OPEN ACCESS

Volume: 13

Special Issue: 1

Month: May

Year: 2026

P-ISSN: 2321-4643

E-ISSN: 2581-9402

Citation:
Valan, Sudha, and Vikas Kumar. "An Empirical Study on Trust, Ethical Concerns, and Bias in AI-Powered Educational Chatbots." *Shanlax International Journal of Managment,* vol. 13, no. S1, 2026, pp. 13-28.

DOI:
https://doi.org/10.34293/management.v13iS1-i3-may.11026

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

Mrs. Sudha Valan
Student, Department of Computer Science & IT
Chhatrapati Shivaji Maharaj University

Dr. Vikas Kumar
Head, Department of Computer Science & IT
Chhatrapati Shivaji Maharaj University

Abstract

In contemporary ages, Artificial Intelligence has swiftly changed all the paces of education. AI tools, such as ChatGPT, and other chatbots are being widely harnessed to support students in their learning. As AI-driven chatbots become more prevalent and engaged in the education landscape, key issues of trust, ethics, and anticipated bias in automated interactions are becoming a major focus. 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. To fill this gap, this study used a mixed-method research approach, where both systematic literature analysis and primary data collection was performed. A total of 134 research papers were read, from which 29 papers were chosen for detailed analysis to identify educational chatbots' dimensions of trust, ethics, and bias. To build on this, an investigation study was conducted on a sample of n = 120 students. They were divided into two groups. One group was engaged with an AI Chatbot developed by the researchers and another group was trained through conventional teaching methods including giving notes to the group and giving structured assessments. Both groups were given a standardized survey instrument to assess their perception of trust, ethical reliability and bias in learning interactions. It was concluded that there is a positive influence of the use of chatbot in learning on perceived accessibility and perceived engagement, and the level of trust increased by about 18.4% from the traditional group. Likewise, the score for ethical perception rose by about 13 percent, which meant that it became more in line with the principles of fairness, transparency, and accountability. The analysis of bias-related parameters, however, showed that there was an almost 8% mean bias indicating in need for further refining to reduce bias and make it more neutral to ensure the equality of responses. The differences observed was validated statistically by two sample t-test and it was found to be 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.

Introduction

Artificial Intelligence (AI) is now a part of the human being and is very much interwoven in the day to day 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 offer personalized assistance, immediate feedback, and increased access, which in turn boost student engagement and learning productivity (Holmes et al., 2019; Zawacki-Richter et al., 2019). Chatbot technologies are becoming more popular in modern times for educational institutions to use as an additional teaching tool to support their learning processes on a scale of many students (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). Therefore, it is imperative to have an understanding of its concept, history, and studies already carried out on chatbots.

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). Here is a list of AI chatbots that are widely used in education.Below is a list of the most popular AI chatbots used in education–

Table 1 Common AI Chatbots used in Education

Year	Chatbot	Source	Used for
2023	Bard	(Rudolph, 2023)	• Research assistance • Producing different kinds of contents • Provide responses to queries
2022	ChatGPT	(Dergaa, 2023; Khademi, 2023; Rudolph, 2023)	• Generate text • Produced creative content • Delivered enlightening replies to the queries raised by the learners
2019	IBM Watson Assistant	(Oliveira et al., 2019)	• FAQs • Student Support Chatbot
2017	Ada	(Kabiljo et al., 2020; Konecki et al., 2023)	• Provide coaching to individual students • Promote personalized education to students • Deliver replies and feedback
2017	Replika	(Pentina et al., 2023; Xie & Pentina, 2022)	• 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)	• Student engagement • Student retention
2016	Duolingo Bot	(Loewen et al., 2019)	• Language learning • Interactive dialogue practice
2016	Jill Watson	(Goel & Polepeddi, 2016)	• Act as Teaching Assistant • Answer student queries in online courses
2013	Socratic	(Alsanousi et al., 2023; Moppel, 2018; St-Hilaire et al., 2022)	• Chatbot like property and functionality • Assist students in learning new concepts

2013	Habitica	(Sales & Antunes, 2021; Zhang, 2023)	• Support students acquire good study traditions • Project learning process as a game
1995	Alice	(AbuShawar & Atwell, 2015)	• Rule-based conversational tutoring
1966	Eliza	(Weizenbaum, 1966)	• Early conversational learning model

Table It lists the most popular AI Chatbots for education, from the early days of conventional chatbots like ELIZA to the latest generative AI systems such as ChatGPT.

