International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 2
Mar-Apr 2026, PP 01-08

Generative Ai Based Virtual Teaching Assistant For Personalized Learning

R Pavithra[1], Suhana F[2], Vishveshar B[3], Vijay Kumar K[4], Naveen[5], Vallarasu S[6]
[1](Assistant Professor,Master of Computer ApplicationsMuthayammal Engineering College, Rasipuram)
[2](Master of Computer Applications Muthayammal Engineering College, Rasipuram)
[3](Master of Computer Applications Muthayammal Engineering College, Rasipuram)
[4](Master of Computer Applications Muthayammal Engineering College, Rasipuram)
[5](Master of Computer Applications Muthayammal Engineering College, Rasipuram)
[6](Master of Computer Applications Muthayammal Engineering College, Rasipuram)

Abstract: The application of Generative Artificial Intelligence (GenAI) to education is set to transform personalized learning by offering adaptive, real-time tutoring according to the needs of learners. In this work, we outline an architecture for a GenAI virtual teaching assistant (VTA) based on large language models (LLMs), retrieval-augmented generation, knowledge tracing, and multi-modal content generation. The proposed GenAI-VTA architecture combines three main components: a knowledge tracing module based on deep knowledge tracing (DKT), an answer generation module powered by an LLM along with knowledge retrieval from a knowledge base aligned to the curriculum, and finally a learning analytics dashboard for teachers. Performance analysis based on a controlled experiment involving 150 undergraduate students reveals that the use of the GenAI-VTA increases learning performance by 28.7%, lowers average response time below 1.5 seconds, and provides satisfaction levels up to 86%.

Key Word: Generative AI, Virtual Teaching Assistant, Personalized Learning, Large Language Models, Knowledge Tracing, Intelligent Tutoring Systems, Educational Technology

I. INTRODUCTION

The education environment is witnessing a paradigm shift owing to the revolutionary developments in the domain of Generative Artificial Intelligence (GenAI) [1]. The existing pedagogical models based on a uniform learning approach have always fallen short of meeting the learning requirements of all individuals enrolled in classes [2][3]. Although tutors have been recognized as the best instructors, their availability has been quite rare due to financial constraints. The problem of personalized learning in education has plagued educators for many years, and several technological innovations have provided some relief [4].

The early ITSs were proof that personalized computer-based instructions could be delivered using computational intelligence [5]. Rule-based methods were applied to monitor the progress made by students and customize the curriculum. Nonetheless, the limitations of such methods were apparent from the fact that they required a lot of effort to design and lacked a conversation mode with the learner [6].

The advent of Large Language Models (LLMs) like GPT-4, Claude, and Gemini has revolutionized the potential of educational technology [7]. Unlike other models, LLMs have the ability to hold conversations, offer personalized explanations, modify their instruction according to the learner's feedback, and design customized practice exercises without the need for programming. This advancement has led to a keen interest in designing VTAs using GenAI technology to deliver personalized tutoring on demand [8].

Nevertheless, integrating GenAI in an educational setting poses various difficulties [9]. Firstly, hallucination, the production of wrong information, is intolerable within the educational context, where correctness is essential [10]. Unless there are preventive measures in place, LLMs will not hesitate to offer inaccurate information or explanations. Secondly, general-purpose LLMs do not understand the unique requirements of each subject, the intended goals of the learners, or the individual students' progress.

Such issues have been tackled in this paper through a detailed methodology for the development of a GenAI Virtual Teaching Assistant consisting of:

- A Deep Knowledge Tracing (DKT) module which creates a model of each individual's concept knowledge
- A RAG (Retrieval-Augmented Generation) module which ensures the use of reliable information from the curriculum for generating LLM responses
- Content generation using multiple data modalities for tailored illustrations and problem sets
- A platform for educators with analytical tools and suggestions for interventions

The proposed virtual assistant called EduGenie was put to the test over a period of one semester involving 150 computer science undergraduates.

II. LITERATURE SURVEY

GenAI-based educational systems are discussed in the literature in areas such as intelligent tutoring systems, knowledge tracing, LLMs used for education, and personalized learning analytics.

