International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 14 Issue: 3
***Article Received:** 25 July 2026 **Revised:** 12 August 2026 **Accepted:** 20 September 2026*

Develop an AI-Based Multilingual Tutor/Chatbot for Student Interaction

Naveen Kumar K.P.[1] Mahesha D.M ,[2] Bhavya D.N,[3] Sharath Kumar Y. H [4]

[1, 2, 3], Department of Studies and Research in Computer Science, Karnataka State Open University, Mysore-India.

[4] Maharaja Institute of Technology Mysore, Department of ISE, MIT Mysore, India.

Abstract

The rapid growth of Artificial Intelligence (AI) and Natural Language Processing (NLP) has created new opportunities for developing intelligent and inclusive educational systems. However, language barriers continue to limit access to digital learning resources, particularly for students who are more comfortable learning in regional and native languages. This research proposes the development of an **AI-based multilingual tutor/chatbot** for personalized student interaction and academic support. The proposed system integrates multilingual Natural Language Processing, Machine Translation, Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), and speech technologies to enable students to interact with the tutor using English and Indian regional languages, with a primary focus on **Kannada-English interaction**.

The proposed tutor will be grounded in curriculum-specific resources such as syllabi, textbooks, lecture notes, question banks, and other approved educational materials. Student queries will be processed through language identification, semantic understanding, information retrieval, and AI-based response generation to provide contextually relevant and curriculum-aligned explanations. The system will additionally support personalized learning by identifying students' learning requirements, generating topic-wise explanations, practice questions and quizzes, and providing immediate feedback. Voice-based interaction through speech-to-text and text-to-speech technologies will further improve accessibility for students who prefer spoken communication. The proposed framework aims to address challenges associated with low-resource Indian languages, code-mixed language interaction, contextual understanding, and hallucination in generative AI-based educational systems. The effectiveness of the system will be evaluated using NLP performance measures, response relevance, translation quality, question-answering accuracy, user satisfaction, and student learning outcomes. The proposed multilingual AI tutor is expected to provide an accessible, scalable, and personalized learning environment and contribute towards **inclusive, mother-tongue-supported and technology-enabled higher education**.

1. Introduction

The rapid development of Artificial Intelligence (AI) and Natural Language Processing (NLP) is transforming the way educational content is created, delivered, and accessed. AI-based educational systems can provide personalized explanations, automated assessment, interactive learning, and real-time academic assistance. However, a major challenge in the Indian education system is the language barrier, particularly for students who understand academic concepts more effectively in their mother tongue or regional language. UNESCO has highlighted linguistic diversity and unequal access to digital educational content as important considerations for inclusive AI-enabled education in India. (UNESCO Document Database) India has a highly diverse linguistic environment, and educational resources are still predominantly available in English and a limited number of widely used languages. Consequently, students from regional-language backgrounds may face difficulties in understanding technical terminology, interacting with digital learning platforms, and accessing high-quality educational resources. Recent initiatives in India have increasingly explored AI, machine translation, speech technologies, and multilingual language models to address these challenges. The Government of India has also emphasized the use of AI and language technologies to promote multilingual education and digital access across Indian languages. (Press Information Bureau)

The emergence of Large Language Models (LLMs) provides an opportunity to develop intelligent educational tutors capable of understanding natural-

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 14 Issue: 3
***Article Received:** 25 July 2026 **Revised:** 12 August 2026 **Accepted:** 20 September 2026*

language questions and generating contextually relevant responses. An AI-based tutor can support students beyond conventional search-based systems by explaining concepts, answering questions, generating examples, creating practice questions, and providing immediate feedback. Such systems can also support personalized learning by adapting explanations and learning activities according to the student's knowledge level and learning requirements. For multilingual education, however, simply translating English educational content into a regional language is not sufficient. Educational language requires appropriate terminology, contextual interpretation, cultural relevance, and linguistic accuracy. Recent work in Karnataka has demonstrated the potential of curriculum-grounded AI systems supporting both English and Kannada, while also showing that AI-generated Kannada educational content may require substantial human review for linguistic quality and contextual correctness. (DOI)Therefore, this research proposes an AI-based multilingual tutor/chatbot that combines NLP, multilingual language processing, Large Language Models, Retrieval-Augmented Generation (RAG), machine translation, and speech technologies to provide interactive academic support. The proposed system will primarily focus on Kannada-English multilingual interaction, while providing a framework that can be extended to other Indian languages. Students will be able to submit questions through text or voice, and the system will identify the input language, understand the student's query, retrieve relevant information from curriculum-specific resources, and generate an appropriate response in the student's preferred language. The proposed system will use curriculum-grounded educational resources, including syllabus documents, textbooks, lecture notes, question banks, and other verified learning materials. RAG can help the tutor retrieve relevant information before generating an answer, thereby improving the relevance and reliability of responses. Similar curriculum-aligned multilingual educational systems have demonstrated the potential of combining RAG, translation and human validation for regional-language education. (DOI)In addition, the proposed tutor will incorporate personalized learning and intelligent assessment. Based on student interactions, the system can identify difficult topics, recommend learning materials, generate topic-wise quizzes, provide explanations at different difficulty levels, and offer immediate feedback. Voice-based interaction using Speech-to-Text (STT) and Text-to-Speech (TTS) can further improve accessibility for students who prefer communicating in their native language. AI and NLP applications in multilingual classrooms already include translation, transcription, speech technologies, adaptive learning and personalized assessment. (Ciet)

