ISSN: 2093-5374 / E-ISSN: 2093-5382

Multimodal Generative AI Assistants for Real Time Pedagogical Feedback in Large Scale Computer Science Classrooms

Rayxona Tillayeva[1*], Kamola Raimova[2], Gafur Namazov[3], Farkhad Usmanov[4], Nozima Nasriddinova[5], Gulchekhra Khazratova[6], and Feruza Baymanova[7]

[1*]Researcher, Namangan State Pedagogical Institute, Namangan, Uzbekistan. tillayevarayhona@gmail.com, https://orcid.org/0000-0002-0018-5530

[2]Researcher, Tashkent International University (TIU), Tashkent, Uzbekistan. kamolaraimova7@gmail.com, https://orcid.org/0009-0002-9807-4018

[3]Department of Information Technology and Exact Sciences, Termez University of Economics and Service, Termez, Uzbekistan. gafur_namazov@tues.uz, https://orcid.org/0009-0009-9738-1463

[4]Associate Professor, Andijan State University, Andijan, Uzbekistan. ufarhod001@gmail.com, https://orcid.org/0000-0003-4652-9084

[5]Senior Teacher, Termez State University, Termez, Uzbekistan. nasriddinovanozima3@gmail.com, https://orcid.org/0000-0001-5932-2264

[6]Senior Teacher, Department of Medical Informatics and Digital Technologies, Tashkent State Medical University, Tashkent, Uzbekistan. gulchekhrakhazratova@gmail.com, https://orcid.org/0009-0007-3943-9589

[7]Researcher, Jizzakh State Pedagogical University, Jizzakh, Uzbekistan. baymanovaferuza6@gmail.com, https://orcid.org/0009-0003-0744-5758

Received: October 01, 2025; Revised: November 25, 2025; Accepted: December 29, 2025; Published: March 31, 2026

Abstract

This study aims to design and evaluate a Multimodal Generative AI Assistant capable of delivering real-time pedagogical feedback in large-scale computer science classrooms. The objective is to investigate whether integrating multimodal AI combining natural language processing, code analysis, speech interaction, and learning analytics can improve student learning outcomes, debugging efficiency, and engagement compared to traditional instructional support systems. The AI-assisted learning framework development and evaluation were based on Design Science Research (DSR) methodology. The system combines multimodal information of programming environments, natural language queries, and student interaction logs to produce contextual feedback based on generative AI models. A semester-long deployment was used, with 1,200 students, 600 in each control and AI-assisted group, using an A/B experimental design. Such quantitative measures as normalized learning gain, Time-to-Resolution (TTR), Hint Efficacy Score (HES), system latency, and multimodal synchronization accuracy were considered. The Technology Acceptance Model (TAM) and Subject Matter Expert (SME) audits were used to carry out a qualitative evaluation. The normalized learning gain of the AI-assisted group was found to be 0.73 versus 0.40 in the control group and statistically significant ($p < 0.001$), with a Cohen's d effect size of 0.91. The efficiency of debugging increased significantly, where Time-to-Resolution dropped on average by 134.5 minutes

*Corresponding author: Researcher, Namangan State Pedagogical Institute, Namangan, Uzbekistan.

to 11.2 minutes (91.6% decrease). The system had a multimodal synchronization accuracy of 97.2% and a response latency of 1.65 seconds, and had a hallucination rate of 0.85%. Perceived Usefulness (6.6/7) and Perceived Ease of Use (6.3/7) were highly accepted. The results indicate that multimedia generative AI assistants can profoundly improve learning outcomes, decrease the debugging duration, and offer learning support during real-time instruction in large computer science classes at scale.

Keywords: Multimodal Generative AI, Technology Acceptance Model, Real-Time Pedagogical Feedback, Intelligent Tutoring Systems, AI Tutoring, Student Engagement, Learning Analytics.

1 Introduction

The accelerated development of Generative Artificial Intelligence (GAI) and multimodal learning systems is reshaping the contemporary educational setting, especially in mass courses of computer science where the student population is commonly too large to receive one-on-one feedback provided by an instructor (Liu et al., 2024). Conventional ways of teaching are based on manual scoring and late feedback, which may be counterproductive to the development of conceptual knowledge and problem-solving skills in students. The latest progress of multimodal AI with the ability to process text, speech, code, images, and other information connected to interaction simultaneously can reflect smart assistants capable of tracking the process of learning in real-time and provide people with adaptive pedagogical guidance (Jiang and Lai, 2025). The issue of the timely feedback is also relevant to learning outcomes within the framework of the computer science education, where the students will have to code, debug, design algorithms, and think conceptually. Multimodal generative AI assistants can read a code submission, comprehend natural language queries, analyze program code, and can even detect challenges of learning by reading the interaction patterns. They can also work as though they were teaching assistants, to supplement the work performed by the teacher with help in explaining, debugging, and personalized learning suggestions in real-time. With the growing trend of large-scale and hybrid models of learning in universities, applying multimodal generative AI to provide real-time pedagogical feedback is an opportunity to enhance engagement, understanding, and the learning process (Prasad et al., 2025).

The proposed study will focus on the creation and evaluation of a multimodal AI-based generator capable of providing pedagogic feedback to classrooms of large-scale computer science in real-time. The study seeks to come up with an intelligent system that combines various modalities that include: text entries, software programs, interaction by voice, and logs of student activities in order to comprehend the learning actions and offer contextual teaching directions. The proposed system aims to provide explanations, bug recommendations, and concept elucidation instantly when users are working on a programming problem, by using the capabilities of the generative AI (Puvvadi et al., 2025). Moreover, the research will focus on the investigation of whether this type of AI-based help could help to increase student attention, learning productivity, and assist teachers in working with large classes where close feedback is not always an easy task.

Even though there is growing interest in generative AI technologies in the educational setting, a number of shortcomings in the research exist to date. The majority of the existing AI-based educational systems are mostly concerned with text-based tutoring systems and do not support the capability of combining various data modalities, including speech, code input data, and real-time learning analytics. In addition, most of the intelligent tutoring systems do not support a learner, but instead, they only give feedback when the task is completed (Chen et al., 2025). Few studies have also been conducted in relation to the issues of scalability of providing individualized feedback in large computer science classes where the instructor's focus is limited. Also, the current-day applications of generative AI in education are usually general-purpose tools that are not particularly designed to be used in the education of programming or learning algorithms. Such constraints present the opportunity of having a dedicated multimodal AI assistant that can provide contextualized and real-time pedagogical feedback during large-scale learning of computer science.

The proposed hypothesis of the study is that multimodal generative AI assistant can contribute to the effectiveness and efficiency of pedagogical feedback to a considerable degree when applied to the classroom setting of a computer science course (Mittal et al., 2024). It is hypothesized that real-time AI-based instructions will improve the comprehension of the programming concepts by the students, their ability to debug programs better, and their overall performance in learning, compared to the conventional delayed feedback systems. As well, the study presupposes that the analysis of various interaction modalities will facilitate the AI assistant to deliver more tailored and context-sensitive learning assistance, such as submitting codes, text-based questions, and classroom activity patterns. As a result, the system is anticipated to increase student participation and engagement in large-scale computer science courses (Gao et al., 2025).