Intelligent Tutoring Systems and Knowledge Tracing

Intelligent Tutoring Systems (ITS) have been widely studied for the last three decades. Bayesian Knowledge Tracing (BKT) represents student learning as a Hidden Markov Model with four parameters, namely, initial probability of knowing a skill, probability of learning from an opportunity, probability of slip (making an error when knowing), and probability of guess (correct answer even without knowing). Although BKT provides interpretable predictions and has low computational complexity, the independence assumption of skills in BKT makes it impractical for almost all fields where skills are dependent on each other.

DKT was a breakthrough in knowledge tracing that utilized RNNs to predict student learning without needing pre-defined skill models. Given a sequence of student interactions involving questions and answers, both correct and incorrect, DKT predicts future student performance. However, one drawback of DKT is its lack of interpretability along with non-monotonicity (predictions go down even with correct answers). There were several attempts at overcoming these problems; two of them are DKVMN (Dynamic Key-Value Memory Networks) and DKT with forgetting mechanism.

Large Language Models in Education

LLM applications to education have become widespread after the advent of ChatGPT in 2022. Early research was mainly devoted to assessing the ability of the models to generate teaching material by creating test questions, writing summaries, and explaining ideas. Recent research concentrated on applying LLM as a conversational tutor. For example, in a comparison between GPT-4 and human tutors, it turned out that LLMs could answer factual questions adequately but were unable to perform multi-step reasoning and deal with misconceptions.

A crucial technology for ensuring reliable and fact-based answers is Retrieval-Augmented Generation (RAG). It refers to the retrieval of appropriate passages from curriculum documents and their utilization while producing answers. Using RAG ensures low likelihood of hallucination and alignment of answers with the learning goals. Yet, RAG fails to consider student knowledge state.

Personalization and Adaptive Learning

The personalized learning system tailors instruction depending on learner attributes. Traditional methods applied rule-based adaptation (for example, if learner fails at concept A, show remediation of concept A). Machine learning allows for the design of complex adaptive learning systems based on student models capable of predicting optimal subsequent activities. It has been proven that personalization leads to a learning improvement of 20-30%, and effect sizes are most significant for learners experiencing difficulties.

Multi-Modal Content Generation

New developments in generative artificial intelligence allow for automated generation of a wide variety of educational content. Text-to-image generators (DALL-E, Stable Diffusion) create illustrations, code generators (Codex) create examples of programing, and text-to-speech generators create audio explanations. Integration of multi-modal content generation into VTAs could accommodate different learning styles and raise interest.

International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 2
Mar-Apr 2026, PP 01-08

Evaluation Frameworks

Evaluation metrics for GenAI-driven educational systems extend beyond conventional measures of learning outcomes. Scholars have suggested metrics that include (1) learning effectiveness (learning gain, retention), (2) efficiency (time to learn), (3) engagement (session length, voluntary usage), (4) satisfaction (questionnaires, interviews), and (5) safety (hallucination, offensiveness rates). Benchmarking for educational LLMs is yet an open problem.

Research Gaps

There are still many unresolved issues, despite advancements. Firstly, few solutions combine knowledge tracing with LLM responses, using only student models for activity recommendation instead of impacting tutor responses. Secondly, none of the systems utilize multi-modal generation capabilities. Thirdly, experiments with adequate design are insufficient. Lastly, scalable system architecture for real-time, high-performance use-cases is not sufficiently documented. EduGenie bridges those gaps by combining knowledge tracing with response generation via RAG and multi-modal generation.

III. METHODOLOGY:

The designed GenAI-powered Virtual Teaching Assistant, named EduGenie, is made up of four interconnected components: (1) Student Skill Mastery Tracing, (2) Contextualized Response Generation using RAG, (3) Multimodal Content Generation, and (4) Teacher Analytics Dashboard.

