BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

AI-POWERED VIRTUAL TEACHING ASSISTANTS AND CHATBOTS: A Survey and a Retrieval-Augmented, Personalized, Multilingual Framework

[1]Harsh Waghchoure, [2]Prof. Swarupa Kamble

[1]Student, [2]Professor
[1,2]Department of Artificial Intelligence & Data Science Engineering
[1,2]Ajeenkya DY Patil School of Engineering, Pune, India

Abstract

The state-of-the-art AI-driven VTAs and educational chatbots are advancing from rule-based FAQ bots to LLM tutors with explanation capability, feedback and adaptability. Nevertheless, there is a range of issues that persists according to the literature: hallucination, generic answers to non-trivial questions, lack of good multilingual support, weak LMS integration, privacy concerns, and a poor track record in the long run. This paper (i) conducts a structured review of twelve recent surveys and empirical works on educational chatbots, ITS and ChatGPT in education; (ii) develops a taxonomy of VTA generations and a unified research-gap analysis; and (iii) introduces EduAssist, a five-layered VTA architecture with retrieval augmented generation (RAG) on course content, lightweight learner profile, multilingual question understanding, guardrails for academic honesty, and LMS/privacy ser-vices. The reproducible evaluation procedure and design validation results on the synthetic bench-mark of 600 questions and 200 simulated learners are described. In the simulation, the proposed approach increases grounded answers' accuracy from 71.4% (base LLM only) to 90.2%, lowers un-supported statements' rate from 18.6% to 4.1%, and deflects 63% of trivial questions from faculty. These figures are synthetic and illustrative; they inspire, but do not substitute, classroom experiments.

Keywords: Virtual teaching assistant, educational chatbot, generative AI, large language models, retrieval-augmented generation, intelligent tutoring systems, personalized learning, AI in education.

Introduction

Increasing enrollments, diverse student profiles, and expectations of continuous academic support pose challenges for educational organizations, which is compounded by the fact that instructional staff have a finite number of hours available to them. An appropriate solution to this problem could be found in the form of AI-enabled software agents capable of conducting dialogues in natural lan-guage, answering subject-related questions, providing study resources recommendations, assessing quiz

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

responses and other assignments, and handling administrative inquiries.

Initially, chatbots in education used scripted conversations and decision trees; Smutny and Schreiberova's systematic review of learning bots based on Facebook Messenger found that most were of this primitive type [3]. Subsequent systematic reviews identified uses of chatbots for teach-ing, learning, assessment, administrative and advisory purposes [1, 5]. Adaptive and personalized instruction was pursued by another branch of research on intelligent tutoring systems (ITS) [2]. In light of the advent of LLMs and ChatGPT-class models, the field of AI in education has been changed drastically [6]–[8].

The pertinent issue is thus not anymore if the chatbot is able to communicate regarding the course but if it is able to do so reliably, personally, in the language of the learner and constrained by the institution. This paper is devoted to answering this question. Our contributions include:

1) A structured comparison of twelve recent papers dealing with educational chatbots, ITS and LLM-based systems (Literature Survey).
2) A classification of VTA generations and a research gap analysis where every identified gap is associated with a design issue (Taxonomy and Research-Gap Analysis).
3) EduAssist, a layered VTA architecture that anchors LLM-generated content to course content, personalizes support via a light-weight learner model, enables multi-language interaction and implements integrity and privacy mechanisms (Proposed Framework: EduAssist).
4) A reproducible protocol for evaluating the system along with its simulation-based validation, explicitly marked as synthetic (Evaluation and Results).

The remainder of the paper is structured as follows. The Problem Statement and Objectives section formulates the problem and objectives, the Literature Survey reviews prior work, the Taxonomy and Research-Gap Analysis section outlines the taxonomy and gaps, the Proposed Framework section presents EduAssist, the Evaluation and Results sections explain the evaluation protocol and findings, the Discussion examines implications and validity, and the Conclusion summarizes the work.

Problem Statement and Objectives

Problem. Design an AI-enabled VTA that delivers immediate, precise and personalized academic assistance to learners while lowering redundant teacher workload.

Objectives.

