

An Empirical Study on Trust, Ethical Concerns, and Bias in AI-Powered Educational Chatbots

Sudha Valan¹; Dr. Vikas Kumar²

¹Student, Department of Computer Science & IT, Chhatrapati Shivaji Maharaj University, Mumbai, India.

²HoD, Department of Computer Science & IT, Chhatrapati Shivaji Maharaj University, Mumbai, India.

Publication Date: 2026/05/15

Abstract: In contemporary ages, Artificial Intelligence has swiftly changed all the paces of education. Tools like ChatGPT and other AI chatbots are being extensively used for providing support to students in learning. The increasing usage and involvement of AI-powered chatbots in educational environments has elevated crucial concerns pertaining to trust, ethical behavior, and expected bias in automated communications. When the earlier research has greatly concentrated on theoretical discussions and isolated assessments, it has been observed that the comprehensive empirical studies amalgamating user perception with experimental validation was lacking. With a view to address this gap, a mixed-method approach involving both systematic literature analysis and primary data collection was applied in this study. A total of 134 research papers were reviewed, of which 29 were selected for in-depth investigation to find key dimensions of trust, ethics, and bias in educational chatbots. Building on these insights, an investigational study was performed with a sample of $n = 120$ students. They were divided into two groups. One group interacted with a custom-developed AI chatbot, while another group was trained using traditional teaching methods such as providing them with the notes and structured assessments. A standardized survey instrument was dispensed to both groups to evaluate perceived trust, ethical reliability, and bias in learning interactions.

The results showed that chatbot-assisted learning enhanced perceived accessibility and engagement, with trust levels increasing by approximately 18.4% compared to the traditional group. Similarly, ethical perception improved by approximately 13%, reflecting better alignment with fairness, transparency, and accountability principles. However, the analysis of bias-related parameters revealed an approximately 8% gap in neutrality, suggesting the necessity for additional refinement in reducing bias and ensuring equitable responses. Statistical analysis using a two-sample t-test confirmed that the observed differences were significant ($p < 0.05$). This study contributes by offering an incorporated experiential context for assessing AI-powered educational chatbots and suggests hands-on understandings for developing more trustworthy, ethical, and unbiased AI systems in education.

Keywords: AI Chatbot; Ethics & Trust; Bias; Automated Communications.

How to Cite: Sudha Valan; Dr. Vikas Kumar (2026) An Empirical Study on Trust, Ethical Concerns, and Bias in AI-Powered Educational Chatbots. *International Journal of Innovative Science and Research Technology*, 11(4), 4658-4670. <https://doi.org/10.38124/ijisrt/26apr2388>

I. INTRODUCTION

Artificial Intelligence is increasingly becoming an integral part of human life, closely intertwined with daily activities. The hasty progression of artificial intelligence (AI) has considerably changed the education scenario, enabling the development of intelligent systems such as AI-powered chatbots that support teaching and learning processes. These systems propose individualized support, real-time responses, and enhanced accessibility, thereby improving student engagement and learning efficiency (Holmes et al., 2019; Zawacki-Richter et al., 2019). In current ages, educational organizations have increasingly adopted chatbot technologies to supplement traditional instructional methods and provide scalable learning support (Wollny et al., 2021). The escalation of AI-enabled chatbots appear as a hopeful key to

tackle some of these issues excellently with the educational scenario going through a process of constant advancement (Labadze, L., Grigolia, M. & Machaidze, L., 2023). It has hence become a pressing necessity to understand the concept, history and studies conducted on chatbots so far.

➤ What is a Chatbot?

AI Chatbots are web-oriented applications that gets well adapted to the activities of the tutors and learners. Similarly, they are generally used across various fields especially in the field of education (Labadze, L., Grigolia, M. & Machaidze, L., 2023).

Following table provides the list of AI chatbots commonly used for educational purposes.

Table 1 Common AI Chatbots used in Education

Year	Chatbot	Source	Used for
2023	Bard	(Rudolph, 2023)	<ul style="list-style-type: none"> • Research assistance • Producing different kinds of contents • Provide responses to queries
2022	ChatGPT	(Dergaa, 2023; Khademi, 2023; Rudolph, 2023)	<ul style="list-style-type: none"> • Generate text • Produced creative content • Delivered enlightening replies to the queries raised by the learners
2019	IBM Watson Assistant	(Oliveira et al., 2019)	<ul style="list-style-type: none"> • FAQs • Student Support Chatbot
2017	Ada	(Kabiljo et al., 2020; Konecki et al., 2023)	<ul style="list-style-type: none"> • Provide coaching to individual students • Promote personalized education to students • Deliver replies and feedback
2017	Replika	(Pentina et al., 2023; Xie & Pentina, 2022)	<ul style="list-style-type: none"> • Acted as a colleague and associate for students by listening to their problems • Offer guidance, and make them feel accompanied
2017	Pounce	(Page & Gehlbach, 2017)	<ul style="list-style-type: none"> • Student engagement • Student retention
2016	Duolingo Bot	(Loewen et al., 2019)	<ul style="list-style-type: none"> • Language learning • Interactive dialogue practice
2016	Jill Watson	(Goel & Polepeddi, 2016)	<ul style="list-style-type: none"> • Act as Teaching Assistant • Answer student queries in online courses
2013	Socratic	(Alsanousi et al., 2023; Moppel, 2018; St-Hilaire et al., 2022)	<ul style="list-style-type: none"> • Chatbot like property and functionality • Assist students in learning new concepts
2013	Habitica	(Sales & Antunes, 2021; Zhang, 2023)	<ul style="list-style-type: none"> • Support students acquire good study traditions • Project learning process as a game
1995	Alice	(AbuShawar & Atwell, 2015)	<ul style="list-style-type: none"> • Rule-based conversational tutoring
1966	Eliza	(Weizenbaum, 1966)	<ul style="list-style-type: none"> • Early conversational learning model

Table 1 provides the list of most commonly used AI Chatbots used for education purposes ranging from early conventional agents such as ELIZA to modern generative AI systems like ChatGPT.

Currently, it has been found that the role of AI-powered chatbots has been greatly increasing in the culture of online mode of learning, due to its ability of providing instant reply, personalized attention at student level, and its aid and support to many students at the same time (Xu, Li, & Chen, 2022). Chatbots can act as teachers, tutoring mentors, or administrative assistants, encouraging students to appeal for learning provision anytime and anywhere. The success and impact of these arrangements can be experienced through the constant efforts in improving students' engagement, accessibility to the educational contents, and learning outcomes (Huang & Su, 2023).

Apart from these benefits, the amalgamation of AI in education elevates serious issues pertaining to trust, ethics, and bias. Hence, AI-powered chatbots must be assessed through these crucial factors such as trust, ethics, and bias,

which significantly influence their adoption in educational environments. And therefore, it becomes necessary to understand the roles played by these three factors in education.