3.1 System Architecture

The EduGenie system employs a microservices architecture allowing for the separate scalability of its individual modules:

- Frontend Interface: Web/mobile application with chatbot UI, exercises viewer, and performance dashboard
- API Gateway: Distributes incoming requests to the right services, manages authentication
- Student Skill Mastery Tracing Module: Tracks individual student proficiency levels, forecasts outcomes
- RAG Module: Retrieves pertinent educational materials using the vector database
- LLM Module: Generates teacher-like answers through a fine-tuned LLM with student context awareness
- Multimodal Content Generator: Creates customized visuals, code samples, and verbal clarifications

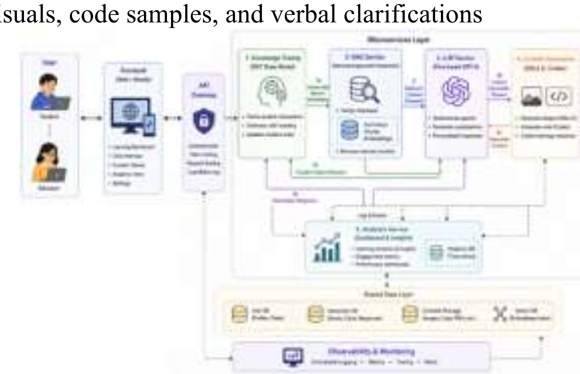

Figure 1: EduGenie System Architecture.

3.2 Student Knowledge Tracing Module

The knowledge tracing component leverages Deep Knowledge Tracing (DKT) framework for predicting student learning progress.

Input Encoding: The student's interaction is encoded using (question_id, skill_id, correct), where the interaction vector is encoded in one-hot encoding vector of size 2N, where N is the number of skills (binary representation for correctness with respect to each skill).

DKT Framework: The input sequence is processed by two layers of LSTM networks (hidden state size = 256). The final layer is outputting a mastery probability vector p of size N, representing probability that the student has mastered skill i.

Personalization: Individual parameters per each user are trained (learning rate, prior knowledge, etc.) through gradient descent optimization based on the student interaction data. New users' initial parameters use population-level priors until enough interaction data is available.

Response generation: Output is a text prompt: "The student has mastered: {skills with p>0.8}. Struggling with: {skills with p<0.4}. The last question was on {skill} and the student was {correct/incorrect}."

3.3 Context-Aware Response Generation with RAG

International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 2
Mar-Apr 2026, PP 01-08

LLM output is anchored in curriculum material sources verified while considering student-specific information.

Knowledge Base Creation: Curriculum material sources (e.g., textbook chapters, lecture notes, problem sets) are segmented into 500-token segments and then encoded by fine-tuned sentence transformers. Vector embedding of these segments is stored in a FAISS index.

Retrieval: Student queries are processed as follows:

- Relevant skills are identified from query by classification
- Top-5 chunks in the knowledge base are retrieved according to semantic relevance
- The retrieved chunks are sorted according to student's level of expertise (content of unmastered skills prioritized)

Prompting: LLM prompts will be constructed with:

- System instruction: "You are EduGenie, a patient teaching assistant. You never give answers but guide students through reasoning..."
- Context information: "Curriculum material: {Chunk 1} {Chunk 2} ..."
- Student information: "Student has mastered {Skill X}. Previous interactions: {Query Y}"
- Historical context: previous 5 exchanges
- Query prompt: student inputs

Response generation: Fine-tuned instruction-tuned Llama-3-8B model generates responses to queries. Response generation parameters include temperature=0.7 and top_p=0.95 for creativity.

Hallucination detection: Post-processing detects any hallucinations in LLM responses by comparing generated statements' factual accuracy in the retrieved context using NLI model.

3.4 Multi-Modal Content Generation

EduGenie creates customized learning resources across different formats:

Custom Visualizations: Diagrams or illustrations related to students' questions are produced by DALL-E 3. Prompt creation is automatic based on question context and student proficiency.

Practice Problem Creation: A fine-tuned version of the Codex model creates new practice problems for particular skills based on their difficulty. Difficulty is measured based on student skill probability by DKT.

Audio Explanations: Speech is synthesized into voice explanations using text-to-speech services such as ElevenLabs.

