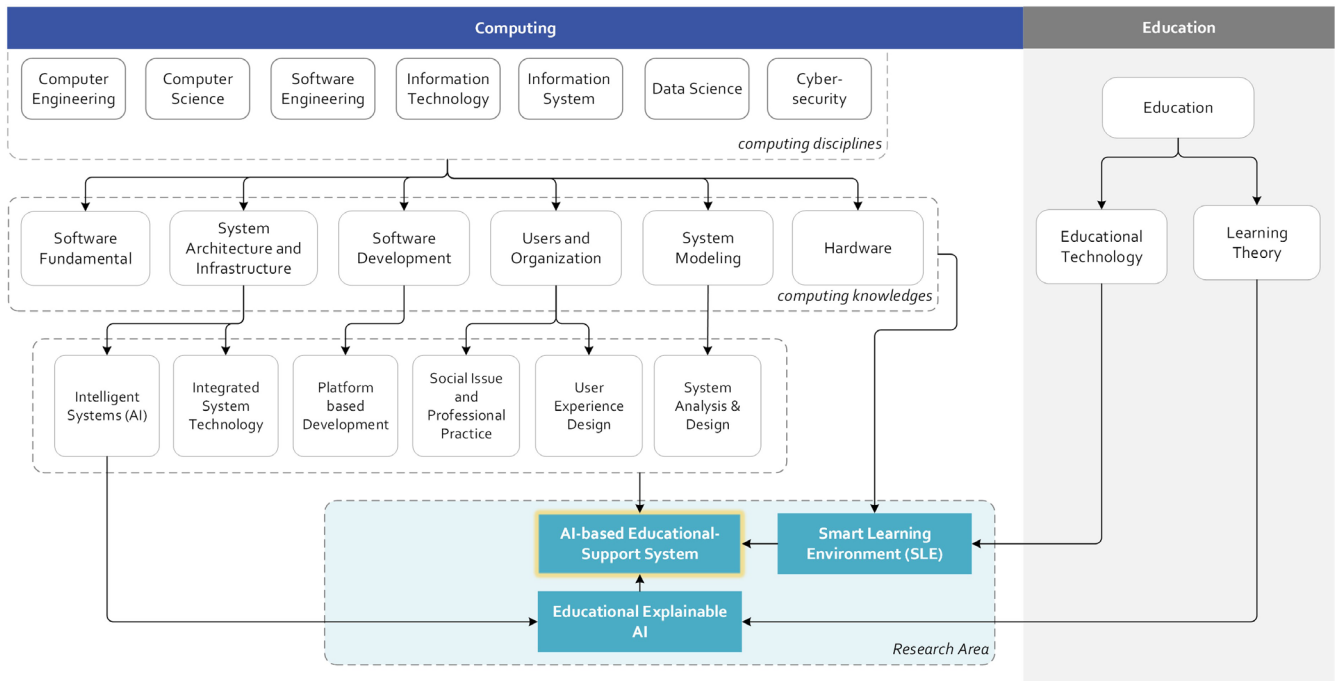


Figure 3 The Ontology and Epistemology of AI-Based Educational Supporting System

The Axiology

Axiology in the context of this research refers to smart learning environments, with the following axiological aspects of the research: (1) **Research ethics:** At each stage of research planning, implementation, and reporting, protocols must be prepared to address issues of privacy, security, and protection of students' personal information. (2) **Goals and Values:** Research supports the achievement of positive educational goals and values such as inclusion, equity, and individual development. (3) **Humanity and social impact:** Research considers the social and humanitarian impact of the technology being developed or tested. This includes questions about how these technologies may affect students, teachers, and society more broadly. (4) **Transparency and Accountability:** Research includes procedures that ensure transparency in methodology, reporting, and data use. Procedures include public access and oversight to ensure accountability. (5) **Impartiality:** Research is impartial and does not favor any group or interest. Ensure that the research is neutral and objective. (6) **Benefit and welfare of students:** The research prioritizes the benefit and well-being of students by improving their learning experience and achievement.