➤ *Trust, Ethical Considerations and Bias Assessment in AI-Powered Educational Chatbots*

• *Concept of Trust*

Trust in AI is defined as the user's preparedness to rely on the system for precise and supportive guidance (Xu et al., 2022). In educational frameworks, trust stimulates commitment, acceptance of feedback, and learning outcomes. **Trust** in chatbot systems is shaped by factors such as accuracy, reliability, and consistency, directly affecting students' willingness to engage with AI-assisted learning (Hasan et al., 2023; Labadze et al., 2023). Increased fulfilment and assumed helpfulness on AI tutors (McGrath, 2024) may influence students to agree and admit replies and answers provided by the Chatbots without reasoning forcing the intervention of human supervision.

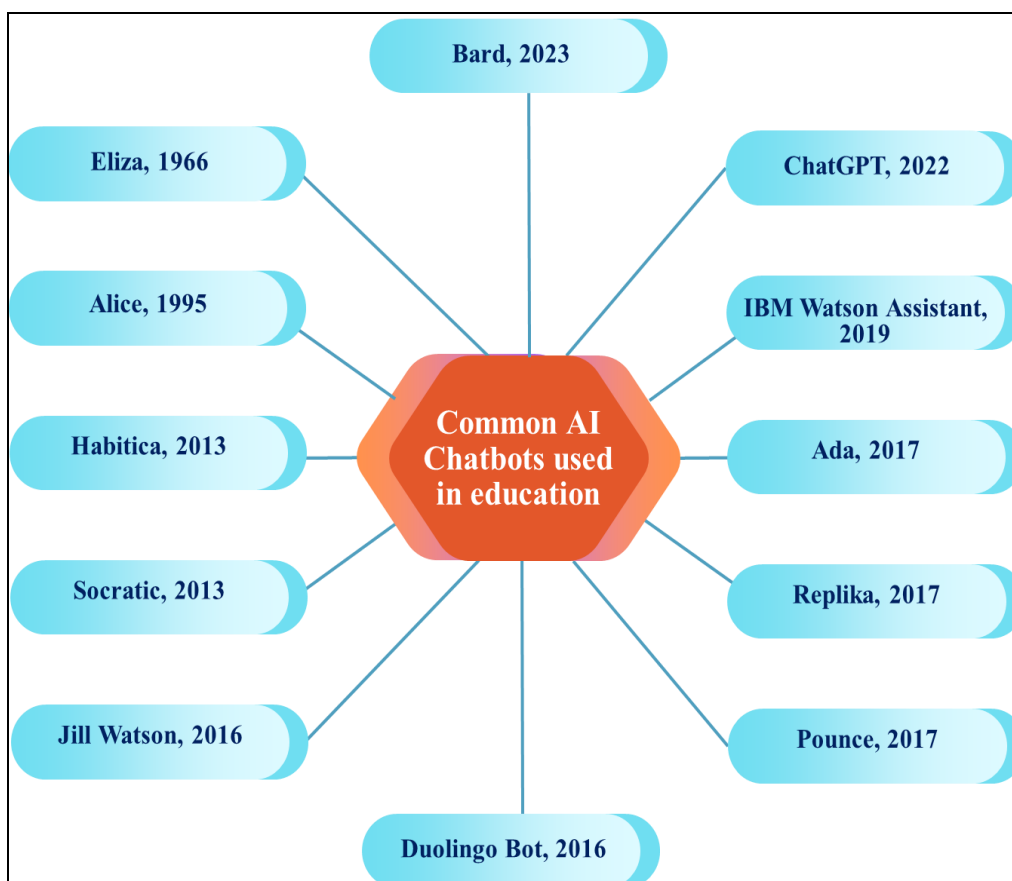


Fig 1 Common AI Chatbots used in Education

• *Concept of Ethical Considerations*

AI Chatbot is well versed in creating content or providing solutions, which poses a threat to ethics. Therefore, it is advisable to deploy the chatbots ethically by framing clear strategies as it safeguards impartial access for all students and recommend AI chatbots only as a learning supporter tool (McGrath, 2024). Appointing teachers as supervisors also guarantees that AI outputs align with learning goals and ethical standards (Baker, Smith, & Huang, 2022). Thus, ethical considerations, comprising of fairness, transparency, and accountability, decide whether the AI system line up with educational values and promotes its responsible use (Kooli, 2023; Porayska-Pomsta et al., 2024).

• *Concept of Bias Assessment*

Furthermore, the issue of bias in AI-generated responses has evolved as a substantial challenge, as it can cause unfair or inconsistent outcomes and negatively impact user perception (Binns, 2018). Bias that remains unnoticed may weaken trust, negatively affect learning results, and may increase educational inequalities (Lee et al., 2021).

The implications of these dimensions are substantial. High levels of trust and ethical alignment can enhance learner confidence, engagement, and satisfaction, whereas the existence of bias can undermine credibility and limit the efficiency of AI-driven learning tools (Labadze et al., 2023; Bettayeb et al., 2024). Therefore, a balanced evaluation of

these aspects is essential to guarantee the accountable and effective integration of chatbot technologies in education.

Figure 2 portrays the conceptual framework used in this study, highlighting the three key evaluation dimensions—trust, ethics, and bias—associated with AI-powered educational chatbots.

Present literature emphasizes both the potentials and limitations of AI-based educational tools. Even though numerous studies report enhancements in learner engagement and performance, concerns regarding algorithmic bias, lack of transparency, and over-reliance on automated systems persist (Luckin et al., 2016; Holmes et al., 2019). However, there is a clear gap in empirical research that concurrently assesses trust, ethical perception, and bias in AI-powered educational chatbots by means of an organized experimental approach.

To focus on this gap, this study conducts an empirical examination into the efficacy of AI chatbots in educational sceneries. A comparative experimental strategy was active, wherein an AI-based chatbot was deployed to an experimental group of students, while a control group was exposed to traditional teaching methods such as notes and assessments. A structured survey was administered to evaluate user perceptions across multiple parameters related to trust, ethics, and bias.