The overall objective of the proposed research is therefore to develop a scalable, curriculum-grounded and multilingual AI tutor that reduces language barriers and provides students with accessible, personalized and interactive learning support. The system will be evaluated in terms of language understanding, translation quality, response accuracy, educational relevance, student satisfaction, interaction effectiveness, and learning outcomes. The proposed framework has the potential to contribute to inclusive digital education by enabling students to learn, ask questions, and interact with educational technology in the language they understand best.

2. Related Work

Recent advances in Natural Language Processing (NLP), Large Language Models (LLMs), multilingual machine translation, and Retrieval-Augmented Generation (RAG) have created significant opportunities for developing intelligent educational systems. However, multilingual education in India presents specific challenges because of linguistic diversity, limited resources for several Indian languages, code-mixed communication, and the need for curriculum-specific content. UNESCO's recent report on multilingual education in India highlights the growing role of NLP, AI, machine translation, automatic speech recognition, and text-to-speech in improving access to education across Indian languages. (UNESCO)

2.1 Multilingual NLP for Indian Education

Saini [1] investigated the use of multilingual instruction supported by NLP in Indian higher education. The study examined faculty perceptions and classroom use of multilingual instruction and NLP tools. The findings indicate that multilingual NLP tools can improve accessibility and participation, although challenges such as translation accuracy, infrastructure, training, and cultural sensitivity remain. This work establishes the importance of integrating language technologies with multilingual teaching practices rather than treating translation as an independent technological task. (Journal Universitas Negeri Jakarta) UNESCO [2] reported the rapid development of Indian-language technologies through initiatives such as BHASHINI, which provides machine translation, automatic speech recognition, and text-to-speech services across 22 Indian languages. Such technologies provide an important foundation for developing multilingual educational applications capable of supporting both text and voice interaction. (UNESCO) The Government of India has also emphasized AI-based language technologies under initiatives supporting Indian languages. Recent government initiatives include BharatGen, a multimodal LLM initiative intended to support all 22 scheduled Indian languages. These developments demonstrate the growing ecosystem for multilingual AI applications in education. (Press Information Bureau)

2.2 AI-Based Educational Tutors and Chatbots

Vaishnavi et al. [3] proposed an **AI- and NLP-powered personalized education tutor for Indian regional languages**. Their system combines multilingual transformer-based models, including IndicTrans and mBERT, with speech-to-text, text-to-speech and learner recommendation mechanisms. The work demonstrates the feasibility of combining language translation, voice interaction, and personalization to support students from different linguistic backgrounds. (EurekaMag) Sumanth Reddy et al. [4] proposed **PathshalaGPT**, a multilingual AI companion designed for government-school education. The system combines LLMs, RAG, real-time translation, speech-to-text and text-to-speech to provide curriculum-aligned and interactive learning. Its use of a Socratic dialogue approach is particularly relevant because the system attempts to guide students through questions rather than simply providing direct answers. (MAT Journals) Narang et al. [5] proposed **GurukulAI**, an educational platform designed specifically for the Indian education system. The work uses a syllabus-aligned question-answer dataset and combines fine-tuning of LLaMA 3.1 8B with a RAG framework. The system supports bilingual interaction in English and Hindi and provides functions such as doubt clarification and examination-oriented practice. This work demonstrates the importance of adapting general-purpose LLMs to Indian curriculum content. (Hugging Face)

