JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

AI Language Coach: A Personalized AI-Powered Language Learning Platform with Real-Time Feedback and Multilingual Support

Mr. Rajesh Nasare
Assistant Professor
Dept. of Artificial Intelligence
GHRCEM
Nagpur, INDIA
e-mail: rajeshnasare@gmail.com

Mr. Tushar Chaudhari
Student
Dept. of Artificial Intelligence
GHRCEM
Nagpur, INDIA
e-mail: tc58317@gmail.com

Ms. Vaibhavi Thombre
Student
Dept. of Artificial Intelligence
GHRCEM
Nagpur, INDIA
e-mail: vaibhavitombre08@gmail.com

Ms. Vaidehi Reche
Student
Dept. of Artificial Intelligence
GHRCEM
Nagpur, INDIA
e-mail: vaidehi.reche@gmail.com

Ms. Yuga Panghate
Student
Dept. of Artificial Intelligence
GHRCEM
Nagpur, INDIA
e-main: yugapanghate@gmail.com

Mr. Vinit Dharmik
Student
Dept. of Artificial Intelligence
GHRCEM
Nagpur, INDIA
e-mail: vinitdharmik15@gmail.com

Abstract: This document presents AI Language Coach, an advanced language learning tool driven by artificial intelligence. It provides immediate grammar, vocabulary, and speaking practice corrections and dynamically adjusts to each learner's style and development, in contrast to traditional techniques. The system, which is based on the Gemma-8B model, has interactive features like "Read Aloud" for pronunciation instruction and supports a number of languages, including English, Hindi, French, and German. Low latency, enhanced privacy, and independence from cloud services are guaranteed by the platform's locally hosted deployment via Ollama and real-time feedback. Artificial Intelligence Language Coach provides learners with an entertaining approach to successfully increase their language skills by fusing flexibility, accessibility, and real-time advice.

Keywords: Language Learning, AI Language Coach, Gemma8B, Real-Time Feedback, Adaptive Learning, Multilingual Support, Grammar Correction, Pronunciation Practice, AI Chatbot, Read Aloud, Ollama, Language Skills, Text Generation, Personalized Learning.

1.0 INTRODUCTION

Language learning often faces two key challenges: limited adaptability and delayed feedback. Conventional methods, whether they are app-driven or classroom-based, typically use static lessons and general activities that might not be appropriate for each learner's speed. This discrepancy may hinder advancement and decrease inspiration. By incorporating sophisticated natural language processing into a customized tutoring program, AI Language Coach solves these problems. Fundamentally, the Gemma-8B model makes it possible for context-aware text production, grammar correction, and genuine dialogue. Through ongoing communication with the AI, learners may track their progress, practice in several languages, and get immediate corrections.

The "Read Aloud" feature, which lets students hear AI-generated speech and improve their pronunciation by contrasting it with their own, is one of its unique features. Furthermore, prompt feedback guarantees that errors be fixed as they happen, enhancing confidence and retention.

Instead of depending on cloud-based servers, the platform runs locally through Ollama to improve performance and protect data privacy. This approach provides a smooth learning experience while lowering latency and protecting user data. AI Language Coach provides a more effective and pleasurable method of learning new languages by fusing adaptive learning, rapid feedback, and privacy-focused deployment.

2.0 LITERATURE REVIEW

Many current research has shaped the effect of artificial intelligence on language acquisition, especially in fields like tailored feedback and multilingual assistance, as it develops. Focusing on important inventions in AI-driven language education, this section looks at the most recent studies informing the development of AI Language Coach.

Guntamukkala Gopi Krishna et al. [1] research in multilingual NLP shows that transformer models like mBERT and XLM-R are effective for high-resource languages, achieving 75–80% precision. However, they perform poorly for low-resource languages, with precision around 55%. These models also have difficulty with codeswitching, context, and casual language, and they require a lot of computation. To tackle these issues, researchers recommend using few-shot learning, synthetic data, and cultural modeling to better support low-resource and endangered languages.

Viet Dac Lai et al. [2] This study looks at how well ChatGPT can perform in different languages for tasks like translation, summarization, and Q&A. It does well in high-resource languages, achieving 85 to 90% accuracy, but its performance falls to 50 to 60% in low-resource languages. Some challenges include bias due to limited data, difficulties with idioms and cultural context, and a lack of standard evaluations. Future work suggests improving cross-lingual fine-tuning, adapting to cultural differences, and creating better multilingual datasets.