Strategies to Avoid Pseudoscience

Since research related to vulnerable artificial intelligence is trapped in pseudo-science, we need to define clear barriers to the conduct of research. Per the pseudo-science demarcation criteria by Mario Bunge (Bunge, 1984), the strategies in Table VII can be applied.

Table VII Strategies to Avoid Pseudoscience

Criteria	Strategies
Cognitive Community	Engage in scientific meetings, undergo specialized training related to science.
Society	Join the science group of technology-enhanced learning and regularly renew the science.
Domain	Study the landscape of knowledge in computing and education through scientific articles
General Outlook	Identify ontology, epistemology, axiology for every new concept learned.
Formal background	Testing using mathematical methods, if possible.
Specific background	Collect theories, hypotheses, and up-to-date data, and confirm the truth.
Problematic	Identify the problem, whether it's cognitive or practical.
Fund of Knowledge	Testing theories, hypotheses, and current data.
Aims	Ensure that the purpose of the research is for cognitive purposes, not practical.
Method	Ensure that the employed methodology has undergone thorough verification and validation to establish the reliability and validity of the generated output.

Limitations

This analysis encompasses a restricted selection of academic work. While endeavoring to encompass philosophical, technological, and educational viewpoints, there remain numerous additional facets that can be explored to attain a more complete comprehension.

Conclusion

Trust is a psychological process that minimizes uncertainty and increases the possibility of a secure, pleasurable, and convenient connection with entities in the environment. With the rise of artificial intelligence, trust in AI is becoming a paramount concern for society, with ethical and existential questions being raised and fear and anxiety generated by applications. Building student trust in AI is important for students' willingness to use the system, reducing cognitive workload, and providing transparent information.

To build student trust, AI systems must meet technical criteria, including ability, consistency, transparency, and fairness. Ability involves delivering sensible and objective responses, while consistency involves providing a stable and predictable response to a given question or condition. Fairness requires the system to provide an impartial response to every user, and transparency involves explaining the process of an answer while maintaining integrity and honesty in actions and communications.

Trust in AI systems is a complex process that requires consideration of both system performance and user perception. The Foundational Trust Framework provides a comprehensive foundation for trust research in AI, highlighting key insights such as how trust development varies by system type, purpose-dependent, being influenced by dispositional, cultural, and psychological factors, and the presence of structural assurances. Incorporating human oversight into AI systems, particularly in sensitive domains like healthcare, can enhance trust by validating decisions and ensuring the AI's confidence.

Explainable AI (XAI) addresses ethical challenges related to AI decision-making by promoting transparency, interpretability, and trust. However, the relationship between transparency, trust, and acceptance of AI outcomes is influenced by various circumstances. Governments play a crucial role in facilitating AI use but must consider the varying levels of comprehension and intellectual capacity among citizens.

Understanding emergent behavior in socio-technical systems is essential for designing adaptive interventions to enhance AI systems' performance in decision-making processes. Designers can anticipate negative outcomes and design adaptive interventions to mitigate risks, ensuring positive socio-technical system contributions. Dynamic environments and human-AI collaboration are also crucial for designing AI systems that effectively collaborate with humans.

Conceptual challenges arise as traditional trust dynamics do not apply to AI systems. Building trust requires addressing technical, ethical, and societal challenges to ensure responsible AI use for individuals and society.

Acknowledgement

We would like to express our gratitude to the School of Electrical Engineering and Informatics, Institut Teknologi Bandung, for organizing a Philosophy of Science course that significantly contributed to strengthening the fundamental aspects of our doctoral research.