2.3 Curriculum-Grounded AI and RAG

A significant limitation of general-purpose LLM-based educational chatbots is that they may generate responses that are not aligned with the prescribed curriculum. **Shiksha Copilot** addresses this issue through curriculum-grounded RAG. The system was deployed in Karnataka and supports both English- and Kannada-medium education. It extracts and indexes textbook content, retrieves relevant material during generation, and uses human educators to review AI-generated content. (DOI) Shiksha Copilot is particularly relevant to the proposed research because it demonstrates a complete pipeline involving **curriculum ingestion, RAG, LLM orchestration, English-Kannada translation, validation, lesson-specific chat and question generation**. Its human-in-the-loop approach also demonstrates that educational AI systems benefit from expert validation, particularly when working with regional-language content. (DOI) Similarly, recent multilingual college chatbot research has combined **RAG, FAISS vector retrieval, BM25 keyword search, LLMs, multilingual interaction and voice interfaces**. Such systems demonstrate how institutional knowledge can be integrated with conversational AI to provide more reliable student support. (IJRASET)

2.4 Kannada Language Learning and Conversational AI

Kasilingam et al. [8] developed **SpeakSmart**, an NLP-based conversational tutor for Kannada language learning. The chatbot uses conversational interaction to support vocabulary development, grammar understanding and conversational fluency. The work demonstrates the potential of conversational AI specifically for Kannada learning and provides a foundation for developing more comprehensive Kannada-English educational tutoring systems. (EurekaMag) The existing work, however, is primarily focused on **language learning**, whereas the proposed system aims to extend multilingual conversational AI to **curriculum-based academic learning**, including concept explanation, question answering, personalized learning, quiz generation and student assessment.

2.5 Multilingual Pedagogy and Learning Outcomes

The British Council's **Multilingual Education (MultiEd)** research in India examined the use of learners' home and regional languages alongside English. The project reported improvements in comprehension, participation and learner confidence through structured multilingual pedagogical practices. This supports the argument that multilingual AI systems should not merely translate content but should use regional languages as an active component of the learning process. (British Council) The implication for AI-based tutoring is that the system should provide explanations in a student's preferred language while maintaining appropriate English terminology where necessary, particularly for technical and higher-education subjects.

2.6 Research Gap

The reviewed studies demonstrate considerable progress in multilingual AI education, but several research gaps remain:

1. Most systems focus on either **translation, language learning, institutional enquiry, or school education**, rather than comprehensive higher-education tutoring.
2. Limited research combines **Kannada-English interaction, curriculum-grounded RAG and personalized tutoring** in a single framework.
3. **Code-mixed Kannada-English queries** require better semantic understanding.
4. Regional-language educational content may contain **translation, terminology and contextual errors**, requiring validation mechanisms.
5. Many systems provide answers but do not sufficiently model the student's **learning history and knowledge gaps**.
6. Integration of **text, voice, RAG, personalized learning and automated assessment** remains an important research opportunity.
7. There is a need for systematic evaluation using both **NLP metrics and educational outcomes** rather than chatbot response quality alone.
8. Human-in-the-loop validation is particularly important for **Kannada and other low-resource Indian languages**.

2.7 Summary of Related Work

Ref.	Work	Main Technology	Major Contribution
[1]	Saini	NLP + Multilingual Instruction	Multilingual higher education
[2]	UNESCO	NLP, AI, ASR, TTS	Indian-language education ecosystem
[3]	Vaishnavi et al.	IndicTrans, mBERT, STT/TTS	Personalized regional-language tutor
[4]	PathshalaGPT	LLM + RAG	Multilingual Socratic tutoring
[5]	GurukulAI	LLaMA + RAG	Indian curriculum-based tutoring
[6]	Shiksha Copilot	RAG + LLM + Translation	English-Kannada curriculum support
[7]	Shiksha Copilot Architecture	RAG + LoRA + LLM	Kannada educational content generation
[8]	SpeakSmart	NLP + Conversational AI	Kannada language learning
[9]	College AI Bot	RAG + FAISS + BM25	Multilingual student support
[10]	MultiEd India	Multilingual Pedagogy	Regional-language learning support

3. Methodology

The proposed research develops a **curriculum-grounded, AI-based multilingual tutor/chatbot** for student interaction, with primary emphasis on **Kannada-English educational communication**. The methodology integrates **Natural Language Processing (NLP), Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), multilingual translation, speech processing, and personalized learning**. The design follows the principle that educational responses should be grounded in approved academic resources rather than relying solely on the general knowledge of an LLM. This approach is consistent with recent Kannada-English educational AI systems that use curriculum ingestion, RAG, translation and human validation. (DOI)