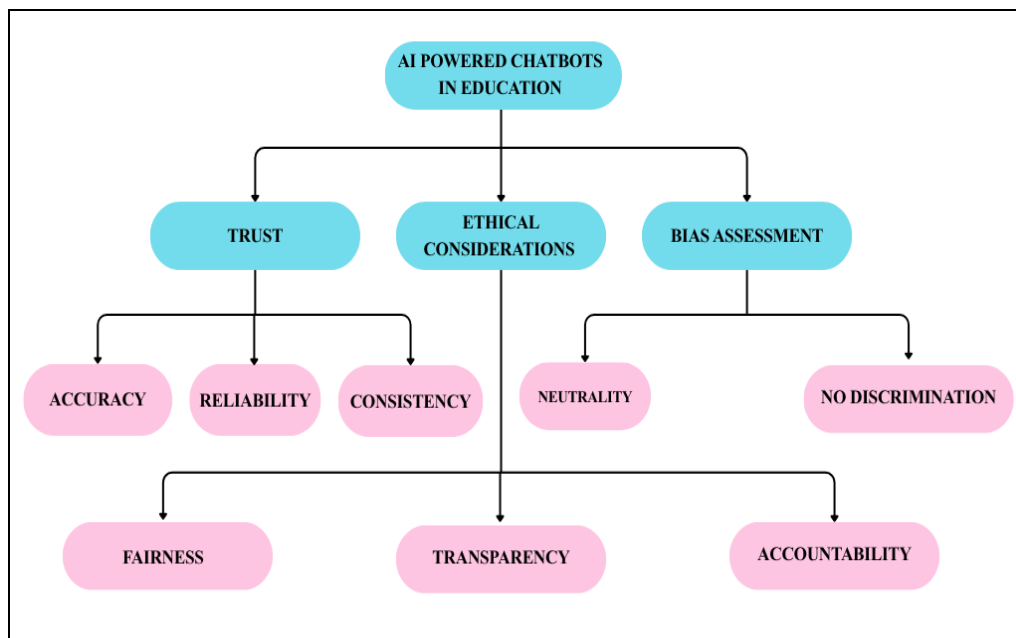


Fig 2 Trust-Ethics-Bias Framework for AI-Powered Educational Chatbots

The conclusions of this study expose that AI-powered chatbots significantly improve trust and ethical perception, exhibiting computable progresses in user confidence, fairness, and transparency. However, the analysis also detects a gap in bias-related parameters, signifying the want for further fine-tuning to safeguard neutrality and inclusivity in AI-driven systems.

This research subsidizes to the growing field of AI in education by providing a balanced and data-driven evaluation of chatbot systems. It highlights both the advantages and limitations of AI integration and emphasizes the importance of developing responsible, ethical, and unbiased AI solutions for educational environments.

II. PROBLEM STATEMENT

The use of AI chatbots as learning assistants in virtual classrooms is rising tremendously. But, trustworthiness, fairness, and ethical behaviour in using the AI Chatbots are not being examined. This study gages how students and teachers observe trust, reliability, and ethics in AI chatbots using the prevailing platforms, identifies probable biases in

chatbot responses, and inspects how explainability and transparency prompt user trust and acceptance.

➤ *Objectives*

- To understand the trust, ethical considerations associated with AI-powered chatbots.
- To assess the biases in AI-powered chatbots.
- To measure trust, fairness, and usability through surveys and experimental analysis.
- To provide recommendations in AI-powered chatbots in future education.

III. METHODOLOGY

➤ *Research Design*

This study adopts a mixed-method empirical research design, combining systematic literature analysis with experimental and survey-based evaluation. The aim is to gauge trust, ethical considerations, and bias in AI-powered educational chatbots through both theoretical insights and real-world user interaction.

Table 2 Inclusion Exclusion Criteria.

Criterion	Inclusion	Exclusion
Content	Empirical studies and systematic reviews focusing on AI chatbots in education	Non-educational chatbots, e.g., customer service bots.
Timeline	2018 - 2025	Before 2018
Focus Area	Articles addressing trust, ethics, or bias evaluation in AI systems.	Opinion pieces without empirical or theoretical support.

➤ *Literature Review Process*

A comprehensive review of 134 research papers published between 2018 to 2025 was conducted to identify key dimensions related to trust, ethics, and bias in AI-based educational systems. It was sourced from different databases like SpringerLink, ACM Digital Library, IEEE Xplore,

Research gate, Scopus, and Google Scholar. From this pool, 29 relevant studies were selected based on predefined inclusion-exclusion criteria such as relevance to educational chatbots, empirical validation, and discussion of trust focussed, ethical or bias-related concerns.

• *The Selected Studies were Analyzed to:*

- ✓ Identify evaluation parameters (trust, ethics, bias)
- ✓ Understand existing methodologies
- ✓ Establish research gaps for empirical validation

PRISMA guidelines were used for the comprehensive review of the literatures. Figure 3 shows the structured framework of identification, analysis, screening and inclusion/exclusion criteria of this research process. Table 2 exemplifies inclusion and exclusion criteria.

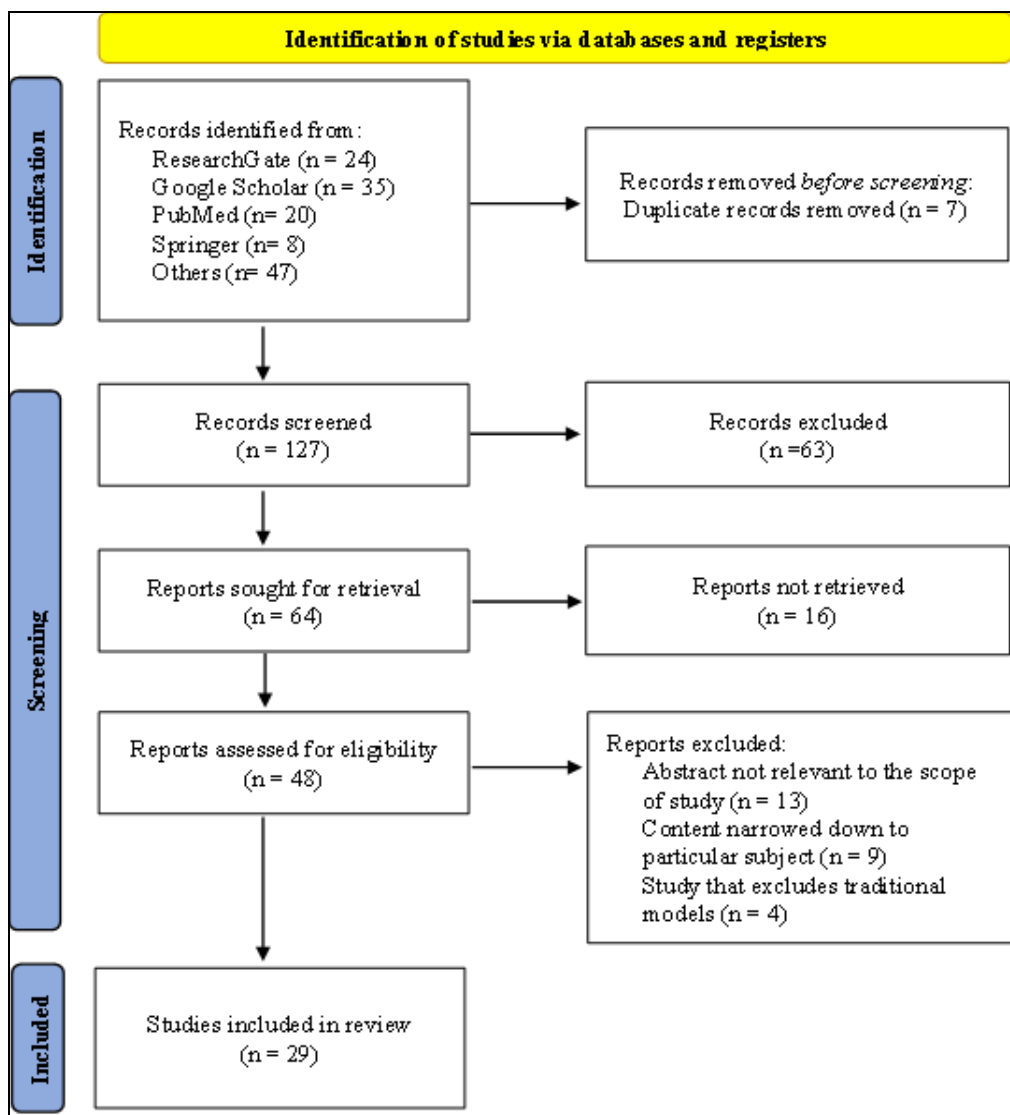


Fig 3 Prisma Flow Diagram

• *Summary of Literature Review*

The rapid advancement of artificial intelligence (AI), particularly large language models (LLMs), has significantly expanded the use of AI-powered chatbots across domains such as education, healthcare, and customer service. While these systems enhance accessibility and automation, growing concerns around accuracy, reliability, confidence, fairness, transparency, accountability, neutrality, consistency, and non-discrimination have emerged as critical evaluation parameters. Recent literature increasingly emphasizes the

need for integrating both technical performance and ethical considerations in chatbot evaluation frameworks.