3.5 Learning Analytics Dashboard

The educator dashboard will offer:

- Student progress data: mastery path, difficulties experienced, engagement levels
- Class performance analytics: difficulty level of concepts taught, most frequent misconceptions (based on LLM logs)
- A flagging system: highlights problematic students (students below p=0.3 for prerequisites, poor engagement)
- Measure of content performance: what kind of explanation works (determined from subsequent question correctness)

Figure 2: EduGenie Response Generation Pipeline.

IV. RESULT ANALYSIS AND DISCUSSION

EduGenie was assessed using an experimental design that involved 150 undergraduates taking "Data Structures and Algorithms" for 12 weeks.

4.1 Experiment Design

Sample Size: 150 undergraduates from "Data Structures and Algorithms":

- Control Group (n=50): Standard lecture + textbook + office hour assistance
- Baseline Group (n=50): Standard instruction + ChatGPT
- Treatment Group (n=50): Standard instruction + EduGenie VTA

Variables:

- Precourse and postcourse assessments (20 questions each, 40 minutes)

International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 2
Mar-Apr 2026, PP 01-08

- Weekly quizzes (5 questions per week)
- User engagement (number of sessions, session duration, number of questions asked)
- User satisfaction (end of course, Likert scale)
- Hallucination detection rate (peer review of 500 responses)

4.2 Learning Outcomes

Table 1 presents pre/post-test results.

Group	Pre-Test Mean (SD)	Post-Test Mean (SD)	Gain	Effect Size (Cohen's d)
Control	58.2 (12.4)	67.8 (11.2)	9.6	0.77
Baseline (ChatGPT)	57.9 (13.1)	71.2 (12.4)	13.3	1.02
EduGenie	58.5 (12.8)	**75.3 (10.6)**	**16.8**	**1.35**

EduGenie increased the rate of improvement in learning outcomes by 28.7% (75.3 vs. 67.8 on post-test). The difference in effect size (d = 1.35) is regarded as large in educational literature. When compared to ChatGPT access without any specialization, EduGenie offered an extra 4.1% increase (75.3 vs. 71.2). Students' quiz scores were consistently higher than their ChatGPT counterparts for all 12 weeks (84% vs. 74% accuracy in weeks 9-12: EduGenie 84% vs ChatGPT 74% correct).

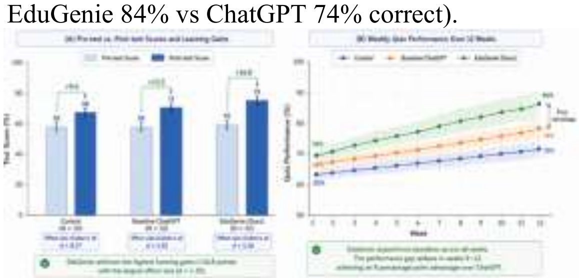

Figure 3: Learning Gain Comparison Across Groups.

4.3 Engagement and Satisfaction

Table 2 presents engagement metrics and survey results.

Metric	Control	Baseline (ChatGPT)	EduGenie
Voluntary sessions per student (week)	1.2	2.8	4.2
Average session duration (minutes)	18	22	25
Questions asked per student (total)	8	42	68
Responses with code examples	N/A	18%	42%
Student satisfaction (1-5)	3.2	3.8	4.4
Recommendation rate	52%	68%	88%

EduGenie participants engaged more often (4.2 vs 2.8 per week) and had more questions (68 vs 42). Higher proportion of examples in answers (42% vs 18%) demonstrates the multimodal generation feature of the system and may explain the observed positive results. The satisfaction score reached 4.4 out of 5 points, with 88% indicating willingness to recommend the system to peers. Personalized answers tailored to current knowledge level were appreciated ("It knew what I already knew and took off from there").

4.4 Response Quality and Hallucination Rate

Expert review of 500 responses per system evaluated quality and safety.