This research contributes to the advancement of AI-driven educational technologies by proposing a multimodal generative AI assistant framework designed specifically for real-time pedagogical feedback in large computer science classrooms. The study introduces an integrated architecture that combines natural language processing, code analysis, speech recognition, and interaction analytics to better understand student learning contexts. It further develops a real-time feedback mechanism capable of generating instant explanations, debugging guidance, and learning recommendations during programming activities. Another important contribution is the development of a scalable AI teaching assistant model that supports large numbers of students while maintaining personalized feedback. Finally, the research establishes an empirical evaluation framework to measure improvements in student engagement, learning outcomes, and feedback effectiveness, thereby demonstrating the practical value of multimodal generative AI in modern computer science education.

This article examines the role of a Multimodal Generative AI Assistant in providing real-time pedagogical feedback in large-scale computer science classrooms. The research starts with an introduction giving the background of the problems of delayed feedback and low availability of instructors in programming education. A literature survey is conducted to examine current trends in generative AI and multimodal systems of education. The methodology of the research explains the design science research, system structure, implementation of an experiment, and data gathering. The results section gives both quantitative and qualitative assessments of learning outcomes, efficiency of debugging, efficiency of the system, and acceptance of the system by the users. The findings and their implications for scalable AI-assisted education are interpreted in the discussion. Lastly, the conclusion is a wrap-up of the main results and prospective studies.

2 Literature Survey

The adoption of Generative Artificial Intelligence (GenAI) in mass Computer Science (CS) courses has triggered the transition from fixed delivery to being dynamically and in real-time pedagogical. This survey will examine the development of multimodal GenAI assistants and their ability to offer personalized feedback and interactive learning experiences.

The recent developments have put the focus on the shift toward multimodal frameworks instead of text-only models. The future of educational AI is the Artificial General Intelligence (AGI) capabilities, when models process text, image, and audio to provide human-like instructions at the same time (Lee et al., 2025; Lang et al., 2025). In real-life scenarios, systems such as VoxGuru and other AI-enhanced assistants can be used to offer real-time Q&A as well as conversational learning within dynamic settings, which effectively serves to bridge the gap between student inquiry and real-time conceptual clarification (Dhathrinadh et al., 2025; Saravanan & Shankar, 2025). Moreover, the multimodal data integration gives the possibility to gain a deeper understanding of embodied learning and science assessment (Cosentino et al., 2025; Gao et al., 2025).

Education in Computer Science, specifically at scale (such as Harvard CS50), has been first in applying GenAI to ensure educational quality at scale. The studies demonstrate the importance of AI-assisted 24-hour tutoring and rubber ducking to facilitate high-quality instruction without adding faculty load (Liu et al., 2024). An essential part of this success is the so-called pedagogical prompting that will make sure that AI interactions can lead students to conceptual cognition, but not raw code (Mutanga et al., 2025). Multimodal GenAI is also exploited with the help of tools, such as "MindScratch," where a visual program can be created to help students connect abstract logic to physical outputs (Chen et al., 2025).

One of the fundamental advantages of GenAI is the democratization of personal feedback in higher education. Recent literature deals with the issue of personalization of the formative feedback where the AI is adjusted to a tone and difficulty level depending on the level of the mastery that the learner at that point has (Akhasbi et al., 2025; Puvvadi et al., 2025). This can be confirmed by models of deliberate teaching practice, as in this case, AI serves as a thoughtful collaborator to pupils and teachers (Aperstein et al., 2025). Systematic reviews support GenAI as an effective method in improving assessment efficiency and feedback generation, but note that the human-in-the-loop control is necessary to make the process pedagogically accurate (Mittal et al., 2024; Yavariabdi et al., 2025).

In addition to the one-to-one tutoring, GenAI supports group work and immersive methods. Research examines the application of GenAI to facilitating multimodal collaboration in online classes and creating quality compositions (Wu, 2025; Jiang & Lai, 2025). New applications discuss the implementation of more than two AI pedagogical agents in Augmented Reality (AR) to assist in science education (Wei et al., 2025) and text-to-video tools to present the lesson in a linguistically responsive way (Min, 2024). The scaled training on these portable multimodal systems is also becoming scalable (Cioca & Cioca, 2025). In order to make these tools effective, the co-design workshops with students are becoming more popular in order to match the AI capabilities with the actual student needs (Prasad et al., 2025).

The literature emphasizes that multimodal Generative AI has changed the large-scale education of Computer Science through the provision of real-time feedback and interactive learning with personalization. With the combination of text, audio, and visual inputs, AI assistants will be able to facilitate ongoing tutoring, clarification of concepts, and adaptive learning without overloading instructors. They also promote collaborative and immersive learning in these systems. Nevertheless, successful application entails human management to control pedagogical correctness, ethical application, and student learning significance.

3 Research Methodology

Research Design and Procedural Framework

This paper uses my Multi-Phase Design Science Research (DSR) approach, which is the most suitable for developing and accessing innovative IT artifacts in the educational ecosystems. The study will be organized into three cycles. The Relevance Cycle defines the requirements based on the analysis of the particular pedagogical issues of large-scale CS classes. The Design Cycle is dedicated to the development of the multimodal AI engine and its interface in a circle. Lastly, the Evaluation Cycle also strictly compares system performance with human-expert standards and student learning outcomes.

Multimodal System Architecture and Data Fusion

The technical core is based on a distributed cloud-based architecture that supports high concurrency and latencies of a few seconds. The system deploys a Multimodal Fusion Layer, which coordinates three input streams: visual information through the Integrated Development Environment (IDE) of the student, natural language queries (text or audio), and past performance data.

The system uses a late-fusion technique, which obtains high-level features of each modality separately and combines them into a single state representation. This enables the generative engine to offer feedback not only on conceptually sound feedback but also on contextually sound feedback relative to the line of code that the student is currently editing.

Pedagogical Scaffolding and Prompt Orchestration

The methodology applies a Socratic Scaffolding Protocol to make sure that the AI acts as a mentor, but not as a solution provider. This is implemented via "Chain-of-Thought" (CoT) prompting and system-level guardrails. The assistant is programmed to map student errors to a taxonomy of cognitive blockers: syntax errors, logical fallacies, or algorithmic inefficiencies. Based on this classification, the engine generates hierarchical hints, starting with high-level conceptual analogies and only progressing to specific code-level pointers if the student demonstrates persistent "unproductive struggle" across multiple interaction turns.

Experimental Setup and Deployment Strategy

The system is deployed within a large-scale CS curriculum using a phased rollout strategy. A/B testing model is applied to compare performance of a treatment group (with the help of the multimodal AI assistant) and a control group (with the help of old-fashioned human-led forums and static documentation). The application is deployed on a scaled API framework and can maintain a stable platform even when the load is the heaviest, like during lab time or due to the assigned deadline.