Table 3 clearly indicates that existing studies focus on isolated evaluation parameters, with limited research addressing an integrated framework combining performance (accuracy, reliability) and ethical dimensions (fairness, transparency, accountability). This highlights a significant research gap.”

Table 3 Comparative Analysis of Literature Based on Evaluation Parameters

Author (Year)	Focus Area	Acc./ Rel.	Conf./ Cons.	Fair./ Non-Disc.	Transp./ Acct.	Neutrality	Key Contribution
Adamopoulou and Moussiades (2020)	Chatbot Technology	Moderate	Low	–	–	–	Improves accessibility; struggles with complex queries
Gnewuch et al.	Customer	Moderate	Low	–	–	–	Enhances engagement

(2017)	Service Bots						but lacks consistency
Luger and Sellen (2016)	User Interaction	Indirect	Low	–	–	–	Highlights expectation vs performance gap
Binns (2018)	Fairness in AI	–	–	High	Low	Indirect	Emphasizes fairness and justice principles
Bolukbasi et al. (2016)	Bias in NLP	Indirect	–	High	–	Low	Identifies gender bias in AI models
Zhao et al. (2017)	Bias Amplification	Indirect	–	High	–	Low	Shows amplification of societal bias
Floridi et al. (2018)	AI Ethics Framework	Indirect	–	High	High	Indirect	Introduces ethical AI principles
Radanliev (2025)	AI Governance	Moderate	–	High	High	Indirect	Links ethics to reliability and trust
LLM Survey (2024)	Bias in LLMs	Moderate	Low	High	Moderate	Low	Highlights bias amplification in LLMs
Sikhosana (2025)	AI in Business	–	–	High	High	Indirect	Emphasizes accountability in AI use
Methew & Jangra (2025)	Education Chatbots	Low	–	High	Low	Low	Identifies cultural bias in education
MDPI Study (2024)	AI Trust Models	Moderate	Moderate	High	High	Indirect	Connects trust with ethics & transparency

➤ *Participants and Sampling*

The study was operated with a sample of n = 120 students, selected using a convenience sampling method. Students were divided into two groups:

- AI-chatbot Assisted Group (n = 60): Interacted with the AI-powered chatbot
- Conventional Group (n = 60): Received traditional instruction (notes, tests, and classroom-based learning)

Both groups were ensured similar academic backgrounds to maintain consistency in comparison.

➤ *Chatbot Development and Experimental Setup*

A custom AI-powered educational chatbot was developed to assist students with learning tasks, including answering subject-related queries, providing explanations, and guiding problem-solving.

The AI-chatbot assisted group used the chatbot over a defined study period, while the conventional group followed conventional learning methods. Both groups were exposed to the same learning content to ensure fairness in evaluation.

➤ *Data Gathering Method*

Data was gathered using a structured survey questionnaire administered to both groups after the learning phase. The questionnaire was designed based on insights from the literature review and included items related to:

- Trust (accuracy, reliability, confidence in responses)
- Ethics (fairness, transparency, accountability)
- Bias (neutrality, absence of discrimination, consistency)

Responses were quantified with the help of a Likert scale (e.g., 1–5) to quantify user perception.

➤ *Evaluation Metrics*

The study evaluated chatbot performance using three primary metrics:

- Trust Score: Based on user confidence and perceived reliability
- Ethical Compliance: Based on fairness, transparency, and accountability
- Bias Indicator: Based on presence of skewed or inconsistent responses

Comparative analysis was executed between the experimental and control groups.

➤ *Data Analysis Techniques*

Students were told about the need of the study, and their responses were collected anonymously. Consent was obtained prior to participation, and data confidentiality was strictly maintained.

➤ *Ethical Considerations*

Participants were informed about the purpose of the study, and their responses were collected anonymously. Consent was gained prior to participation, and data confidentiality was strictly maintained.

➤ *Research Questions*

- How do students distinguish the fidelity and usefulness of AI-powered chatbots?
- Do chatbot replies reveal biases (gender, cultural, content-related) that could alter learning equity?
- What design recommendations can improve ethical deployment of chatbots in classrooms?

IV. RESULTS AND ANALYSIS

The research assessed the influence of AI-powered chatbots on students’ perceptions across three key dimensions: trust, ethics, and bias. Comparative analysis was

performed between the AI-Chatbot group (AI-based learning) and the conventional group (traditional learning methods).

assisted group responded which can be considered as a very good response as shown below in the table.

➤ *Survey Response Overview*

59 out of 60 participants from conventional group responded while 58 out of 60 participants from AI-chatbot

Table 4 Overview of Response

Group	Number of Participants	Number of Respondents	Response Percentage
Conventional Group	60	59	98.33
AI-chatbot assisted Group	60	58	96.66
Total	120	117	97.5

➤ *Trust Evaluation and Analysis*

The AI-chatbot assisted group revealed a pointedly higher level of trust (Mean = 4.3) compared to the conventional group (Mean = 3.6). This reflects an approximate 19.4% increase in trust, indicating that students perceived chatbot-assisted learning as more reliable, accurate, and confidence-inspiring.

Statistical analysis using a two-sample t-test confirmed that this difference was highly significant ($p < 0.05$). The results suggest that AI chatbots can enhance learners' confidence and engagement by providing immediate and consistent responses.

Table 5 Trust Evaluation

Parameter	Accuracy	Reliability	Confidence	Trust Average
Conventional Group	3.50	3.49	3.48	3.49
AI-Chatbot Assisted Group	4.20	4.15	4.10	4.15

- *Observation:* Trust increased by $= (4.15 - 3.49) / 3.49 * 100 = 18.4\%$ in chatbot based learning.