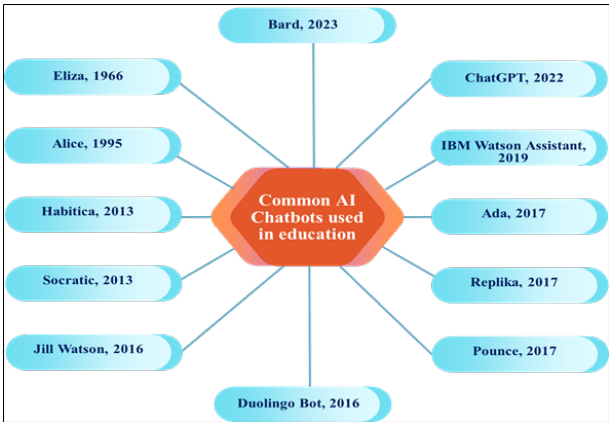

Figure 1 Common AI Chatbots used in Education

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 serve as a teacher, tutor mentor or administrative assistant, inviting students to appeal for learning provision anytime and anywhere. Through constant efforts to enhance students' engagement, access to the educational contents, learning outcomes, and the success and impact of these arrangements can be felt (Huang & Su, 2023). Apart from these benefits, the amalgamation of AI in education elevates serious issues pertaining to trust, ethics, and bias. Therefore, it is essential to evaluate AI-powered chatbots in education based on these critical metrics, including trust, ethics, and bias, that impact their adoption. Consequently, it is important to grasp their importance in education as the three factors.

Concept of Trust: Xu et al. (2022) describes the concept of trust as the willingness of the user to depend on the AI system to provide accurate and helpful guidance.Ethical Considerations and Bias Assessment in AI-Powered Educational Chatbots: The willingness of the user to trust the AI system for accurate and helpful guidance (Xu et al., 2022). Trust in learning systems promotes commitment, acceptance of feedback, and learning outcomes. The accuracy, reliability, and consistency of chatbot systems influence students' trust and engagement in the learning process, impacting their willingness to interact with AI-supported education (Hasan et al., 2023; Labadze et al., 2023). The AI tutors are more likely to be agreeable with the student's actions, often seeming to be helpful, which could lead the student to accept and acknowledge responses and answers from the AI without needing to justify their actions to the teacher (McGrath, 2024).

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). Having teachers supervise also ensures

the AI's output will meet learning objectives and moral 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). An overlooked bias can erode trust, impact educational outcomes, and exacerbate educational disparities (Lee et al., 2021). These dimensions have far reaching implications. Trust and ethical alignment can positively influence students' confidence, engagement, and satisfaction, while bias can adversely impact credibility and hinder the effectiveness of AI tools in education (Labadze et al., 2023; Bettayeb et al., 2024). Thus, it is imperative to consider these aspects balanced to ensure the responsible and efficient use of chatbots in education.