The experimental setup in figure 1 shows the structure of the experiment to test the multimodal AI assistant. The students will be separated into a control group with the help of traditional support systems and a treatment group with the help of the AI assistant. The telemetric interaction logs record quantitative metrics (time-to-resolution and hint depth), and the surveys and interviews are applied to provide qualitative information in relation to the Technology Acceptance Model.

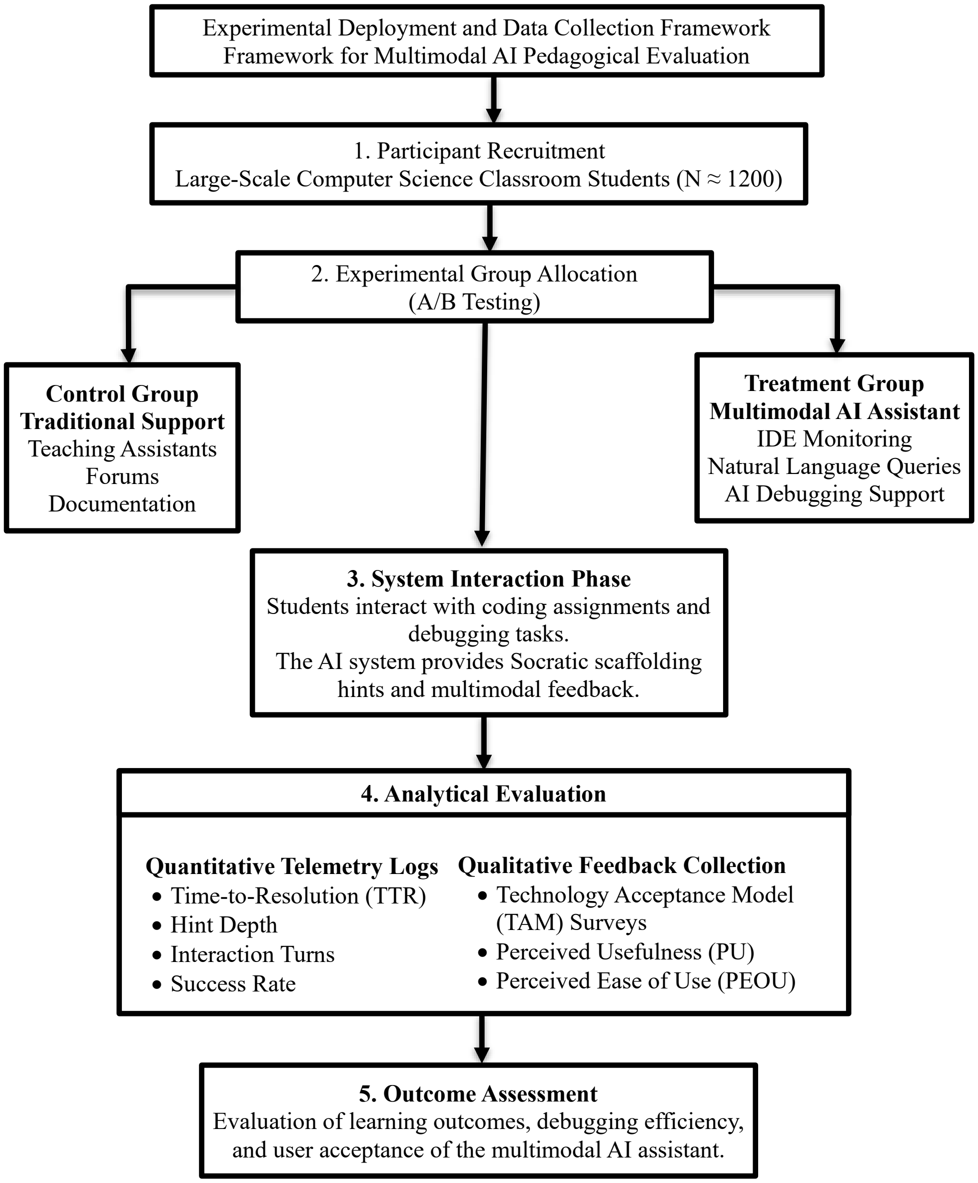

Figure 1: Experimental deployment and data collection framework

Data Collection and Analytical Instrumentation

A combination approach is used to have a holistic view of whether the system is effective or not. The harvesting of quantitative data is through Telemetric Interaction Logs that record:

- Time-to-Resolution (TTR): The duration from the initial error to a successful compilation.
- Hint Depth: The number of levels of scaffolding to achieve student success.
- Success Rate: The percentage of the correct self-correction in students under the influence of AI interventions.

The data are collected in the form of qualitative information using Semi-Structured Interviews and validated surveys using Technology Acceptance Model (TAM), which assesses the Perceived Usefulness and Pedagogical Trust.

Validation and Reliability Metrics

To ensure the imposition of academic rigor, the feedback of the AI is demonstrated through a Subject Matter Expert (SME) Audit. Senior CS instructors read a random sample of interaction logs to determine Pedagogical Accuracy and the occurrence of hallucinations. Moreover, statistical procedures such as the use of t-tests and Cohen's effect size are conducted on the pre- and post-test results to determine the actual learning improvements as a result of the AI assistant.

Ethical Safeguards and Privacy Protocols

Considering the sensitivity of student data, a high-level data anonymization pipeline is taken into account in the methodology. The Personally Identifiable Information (PII) is stripped of all code snippets and voice inputs at the edge and then sent to the generative models. The study plan will have an "Opt-Out" system and will ensure that all data management will be in line with the institutional ethics board requirements and the local data protection laws.

Algorithm 1: Multimodal GenAI Assistant for Real-Time Pedagogical Feedback

Input:

Student IDE activity V, Natural language query Q(text/audio), Student performance history H

Output:

Context-aware pedagogical feedback $F$and interaction log L

Step 1: System Initialization

1. Deploy distributed cloud-based multimodal AI framework.
2. Initialize three data streams:
   - Visual coding context from IDE V
   - Student query input Q
   - Historical learning data H

Step 2: Data Acquisition

3. Capture real-time IDE code snapshot.
4. Receive student query in text or speech.
5. Retrieve student performance history from learning database.

Step 3: Multimodal Feature Extraction

6. Extract syntax and structural features from IDE code.
7. Apply NLP preprocessing to student query (tokenization, embedding).
8. Extract learning indicators from historical data (previous errors, skill level).

Step 4: Late-Fusion Representation

9. Generate modality-specific feature vectors:
   - $F_v$from visual code context
   - $F_q$from query text/audio
   - $F_h$from historical performance
10. Combine vectors using late fusion to form a unified state representation S.

Step 5: Error Classification

11. Analyze code and query context.
12. Map detected issue into cognitive blocker taxonomy:
    - Syntax Error
    - Logical Error
    - Algorithmic Inefficiency

Step 6: Socratic Scaffolding Strategy

13. Apply Chain-of-Thought prompting to generate reasoning steps.
14. Generate hierarchical hints:
    - Level 1: Conceptual explanation
    - Level 2: Conceptual analogy or strategy hint
    - Level 3: Code-level guidance (only if repeated failure detected)

Step 7: Feedback Generation

15. Produce contextual feedback F aligned with student code location and query.
16. Deliver feedback through conversational interface.