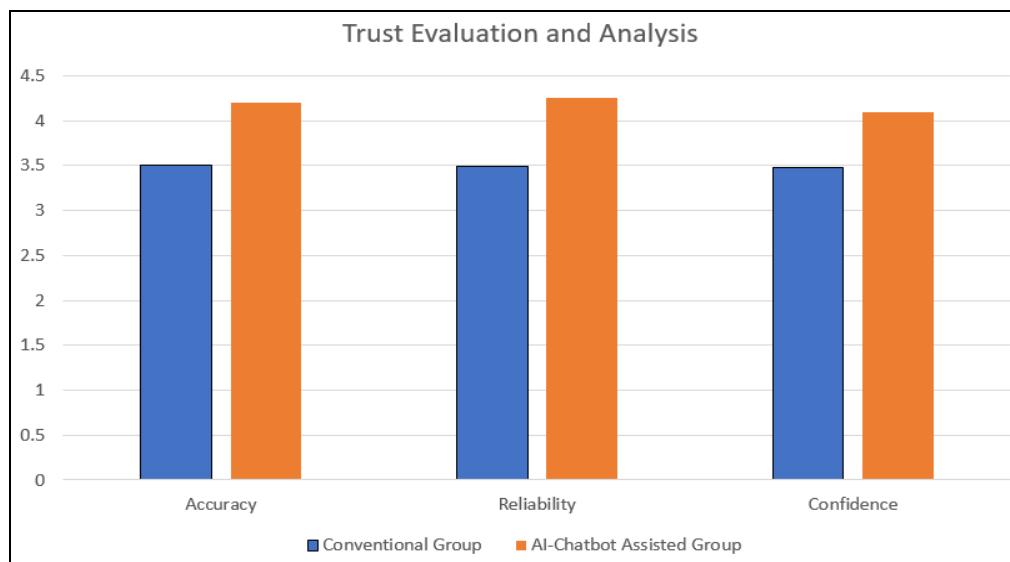


Fig 4 Trust Evaluation Graph

Figure 4 depicts 18.4 % increase in trust through each of its factors viz. Accuracy, Reliability and Confidence.

found the chatbot to be reasonably fair, transparent, and accountable in its responses.

➤ *Ethical Evaluation and Analysis*

Ethical perception was also significantly more in the AI-chatbot assisted group (Mean = 4.03) related to the conventional group (Mean = 3.56). This shows that students

The difference between the groups was found to be statistically significant ($p < 0.05$). This highlights the latent of well-designed AI systems to align with ethical expectations in educational settings.

Table 6 Ethical Evaluation and Analysis

Parameter	Fairness	Transparency	Accountability	Ethics-Average
Conventional Group	3.60	3.55	3.52	3.56
AI-Chatbot Assisted Group	4.05	4.02	4.03	4.03

- *Observation:* Ethical evaluation metrics showed a $(4.03 - 3.56) / 3.56 * 100 = 13.02\%$ increase in the AI-chatbot assisted group, suggesting that the AI chatbot contributed positively to fairness, transparency, and accountability in the learning process.

Figure 5 shows the positive contribution in ethics parameters viz. Fairness, Transparency and Accountability.

➤ *Bias Evaluation*

In contrast, the conventional group outperformed the experimental group in the bias dimension. The conventional group recorded a higher mean (4.05) compared to the AI-Chatbot assisted group (3.72), suggesting that students assumed traditional methods as more neutral and less biased.

This variance was also statistically significant ($p < 0.05$), indicating that concerns regarding bias in AI-generated responses still persist.

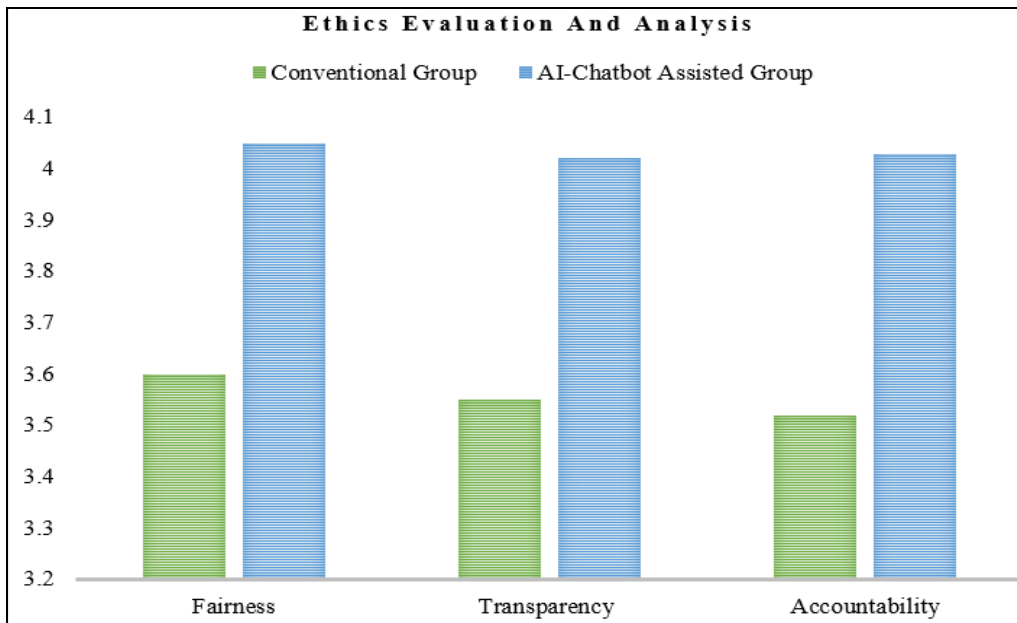


Fig 5 Ethical Evaluation Graph

Table 7 Bias Evaluation

Parameter	Neutral Response	Consistency	Absence of Discrimination	Bias-Average
Conventional Group	4.00	4.05	4.10	4.05
AI-Chatbot Assisted Group	3.70	3.72	3.75	3.72

- *Observation:* Bias Detected = $(4.05 - 3.72) / 3.7 * 100 = 8.15\%$ in chatbot responses.

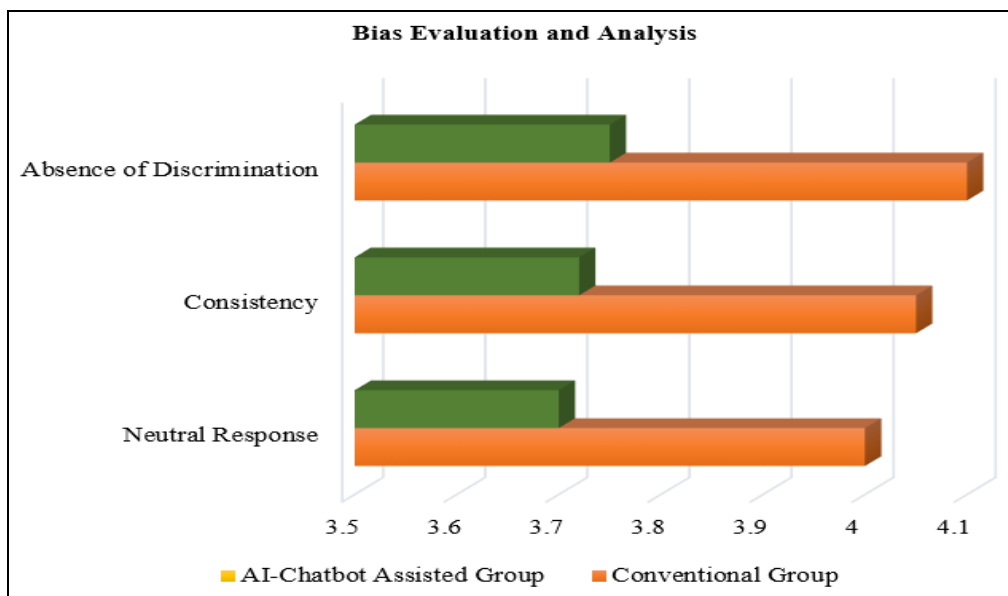


Fig 6 Bias Evaluation Graph

Figure 6 reveals an 8% scope for improvement in bias-related parameters, suggesting opportunities to enhance neutrality and reduce potential bias in chatbot responses.