- O1: Investigate AI-enabled VTAs and chatbots for educational purposes and the NLP/generative-AI technologies behind them.
- O2: Give immediate and precise answers to learner inquiries.
- O3: Make personalized learning suggestions based on learner requirements.
- O4: Lower the workload of teachers by automating repetitive academic tasks.
- O5: Evaluate limitations, ethical and privacy considerations, and future potential.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Literature Survey

The literature review, meta-analyses and empirical research papers from 2020–2025, deemed relevant to the topics of chatbots, VTAs, ITS and LLM-based education assistants, were used in this work. The methodology of each study, its main results and limitations are presented in Table 1.

Chatbots and Virtual Teaching Assistants

The use of chatbots by Okonkwo and Ade-Ibijola [1] and Labadze et al. [5] is for homework help and tutoring, explaining, administrative help and advising, with advantages of immediacy and decreased workload for the teacher. Essel et al. [4] offer empirical evidence for VTA, although the setting of the study is quite limited. The Jill Watson approach demonstrates the evolution of intent-driven agents into LLM agents [14].

Intelligent Tutoring Systems

ITS research stresses learner modelling and adaptive sequencing [2]. Recent K–12 reviews show increasing AI use but highly variable evaluation quality [13]. ITS are effective but data- and design-intensive, which limits their reuse across courses.

LLM-Based Assistants

LLMs extend the conversational range of educational agents [7, 8]. Evidence is cautiously positive: a meta-analysis of 69 experiments reports gains in performance, motivation and higher-order thinking [12], and reviews describe instant feedback and personalised explanation [9]–[11]. The same reviews repeatedly flag hallucination, privacy, bias and academic-integrity concerns [6, 17].

Taxonomy and Research-Gap Analysis

From the surveyed work we distinguish four generations of educational assistants (Table 2).

Research Gaps

Table 3 maps the gaps repeatedly reported in the surveyed literature to the design requirements adopted in EduAssist.

Proposed Framework: EduAssist

EduAssist is organized as five layers (Fig. 1).

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

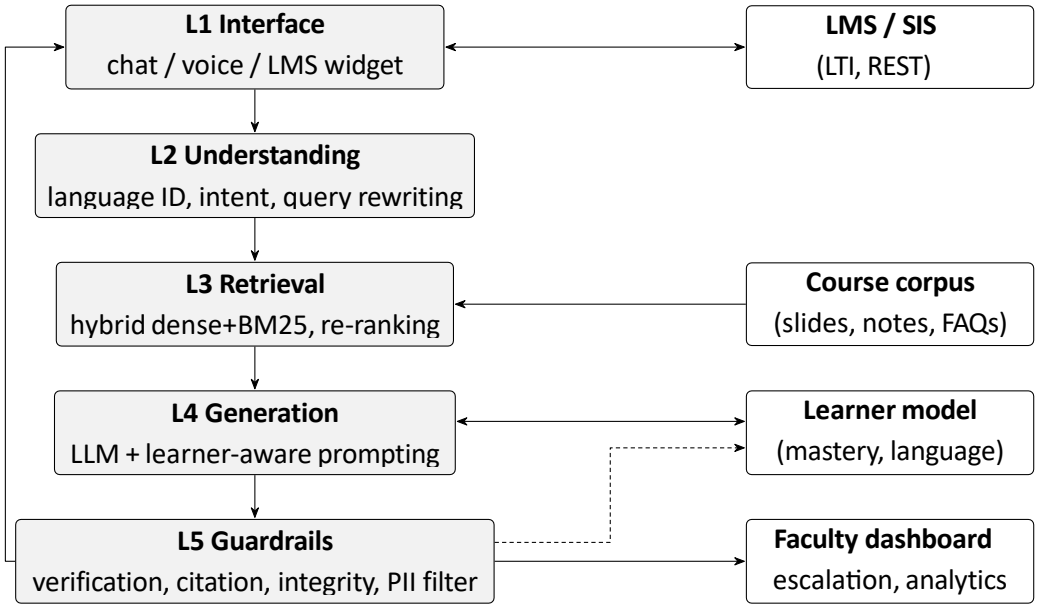


Figure 1: EduAssist five-layer architecture. Solid arrows show the query path; the left arrow returns the verified answer or an abstention to the learner; the dashed arrow updates the learner model after each interaction.