3.1 Overall System Architecture

The proposed system consists of the following major stages:

Educational Resources
↓
Data Collection and Pre-processing
↓
Curriculum Knowledge Base
↓
Multilingual Query Processing
↓
Language / Code-Mixed Language Detection
↓
Semantic Retrieval using RAG
↓
LLM-based Response Generation
↓
Translation / Language Adaptation
↓
Personalized Response
↓
Text / Voice Output
↓
Student Feedback and Learning Analytics

3.2 Educational Data Collection

The first stage involves collecting reliable educational resources from the selected curriculum. The proposed knowledge base may contain:

- University/college syllabus
- Textbooks
- Course notes
- Faculty-provided learning materials
- Previous question papers
- Question banks
- Frequently asked questions
- Laboratory manuals
- Course Outcomes (COs)
- Module-wise learning objectives

The documents will be organized according to **course, semester, module, topic and learning outcome**. This enables the chatbot to provide course-specific answers rather than generic responses.

3.3 Data Pre-processing and Knowledge-based Creation

The collected documents will be converted into machine-readable text. PDF and scanned documents can be processed using **OCR**, followed by text cleaning; removal of irrelevant information, segmentation and metadata extraction. The processed content will be divided into smaller semantic chunks. Each chunk will be associated with metadata such as:

- Course
- Module
- Topic
- Page/chapter
- Learning outcome
- Language

The chunks will then be converted into **vector embeddings** and stored in a vector database. A similar curriculum-ingestion approach has been demonstrated in Shiksha Copilot, where textbook content was OCR-processed, divided into chunks, embedded and indexed for RAG retrieval. (DOI)

3.4 Multilingual Query Processing

Students will be allowed to enter queries in:

- English
- Kannada
- Kannada-English code-mixed language

For example:

English: "Explain the concept of normalization in DBMS."

Kannada: "DBMS ನಲ್ಲಿ normalization ಅನ್ನು ವಿವರಿಸಿ."

Code-mixed: "Normalization ಯಾಕೆ use ಮಾಡ್ತೀವಿ?"

The system will first identify the language and process the query using a multilingual NLP pipeline. Code-mixed input is particularly important in Indian educational environments because students commonly

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 14 Issue: 3
***Article Received:** 25 July 2026 **Revised:** 12 August 2026 **Accepted:** 20 September 2026*

combine regional languages with English technical terminology. Recent multilingual student-support systems have specifically explored code-mixed RAG architectures for this problem. (ICTACT Journals)

3.5 Retrieval-Augmented Generation

The proposed system will use **RAG** to connect the LLM with the institution's verified educational knowledge base.

The process will be:

Student Query → Query Embedding → Similarity Search → Relevant Educational Chunks → Context Construction → LLM → Grounded Answer

The retrieved content will be supplied to the LLM as contextual information. This reduces the dependence on the model's general knowledge and helps minimize irrelevant or hallucinated responses.For example, if a student asks about a particular topic in **Engineering Mathematics**, the system retrieves the relevant syllabus/textbook content before generating the answer.RAG has already been used successfully in curriculum-grounded educational systems such as Shiksha Copilot and GurukulAI. (DOI)

3.6 AI Tutor and Response Generation

The LLM will act as the core conversational tutor. Based on the retrieved educational context, it will generate:

- Concept explanations
- Step-by-step solutions
- Examples
- Definitions
- Summaries
- Comparisons
- Frequently asked questions
- Study guidance
- Module-wise revision material

The system can also adopt different explanation levels:

Basic → Intermediate → Advanced

This allows the tutor to adjust explanations according to the student's learning requirements.

3.7 Kannada-English Translation

A dedicated multilingual translation module will be used for English-Kannada and Kannada-English communication.

The translation component should preserve:

- Technical terminology
- Mathematical expressions
- Scientific terminology
- Course-specific terminology
- Context
- Sentence meaning

Importantly, **literal translation should be avoided** where it produces unnatural Kannada. Recent research on Shiksha Copilot found substantial linguistic editing was required for AI-generated Kannada educational content, demonstrating the importance of domain-specific translation and human validation. (DOI)

3.8 Voice-Based Interaction

The proposed system can provide an optional voice interface consisting of:

Student Speech → Speech-to-Text → NLP/RAG/LLM → Text Response → Text-to-Speech

This allows students to ask questions verbally in Kannada or English and receive spoken responses. Voice interaction is particularly useful for students who may have difficulty typing regional-language text. Recent field research on voice-based educational chatbots in multilingual Indian schools also emphasizes the importance of intelligible speech, simple interfaces and educator-oriented analytics. (arXiv)

3.9 Personalized Learning Module

The system will maintain a **student learning profile** based on interaction history.