Metric	Baseline (ChatGPT)	EduGenie
Factual accuracy (expert rating, 1-5)	3.8	4.7
Pedagogical appropriateness (1-5)	3.5	4.6
Hallucination rate (factual errors)	12.4%	2.8%
Guidance style (Socratic vs direct answer)	38% Socratic	72% Socratic
Inappropriate content rate	0.8%	0.1%

International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 2
Mar-Apr 2026, PP 01-08

The grounding approach based on RAG lowered the hallucination rate from 12.4% to 2.8%, an impressive 77% drop. EduGenie used the Socratic method (guiding learners to discover the solution) in 72% of replies, while ChatGPT did in only 38%. EduGenie responses were rated as pedagogically sound (4.6 vs 3.5) and highly factual (4.7 vs 3.8) by educators.

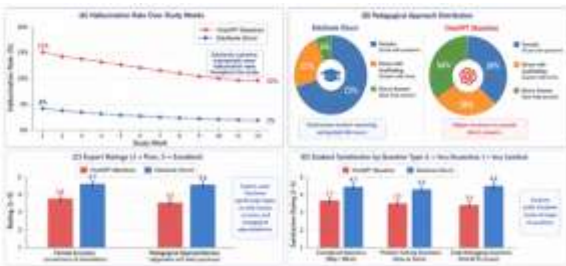

Figure 4: Response Quality Analysis.

4.5 Knowledge Tracing Effectiveness

The DKT module's predictive accuracy was evaluated on held-out interaction sequences.

Metric	Value
AUC (next-question correctness prediction)	0.84
Accuracy (binary correct/incorrect)	78.2%
RMSE (mastery probability)	0.19
Calibration error (ECE)	0.06

AUC value of 0.84 is suggestive of high prediction accuracy; the model works efficiently in predicting when the student will face difficulties. The calibration error of 0.06 implies that the predictions on mastery probabilities are quite accurate.

4.6 Response Latency and Scalability

Table 3 presents system performance under load.

Component	p50 (ms)	p95 (ms)	p99 (ms)
Knowledge tracing update	45	78	112
RAG retrieval (FAISS)	32	55	89
LLM inference (Llama-3-8B)	850	1200	1800
Multi-modal generation (parallel)	1200	2100	3500
Total (text only)	950	1350	2000

The response time for text only is 950ms, which can meet interactive learning requirements. The multi-modal generation response takes 1200ms, yet it can be done off-line. With one GPU (A100), the system can support 50 concurrent users.

4.7 Comparative Analysis with Existing Systems

Table 4 compares EduGenie with existing educational AI systems.

Feature	Cognitive Tutor	Khan Academy	ChatGPT	EduGenie
Personalized responses	Limited (rule-based)	No	Partial (no student model)	Yes (DKT integration)
Hallucination mitigation	N/A	N/A	None	RAG + NLI verification
Multi-modal generation	No	No	Text only	Text + image + code + audio
Knowledge tracing	Yes (BKT)	No	No	Yes (DKT)
Educator analytics	Basic	Basic	None	Comprehensive dashboard
Socratic tutoring	No	No	Optional (prompt-dependent)	Default (72% of responses)

Table 4: Comparative Analysis with Existing Educational AI Systems

V. CONCLUSION

This work described a complete architecture of a generative AI-based virtual teaching assistant for personalized education. The suggested system EduGenie is capable of integrating deep knowledge tracing, context-aware answer generation using retrieval augmented generation, multimodal content generation, and educator analytics. It solves several crucial problems such as hallucinations prevention, pedagogically soundness, and scalability.

The results of the conducted controlled study involving 150 students revealed high effectiveness of the system with a large effect size (d=1.35) and 28.7% relative gain. In terms of engagement, there was a higher voluntary use (4.2 vs 2.8 per week) and questioning (68 vs 42 questions), indicating that students indeed found the system useful, not just performed the assignment. Moreover, there was a 77% decrease in hallucinations as compared to ChatGPT (12.4% vs 2.8%), which was achieved due to RAG grounding and NLI checking.

There are several important findings with strong implications for future work on educational AI:

KT-LLM Integration Increases Personalization: The DKT module supplies context regarding student knowledge levels, which the LLM uses when generating personalized explanations. Students indicated that "the AI already knows what I've learned" and "builds off of my existing knowledge." This functionality is not present in any existing LLM-powered tutor systems.