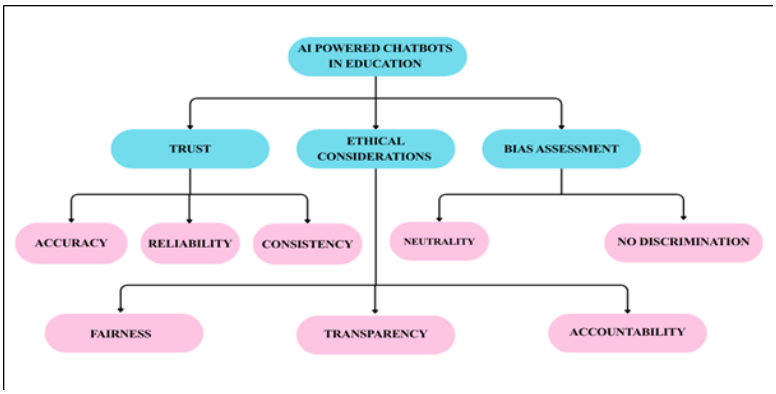

Figure 2 Trust-Ethics-Bias Framework for AI-powered Educational Chatbots

The conceptual model employed in this study is shown in Figure 2, which depicts the three dimensions of evaluation of interest in this study: trust, ethics, and bias in relation to AI-based educational chatbots.

The literature highlights the capabilities and constraints of AI-powered educational technologies. Despite the various studies that have been published that have shown improvements in learner engagement and performance, there are still concerns around algorithmic bias, lack of transparency, and over-reliance on automated systems (Luckin et al., 2016; Holmes et al., 2019). But, there is a clear research gap in empirical research which examines trust, ethical perception, and bias in AI-powered educational chatbots simultaneously using an organized experimental methodology. To investigate this gap, this study empirically examines the effectiveness of the use of AI chatbots in educational scenarios. A comparative experimental method was adopted, which involves the use of an AI-based chatbot in an experimental group of students and the use of other traditional teaching methods such as notes and assessment among the control group. A questionnaire was used to assess users' perception on several trust, ethics and bias parameters in a structured way. Findings from this research reveal that AI-driven chatbots have a definite positive impact on trust and ethical perception, making measurable strides in users' trust, fairness, and transparency. The study also notes a lack of parameters related to the bias, indicating a need for more refinement to ensure neutrality and inclusivity in AI systems. This study contributes to the expanding field of AI in education by offering a well-rounded and fact-based assessment of chatbot systems.

It outlines the benefits and challenges involved in implementing AI technology and underscores the need for creating AI tools that are responsible, ethical, and unbiased in educational settings.

Problem Statement

AI chatbots have gained a lot of ground as virtual classroom learning assistants. However, there is no look into the trustworthiness, fairness and ethical use of the AI Chatbots. The purpose of this study is to measure students' and teachers' attitudes towards trust, reliability and ethics in the use of AI-powered chatbots in the dominant platforms, to explore potential biases in chatbot responses, and to examine the role of explainability and transparency in building trust and acceptance among users.

Objectives

- To understand the trust, ethical considerations associated with AI-powered chatbots.
- To evaluate AI chatbots' biases.
- To assess the trust, fairness and usability by conducting surveys and experimental analysis.
- To recommend the use of chatbots in the field of education using artificial intelligence in the future.

Methodology

Research Design

The current study uses an empirical combination of systematic literature analysis, experimental research and the survey-based evaluation. The objective is to measure the level of trust, ethical concerns, and bias in AI-powered educational chatbots, both theoretically and practically, by engaging with them.The goal is to assess how trustworthy, ethical and biased these learning chatbots are, both theoretically and practically, through interaction with users.

Literature Review Process

To identify key dimensions relating to trust, ethics, and bias in AI-based educational systems, a comprehensive review was conducted with 134 research papers published from 2018 to 2025. It was retrieved from various databases such as SpringerLink, ACM Digital Library, IEEE Xplore, Research gate, Scopus and Google Scholar. Among the 29 studies identified, 21 were selected, based on the inclusion-exclusion criteria, which included relevance to educational chatbots, empirical validation, and the discussion of trust focussed concerns, ethical and/or bias-related. The following analysis of the selected studies aimed to reach the following goals: Identify parameters for evaluation (trust, ethics, bias), Understand the existing methodologies and Establish research gaps for empirical validation

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.