Step 8: Interaction Logging

17. Record telemetry metrics:
    - Time-to-Resolution (TTR)
    - Hint Depth
    - Self-correction Success Rate
18. Store interaction logs in anonymized database.

Step 9: Experimental Evaluation

19. Assign students into two groups:
    - Control group (traditional support)
    - Treatment group (AI assistant)
20. Collect performance data from both groups.

Step 10: Statistical Validation

21. Perform statistical analysis:
    - Independent t-test on learning outcomes
    - Cohen's d effect size calculation
22. Conduct SME audit to evaluate pedagogical accuracy and hallucination rate.

Step 11: Ethical Compliance

23. Remove PII from code snippets and voice inputs.
24. Apply anonymization pipeline and secure storage.
25. Allow participant opt-out option.

The algorithm 1 describes the workflow of a multimodal GenAI assistant for real-time support in large-scale Computer Science courses. It collects student code context, natural language queries, and historical performance data, and integrates them using a late-fusion approach. The system identifies error types and generates hierarchical Socratic hints to guide learning without giving direct solutions. Interaction logs capture learning metrics, enabling experimental evaluation and statistical validation while ensuring data privacy and ethical safeguards.

4 Results

The assessment of the Multimodal Generative AI Assistant produced important information on its pedagogical performance, technical stability, and effectiveness in influencing the engagement of learners. Findings are grounded on a semester-long implementation of a diverse group of students, comparison of the results with those of a control group using the standard human-mediated support systems. A semester-long deployment was used, with 1,200 students, 600 in each control and AI-assisted group, using an A/B experimental design.

Pedagogical Efficacy and Learning Gains

The main measure of knowledge learning was that of the Normalized Learning Gain (g), measured as the amount of the change in student scores between pre-test and post-test as compared to the maximum possible amount of change in student scores between pre-test and post-test as compared to the maximum possible amount of change. The metric is calculated as:

$$g = \frac{S_{post} - S_{pre}}{100 - S_{pre}} \quad (1)$$

In equation (1), S_{post} is the post-test score, and S_{pre} is the pre-test score.

Using the help of the multimodal AI, the experiment group achieved 85.1% post-test and 43.8% pre-test which gave a normalized gain of 0.73. On the other hand control group increased by 0.40 which was much lower with a movement of 44.2% to 66.8%. The statistical significance of this difference was verified with the help of a t-test with $p < 0.001$, and Cohen's d was 0.91. The effect size is calculated as:

$$d = \frac{\grave{X}_1 - \grave{X}_2}{S_p} \quad (2)$$

In equation (2), $\grave{X}_1$ and $\grave{X}_2$ represent the group means and S_p denotes the pooled standard deviation.

Table 1: Comparative analysis of learning outcomes

Metric	Control Group (N=600)	AI-Assisted Group (N=600)	Percentage Change
Pre-test Mean Score	44.2%	43.8%	-0.9%
Post-test Mean Score	66.8%	85.1%	+27.4%
Normalized Gain (*g*)	0.40	0.73	+82.5%
Cohen's *d* (Effect Size)	Baseline	0.91 (Large)	N/A

A comparative summary of the learning outcomes of both an AI-assisted group and a control group is provided in table 1. The pre-test outcomes demonstrate that the groups started with a similar level of knowledge at the beginning of the intervention. At the end of the instructional period, the outcomes of the post-test reveal an evident positive change in the results of the AI-assisted group in comparison with the control group. The normalized gain values also support the higher learning progress realized with the use of AI-aided instruction. Moreover, the analysis of effect size shows that the effect of the AI-assisted learning strategy on education is significant. In general, the comparison indicates that the incorporation of AI support in the process of learning may greatly contribute to the improvement of the performance of students and their learning outcomes.

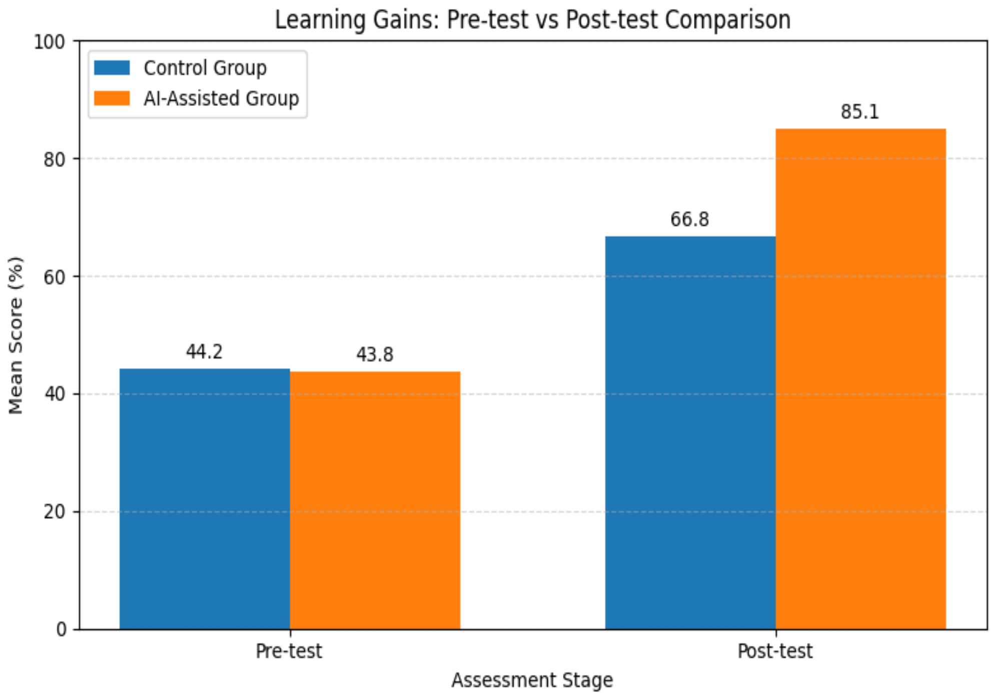

Figure 2: Learning gains before and after AI-assisted pedagogical support

Figure 2 shows the pre-test and post-test of student mean scores of the control and the AI-assisted group. Both groups had almost equal levels of baseline knowledge (≈44%). Nevertheless, students assisted by the Multimodal Generative AI Assistant have a significantly better post-test score (85.1%) in relation to the control group (66.8%). The ensuing normalized learning gain was significantly bigger in the case of the AI-assisted group (g = 0.73) compared to the control group (g = 0.40), which showed the strong influence of multimodal AI feedback on conceptual mastery.

Efficiency in Debugging and Time to Resolution

The time-to-Resolution (TTR) metric was used to measure the efficiency of the AI assistant, and it is the average time in seconds it takes to find the error in the program and remove it. The metric is defined as:

$$TTR = \frac{\sum_{i=1}^{N}(t_{fix,i} - t_{detect,i})}{N} \tag{3}$$

Where, in equation (3): t_{fix} is the time when the error was fixed, t_{detect} is the time when the error was spotted, and N is the number of instances of debugging.