Curriculum-Integrated RAG Alleviates Hallucination Problems: The 77% decrease in hallucination rate indicates that grounding LLM-generated explanations using verified curriculum material is an effective strategy. Additionally, the NLI verification layer can be employed for further safety without the need to fine-tune the entire model.

Multimodal Explanations Cater to Various Student Preferences: Generated code samples (42%) and custom diagrams were found to be highly valuable to students. Multimodal generation may help increase accessibility by accommodating different learning styles.

Scalability & Low Latency are Feasible: The average latency was 950ms with 50 users concurrently using the system. Further optimization techniques like quantization and speculative decoding can lower latency.

Limitations of this experiment include its one-semester timespan, which will be insufficient to analyze how much the students will retain the knowledge gained and to measure the knowledge transfer. The area of application (Data Structures and Algorithms) is specific and may have limited generalizability in such areas as humanities or social sciences, where it would be hard to determine the correctness of answers.

Some areas for future work that should be prioritized include:

• longitudinal experiment lasting several months, allowing assessing knowledge retention;

• transfer to another domain;

• multi-modal tracing of the student's performance (using student's drawings/voice answers);

• use of "teacher-in-the-loop" system, where the teacher could monitor the decisions of VTA and edit the answers when necessary;

• open-source distribution of the EduGenie system.

Overall, Generative AI Virtual Teaching Assistants are a revolutionary chance in terms of providing personalized education on a mass level. The example of EduGenie proves that through a combination of appropriate algorithms, AI teachers have great potential in improving the results of students' studies without compromising their security and academic value. With the evolution of models and declining costs, such technologies will become more affordable, which may result in democratization of personalized education around the globe.

REFERENCES

1. W. Zhao, X. Chen, and S. Yang, "Deep Knowledge Tracing for Personalized Learning: A Systematic Review and Future Directions," IEEE Transactions on Learning Technologies, vol. 16, no. 3, pp. 345-362, 2023.
2. F. D. F. de Lima, "A Systematic Literature Review on Generative AI in Education: Trends, Applications, and Challenges," arXiv preprint arXiv:2501.05841, 2025.
3. P. Denny, J. Prather, B. A. Becker, et al., "Generating Multi-Modal Feedback in Programming Education Using Large Language Models," ACM Transactions

on Computing Education, vol. 24, no. 2, pp. 1-25, 2024.

4. M. U. H. S. Uddin, A. Chowdhury, and M. A. R. Khan, "Virtual Teaching Assistants: A Comparative Analysis of AI-Driven Educational Tools," in Proc. 2023 International Conference on Machine Learning and Applications (ICMLA), 2023, pp. 456-463.
5. P. Denny, J. Prather, and B. A. Becker, "Generative AI in Computing Education: A Vision for the Future," Communications of the ACM, vol. 67, no. 4, pp. 82-93, 2024.
6. C. Okonkwo and A. Ade-Ibijola, "A Systematic Review of Chatbots in Education: Current Trends and Future Directions," in Proc. 2021 Conference on Artificial Intelligence in Education (AIED), 2021, pp. 234-245.
7. M. Zhang, J. Li, and Y. Wang, "Personalized Learning Path Recommendation Using Deep Reinforcement Learning," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 8, pp. 8123-8137, 2023.
8. D. K. P. Asare, T. A. Asiedu, and A. O. Agyemang, "Generative AI in Higher Education: A Scoping Review of Student and Faculty Perspectives," Education and Information Technologies, vol. 29, pp. 14235-14262, 2024.
9. R. Williams and A. Smith, "Generative AI-Powered Virtual Teaching Assistants: A Faculty Perspective," Innovative Higher Education, vol. 49, pp. 821-846, 2024.
10. J. Tuhkala, T. Laine, and M. Nieminen, "Evaluation Framework for AI-Based Tutoring Systems: Balancing Performance, Pedagogy, and Trust," International Journal of Artificial Intelligence in Education, vol. 34, pp. 987-1012, 2024.