Table Includes the inclusion and exclusion criteria that outline the conditions under which relevant and/or irrelevant studies are included in and excluded from the review process.

Literatures were thoroughly reviewed in accordance with the guidelines of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Flow Diagram 2020. A visual chart of the process of identifying, screening, eligibility checking, and finally selecting research articles for the review.

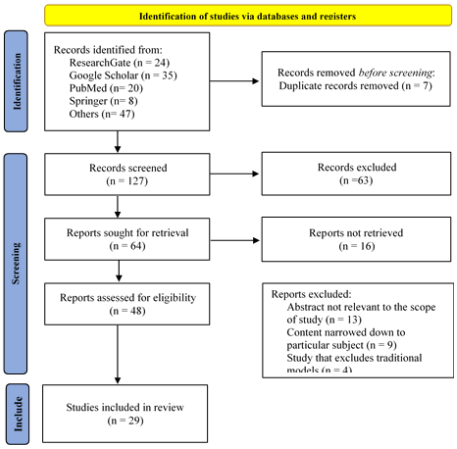

Figure 2 Prisma Flow Diagram (adapted from PRISMA 2020)

The structured framework of identification, analysis, screening and inclusion/exclusion criteria of this research process is shown in Figure 2.

A Summary of the Literature Review

AI-powered chatbots are being used in various sectors, including education, healthcare, and customer support, due to the rapid development of artificial intelligence (AI) and large language models (LLMs). These systems improve accessibility and automation, but there are increasing concerns about their accuracy, reliability, confidence, fairness, transparency, accountability, neutrality, consistency, and non-discrimination. In recent years, the importance of incorporating technical performance and ethical aspects into chatbot assessment frameworks has been highlighted more and more

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. (2017)	Customer Service Bots	Moderate	Low	–	–	–	Enhances engagement 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

Table 3 In the existing literature, it is evident that the evaluation parameters are examined separately, and very few studies explore an integrated approach to a combination of parameters, such as performance (accuracy, reliability) and ethical (fairness, transparency, accountability). This illustrates an important research need.”.”

Participants and Sampling

Participants and Sampling

The number of students in the study was n = 120 students, which was obtained by convenience sampling technique. The students were split into two groups:

The experimental group (n = 60) had access to the AI-driven chatbot.The experimental group (n = 60) received the AI-driven chatbot.

Experimental Group (n = 60): Received traditional instruction with an additional component of online learning (notes, tests, and online learning support)

The two groups were guaranteed the same educational background so that they could be compared and contrasted.

Chatbot Development and Experimental Setup

It has been developed by using the Botpress framework, an open-source conversational AI platform that supports natural language understanding and workflow-centric dialog management helping students complete learning tasks such as solving problems, explanations, and answering questions about the subject. The knowledge base was sifted from specific knowledge sources to ensure its significance and reliability. The group that was assisted by the AI-chatbot used the chatbot within a given study time, and the conventional group used traditional learning approaches. The same learning content was presented to both groups to provide a fair evaluation of both.

Data Gathering Method

A structured survey questionnaire was used to collect data for both groups following the learning phase. Items in the questionnaire were developed based on the literature review and covered aspects of: Trust (accuracy, reliability, confidence in the answers), Ethics (fairness, transparency, accountability) and Bias (neutrality, absence of discrimination, consistency). A Likert scale (e.g., 1 – 5) was used to quantify user perception for the responses.

Evaluation Metrics

Trust Score: Depends on the user's confidence and the reliability of the website.
Ethical Compliance: Fairness, transparency and accountability
Bias Indicator: Based on skewed responses and/or inconsistent responses
Comparative analysis was carried out between experimental and control group.

Data Analysis Techniques

Students were informed about the needs of the study and their responses were taken anonymously. All participants gave informed consent and data confidentiality was ensured.

Ethical Considerations

The participants were told the purpose of the study and their answers were anonymous. Participants gave their informed consent and data confidentiality was ensured.