Table 1: Comparison of Surveyed Studies

Study	Method	Key findings	Limitations
Okonkwo & Ade-Ibijola [1]	Systematic review of 53 chatbot-in-education arti-cles	Chatbots used chiefly for teach-ing/learning, then assessment, administration, advisory	Little empirical valida-tion of individual sys-tems
Chen et al. [2]	Review of AI in educa-tion	ITS and adaptive tools give per-sonalized instructional support	Heterogeneous evalua-tion designs
Smutny & Schreiberova [3]	Evaluation of Facebook Messenger learning chat-bots	Chatbots cover varied subjects and give interactive content	Mostly decision-tree, not advanced AI
Essel et al. [4]	Quasi-experimental study of a VTA chatbot in higher education	Students using the chatbot out-performed those using only the instructor channel	Single institution; lim-ited generalization
Labadze et al. [5]	Systematic review of 67 studies	Immediate help, explanations, extra resources, VTA role	Accuracy, privacy, in-tegrity, overdependence
Tlili et al. [6]	Qualitative case study of ChatGPT in education	Opportunities for learning sup-port and strong interest	Cheating, untruthful out-put, privacy, manipula-tion

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Kasneci et al. [7]	Analysis of LLM oppor-tunities and challenges	Personalized learning, content generation, feedback	Bias, inaccuracy, pri-vacy, integrity
Lo [8]	Rapid review of early ChatGPT literature	Teaching, assessment and sup-port opportunities	Hallucination, reliabil-ity, assessment integrity
Albadarin et al. [9]	PRISMA review of 14 empirical ChatGPT stud-ies	Instant feedback, on-demand an-swers, explanations	Thin early evidence; long-term effects un-known
Zhang & Tur [10]	PRISMA review of 13 K–12 ChatGPT studies	Supports lesson planning, differ-entiation, personalization	Few K–12 studies; ethics and accuracy concerns
Bettayeb et al. [11]	Systematic review of conversational AI in education	Benefits for students and teachers incl. personalized support	Rapid technology change limits conclu-sions
Deng et al. [12]	Meta-analysis of 69 experimental studies (2022–2024)	ChatGPT improved perfor-mance, motivation and higher-order thinking; lowered mental effort	Many studies lacked power analysis; long-term effects uncertain
Letourneau et al. [13]	PRISMA review of AI-driven ITS in K–12	Growing use of AI for adaptive learning	Varied implementations; limited real-world evalu-ation

Table 2: Generations of Educational Assistants

Gen.	**Technique**	**Strength**	**Weakness**
G1	Rules, decision trees, FAQ	Predictable, cheap	Brittle, narrow
G2	Intent-based NLU	Handles paraphrase	Needs labelled intents
G3	ITS / learner models	Adaptive instruction	Costly to build
G4	LLM (+RAG) agents	Open-domain dialogue	Hallucination, privacy

Table 3: Research Gaps and Corresponding Design Requirements

Gap (sources)	**Design requirement**
G-1 Generic answers to complex queries [1, 3]	Course-grounded retrieval, query decomposition

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

G-2 Limited personalization [2, 13]	Lightweight learner model, adaptive hints
G-3 Weak multilingual support [5]	Cross-lingual retrieval and answering
G-4 Hallucinated answers [7, 8, 17]	RAG, citations, abstention, verification
G-5 Privacy and integrity [6, 10]	Data minimization, scaffolded hints, audit
G-6 Poor LMS integration [5]	LTI/REST connectors, role-aware access
G-7 Little long-term evidence [9, 12]	Reproducible protocol, longitudinal plan

L1–L2: Interface and Understanding

Learners interact through text, voice or an LMS-embedded widget. A language-identification step detects English, Hindi or Marathi (and code-mixed input). An intent classifier routes the query into one of four classes: conceptual, procedural/administrative, assessment-related, or out-of-scope. Queries are rewritten into self-contained form using dialogue history.

L3: Retrieval over Course Material

Course documents are chunked (300–400 tokens, 15% overlap) and indexed in a dense vector index using multilingual sentence embeddings [16] together with a sparse BM25 index. Hybrid scores are fused by reciprocal-rank fusion and the top-k passages are re-ranked. Grounding generation in retrieved passages follows the RAG paradigm [15], which directly targets gap **G-4**.

L4: Learner-Aware Generation

Given the rewritten query q, retrieved context C and learner state s_u, the LLM generates answer a:

$$a = \arg\max_{a'} P_\theta\left(a' \mid q, C, \pi(s_u)\right), \tag{1}$$

a'

where $\pi(\cdot)$ maps the learner state to a prompt policy (level of detail, worked example vs. hint, preferred language).