References

- Al-Shaikhli, D. (2023). The effect of the tracking technology on students' perceptions of their continuing intention to use a learning management system. *Education and Information Technologies*, 28(1), 343–371. <https://doi.org/10.1007/s10639-022-11156-8>
- Bağrıacık Yılmaz, A., & Karataş, S. (2022). Why do open and distance education students drop out? Views from various stakeholders. *International Journal of Educational Technology in Higher Education*, 19(1), 28. <https://doi.org/10.1186/s41239-022-00333-x>
- Baker, R., Aleven, V., Blomberg, N., Andres-Bray, M., Sewall, J., Wang, Y., & Popescu, O. (2017, January). *Integrating MOOCs and Intelligent Tutoring Systems: edX, GIFT, and CTAT*.
- Bareis, J., & Katzenbach, C. (2022). Talking AI into being: the narratives and imaginaries of national AI strategies and their performative politics. *Science, Technology, & Human Values*, 47(5), 855–881. <https://doi.org/10.1177/01622439211030007>
- Boonroungrut, C., Saroinsong, W. P., & Thamdee, N., (2022). Research on students in COVID-19 pandemic outbreaks: A bibliometric network analysis. *International Journal of Instruction*, 15(1), 457–472. <https://doi.org/10.29333/iji.2022.15126a>
- Bostrom, N. (2017). *Superintelligence: Paths, dangers, strategies* (Reprinted with corrections 2017). Oxford University Press.
- Bunge, M. (1984). *Pseudoscience can be clearly distinguished from science only if a number of features are checked*.
- Chen, F., & Cui, Y. (2020). Utilizing student time series behaviour in learning management systems for early prediction of course performance. *Journal of Learning Analytics*, 7(2), 1–17. <https://doi.org/10.18608/jla.2020.72.1>
- de Bruijn, H., Warnier, M., & Janssen, M. (2022). The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision-making. *Government Information Quarterly*, 39(2), 101666. <https://doi.org/10.1016/j.giq.2021.101666>
- Ding, W., Abdel-Basset, M., Hawash, H., & Ali, A. M. (2022). Explainability of artificial intelligence methods, applications and challenges: A comprehensive survey. *Information Sciences*, 615, 238–292. <https://doi.org/10.1016/j.ins.2022.10.013>
- Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys*, 51(5), 93:1-93:42. <https://doi.org/10.1145/3236009>
- Hamid, S., Nizam Ismail, S., Hamzah, M., & W. Malik, A. (2022). Developing engagement in the learning management system supported by learning analytics. *Computer Systems Science and Engineering*, 42(1), 335–350. <https://doi.org/10.32604/csse.2022.021927>
- Ifenthaler, D., & Yau, J. Y.-K. (2020). Utilising learning analytics to support study success in higher education: A systematic review. *Educational Technology Research and Development*, 68(4), 1961–1990. <https://doi.org/10.1007/s11423-020-09788-z>