Research Questions

Why do students differentiate fidelity and usefulness of AI-powered chatbots?

Are the responses from the chatbots biased (gender, culture, content) and could this affect the equity of learning?

What do you think are some of the design suggestions that help you to promote ethical use of chatbots in classrooms?

Results and Analysis

The study evaluated the impact of AI-driven chatbots on student trust, ethics, and biases in three important factors. Comparative analysis was done between AI-Chatbot group (AI based learning) and conventional group (conventional learning methods).

Survey Response Overview: 59 out of 60 participants from conventional group responded while 58 out of 60 participants from AI-chatbot assisted group responded which can be considered as a very good response as shown below in the table –

Table 3 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

Source: Primary Survey

Trust Evaluation and Analysis

The group that interacted with the AI-chatbot indicated a significantly greater level of trust (Mean = 4.3) than the conventional group (Mean = 3.6). The value of 19.4% shows that the students felt that the chatbot-based learning was more reliable, accurate, and more inspiring to

give trust in the learning process. The difference was highly significant ($p < 0.05$) as confirmed by statistical analysis by a two-sample t-test. The results suggest that AI chatbots can enhance learners' confidence and engagement by providing immediate and consistent responses.

Table 4 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

Source: Primary Survey

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

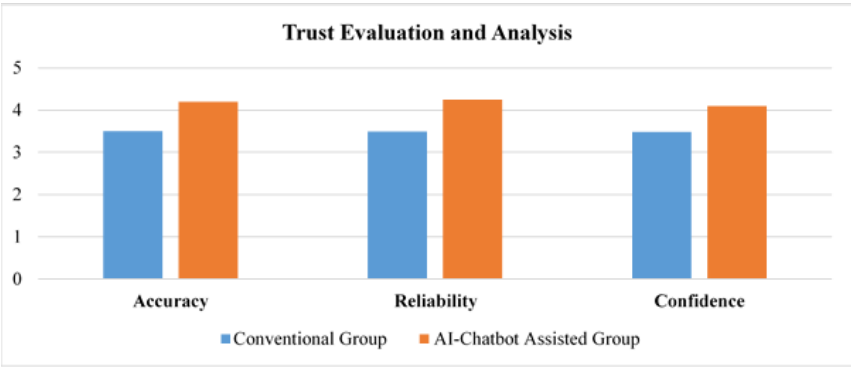

Figure 3. Trust Evaluation Graph

Figure 3. depicts 18.4 % increase in trust through each of its factors viz. Accuracy, Reliability and Confidence

Ethical Evaluation and Analysis

The group that was assisted by the AI-chatbot was also significantly higher in ethical perception (Mean = 4.03) than the conventional group (Mean = 3.56). This indicates that students perceived the chatbot's responses as being fair, transparent, and accountable to some extent. The difference between the groups was found to be statistically significant ($p < 0.05$). This underscores the capacity of good AI systems to conform to moral expectations in the field of education.

Table 5 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.

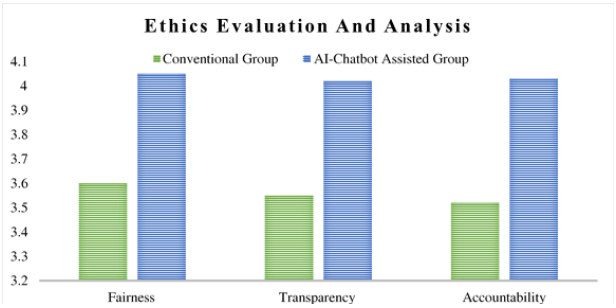

Figure 3 Ethical Evaluation Graph

Figure 3 shows the positive contribution in ethics parameters viz. Fairness, transparency and accountability.

Bias Evaluation

The conventional group on the other hand, performed better than the experimental group in the bias dimension. The scores of the traditional group are higher than those of the AI-Chatbot assisted group, with a mean score of 4.05 vs. 3.72, indicating that students perceived traditional approaches as less biased and more neutral. This difference was also statistically significant ($p < 0.05$), and suggests continued concerns about the potential for bias in AI-generated responses.