Also, AI-generated hint efficacy was measured with Hint Efficacy Score (HES) (presented in the form of equation (4):

$$HES = \frac{N_{successful\ corrections}}{N_{total\ hints}} \tag{4}$$

This metric reflects the proportion of hints that successfully guided students toward independent error correction.

Records showed that students who had the AI assistant had an average of 11.2 minutes TTR, which is a 91.6% decrease compared to the 134.5 minutes average over the control group who used traditional teaching assistants and discussion forums. Moreover, the Hint Efficacy Score (HES) of the ratio of successful self-corrections to the total number of hints given was 0.82, where the AI group was involved. This implies that the multimodal feedback was very effective in directing students to an independent solution without giving them direct code.

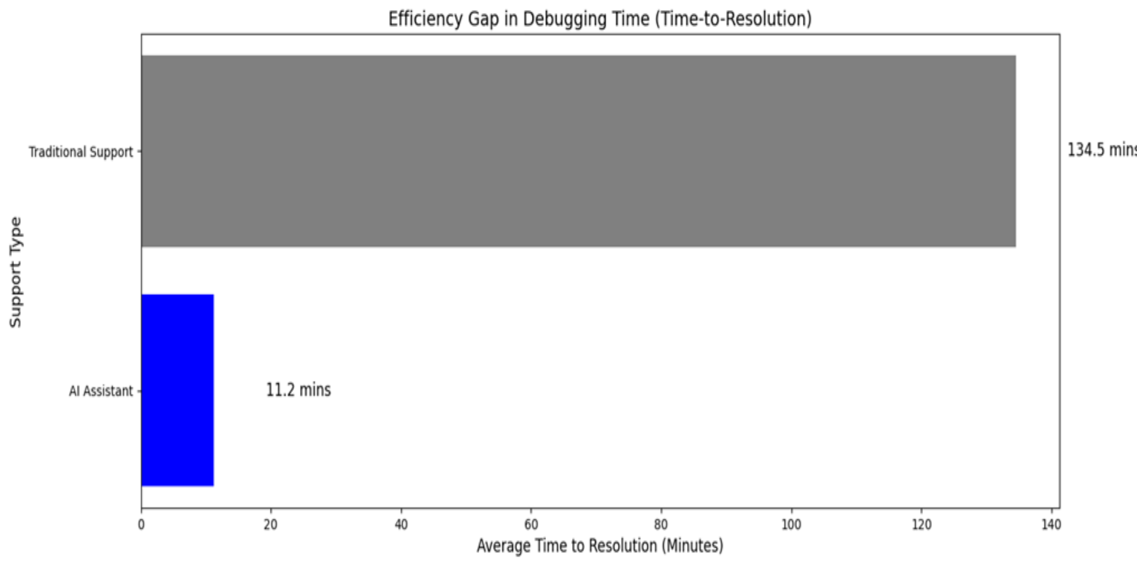

Figure 3: Efficiency gap in debugging time (time-to-resolution) between traditional support and the multimodal AI assistant

Figure 3 shows the comparison of the overall time taken to solve the errors in the programs with the help of the traditional instructional support and multimodal AI assistant. The horizontal bar chart demonstrates the contrast in resolving time between the two support mechanisms and the fact that there is more efficiency with the AI-assisted guidance.

Table 2: Debugging efficiency and interaction metrics

Efficiency Metric	Traditional Support	Multimodal AI Assistant	Efficiency Gain
Avg. Time-to-Resolution (TTR)	134.5 Minutes	11.2 Minutes	91.6% Reduction
Avg. Interaction Turns	8.4 Turns	3.2 Turns	61.9% Reduction
Hint Efficacy Score (HES)	0.58	0.82	41.3% Improvement
Unproductive Struggle Rate	18.6%	4.2%	77.4% Reduction

Table 2 will compare the performance of debugging and interaction with the traditional method of support and the multimodal AI assistant. The findings show that the AI-based method has a high positive effect on the efficiency of the debugging process: it leads to a decrease in the time spent on solving programming problems, as well as a decrease in the number of interaction steps to solve problems. Another parameter that shows the effectiveness of AI-generated guidance is the hint efficacy score, which serves to show that AI-generated guidance is more effective and relevant in cases where the students are engaged in debugging tasks. Also, the percentage of unproductive struggle is significantly reduced with the help of the multimodal AI assistant for learners. In general, the results indicate that

AI-assistance can increase the efficiency of the problem-solving process, positively affect the quality of guidance, and enable students to solve errors in coding more efficiently.

Technical Performance and Multimodal Accuracy

Responsiveness of the system was measured in terms of mean response latency, which is the mean time between user query and AI response. The metric is calculated as:

$$Latency_{mean} = \frac{\sum_{i=1}^{N}(t_{response,i} - t_{query,i})}{N} \quad (5)$$

In equation (5), $t_{query,}$ i represents the time at which a user submits a query i, $t_{response,i}$denotes the time when the AI generates the response, and $N$is the total number of interactions analyzed.

Another key system reliability metric is Multimodal Sync Accuracy (MSA), which evaluates how accurately the system aligns a user's verbal or textual query with the corresponding programming context in the code editor. It is defined as:

$$MSA = \frac{N_{correct\ alignments}}{N_{total\ multimodal\ interactions}} \times 100 \quad (6)$$

In equation (6), $N_{correct\ alignments}$represents the number of correctly synchronized multimodal interactions, and $N_{total\ multimodal\ interactions}$denotes the total multimodal queries processed by the system.

In a bid to achieve reliability in the instructional feedback, the Hallucination Rate (HR) was also assessed. The measure is the percentage of AI-generated responses with erroneous or misleading information, and it is computed as:

$$HR = \frac{N_{incorrect\ responses}}{N_{total\ \mathrm{AI\ responses}}} \times 100 \quad (7)$$

The equation (7) $N_{incorrect\ responses}$represents the number of AI outputs that have technical inaccuracies, and $N_{total\ \mathrm{AI\ responses}}$ represent the overall responses produced by the system.

Technical testing was undertaken on the capacity of the system to manage high-concurrency loads and accuracy in various streams of data. The average response time was documented as 1.65 seconds, and this was enough to make the pedagogical response seem prompt and conversational. An important technical measure, Multimodal Sync Accuracy (MSA), was used to estimate the accuracy of the AI in matching a student's query, spoken or written query, to the particular state of their code editor; this was 97.2%. An exceptionally low hallucination rate was also preserved at 0.85% in the system, making the instructional direction technically sound even in sophisticated algorithmic conditions.