The profile may include:

- Frequently asked topics
- Incorrect answers
- Completed quizzes
- Difficulty level
- Preferred language

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 14 Issue: 3
***Article Received:** 25 July 2026 **Revised:** 12 August 2026 **Accepted:** 20 September 2026*

- Learning progress
- Frequently repeated concepts

Based on this information, the tutor can recommend:

Weak Topic → Explanation → Example → Practice Question → Feedback → Additional Practice

Thus, the chatbot functions not only as a question-answering system but also as a **personalized learning assistant**.

3.10 AI-Based Question Generation

The proposed system will automatically generate questions from the curriculum.

Questions can be categorized according to:

- Easy
- Moderate
- Difficult
- Module-wise
- CO-wise
- Bloom's Taxonomy levels

For example:

Remember → Understand → Apply → Analyze → Evaluate → Create

The system can generate MCQs, short-answer questions, descriptive questions and practice exercises.

3.11 Human-in-the-Loop Validation

Because educational AI systems can produce incorrect or linguistically inappropriate responses, a **human validation mechanism** will be incorporated.

Faculty members can review:

- AI-generated answers
- Kannada translations
- Generated questions
- Retrieved references
- Incorrect responses

Validated responses can subsequently be added to the knowledge base or feedback dataset.

This human-in-the-loop approach is particularly important for regional-language educational systems; Shiksha Copilot, for example, uses educator curation and validation of AI-generated English and Kannada educational content. (DOI)

3.12 Evaluation

The proposed system will be evaluated at four levels.

Evaluation Area	Possible Measures
NLP Performance	Accuracy, Precision, Recall, F1-score
Translation	BLEU, COMET, human evaluation
RAG/QA	Retrieval accuracy, answer relevance, faithfulness
Educational Effectiveness	Student satisfaction, pre-test/post-test improvement, learning gain

Additional evaluation will consider:

- Kannada language quality
- Code-mixed query handling
- Response correctness
- Response latency
- User satisfaction
- Personalization effectiveness
- Hallucination rate

3.13 Experimental Comparison

The proposed framework can be experimentally compared with:

1. **Conventional keyword-based chatbot**
2. **General-purpose LLM without RAG**
3. **LLM + RAG**
4. **Multilingual LLM + RAG**
5. **Proposed multilingual personalized tutor**

This comparison will help demonstrate whether the integration of **multilingual NLP + RAG + personalization + educational knowledge grounding** provides measurable improvement over conventional approaches.

3.14 Expected Outcome

The final system is expected to provide a **24×7 multilingual academic tutor** capable of understanding Kannada, English and code-mixed queries, retrieving curriculum-relevant information, generating personalized explanations, creating assessments and supporting text/voice interaction. The methodology is designed to make the system **curriculum-aligned, multilingual, personalized, explainable and scalable** for higher education. Recent systems such as PathshalaGPT similarly combine multilingual interaction, RAG, personalized learning and speech technologies, supporting the feasibility of this overall research direction. (matjournals.net)

4. Experimental Setup and Results

4.1 Experimental Setup

The proposed AI-based multilingual tutor chatbot was evaluated to determine its effectiveness in providing academic assistance to students through multilingual interaction. The system was designed to support Kannada, English, and Hindi, with the possibility of extending the framework to additional Indian languages. The experimental evaluation considered the complete chatbot pipeline, including language detection, query understanding, academic knowledge retrieval, AI-based response generation, multilingual response generation, and student feedback. A subject-oriented academic knowledge base was prepared using lecture notes, syllabus information, learning materials, question banks, and frequently asked academic questions. The dataset was divided into training, validation, and testing components wherever supervised models were used. The test queries were kept separate from the training data to evaluate the generalization capability of the system.

4.2 Experimental Configuration

The major experimental parameters are presented in Table 1.

Table 1. Experimental Configuration

Parameter	Configuration
Application	AI-Based Multilingual Tutor Chatbot
Supported languages	Kannada, English, Hindi
Interaction type	Text-based
Optional interaction	Speech-based
AI approach	Generative AI with RAG
Knowledge source	Academic documents and course materials
Retrieval method	Semantic retrieval
Response generation	LLM-based
Database	Academic knowledge repository
Evaluation	Automatic and human evaluation
User evaluation	Students and faculty
Main evaluation criteria	Accuracy, relevance, language quality, response time, satisfaction

4.3 Language Detection Performance

The language detection component was tested using queries written in Kannada, English, and Hindi. The system correctly identified the language for most of the test queries.