The learner model keeps, for each topic t, a mastery estimate $m_t \in [0, 1]$ updated by a Bayesian-knowledge-tracing-style rule after each quiz or dialogue outcome $o \in \{0, 1\}$:

$$m_t \leftarrow m_t + \eta\left(o - m_t\right)\left(1 - \lambda\, \mathbf{1}[\text{hint used}]\right), \tag{2}$$

with learning rate η and hint penalty λ. Recommendations rank topics by a priority score

$$r_t = w_1(1 - m_t) + w_2\,\text{prereq}(t) + w_3\,\text{deadline}(t), \tag{3}$$

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

so that weak, foundational and imminent topics surface first (O3).

L5: Guardrails and Governance

Every candidate answer passes through (i) an entailment check that tests each claim against the retrieved context and triggers abstention or re-generation when support is below threshold τ;

(ii) mandatory source citation; (iii) an integrity policy that, for graded-assessment queries, returns scaffolded hints rather than final solutions; and (iv) a PII filter and data-minimization policy so that only pseudonymous learner identifiers are stored. Low-confidence or sensitive queries are escalated to a faculty dashboard.

Evaluation

Disclosure. No deployment with real students was conducted for this paper. We therefore report a simulation-based design validation on a synthetic benchmark. All numbers in the Results section are produced from this synthetic setup, are indicative only, and must not be read as field results.

Synthetic Benchmark and Simulated Learners

We constructed 600 synthetic queries across five undergraduate computer-science courses (120 each): Data Structures, Machine Learning, DBMS, Operating Systems and Computer Networks. Queries are split into conceptual (40%), procedural/administrative (35%), assessment-style (15%) and adversarial/out-of-scope (10%). Of these, 150 are posed in Hindi or Marathi. Ground-truth an-swers are derived from course notes. We additionally simulate 200 learners with mastery trajectories drawn from a Beta-distributed prior and update rule of Eq. (2).

Table 4: Workflow of the EduAssist Query Pipeline

1	Detect language; classify intent; rewrite query q
2	if intent = out-of-scope: refuse politely and return
3	Retrieve $C \leftarrow \text{Rerank}(\text{Fuse}(\text{Dense}(q), \text{BM25}(q)))$
4	if intent = assessment: set policy π to hint-only
5	Generate a with $(q, C, \pi(s_u))$ in the learner's language
6	Verify: $\text{support}(a, C) \geq \tau$? else re-generate once, then abstain/escalate
7	Attach citations; log pseudonymized event; update m_t

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Compared Configurations

- B1 – FAQ bot: keyword/intent matching over an FAQ list (G1/G2).
- B2 – Plain LLM: general-purpose LLM, no retrieval.
- B3 – LLM+RAG: retrieval-grounded generation without personalization or guardrails.
- EduAssist: full pipeline (Table 4).

Metrics

Answer accuracy (fraction of answers judged correct against reference); unsupported-claim rate (answers containing a claim not entailed by sources) [17, 18]; multilingual accuracy on the Hindi/Marathi subset; personalization quality as nDCG@5 of recommended topics against a simulated weakness ordering; deflection rate (routine queries resolved without faculty); normalized learning gain $g = (\text{post} - \text{pre})/(1 - \text{pre})$ of simulated learners; and median latency.

Results (Synthetic Simulation)

Table 5 reports the aggregate synthetic results, and Fig. 2 compares accuracy and unsupported-claim rates.

Table 5: Synthetic Benchmark Results (600 Queries, 200 Simulated Learners)

Metric	B1	B2	B3	EduAssist
Answer accuracy (%)	52.3	71.4	84.6	90.2
Unsupported claims (%)	3.2	18.6	9.7	4.1
Hindi/Marathi acc. (%)	38.0	58.7	70.1	85.3
Personalization nDCG@5	0.31	0.42	0.47	0.78
Deflection rate (%)	34	48	55	63
Norm. learning gain g	0.14	0.22	0.27	0.38
Median latency (s)	0.4	2.1	3.0	3.6