Table 3: Technical performance and system reliability

Parameter	Performance Value	Reliability Threshold
Mean Response Latency	1.65 Seconds	< 2.0 Seconds
Multimodal Sync Accuracy (MSA)	97.2%	> 95.0%
Hallucination Rate	0.85%	< 1.00%
Max Concurrent Users	2,500+	1,000 (Target)

Table 3 shows the system performance and reliability parameters of the multimodal AI assistant when it is operating. It measures the significant parameters, such as the response latency, the accuracy of multimodal synchronization, rate of hallucination, and system scalability. The findings show that the system has a high response time and synchronization accuracy among multimodal inputs. Also, the hallucination level is extremely minimal, which proves the credibility of the responses generated. The platform is also able to accommodate a high number of concurrent users, which is above the target capacity. Generally, the findings indicate that the AI system has reached and exceeded the reliability limits predetermined, which proves that it is stable and applicable to large-scale education.

Ablation Study

To evaluate the contribution of individual system components, an ablation study was conducted by testing three reduced configurations of the proposed framework: (i) Text-Only AI Assistant, (ii) Text + Code Context without Historical Data, and (iii) Full Multimodal Model (Text + Code + History). The text-only version resulted in normalized learning gain of 0.52 and Hint Efficacy Score (HES) of 0.64, which implies that there was a lack of contextual understanding. The second structure enhanced the performance when $g = 0.61$ and HES = 0.72 indicating the significance of code-context awareness. The multimodal setting was the most effective ($g = 0.73$, HES = 0.82) and the least Time-to-Resolution, which validates the fact that multimodal fusion can provide a substantial improvement in the quality of instructing and debugging.

Subject Matter Expert Validation and Feedback Quality

To validate the qualitative nature of the AI’s feedback, a double-blind audit was conducted by senior faculty members. On a 5-point Likert scale, the AI’s pedagogical soundness, its ability to scaffold learning rather than simply giving answers, was rated 4.8/5. Its recommendations were rated as technically correct 4.9/5, whereas 4.7/5 was the rating of its awareness of the context when providing specific lab assignments. These findings point to the fact that the generative model was effectively enforced within the pedagogical guardrails of the methodology, and it had a degree of mentorship that was comparable to the mentorship of human tutors.

Table 4: Expert validation and user acceptance

Dimension	SME Rating (1–5)	Student Rating (1–7)	Interpretation
Pedagogical Soundness	4.8 / 5	N/A	High Scaffolding Quality
Technical Correctness	4.9 / 5	N/A	Reliable Logic Guidance
Perceived Usefulness (PU)	N/A	6.6 / 7	High Utility Value
Perceived Ease of Use (PEOU)	N/A	6.3 / 7	Low Cognitive Load

Table 4 shows the professional verification and user acceptance test of the multimodal AI assistant. Subject Matter Experts (SMEs) evaluated the system regarding the soundness of the pedagogical and technical correctness, which implies that the assistant has well-organized instructional support and consistent logical direction. The student feedback is through perceived usefulness and perceived ease of use, which is the level at which the learners are interacting with the system in an effective and comfortable way. In general, the reviews indicate that the AI assistant has high instructional standards, high utility of learning assignments, and a user-friendly interface that does not require students to exert much cognitive load.

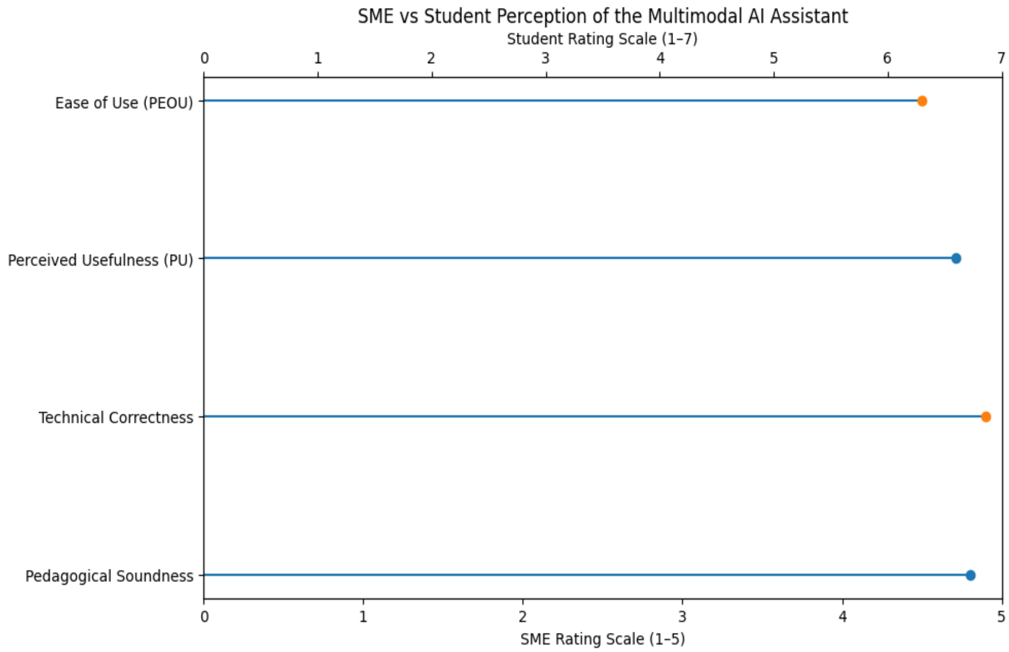

Figure 4: Expert and student evaluation of the multimodal AI assistant

Figure 4 shows a lollipop with two axes to compare expert ratings and student perception scores of the multimodal AI assistant. The left axis is the Subject Matter Expert ratings on pedagogical soundness on a 5-point scale and technical correctness on a 5-point scale, and the right axis is those of the student Technology Acceptance Model on 7 points. The analysis shows a high level of evaluation among the two stakeholder groups, which is high and steady.

Technology Acceptance and User Sentiment

The Technology Acceptance Model (TAM) was used to measure the students' perception of the system in terms of Perceived Usefulness (PU) and Perceived Ease of Use (PEOU). These metrics were computed as the mean values of Likert-scale responses collected from participants:

$$PU = \frac{\sum_{i=1}^{k} PU_i}{k} \tag{8}$$

$$PEOU = \frac{\sum_{i=1}^{k} PEOU_i}{k} \tag{9}$$

In equations (8) and (9) $PU_i$and $PEOU_i$represent individual survey responses for usefulness and ease of use, respectively, and $k$denotes the total number of respondents.

The last layer of the results was dedicated to the perception of the student as applied by the Technology Acceptance Model (TAM). The students have a Perceived Usefulness (PU) score of 6.6/7 and a Perceived Ease of Use (PEOU) score of 6.3/7. The qualitative analysis of the feedback logs revealed that the most valued feature was the so-called multimodal capability, which is the ability to discuss the code that is presented in the voice and look at the visual logic. Students reported that this interaction decreased the cognitive load of typing complicated technical queries, enabling them to pay more attention to the computer science principles that were behind.