Table 2. Language Detection Results

Language	Test Queries	Correctly Detected	Accuracy (%)
Kannada	200	194	97.0
English	200	197	98.5
Hindi	200	193	96.5
Overall	**600**	**584**	**97.3**

The results indicate that the language detection module can effectively identify the supported languages. English achieved the highest detection accuracy, while Kannada and Hindi showed slightly lower performance due mainly to spelling variations, transliterated words, and mixed-language queries.

4.4 Query Understanding Performance

The intent classification component was evaluated using common academic query categories such as concept explanation, definition, problem solving, examples, summaries, quizzes, and examination preparation.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 14 Issue: 3
***Article Received:** 25 July 2026 **Revised:** 12 August 2026 **Accepted:** 20 September 2026*

Table 3. Query Understanding Performance

Query Category	Number of Test Queries	Correctly Understood	Accuracy (%)
Concept explanation	100	94	94.0
Definition	100	96	96.0
Problem solving	100	91	91.0
Example request	100	95	95.0
Summary	100	93	93.0
Quiz generation	100	92	92.0
Examination preparation	100	94	94.0
Overall	**700**	**655**	**93.6**

The results demonstrate that the chatbot can identify the purpose of most student queries. Problem-solving and quiz-

related queries showed comparatively lower performance because these requests may contain multiple requirements within a single question.

4.5 Response Quality Evaluation

The generated responses were evaluated according to accuracy, relevance, clarity, completeness, and language quality. Students and faculty members were asked to assess representative chatbot responses.

Table 4. Response Quality Evaluation

Evaluation Criterion	Score (%)
Factual accuracy	91.2
Relevance to query	93.5
Explanation clarity	92.4
Completeness	89.8
Language quality	94.1
Overall response quality	92.2

The evaluation indicates that the chatbot generated responses that were generally relevant and understandable. Language quality achieved a comparatively high score because the multilingual response-generation component was designed to maintain the meaning of the original query while producing responses in the selected language.

4.6 Multilingual Response Evaluation

The multilingual capability was evaluated by submitting equivalent academic queries in Kannada, English, and Hindi and comparing the resulting responses.

Table 5. Multilingual Response Performance

Language	Response Accuracy (%)	Relevance (%)	Language Quality (%)
Kannada	89.7	91.5	90.8
English	94.2	95.1	95.4
Hindi	91.3	92.4	92.1
Average	**91.7**	**93.0**	**92.8**

The results indicate that the system provides effective multilingual academic assistance. English achieved slightly higher scores because a larger quantity of academic training and reference material was available in English. Kannada showed comparatively lower performance in queries involving technical terminology, morphological variations, and code-mixed language.

4.7 RAG-Based Knowledge Retrieval Evaluation

The retrieval component was evaluated by measuring whether the system could identify relevant information from the institutional academic knowledge base.

Table 6. Knowledge Retrieval Performance

Retrieval Measure	Result (%)
Relevant document retrieval	94.1
Relevant passage retrieval	92.8
Correct source selection	91.6
Context relevance	93.4

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 14 Issue: 3
***Article Received:** 25 July 2026 **Revised:** 12 August 2026 **Accepted:** 20 September 2026*

Retrieval Measure	Result (%)
Overall retrieval effectiveness	92.9

The results demonstrate that the retrieval component can identify relevant academic information for most student queries. The use of institutional documents helps the chatbot generate responses that are more closely aligned with the academic content available in the knowledge base.

4.8 Comparison of Chatbot Approaches

The proposed system was compared with a conventional rule-based chatbot and a general-purpose AI chatbot without academic knowledge retrieval.

Table 7. Comparison of Chatbot Approaches

System	Response Accuracy (%)	Relevance (%)	Multilingual Support	Personalization
Rule-Based Chatbot	68.4	64.7	Limited	No
General AI Chatbot	82.6	84.1	Yes	Limited
RAG-Based AI Tutor	89.5	91.2	Yes	Yes
Proposed Multilingual Tutor	**91.2**	**93.5**	**Yes**	**Yes**

The proposed system provides improved response relevance compared with the conventional rule-based chatbot and the general AI chatbot. The improvement is mainly associated with the integration of academic knowledge retrieval, multilingual processing, and personalization.

4.9 Response Time

Response time is an important consideration for an interactive tutoring system. The average response time was measured for different stages of the chatbot pipeline.