- Jin, S.-H., Im, K., Yoo, M., Roll, I., & Seo, K. (2023a). Supporting students' self-regulated learning in online learning using artificial intelligence applications. *International Journal of Educational Technology in Higher Education*, 20(1), 37. <https://doi.org/10.1186/s41239-023-00406-5>
- Jin, S.-H., Im, K., Yoo, M., Roll, I., & Seo, K. (2023b). Supporting students' self-regulated learning in online learning using artificial intelligence applications. *International Journal of Educational Technology in Higher Education*, 20(1), 37. <https://doi.org/10.1186/s41239-023-00406-5>
- Juhaňák, L., Zounek, J., & Rohlíková, L. (2019). Using process mining to analyze students' quiz-taking behavior patterns in a learning management system. *Computers in Human Behavior*, 92, 496–506. <https://doi.org/10.1016/j.chb.2017.12.015>
- Karapiperis, D., Tzafilkou, K., Tsoni, R., Feretzakis, G., & Verykios, V. S. (2023). A probabilistic approach to modeling students' interactions in a learning management system for facilitating distance learning. *Information*, 14(8), 440. <https://doi.org/10.3390/info14080440>
- Kärner, T., Warwas, J., & Schumann, S. (2021). A learning analytics approach to address heterogeneity in the classroom: the teachers' diagnostic support system. *Technology, Knowledge and Learning*, 26(1), 31–52. <https://doi.org/10.1007/s10758-020-09448-4>
- Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 3, 100074. <https://doi.org/10.1016/j.caeai.2022.100074>
- Kuhn, T. S., & Hacking, I. (2012). *The structure of scientific revolutions* (4th ed). University of Chicago press.
- Lee, L.-K., Cheung, S. K. S., & Kwok, L.-F. (2020). Learning analytics: Current trends and innovative practices. *Journal of Computers in Education*, 7(1), 1–6. <https://doi.org/10.1007/s40692-020-00155-8>
- Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable AI: A review of machine learning interpretability methods. *Entropy*, 23(1), 18. <https://doi.org/10.3390/e23010018>
- 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. <https://doi.org/10.1007/s12525-022-00605-4>
- Martinez-Garcia, A., Horrach-Rosselló, P., & Mulet-Forteza, C. (2023). Evolution and current state of research into E-learning. *Heliyon*, 9(10), e21016. <https://doi.org/10.1016/j.heliyon.2023.e21016>
- Müller, F. A., & Wulf, T. (2020). Technology-supported management education: A systematic review of antecedents of learning effectiveness. *International Journal of Educational Technology in Higher Education*, 17(1), 47. <https://doi.org/10.1186/s41239-020-00226-x>
- Nordheim, C. B., Følstad, A., & Bjørkli, C. A. (2019). An initial model of trust in chatbots for customer service—findings from a questionnaire study. *Interacting with Computers*, 31(3), 317–335. <https://doi.org/10.1093/iwc/iwz022>
- Politi, V. (2019). Thomas Kuhn, The American philosopher. *Metascience*, 28(1), 69–72. <https://doi.org/10.1007/s11016-018-0348-5>
- Qi, X., Sun, G., & Yue, L. (2023). Applying self-optimised feedback to a learning management system for facilitating personalised learning activities on massive open online courses. *Sustainability*, 15(16), 12562. <https://doi.org/10.3390/su151612562>
- Renshaw, C. E., & Taylor, H. A. (2000). The educational effectiveness of computer-based instruction. *Computers & Geosciences*, 26(6), 677–682. [https://doi.org/10.1016/S0098-3004\(99\)00103-X](https://doi.org/10.1016/S0098-3004(99)00103-X)
- Rodrigues, H., Almeida, F., Figueiredo, V., & Lopes, S. L. (2019). Tracking e-learning through published papers: A systematic review. *Computers & Education*, 136, 87–98. <https://doi.org/10.1016/j.compedu.2019.03.007>
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), Article 5. <https://doi.org/10.1038/s42256-019-0048-x>

- Russell, S. J., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (Fourth edition). Pearson.
- Sáiz-Manzanares, M. C., Rodríguez-Díez, J. J., Díez-Pastor, J. F., Rodríguez-Arribas, S., Marticorena-Sánchez, R., & Ji, Y. P. (2021). Monitoring of student learning in learning management systems: an application of educational data mining techniques. *Applied Sciences*, 11(6), 2677. <https://doi.org/10.3390/app11062677>
- Sukhera, J. (2022). Narrative Reviews: flexible, rigorous, and practical. *Journal of Graduate Medical Education*, 14(4), 414–417. <https://doi.org/10.4300/JGME-D-22-00480.1>
- Viehoff, J. (2023). Making trust safe for AI? Non-agential trust as a conceptual engineering problem. *Philosophy & Technology*, 36(4), 64. <https://doi.org/10.1007/s13347-023-00664-1>
- Wei, K., Li, Y., Zha, Y., & Ma, J. (2018). Trust, risk and transaction intention in consumer-to-consumer e-marketplaces: An empirical comparison between buyers' and sellers' perspectives. *Industrial Management & Data Systems*, 119(2), 331–350. <https://doi.org/10.1108/IMDS-10-2017-0489>
- Xin, O. K., & Singh, D. (2021). Development of learning analytics dashboard based on moodle learning management system. *International Journal of Advanced Computer Science and Applications*, 12(7). <https://doi.org/10.14569/IJACSA.2021.0120793>
- Yamagishi, T. (2011). *Trust: The Evolutionary Game of Mind and Society*. Springer. <https://doi.org/10.1007/978-4-431-53936-0>
- Yang, R., & Wibowo, S. (2022). User trust in artificial intelligence: A comprehensive conceptual framework. *Electronic Markets*, 32(4), 2053–2077. <https://doi.org/10.1007/s12525-022-00592-6>