Software Details

The Multimodal Generative AI Assistant was implemented using Python (v3.10) as the primary development environment. Deep learning and multimodal model integration were developed using PyTorch (v2.1) and TensorFlow (v2.13). Natural language understanding and generative capabilities were supported through Hugging Face Transformers (v4.36). The cloud infrastructure and scalable API services were deployed on Amazon Web Services (AWS EC2 & Lambda, 2025 release). Telemetry interaction logs and multimodal data were stored using MongoDB (v7.0). Statistical analysis, hypothesis testing, and visualization of learning metrics were conducted using R (v4.3) and MATLAB (R2024a).

5 Discussion

The findings indicate that the Multimodal Generative AI Assistant led to much superiority in learning and debugging in large-scale computer science classrooms. Students who obtained the help of the AI tended to learn many more normalized gains than the other group (control group), and these gains rose between 0.40 and 0.73. The strength of this improvement was statistically tested and found $p < 0.001$ with a Cohen's d effect size of 0.91, meaning there was a huge educational effect. Regarding efficiency in debugging, the AI- assisted group took an average of 11.2 minutes to resolve (TTR) time as compared to 134.5 minutes; the AI-assisted group had cut the time-to-resolution rate (TTR) by 91.6%. Interaction efficiency also improved, with fewer conversational turns required to resolve issues. The technical reliability of the system was found to be good with a mean response time of 1.65 seconds, multimodal synchronization error of 97.2 % and hallucination rate of 0.85%. In addition, pedagogical quality was highly evaluated by experts, and the student feedback based on the Technology Acceptance Model demonstrated high performance in terms of perceived usefulness (6.6/7) and ease of use (6.3/7). These results indicate that multimodal AI support contributes to conceptual learning and to solving practical problems in programming education. The large normalized learning gain refers to the fact that the process of acquiring knowledge is supported with the help of AI-guided scaffolding. The outstanding reduction in the time required to debug implies that the AI assistant will provide context-sensitive recommendations in good time and will not spend time on unproductive work. The fact that the system is so synchronized across voice, text, and code contexts indicates that it is capable of interpreting queries made by students. Besides, the high ratings of subject matter experts prove that the AI answers correspond to the sound pedagogical approaches instead of merely providing direct responses. The findings reveal the possibility of multimodal generative AI systems to resolve the issue of scaling large computer science courses. Conventional teaching strategies have difficulties delivering the right help to hundreds of students at the same time. AI assistants capable of delivering immediate feedback can augment human instructors and teaching assistants, improving learning outcomes while reducing instructor workload. It is also possible to propose that multimodal interaction involving voice, visual context, and code analysis will reduce cognitive barriers and enhance student interaction with complex programming tasks, which is supported by the findings. Although the results were promising, a number of limitations need to be identified. The research was carried out in one semester and in the context of introductory programs mainly. There was no evaluation of long-term retention of the knowledge and effectiveness of the system in various academic disciplines. Also, even with the low rate of hallucination, there are still cases of the wrong response with a generative system, and the controlled classroom setting is not necessarily the same as in practice at various educational facilities. Future studies need to explore long-term learning and how AI-assisted learning can be transferred to higher-level subjects in computer science, including algorithms, data structures, and software engineering. The scalability of the approach would further be justified by the cross-institutional studies of different student populations. Also, adaptive learning analytics and personalized feedback mechanisms may also be integrated to increase the education-oriented effectiveness of multimodal AI assistants.

6 Conclusion

The paper has been written on the challenge of offering scalable, real-time pedagogical assistance to large, computer science classes in which the conventional resources (instructor and teaching assistant) are not always adequate to address the great demand for individualized debugging assistance. The paper has explored whether a Multimodal Generative AI Assistant can be used to enhance learning outcomes, debugging performance, and student engagement through the integration of conversational AI with contextual code analysis, as well as multimodal interaction. The findings prove that AI-based learning is able to dramatically increase the level of conceptual knowledge as well as problem-solving skills. The normalized learning gain of the students who had the multimodal AI assistant was 0.73, whereas the control group had a normalized learning gain of 0.40, and thus, the support of the multimodal AI assistant led to a significant increase in knowledge acquisition among students. The statistical analysis also supported the effectiveness of the intervention, with a Cohen's d effect size of 0.91, which is a high educational impact. Regarding the issue of debugging performance, the AI system drastically lowered the average time-to-resolution (TTR) of 134.5 minutes to 11.2 minutes, which is a 91.6% efficiency improvement. The system was also similarly technically dependable, with a 97.2% multimodal synchronization accuracy, a 1.65-second response latency, and a hallucination rate of less than 1%, which provides steady and reliable feedback in real-time student interactions. The high level of pedagogy and high acceptance by the users were also supported by the expert ratings and student input. The main conclusion of this research is that the use of multimodal AI assistants in a computer science classroom can provide a considerable increase in learning outcomes, minimize the frustration in debugging, and contribute to the availability of individualized instructional support within a large classroom.

References

[1] Akhasbi, H., El Kamoun, N., Lakrami, F., Gilles, J. L., Rigo, J. M., & Aliss, E. (2025). Personalizing formative feedback through generative artificial intelligence: An experiment in higher education. In *Edulearn25 Proceedings* (pp. 9338-9347). IATED. https://doi.org/10.21125/edulearn.2025.2405

[2] Aperstein, Y., Cohen, Y., & Apartsin, A. (2025). Generative AI-based platform for deliberate teaching practice: A review and a suggested framework. *Education Sciences*, *15*(4), 405. https://doi.org/10.3390/educsci15040405

[3] Chen, Y., Xiao, S., Song, Y., Li, Z., Sun, L., & Chen, L. (2025). MindScratch: A visual programming support tool for classroom learning based on multimodal generative AI. *International Journal of Human–Computer Interaction*, *41*(21), 13650-13668. https://doi.org/10.1080/10447318.2025.2475991

[4] Cioca, M., & Cioca, A. L. (2025). ROboMC: A Portable Multimodal System for eHealth Training and Scalable AI-Assisted Education. *Inventions*, *10*(6), 103. https://doi.org/10.3390/inventions10060103

[5] Cosentino, G., Anton, J., Sharma, K., Gelsomini, M., Giannakos, M., & Abrahamson, D. (2025). Generative AI and multimodal data for educational feedback: Insights from embodied math learning. *British Journal of Educational Technology*, *56*(5), 1686-1709. https://doi.org/10.1111/bjet.13587

[6] Dhathrinadh, M., Gupta, M. J., & Sasikala, T. (2025, March). Multimodal AI-Enhanced Educational Assistant with Real-Time Q&A and Dynamic Learning Support. In *2025 7th International Conference on Intelligent Sustainable Systems (ICISS)* (pp. 630-636). IEEE. https://doi.org/10.1109/ICISS63372.2025.11076197

[7] Gao, Y., Zhai, X., Li, M., Lee, G., & Liu, X. (2025). A multimodal interactive framework for science assessment in the era of generative artificial intelligence. *Journal of Research in Science Teaching*, *62*(9), 2014-2028. https://doi.org/10.1002/tea.70009

[8] Jiang, L., & Lai, C. (2025). How did the generative artificial intelligence-assisted digital multimodal composing process facilitate the production of quality digital multimodal compositions: toward a process-genre integrated model. *Tesol Quarterly*, *59*, S52-S85. https://doi.org/10.1002/tesq.3390