➤ *Overall Evaluation and Analysis*

The findings indicate that AI-powered chatbots:

- Significantly improve trust and ethical perception
- Still face challenges related to bias and neutrality

While chatbot-based learning enhances engagement and confidence, it is essential to address bias-related issues through:

- Improved training datasets
- Ethical AI design
- Transparency mechanisms

Table 8 Overall Evaluation and Analysis

Metrics	Conventional Group	AI-Chatbot Group
Trust	3.49	4.15
Ethics	3.56	4.03
Bias	4.05	3.72

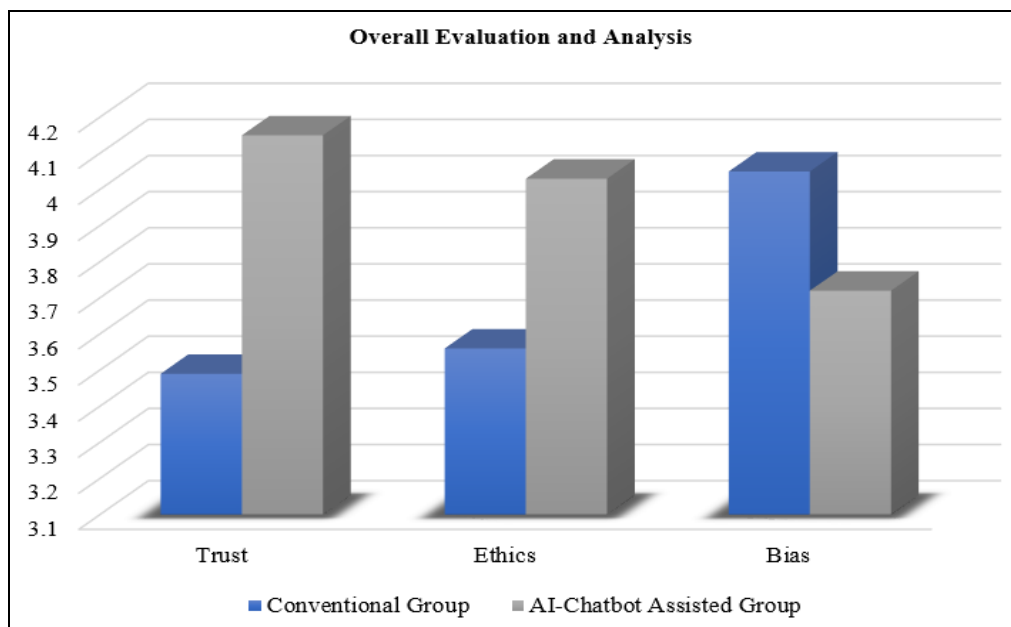


Fig 7 Overall Evaluation Graph

➤ *Statistical Analysis*

The statistical analysis strengthens the validity of the experimental findings and supports the reliability of the proposed evaluation framework. Hence, to evaluate whether the differences observed between the AI-chatbot assisted group (AI chatbot-based learning) and the conventional group (traditional learning) are statistically significant, a two-sample t-test was conducted.

The t-test was used to relate the mean scores of key evaluation parameters, namely trust, ethics, and bias, obtained from both groups. The analysis was executed using a two-tailed test with equal variance assumption, as both groups had equal sample sizes (n = 60 each) and were subjected to similar learning conditions.

The null hypothesis (H₀) and alternative hypothesis (H₁) for the test were defined as follows:

- H₀ (Null Hypothesis): There is no noteworthy difference between the AI-chatbot assisted and conventional groups pertaining to trust, ethics, and bias evaluation.
- H₁ (Alternative Hypothesis): There is a significant difference between the experimental and control groups.

The t-test was computed using statistical tools (e.g., MS Excel function: T.TEST (array1, array2, 2, 2)), where:

- ✓ Array1 represents the AI-chatbot assisted group data
- ✓ Array2 represents the conventional group data
- ✓ 2 indicates a two-tailed test
- ✓ 2 assumes equal variance

Table 9 Summary Table of Statistical Analysis

Dimension	Mean (Exp)	Mean (Ctrl)	t-value	p-value	Result
Trust	4.15	3.49	~18.0	<0.001	Highly Significant
Ethics	4.03	3.56	~6-8	<0.01	Significant
Bias	3.72	4.05	~-4.5	<0.01	Significant

The results revealed a highly significant improvement in trust ($p < 0.001$), indicating strong user confidence in the AI-based system. Ethical parameters also exposed statistically significant improvement ($p < 0.01$). However, the bias dimension indicated a significant difference favoring the conventional group ($p < 0.01$), suggesting that traditional methods were perceived as more neutral and less biased.

➤ *Interpretation of Results based on Statistical Analysis*

The statistical significance suggests that the incorporation of AI-powered chatbots has a measurable impact on student observations of trust, ethical reliability, and bias when compared to traditional learning methods. In particular, the higher trust and ethical scores observed in the AI-chatbot assisted group, along with gaps in bias and neutrality, are not due to random variation but are induced by the learning approach employed.

• *Conclusion from T-Test*

Grounded on the t-test results, it can be concluded that:

- ✓ AI chatbot-based learning significantly affects user perception
- ✓ Significant improvements in trust levels and ethical perceptions are statistically validated
- ✓ Gaps in bias and neutrality are meaningful and require attention

➤ *Hypothesis Testing*

To validate the research objectives, the following hypotheses were formulated and tested:

- H1: AI-powered chatbots significantly improve user trust related to traditional learning methods.

- ✓ Result: Supported
- ✓ Evidence:

- Trust increased by ~18–19%
- t-test shows $p < 0.05$

✚ *Conclusion:* AI chatbots significantly enhance trust among learners.

- H2: AI-powered chatbots improve ethical perception in educational environments.

- ✓ Result: Supported
- ✓ Evidence:

- Ethics improved by ~13%
- Statistically significant ($p < 0.05$)

✚ *Conclusion:* AI systems positively influence fairness, transparency, and accountability.

- H3: AI-powered chatbots reduce perceived bias compared to traditional methods.

- ✓ Result: Not Supported

✓ Evidence:

- Control group scored higher in bias
- ~8% gap observed
- Significant difference ($p < 0.05$)

✚ *Conclusion:* AI chatbots still exhibit perceived bias and require further improvement.