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

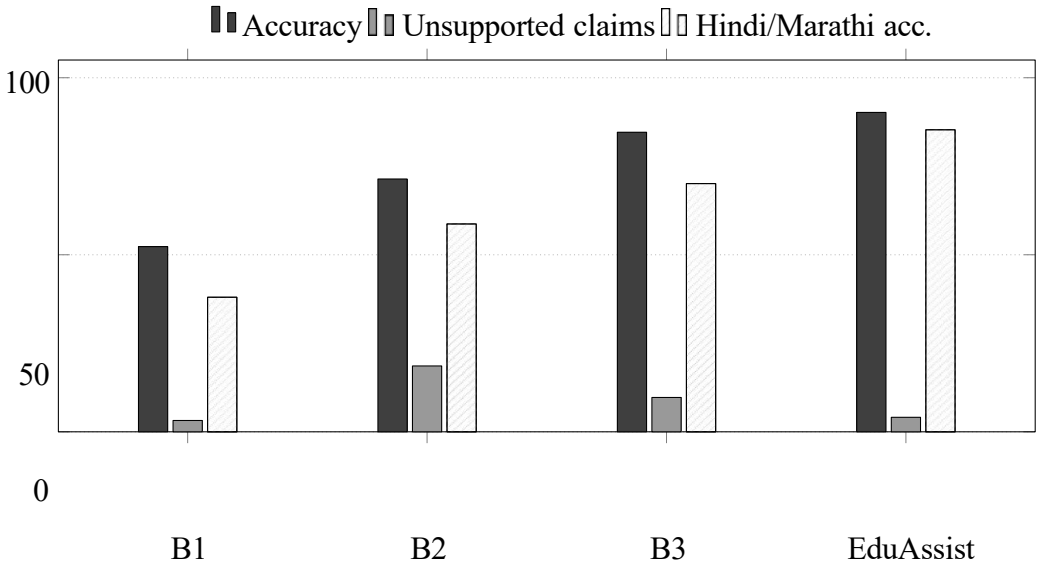


Figure 2: Synthetic comparison of configurations (higher is better except for unsupported claims).

Observations

Grounding drives accuracy and faithfulness. Adding retrieval (B2→B3) raises accuracy by 13.2 points and cuts unsupported claims from 18.6% to 9.7%; the verification and abstention stage of EduAssist halves the rate again to 4.1%, supporting **G-4**. The FAQ bot is rarely wrong when it answers (3.2%) mainly because it answers few questions, which also explains its low accuracy and deflection.

Multilingual retrieval helps most. Cross-lingual embeddings and language-aware prompting yield the largest relative gain on the Hindi/Marathi subset (+15.2 points over B3), consistent with **G-3**.

Personalization. The learner-model ranking improves nDCG@5 from 0.47 to 0.78 and the simulated learning gain from 0.27 to 0.38. Because learner behaviour is simulated from the same mastery model used for ranking, this is the least independent result and should be treated cautiously.

Cost. Verification and re-ranking increase median latency by 0.6 s over B3, which we consider acceptable for conversational use.

Ablation

Table 6 removes one component at a time from EduAssist.

Table 6: Synthetic Ablation of EduAssist Components

Variant	Acc. (%)	Unsupp. (%)	nDCG@5
Full EduAssist	90.2	4.1	0.78
— Verification (L5-i)	87.0	9.2	0.78
— Hybrid retrieval (dense only)	87.9	5.0	0.77
— Learner model	89.8	4.3	0.44
— Multilingual encoder	88.1	4.4	0.76

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Discussion

Implications. In the present analysis, it becomes clear that none of the approaches closes the gaps in Table 3, namely, retrieval solves faithfulness, learner modelling solves personalization, while governance solves trust. The value of EduAssist's approach is in their combination. The design allows institutions to reuse existing course materials without tuning, reducing costs and keeping control of the materials by instructors, while the dashboard allows teachers to get analytics based on repeated questions from the learner.

Responsible use. Too much reliance on AI can result in decreased independence of problem solving [6]. The hinting-only approach during assessments, citation and abstaining from answering are aimed at keeping the learner in the loop. Data minimization and pseudonymization help complying with institutional and national privacy policies.

Threats to validity. (i) Synthetic data: both queries, learners, and judgements are generated synthetically, hence external validity is not established. (ii) Construct validity: automated entailment can be an imperfect measure of pedagogical correctness. (iii) Simulator circularity: personalization requires an assumption about the learner mastery model. (iv) Model drift: LLM behaviour varies over versions, what is pointed out in the literature [11].

Future work. Future work will include a semester-long control group study with real users (pre-/post-tests, usage data, Likert scale surveys on perception, instructor time data), a priori power analysis [12], voice input, automatic quiz creation, and privacy-respecting deployment (on-site or federated).