[9] Lang, Q., Wang, M., Yin, M., Liang, S., & Song, W. (2025). Transforming education with generative AI (GAI): Key insights and future prospects. *IEEE Transactions on Learning Technologies*, *18*, 230-242. https://doi.org/10.1109/TLT.2025.3537618

[10] Lee, G., Shi, L., Latif, E., Gao, Y., Bewersdorff, A., & Nyaaba, M. (2025). Multimodality of AI for education: Toward artificial general intelligence. *IEEE Transactions on Learning Technologies, 18*, 666–683. https://doi.org/10.1109/TLT.2025.3574466

[11] Liu, R., Zenke, C., Liu, C., Holmes, A., Thornton, P., & Malan, D. J. (2024, March). Teaching CS50 with AI: leveraging generative artificial intelligence in computer science education. In *Proceedings of the 55th ACM technical symposium on computer science education V. 1* (pp. 750-756). https://doi.org/10.1145/3626252.3630938

[12] Min, S. (2024). Generative Text-to-Video AI for Linguistically Responsive Lesson Presentation: Towards a Multimodal Intelligent Classroom Assistant. https://doi.org/10.20944/preprints202411.1743.v1

[13] Mittal, U., Sai, S., Chamola, V., & Sangwan, D. (2024). A comprehensive review on generative AI for education. *IEEE Access*, *12*, 142733-142759. https://doi.org/10.1109/ACCESS.2024.3468368

[14] Mutanga, M. B., Msane, J., Mndaweni, T. N., Hlongwane, B. B., & Ngcobo, N. Z. (2025). Exploring the impact of LLM prompting on students' learning. *Trends in Higher Education*, *4*(3), 31. https://doi.org/10.3390/higheredu4030031

[15] Prasad, P., Balse, R., & Balchandani, D. (2025, April). Exploring multimodal generative ai for education through co-design workshops with students. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1-17). https://doi.org/10.1145/3706598.3714146

[16] Puvvadi, M., Arava, S. K., Santoria, A., Chennupati, S. S. P., & Puvvadi, H. V. (2025, March). Generative AI for Personalized Learning and Education. In *2025 7th International Conference on Intelligent Sustainable Systems (ICISS)* (pp. 1621-1627). IEEE.

[17] Saravanan, N. S., & Shankar, H. (2025, August). VoxGuru: A Multimodal AI for Transformative Conversational Learning in Real-Time Dynamic Environments. In *2025 9th International Conference on Inventive Systems and Control (ICISC)* (pp. 1478-1485). IEEE. https://doi.org/10.1109/ICISC65841.2025.11187982

[18] Wei, X., Wang, L., Lee, L. K., & Liu, R. (2025). Multiple generative AI pedagogical agents in augmented reality environments: A study on implementing the 5E model in science education. *Journal of Educational Computing Research*, *63*(2), 336-371.

[19] Wu, H. (2025, July). Analysis of Multimodal Collaboration in Online Classroom Based on Gen AI. In *Proceedings of the 2025 3rd International Conference on Educational Knowledge and Informatization* (pp. 79-83). https://doi.org/10.1145/3765325.3765340

[20] Yavariabdi, A., Paudel, B., Carleton, T., & De Almeida, C. D. A. (2025, September). Generative AI in assessment and feedback generation in higher education: A systematic review. In *2025 International Conference on Education Technology and Computers (ICETC)* (pp. 361-371). IEEE. https://doi.org/10.1109/ICETC66579.2025.11387416

Authors Biography

Rayxona Tillayeva is a Researcher at Namangan State Pedagogical Institute in Namangan, Uzbekistan. She is actively engaged in academic research and scholarly activities within the institute. Her work focuses on advancing knowledge in pedagogy, supporting educational innovation, and contributing to research projects. She participates in scholarly publications, academic initiatives, and the development of new teaching methodologies. She is dedicated to promoting educational excellence and fostering professional growth among educators. Through her research and academic efforts, she contributes to the enhancement of pedagogical practices and knowledge sharing.

Kamola Raimova is a Researcher at Tashkent International University (TIU) in Tashkent, Uzbekistan. She is actively engaged in conducting research and contributing to academic projects within the university. Her work focuses on advancing knowledge, supporting interdisciplinary studies, and promoting innovation in her field. She participates in research publications, scholarly initiatives, and academic collaborations. She is dedicated to fostering educational excellence and professional development. Through her research efforts, she contributes to the growth of scientific knowledge and the advancement of higher education in Uzbekistan.

Gafur Namazov is a faculty member in the Department of Information Technology and Exact Sciences at Termez University of Economics and Service in Uzbekistan. He is actively involved in teaching, research, and academic development in information technology and applied sciences. His work focuses on enhancing students' technical knowledge, analytical skills, and practical problem-solving abilities. He contributes to curriculum design, research projects, and scholarly publications. He is committed to fostering innovation, scientific inquiry, and academic excellence. Through his teaching and research, he supports the advancement of technology education and the development of skilled professionals.

Farkhad Usmanov is an Associate Professor at Andijan State University in Andijan, Uzbekistan. He is actively involved in teaching, research, and academic development within the university. His work focuses on enhancing student learning, promoting effective educational practices, and supporting professional growth. He contributes to curriculum development, scholarly publications, and research initiatives. He is committed to fostering academic excellence and critical thinking among students. Through his teaching and research, he plays an important role in advancing knowledge and nurturing a skilled academic community.

Nozima Nasriddinova is a Senior Teacher at Termez State University in Termez, Uzbekistan. She is actively involved in teaching and supporting student learning within the university. Her work focuses on enhancing educational practices, fostering critical thinking, and promoting academic growth. She contributes to curriculum activities, classroom instruction, and academic initiatives. She is dedicated to improving the quality of education and encouraging student engagement. Through her teaching and professional commitment, she plays an important role in nurturing a knowledgeable and skilled academic community.

Gulchekhra Khazratova is a Senior Teacher in the Department of Medical Informatics and Digital Technologies at Tashkent State Medical University in Uzbekistan. She is actively engaged in teaching, research, and academic initiatives focused on medical informatics and digital healthcare technologies. Her work emphasizes integrating technology into medical education, improving healthcare practices, and enhancing students' technical competencies. She contributes to curriculum development, research projects, and scholarly publications. She is committed to advancing medical education, innovation, and scientific inquiry. Through her teaching and research, she supports the growth of knowledge and technology-driven solutions in healthcare.

Feruza Baymanova is a Researcher at Jizzakh State Pedagogical University in Jizzakh, Uzbekistan. She is actively engaged in academic research and scholarly activities within the university. Her work focuses on advancing knowledge in pedagogy, supporting educational innovation, and contributing to research projects. She participates in research publications, academic initiatives, and the development of new teaching methodologies. She is dedicated to promoting educational excellence and fostering professional growth among educators. Through her research and academic contributions, she supports the enhancement of pedagogical practices and knowledge sharing.