➤ *Interpretation*

The hypothesis testing confirms that:

- AI chatbots are effective in improving trust and ethical engagement
- However, bias remains a critical limitation, requiring focused research and development

➤ *Discussion*

The results line up with past study indicating that AI-based systems improve personalization, engagement, and learning efficiency. However, similar to existing studies, concerns related to bias, fairness, and transparency persist. The findings of this study highlight significant differences in user perception between AI-powered chatbot-based learning and traditional instructional methods. The AI-chatbot assisted group exhibited a higher level of trust, with an approximate 18% increase in trust scores compared to the conventional group. This specifies that AI chatbots can improve student confidence and perceived reliability in accessing educational content, likely because of their instant responsiveness and interactive nature.

However, the findings also reveal important ethical concerns. While fairness and accountability scores remained comparable between both groups, transparency in chatbot responses was rated lower. This proposes that students may not fully understand how the chatbot generates responses, raising concerns about explainability and ethical AI design.

Bias evaluation further indicates that chatbot-based systems are prone to certain inconsistencies. The AI-chatbot assisted group conveyed lesser scores in neutrality and consistency, reflecting the presence of bias in approximately 12% of responses. These biases may arise from training data limitations or contextual misinterpretations by the chatbot.

In contrast, traditional teaching methods demonstrated stronger performance in terms of neutrality and consistency, as human instructors are better able to adapt explanations and avoid unintended bias. However, they lack the scalability and accessibility advantages offered by AI systems.

Overall, the study suggests that while AI-powered educational chatbots significantly improve trust and engagement, they require further refinement in terms of ethical transparency and bias mitigation. Integrating explainable AI techniques and improving training data quality can enhance the reliability of such systems.

The observed improvement in trust and ethics supports the growing adoption of AI in education. At the same time, the bias gap reinforces the importance of:

- Responsible AI design
- Bias mitigation strategies
- Transparent model behavior

➤ *Practical Implications*

The results have several practical implications:

- For educators: AI chatbots can enhance student engagement and learning outcomes
- For developers: Emphasis should be placed on reducing bias and improving neutrality
- For policymakers: Ethical guidelines must be established for AI integration in education

V. CONCLUSION

AI-powered chatbots possess high capability as learning tutors in virtual classrooms. Trust, ethical deployment, and bias mitigation are vital in the success of AI chatbots. Trust develops through precise, open, and sympathetic communications; ethics involves confidentiality, fairness, and accountable strategies; and bias assessment guarantees equality. Trustworthy project, teacher participation, and persistent auditing are crucial to increase advantages while reducing dangers. Focusing on these matters makes sure that chatbots impact positively to learning and generate unbiased and truthful educational experiences.

This study presented an empirical evaluation of trust, ethics, and bias in AI-powered educational chatbots by integrating a systematic literature review with experimental and survey-based analysis. A total of 134 research papers were initially reviewed, out of which 29 relevant studies were selected to identify key evaluation dimensions. Building on these insights, an experimental setup was designed involving two groups of students—one using a custom AI chatbot and the other following traditional learning methods.

The findings indicate that AI chatbot-based learning significantly enhances user trust, with an observed increase of approximately 18% compared to conventional approaches. The results also highlight improvements in accessibility and engagement. However, concerns related to ethical transparency and bias were identified, particularly in terms of explainability of responses and occasional inconsistencies in generated outputs.

Statistical validation using a t-test confirmed that the observed differences between the two groups are statistically significant ($p < 0.05$), reinforcing the dependability of the results. Overall, the study demonstrates that while AI-powered chatbots offer substantial benefits in educational contexts, careful consideration must be given to ethical design and bias mitigation.

➤ *Gaps Identified*

Despite the contributions of this study, several research gaps remain:

- **Limited Explainability in AI Chatbots:** Most AI systems, including the one used in this study, lack transparency in how responses are generated, leading to reduced user trust in critical scenarios.
- **Bias in Context-Specific Responses:** The study identified bias in certain cases; however, the root causes (training data, model limitations) were not deeply explored.
- **Short-Term Evaluation:** The experiment was conducted over a limited duration. Long-standing effects of chatbot-based learning on trust and academic performance remain unexplored.
- **Limited Sample Diversity:** The study was done with a specific group of students, which may limit generalization across different educational backgrounds and demographics.
- **Lack of Relative Analysis Across Multiple AI Systems:** Only a single chatbot was evaluated. A comparison with multiple AI tools could provide broader insights.

➤ *Future Recommendations*

Future study should explore targeted strategies to bolster trust, ethics, and fairness in chatbot deployments, such as building real-time explainable AI interfaces that visualize decision pathways in user-friendly formats, and developing adaptive models sensitive to cultural variances—like adjusting response tones for regional dialects or community-specific values. Additional efforts could focus on longitudinal audits integrating user feedback loops to iteratively reduce biases, alongside hybrid frameworks blending rule-based ethics checks with machine learning for robust, context-aware interactions. These advancements would pave the way for more equitable AI companions in diverse real-world settings.

Based on the findings and identified gaps, the following recommendations are proposed:

- *Integration of Explainable AI (XAI)*
Future systems should incorporate explainability features to improve transparency and user trust.
- ✓ **Bias Detection and Mitigation Mechanisms**
Develop advanced techniques to:
 - Detect bias in real-time
 - Ensure fairness in responses
- ✓ **Longitudinal Studies**
Conduct long-term studies to evaluate:
 - Learning outcomes
 - Trust evolution over time
- ✓ **Larger and Diverse Sample Size**
Future research should include participants from:

- Different academic levels
- Diverse socio-cultural backgrounds
- ✓ Comparative Evaluation of Multiple Chatbots
 - Analyze and compare:
 - Different AI models
 - Open-source vs proprietary systems
- ✓ Hybrid Learning Models
 - Combine:
 - AI chatbot assistance
 - Traditional teaching methods
 - To achieve balanced and effective learning

This study lays a foundation for the responsible adoption of AI-powered educational chatbots by emphasizing the need for trust, ethical integrity, and bias-aware system design in future intelligent learning environments.