Table 6 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.

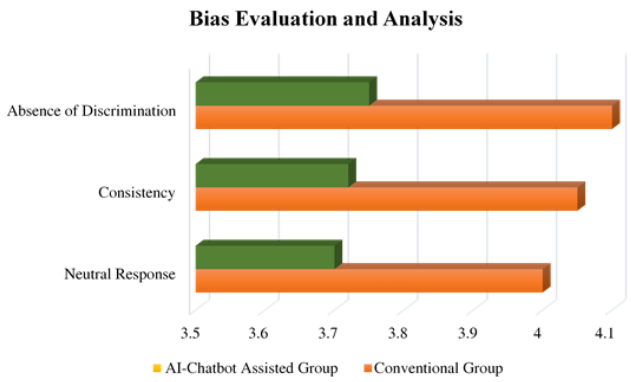

Figure 4 Bias Evaluation Graph

According to Figure 4, there is an 8% room for improvement with regard to the bias-related parameters, indicating opportunities for improving the neutrality and the potential for reducing the biasness of chatbot responses.

Overall Evaluation and Analysis

The results show that AI-driven chatbots: Dramatically enhance trust and ethical perception while still needing to address issues of bias and neutrality. To mitigate bias-related concerns, there are strategies for chatbot-based learning that can be employed, including: Improved training data sets, Ethical AI design, and Transparency mechanisms.

Table 7 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

Source: Primary Survey

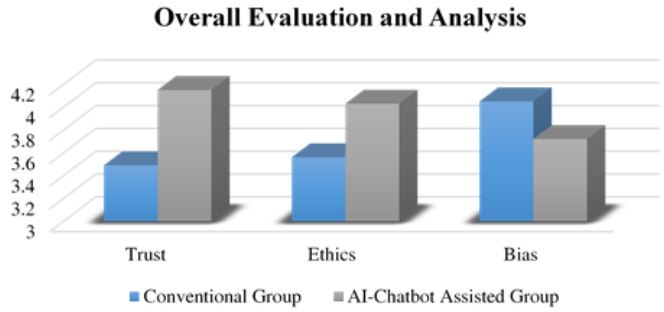

Figure 5 Overall Evaluation Graph

Statistical Analysis

The statistical analysis renders the experimental results more valid and builds up the reliability of the proposed assessment system. Therefore, to determine if the differences noted 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 performed. Both groups were compared with respect to the mean scores of evaluation parameters such as trust, ethics and bias using t-test. The analysis was conducted with a two tailed test with equal variance assumption because the sample size is equal in both groups (n = 60 per group), and they were treated with the same learning conditions.

The null hypothesis (H_0) and alternative hypothesis (H_1) for the test were defined as follows:

H_0 (Null Hypothesis): There is no noteworthy difference between the AI-chatbot assisted and conventional groups pertaining to trust, ethics, and bias evaluation.

A statistical method (such as MS Excel function: T) was used to calculate the t-test.Where, array1 is the data of the group assisted by the AI-chatbot, and array2 is the data of the conventional group, 2 means a two-tailed test, and 2 means equal variance.

Table 8 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

Source: Primary Survey

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 and Gaps in bias and neutrality are meaningful and require attention

Hypothesis Testing

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

H_1: AI-powered chatbots significantly improve user trust related to traditional learning methods. Result: Supported, Evidence: Trust increased by ~18–19% and t-test shows $p < 0.05$. Conclusion: AI chatbots significantly enhance trust among learners.

H_2: AI-powered chatbots improve ethical perception in educational environments.

Result: Supported, Evidence: Ethics improved by ~13% and Statistically significant ($p < 0.05$) and concluded that AI systems positively influence fairness, transparency, and accountability.

H_3: AI-powered chatbots reduce perceived bias compared to traditional methods.