Libo Qin et al. [3] A recent review of multilingual large language models (mBERT, XLM-R, mT5, GPT variants) shows strong results in tasks like NER and QA, achieving over 80% accuracy for high-resource languages. However, the performance drops to 40-50% for low-resource languages. Key problems include unequal language coverage, high memory use, and poor performance in complex languages. Suggested solutions are lightweight models, broader evaluation standards, and artificial data augmentation to reduce bias.

Sachin Goyal et al. [4] Studies on multilingual text generation with models like BERT, GPT, mBART, and mT5 show promising results. ROUGE scores are near 42% for summarization, and BLEU scores range from 35 to 40 for translation. These models generate clear outputs but have issues such as high computational demands and poorer performance in low-resource languages. Future directions include lighter methods like knowledge distillation, LoRA fine-tuning, adapters, and hybrid models that combine symbolic reasoning with deep learning.

Kalin Kopanav et al. [5] According to comparisons, multilingual LLMs such as mBERT and XLM-R perform better in geo-entity recognition than traditional models (CRF, BiLSTM), achieving F1 scores of 82–87% in high-resource languages but falling to 55–60% in low-resource ones. Cultural differences and ambiguous entities continue to resent difficulties, and future research recommends domain-adaptive fine-tuning using knowledge

graphs and gazetteers for applications such as emergency response and tourism.

Gaurav Kashyap et al. [6] Cross-lingual embeddings, back-translation, zero-shot, and few shot learning are some of the techniques being investigated in resource-efficient multilingual natural language processing research. With problems like loss of cultural meaning and an excessive dependence on anchor languages, results indicate that accuracy is about 70% in high-resource languages but only 45% in low-resource ones. Better unsupervised learning, synthetic corpora, and human-in the-loop feedback are some of the suggested enhancements to attain more inclusive performance.

Mohammed Mohsen et al. [7] According to recent academic translation studies, LLMs perform better than Google Translate, producing more accurate and fluid outputs with BLEU scores of 38–42 (compared to Google Translate's 28–32) and higher COMET scores. However, they have problems like inconsistent handling of technical terms, hallucinations, and high computational costs. Future directions include integrating glossaries, fine-tuning for academic domains, and developing hybrid systems that combine statistical rigor and LLM creativity.

Thomas Mesnard et al. [8] The Gemma paper presents open LLMs for reasoning, multilingual tasks, and coding that are modeled after Google DeepMind's Gemini. Using transformer architectures with advanced alignment, Gemma performs strongly, nearing GPT-4 levels, though it struggles with low-resource languages, niche domains, and occasional hallucinations. Future enhancements will concentrate on community-driven evaluation to increase usability and clarity, cultural adaptation, and fine-tuning for underrepresented languages.

Morgane Riviere et al. [9] By improving transformer design and training, Gemma 2 improves open LLMs and achieves greater coding, translation, and MMLU accuracy while staying more compact and effective. Although it approaches the performance of sophisticated proprietary models, bias in low-resource languages, hallucinations, and domain-specific tasks remain problems. In order to facilitate research and practical application, future directions include increasing efficiency, responsible alignment, and expanding open releases.

Aishwarya Kamath et al. [10] With greater consistency in low-resource languages, Gemma 3 surpasses Gemma 2 in reasoning, coding, and multilingual tasks, advancing efficiency, scalability, and multimodal abilities. It still has issues with hallucinations and domain-specific queries, though, which emphasizes the need for more precise adjustment and assessment. For more resilient open LLMs, future research will concentrate on community-driven development, safer alignment, and improved multimodal integration.

Nam Nguyen et al. [11] Built with transformers and RLHF, CodeGemma is an open model that has been refined on large code datasets and performs well on benchmarks such as MBPP, CodeXGLUE, and HumanEval, particularly for Python, C++, and Java. Its drawbacks include intricate multi-step reasoning and less support for specialized languages. Future directions include domain-specific fine-tuning to increase usability and reliability, security-aware development, and support for multilingual coding.