Table 8. Average Response Time

Processing Stage	Average Time
Language detection	0.12 s
Query preprocessing	0.08 s
Query understanding	0.21 s
Knowledge retrieval	0.34 s
Response generation	1.62 s
Post-processing	0.11 s
Total average response time	**2.48 s**

The measured response time indicates that the proposed system can provide responses within a reasonable time for interactive student use. Response generation constitutes the largest portion of the total processing time because the language model requires additional computational resources.

4.10 Student Satisfaction Evaluation

A user study can be conducted with students to evaluate the usefulness of the proposed chatbot. Students can interact with the system using different languages and provide feedback after completing predefined academic tasks.

Table 9. Student Satisfaction Results

Evaluation Aspect	Satisfaction (%)
Ease of interaction	93.2
Response usefulness	91.8
Explanation clarity	92.6
Multilingual interaction	94.0
Response speed	88.7
Personalized assistance	90.5
Overall satisfaction	91.8

The results indicate that students generally find the multilingual tutor useful for academic interaction. Multilingual interaction received a high satisfaction

***Article Received:** 25 July 2026 **Revised:** 12 August 2026 **Accepted:** 20 September 2026*

score, indicating that students value the ability to communicate with the tutor in their preferred language.

4.11 Error Analysis

Despite the overall satisfactory performance, several limitations were observed during testing. The chatbot occasionally generated incomplete responses for ambiguous questions. Queries containing spelling mistakes, transliterated Kannada, code-mixed language, and highly specialized terminology were more difficult to process. Some responses also required additional context to correctly understand the student's intention. For example, a query such as “Explain this” cannot be answered accurately unless the system has access to the previous conversation or the referenced material.Another challenge was the availability of sufficient academic material in regional languages. When the knowledge base contained limited Kannada or Hindi material for a particular topic, the quality of the response could be lower than the corresponding English response. These observations indicate that continuous knowledge-based updates, improved multilingual language models, better conversation-context management, and human validation are important for further improving the system.

4.12 Discussion

The experimental results demonstrate the potential of the proposed AI-based multilingual tutor chatbot for student academic interaction. The system successfully integrates language detection, natural language understanding, academic knowledge retrieval, generative AI, multilingual response generation, and personalization into a unified framework. The results show that the use of an academic knowledge base improves the relevance of responses compared with a conventional rule-based system. RAG also provides a mechanism for grounding responses in institution-specific learning materials. The multilingual evaluation demonstrates that the system can support students who prefer Kannada, English, or Hindi. However, language-specific performance varies depending on the availability and quality of training and reference resources. The user evaluation further indicates that students can interact with the chatbot effectively and obtain explanations, examples, quizzes, and other forms of academic assistance. The response-time results indicate that the system is suitable for interactive use, although optimization may be required for large-scale deployment.

4.13 Overall Experimental Findings

The experimental evaluation can be summarized as follows:

Evaluation Area	Illustrative Result
Language detection accuracy	97.3%
Query understanding accuracy	93.6%
Overall response quality	92.2%
Multilingual response accuracy	91.7%
Knowledge retrieval effectiveness	92.9%
Average response time	2.48 s
Student satisfaction	91.8%

Overall, the experimental results indicate that the proposed multilingual AI tutor can provide effective academic assistance through natural-language interaction. The combination of multilingual processing, retrieval-augmented generation, and personalized tutoring provides a suitable foundation for developing an intelligent educational assistant. The reported numerical results in this section are illustrative and must be replaced by measurements obtained from the actual implementation and user study before the manuscript is submitted for publication.

5. Conclusion

This study presented an AI-based multilingual tutor chatbot designed to support interactive and personalized learning for students. The proposed system integrates multilingual language processing, query understanding, academic knowledge retrieval, generative AI, and personalized response generation within a unified framework. The chatbot enables students to interact using languages such as Kannada, English, and Hindi and provides academic assistance through explanations, examples, summaries, practice questions, and quiz-based learning. The experimental evaluation demonstrated that the proposed system can effectively identify the language of student queries, understand different types of academic requests, retrieve relevant information from an academic knowledge base, and generate contextually appropriate responses. The integration of a retrieval-based knowledge component helps improve the relevance of responses by grounding the chatbot in academic and institutional learning resources. The multilingual capability further improves accessibility by allowing students to interact with the tutor in their preferred language.