Result: Not Supported, Evidence: Control group scored higher in bias ~8% gap observed and Significant difference ($p < 0.05$) and is concluded that 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 and Transparent model behavior.

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

Conclusion

AI-powered chatbots have great power as learning tutors in virtual classrooms. The successful implementation of AI chatbots hinges on the principles of trust, ethical usage, and mitigating biases. Bias assessment ensures equality, trust is established through precise, open, and sympathetic communications, ethics through confidentiality, fairness, and responsible strategies. Essential to enhance gains and minimize risks is trustworthy project, teacher involvement and ongoing auditing. By prioritizing these considerations, chatbots can positively contribute to learning and create unbiased and accurate educational experiences.

This study combined the systematic literature review and an experimental study with a survey analysis to empirically assess trust, ethics, and bias in AI-based educational chatbots. Of the 134 research papers initially reviewed, 29 studies were selected for their relevance, and their evaluation dimensions were identified. Based on these findings, an experimental design was developed with two sets of students: one set with an AI chatbot and one set with the traditional method of learning. The results show that AI Chatbot learning is highly effective in improving user trust, which increased by around 18% compared to the traditional learning model. The results also show improvements with accessibility and engagement. But some issues were raised about ethical transparency and possible bias, such as explainability of responses and some inconsistencies in generated outputs. The statistical validation of the observed differences between the two groups was conducted using a t-test and the results were found to be statistically significant ($p<0.05$), thus ensuring the reliability of the obtained results.

The study's findings suggest that AI chatbots can be a significant asset in educational environments, but they must be carefully designed and developed to minimize potential biases and ethical concerns.

Gaps Identified

Although this study has made contributions, there still are some gaps in the research:

The lack of explainability in AI Chatbots: AI systems, such as the one utilized in this study, have limited explainability and user trust during critical situations.

Bias in Context-Specific Responses: The study indicated some degree of bias in some instances; however, the underlying issues (training data, model limitations etc.) were not thoroughly examined.

Short-Term Evaluation: The experiment was carried out for a short time. The impact of chatbot-based learning on trust and academic outcomes are yet to be studied and researched.

Subjective Data: Some of the conclusions in the study may be subjective and are limited to the participants' experiences.

Weaknesses: Not enough Relative Analysis Across Multiple AI Systems (only one chatbot was evaluated). Multiple AI tools comparison might give more general insights.

Future Recommendations

Future research should investigate specific measures to reinforce trust, ethics, and fairness in the implementation of chatbots, including the construction of real-time EAI interfaces that create a graphical display of decision paths in easily understood formats, as well as the creation of models that are culturally adapted to reflect variations in region, such as tone adjustments for regional dialects or community-specific values. Further work could involve longitudinal evaluation, incorporating user feedback cycles to incrementally mitigate biases, and hybrid models combining rule-based ethical safeguards with machine learning to create more resilient and context-aware user interactions. These developments would pave the way for fairer AI companions in various real-world scenarios. The following recommendations are made based on the findings and gaps identified:

Application of Explainable AI (XAI)

Future systems should include features that are more explainable, thereby enhancing transparency and trust by the users.

Bias Detection and Mitigation Mechanisms: Create enhanced methods to Detect bias in the real-time and Ensure fairness of the response

Long-Term Studies: Carry out longitudinal studies to assess Trust evolution and Learning outcomes over time

Different academic levels (Larger and Diverse Sample Size): Future research needs to incorporate different academic levels and Diverse socio-cultural levels of participants.

Comparative Evaluation of Multiple Chatbots: Analyze and Compare Different AI models and Open source vs proprietary systems;

Hybrid Learning Models: Combine AI chatbot assistance and Traditional teaching methods

The findings of this study provide a basis for creating a responsible future learning environment by emphasizing the importance of trust, ethical integrity, and bias-aware system design in the responsible adoption of AI-powered educational chatbots.

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., Albesher, 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.
46. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71