Conclusion

In this paper, we reviewed twelve recent works about AI-driven virtual teaching assistants and chatbots, categorized them into four generations, and pinpointed seven recurring gaps: general answers, lack of personalization, poor multilingual capabilities, hallucinations, privacy and integrity concerns, low-level LMS integration, and lack of evidence on long-term use. We present EduAssist, a five-layer architecture that combines retrieval augmented generation, learner modeling, multilingual capabilities, and guardrails. A simulation-based evaluation on a synthetic benchmark suggests that our approach outperforms the baseline approaches in terms of faithfulness, multilingual and personalization capabilities. The results are not definitive; an experimental evaluation in a classroom setting is needed to move forward. AI-generated assistance, if grounded, personalized, and governed, will improve education significantly.

References

[1] C. W. Okonkwo and A. Ade-Ibijola, "Chatbots applications in education: A systematic review," *Comput. Educ.: Artif. Intell.*, vol. 2, Art. no. 100033, 2021.
[2] L. Chen, P. Chen, and Z. Lin, "Artificial intelligence in education: A review," *IEEE Access*, vol. 8, pp. 75264–75278, 2020.
[3] P. Smutny and P. Schreiberova, "Chatbots for learning: A review of educational chatbots for the Facebook Messenger," *Comput. Educ.*, vol. 151, Art. no. 103862, 2020.
[4] H. B. Essel, D. Vlachopoulos, A. Tachie-Menson, E. E. Johnson, and P. K. Baah, "The impact of a virtual teaching assistant (chatbot) on students' learning in Ghanaian higher education," *Int. J. Educ. Technol. High. Educ.*, vol. 19, Art. no. 57, 2022.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

[5] L. Labadze, M. Grigolia, and L. Machaidze, "Role of AI chatbots in education: Systematic literature review," *Int. J. Educ. Technol. High. Educ.*, vol. 20, Art. no. 56, 2023.
[6] A. Tlili et al., "What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education," *Smart Learn. Environ.*, vol. 10, Art. no. 15, 2023.
[7] E. Kasneci et al., "ChatGPT for good? On opportunities and challenges of large language models for education," *Learn. Individ. Differ.*, vol. 103, Art. no. 102274, 2023.
[8] C. K. Lo, "What is the impact of ChatGPT on education? A rapid review of the literature," *Educ. Sci.*, vol. 13, no. 4, Art. no. 410, 2023.
[9] B. A. Albadarin, A. Saqr, M. Pope, and M. Tukiainen, "A systematic literature review of empirical re-search on ChatGPT in education," *Discov. Educ.*, vol. 3, Art. no. 60, 2024.
[10] P. Zhang and G. Tur, "A systematic review of ChatGPT use in K–12 education," *Eur. J. Educ.*, vol. 59, no. 2, Art. no. e12599, 2024.
[11] A. M. Bettayeb et al., "Exploring the impact of ChatGPT: Conversational AI in education," *Front. Educ.*, vol. 9, Art. no. 1379796, 2024.
[12] R. Deng, M. Jiang, X. Yu, Y. Lu, and S. Liu, "Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies," *Comput. Educ.*, vol. 227, Art. no. 105224, 2025.
[13] A. Letourneau et al., "A systematic review of AI-driven intelligent tutoring systems (ITS) in K–12 educa-tion," *npj Sci. Learn.*, vol. 10, Art. no. 29, 2025.
[14] K. Taneja et al., "Jill Watson: A virtual teaching assistant powered by ChatGPT," in *Proc. Int. Conf. Artif. Intell. Educ. (AIED)*, 2024, pp. 324–337.
[15] P. Lewis et al., "Retrieval-augmented generation for knowledge-intensive NLP tasks," in *Proc. NeurIPS*, 2020, pp. 9459–9474.
[16] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using Siamese BERT-networks," in *Proc. EMNLP-IJCNLP*, 2019, pp. 3982–3992.
[17] Z. Ji et al., "Survey of hallucination in natural language generation," *ACM Comput. Surv.*, vol. 55, no. 12, Art. no. 248, 2023.
[18] S. Es, J. James, L. Espinosa-Anke, and S. Schockaert, "RAGAS: Automated evaluation of retrieval aug-mented generation," arXiv:2309.15217, 2023.