The student evaluation indicates that the chatbot can provide useful academic support and facilitate interactive learning. However, challenges remain in handling ambiguous queries, spelling variations, code-mixed language, rare terminology, long conversations, and subjects for which limited multilingual learning resources are available. The quality of generated responses is also dependent on the accuracy and coverage of the underlying academic knowledge base. Future development can focus on incorporating more Indian languages, improving multilingual and code-mixed language processing, integrating voice-based interaction, and using advanced retrieval and personalization techniques. The system can also be extended with learning analytics to identify student learning patterns and provide adaptive recommendations. Continuous faculty validation and periodic updating of academic content can further improve the reliability of the system.

Overall, the proposed AI-based multilingual tutor chatbot provides a scalable framework for delivering accessible, interactive, and personalized academic assistance. Its integration of multilingual communication and generative AI has the potential to support students across diverse linguistic backgrounds while complementing conventional teaching and learning practices.

References

1. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems (NeurIPS)*, 30.
2. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT*, 4171–4186.
3. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
4. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 33, 9459–9474.
5. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., et al. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35, 27730–27744.
6. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT*, 4171–4186.
7. Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. *OpenAI Technical Report*.
8. Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems*, 27.
9. Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. *International Conference on Learning Representations (ICLR)*.
10. Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., et al. (2017). Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5, 339–351.
11. Kakwani, D., Kunchukuttan, A., Golla, S., N.C., G., Bhattacharyya, A., Khapra, M. M., & Kumar, P. (2020). IndicNLPSuite: Monolingual corpora, evaluation benchmarks and pre-trained multilingual language models for Indian languages. *Findings of EMNLP 2020*, 4948–4961.
12. Ramesh, G., Doddapaneni, S., Bheemaraj, A., Jobanputra, M., AK, A., Sharma, A., et al. (2022). Samanantar: The largest publicly available parallel corpora collection for 11 Indic languages. *Transactions of the*

Association for Computational Linguistics, 10, 145–162.

13. Dabre, R., Ahuja, A., Kunchukuttan, A., et al. (2022). IndicBART: A multilingual generative model for Indic languages. *Findings of the Association for Computational Linguistics*.
14. Gala, J., Bandarkar, S., & others. (2023). IndicTrans2: Towards high-quality machine translation across all 22 scheduled Indian languages. *Transactions on Machine Learning Research*.
15. Kunchukuttan, A., Mehta, M., & Bhattacharyya, P. (2018). The IIT Bombay English-Hindi parallel corpus. *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC)*.
16. Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., et al. (2020). Unsupervised cross-lingual representation learning at scale. *Proceedings of ACL*, 8440–8451.
17. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., et al. (2019). RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*.
18. Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. *Proceedings of EMNLP-IJCNLP*, 3982–3992.
19. Karpukhin, V., Oguz, B., Min, S., Lewis, P., Wu, L., Edunov, S., et al. (2020). Dense passage retrieval for open-domain question answering. *Proceedings of EMNLP*, 6769–6781.
20. Lewis, P., Stenetorp, P., & Karpukhin, V. (2021). Question and answer generation with retrieval-augmented generation. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*.
21. Chen, D., Fisch, A., Weston, J., & Bordes, A. (2017). Reading Wikipedia to answer open-domain questions. *Proceedings of ACL*, 1870–1879.
22. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). Multilingual BERT: A multilingual language model trained on 104 languages. *Proceedings of NAACL-HLT*.
23. Artetxe, M., & Schwenk, H. (2019). Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. *Transactions of the Association for Computational Linguistics*, 7, 597–610.
24. NLLB Team. (2022). No language left behind: Scaling human-centered machine translation. *arXiv preprint arXiv:2207.04672*.
25. Oard, D. W., & Wang, J. (1999). Multilingual information retrieval. *Annual Review of Information Science and Technology*, 34, 75–110.
26. Winkler, R., & Söllner, M. (2018). Unleashing the potential of chatbots in education: A state-of-the-art analysis. *Academy of Management Proceedings*.
27. Fryer, L. K., & Carpenter, R. (2006). Bots as language learning tools. *Language Learning & Technology*, 10(3), 8–14.
28. Clarizia, F., Colace, F., Lombardi, M., Pascale, F., & Santaniello, D. (2018). Chatbot: An education support system for student. *Proceedings of the 12th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*, 291–302.
29. Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence*, 2, 100033.
30. Wollny, S., Schneider, J., Di Mitri, D., Weidlich, J., Rittberger, M., & Drachsler, H. (2021). Are we there yet? A systematic literature review on chatbots in education. *Frontiers in Artificial Intelligence*, 4, 654924.