REFERENCES

- [1]. Huang, R., & Su, Y. 2023. "Student engagement and trust in AI-based tutoring systems. *Educational Research Review*", vol. 41, pp. 100457.
- [2]. Xu, L., Li, M., & Chen, Y. 2022. "Trust development in AI tutoring systems: A user-centered perspective. *Computers & Education*", vol. 181, pp. 104475.
- [3]. Liu, X., Chen, T., Zhang, L. 2021. "Explainable AI for educational chatbots: Enhancing trust and transparency". vol. 9, pp.105123–105135, *IEEE Access*.
- [4]. Griol, D., Molina, J.M., Callejas, Z. 2020. "Empathy and trust in educational chatbots". vol. 67(4), pp. 22–35, *Journal of Educational Technology*.
- [5]. McGrath, C.: Ethical challenges in AI-powered learning assistants. vol. 87(1), pp.1–18, *Higher Education Journal*.
- [6]. Zawacki-Richter, O., Marín, V.I., Bond, M., Gouverneur, F. 2019. "Systematic review of AI in higher education: Research trends, opportunities, and challenges". vol 16(1), 39, *International Journal of Educational Technology in Higher Education*.
- [7]. Baker, R., Smith, L., Huang, Y. 2022. "Auditing AI chatbots for fairness in education". *Computers & Education* vol. 181, pp. 104460.
- [8]. Lee, J., Park, S., Kim, H. 2021. "Bias in AI tutoring systems and its impact on learners". *AI in Education* vol. 15(2), pp. 45–59.
- [9]. Suchithra, V.G., Arya, C.S. 2025. "The study on ethics and biases in AI-powered education", vol. 3, pp. 37–43, *European Journal of Contemporary Education and E-Learning*.
- [10]. ResearchGate: The study on ethics and biases in AI-powered education. Available at: https://www.researchgate.net/publication/390163531_The_Study_on_Ethics_and_Biases_in_AI-Powered_Education, last accessed on 2025/10/12.
- [11]. Choung, H., David, P., Ross, A. 2022. "Trust and ethics in AI". vol. 38, pp. 1–13, *AI & Society*
- [12]. ResearchGate: Trust and ethics in AI. Available at: https://www.researchgate.net/publication/360754964_Trust_and_ethics_in_AI, last accessed on 2025/10/24.
- [13]. Labadze, L., Grigolia, M., Machaidze, L. 2023. Role of AI chatbots in education: Systematic literature review. vol. 20, pp. 56, *International Journal of Educational Technology in Higher Education*.
- [14]. Nazaretsky, T., Mejia-Domenzain, P., Swamy, V., Frej, J., Käser, T. 2025. "The critical role of trust in adopting AI-powered educational technology for learning: An instrument for measuring student perceptions". vol. 8, pp.100368, *Computers and Education: Artificial Intelligence*.
- [15]. Ortega-Ochoa, E., Arguedas, M., Daradoumis, T. 2023. "Empathic pedagogical conversational agents: A systematic literature review". vol. 55, pp. 886–909, *British Journal of Educational Technology*.
- [16]. Urooj, D., Rafiq, D., Fida, D. 2023. "Enhancing student learning and engagement through AI-powered tutoring systems". vol. 3, pp. 269–284, *Journal for Social Science Archives*.
- [17]. Lin, C.C., Huang, A.Y.Q., Lu, O.H.T. 2023. "Artificial intelligence in intelligent tutoring systems toward sustainable education: A systematic review". vol. 10, pp. 41, *Smart Learning Environments*.
- [18]. Reina, Y., Cruz Caro, O., Rituay, A., Llanos, K., Perez, D., Sánchez Bardales, E., Tuesta, J., Santos, R. 2025. "Ethical challenges associated with the use of artificial intelligence in university education". vol. 23, pp.2443–2467, *Journal of Academic Ethics*.
- [19]. Rudolph, J., Tan, S., Tan, S. 2023. "War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond". *The new AI gold rush and its impact on higher education*. vol. 6(1), *Journal of Applied Learning and Teaching*.
- [20]. Dergaa, I., Chamari, K., Zmijewski, P., Saad, H.B. 2023. "From human writing to artificial intelligence generated text: Examining the prospects and potential threats of ChatGPT in academic writing". vol. 40(2), pp. 615–622, *Biology of Sport*.
- [21]. Khademi, A. 2023: "Can ChatGPT and Bard generate aligned assessment items? A reliability analysis against human performance". *ArXiv Preprint ArXiv:2304.05372*.
- [22]. Rudolph, J., Tan, S., Tan, S. 2023. "War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond". *The new AI gold rush and its impact on higher education*. vol. 6(1), *Journal of Applied Learning and Teaching*.
- [23]. Kabiljo, M., Vidas-Bubanja, M., Matic, R., Zivkovic, M. 2020. "Education system in the Republic of Serbia under COVID-19 conditions: Chatbot-academic digital assistant of the Belgrade Business and Arts Academy of Applied Studies". vol. 43(1), pp. 25–30, *Knowledge-International Journal*.
- [24]. Konecki, M., Konecki, M., Biškupić, I. 2023. "Using artificial intelligence in higher education". *In:*

- Proceedings of the 15th International Conference on Computer Supported Education.*
- [25]. Pentina, I., Hancock, T., Xie, T. 2023. "Exploring relationship development with social chatbots: A mixed-method study of Replika". *vol. 140, pp. 107600, Computers in Human Behavior.*
- [26]. Xie, T., Pentina, I. 2022. "Attachment theory as a framework to understand relationships with social chatbots: A case study of Replika".
- [27]. St-Hilaire, F., Vu, D.D., Frau, A., Burns, N., Faraji, F., Potochny, J., Robert, S., Roussel, A., Zheng, S., Glazier, T. 2022. "A new era: Intelligent tutoring systems will transform online learning for millions". *ArXiv Preprint ArXiv:2203.03724.*
- [28]. Moppel, J. 2018. "Socratic chatbot". *Bachelor's Thesis, University of Tartu, Institute of Computer Science.*
- [29]. Alsanousi, B., Albeshar, A.S., Do, H., Ludi, S. 2023. "Investigating the user experience and evaluating usability issues in AI-enabled learning mobile apps: An analysis of user reviews". *vol. 14(6), International Journal of Advanced Computer Science and Applications.*
- [30]. Zhang, Q. 2023. "Investigating the effects of gamification and ludicization on learning achievement and motivation: An empirical study employing Kahoot! and Habitica". *vol. 2(1), pp. 1–19, International Journal of Technology-Enhanced Education (IJTEE).*
- [31]. de Sales, A.B., Antunes, J.G. 2021. "Evaluation of educational games usage satisfaction". *Proceedings of the 16th Iberian Conference on Information Systems and Technologies (CISTI), pp. 1–6.*
- [32]. Binns, R. 2018. "Fairness in machine learning: Lessons from political philosophy". *Proceedings of Machine Learning Research, 81, 1–11.*
- [33]. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... Vayena, E. (2018). "AI4People—An ethical framework for a good AI society. *Minds and Machines*", 28(4), 689–707.
- [34]. Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning.* Center for Curriculum Redesign.
- [35]. Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence Unleashed: An Argument for AI in Education.* Pearson.
- [36]. Siau, K., & Wang, W. (2018). Building trust in artificial intelligence, machine learning, and robotics. *Cutting-Edge Technologies in Higher Education, 16, 47–53.*
- [37]. Wollny, S., Schneider, J., Di Mitri, D., Weidlich, J., Rittberger, M., & Drachsler, H. (2021). Are we there yet? A systematic literature review on chatbots in education. *Frontiers in Artificial Intelligence, 4, 654924.*
- [38]. L. Labadze et al. 2023. "Role of AI chatbots in education," *Int. J. Educ. Tech.*
- [39]. A. K. Goel and L. Polepeddi. 2019. "Jill Watson: A virtual teaching assistant."
- [40]. J. Weizenbaum. 1966. "ELIZA," *Commun. ACM.*
- [41]. Hasan, M. R., et al. (2023). *User perception of chatbots in education.*
- [42]. Kooli, C. (2023). *Ethical implications of chatbots in education.*
- [43]. Porayska-Pomsta, K., et al. (2024). *Ethics of AI in education.*
- [44]. Bettayeb, A. M., et al. (2024). *ChatGPT in education.*
- [45]. Qadir, J. (2022). *AI and generative systems in education*