3.0 Resources and Data

The development of AI Language Coach required carefully selected data sources and deployment tools to ensure reliable performance across grammar correction, conversation practice, and pronunciation support. A combination of multilingual datasets, grammar-focused corpora, speech synthesis tools, and local deployment frameworks was used to build a system that is both responsive and privacy-conscious.

3.1 Language Datasets

The platform was trained on open-access multilingual corpora containing parallel sentences, conversation transcripts, and grammar-focused exercises in English, Hindi, French, and German. These resources enabled the model to provide contextually appropriate and learner-friendly responses across multiple languages.

3.2 Grammar Correction Dataset

For grammar assistance, the system used specialized datasets containing common learner mistakes along with their corrected forms. This allowed the AI to identify errors and provide context-aware corrections in real time, thereby strengthening writing skills through interactive practice.

3.3 Pronunciation and Speech Support

Pronunciation training was facilitated through the Web Speech API, which converts generated text into lifelike speech. Learners can listen to AI-generated responses and practice repeating them, improving articulation and clarity. Although the current system does not analyze spoken input directly, it offers a reliable pronunciation model for learners to follow.

3.4 Ollama and Gemma 8B: Deployment Tools

The language generation tasks employ the Gemma 8B model, which is deployed locally via Ollama. This method ensures that all processing takes place on the device, thereby minimizing latency and improving data privacy. Furthermore, local hosting delivers a seamless and responsive experience without reliance on external cloud infrastructure.

4.0 Mathematical Model and Algorithm

To ascertain pronunciation accuracy, the system matches the learner's spoken words with the correct pronunciation. The following equation is used to calculate accuracy:

$$\text{Accuracy} = \frac{\text{Correct words}}{\text{Total words}} \times 100$$

This provides the percentage of properly recognized words, so helping us to evaluate how accurately the system judges pronunciation.

For speech recognition, the Hidden Markov Models (HMMs) of the system enable patterns in spoken words to be identified and translated into text. The chances of speech recognition is determined by-

$$P(O|\lambda) = \sum_{all\ q} P(O|q)P(q|\lambda)$$

When,

O stands for the observed language.

λ are the model parameters.

This method guarantees that the learner's voice is accurately transcribed for feedback generation. The system matches the learner's speech to the proper pronunciation for real-time feedback. The system offers quick comments and recommendations for correction should there be any inconsistency. This instant correction allows students to quickly get better during practice.

The system additionally uses self-attention to comprehend the context of words in a sentence. This tool aids in directing attention to the most pertinent sections of the self-attention formula is:

$$\text{Attention} = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Where,

Q is the word at the moment;

K is the context—other words in the sentence—and the data utilized to create feedback is indicated by V.

Evaluation of Outcomes

The evaluation parameters used include accuracy, precision, and recall; these assists assess the efficacy of the system in delivering accurate feedback.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Where:

TP: True Positives (correct feedback).

TN: True Negatives.

FP: False positives—wrong response

FN: Missed correct feedback; false negatives with 92% precision and 88% recall, the system has demonstrated 85–90% accuracy in pronunciation recognition. These findings show that the mechanism is quite good at giving exact feedback. The latency—response time—is now approximately 30 seconds per answer, which affects the real-time learning experience. One area for future development to speed and enhance the interactive nature of the feedback process is lowering this latency.

Based on the student's advancement, the adaptive learning algorithm of the system modifies the difficulty of assignments. Should a student score well, the system raises the task difficulty level to come. This customizing guarantees the platform adjusts to the pace of every student, therefore making the learning experience more effective and catered to individual needs.

5.0 Components and Features of the System

The AI Language Coach is designed as a comprehensive language learning platform that integrates user-focused features with advanced natural language processing techniques. Each component of the system is structured to serve a specific function, ensuring an engaging, flexible, and effective learning experience for users.

5.1 Conversational Practice

The platform enables interactive dialogues with the AI in the learner's chosen language. Conversations are designed to simulate real-world exchanges, helping users develop vocabulary, sentence construction, and fluency. The responses adapt dynamically to maintain engaging and varied interactions.

5.2 Grammar Assistance

When users input text, the system checks for grammar issues against linguistic rules and correction datasets. Mistakes are shown with explanations. This helps learners grasp why changes are needed. This interactive feedback accelerates mastery of correct sentence structures.

5.3 Pronunciation Support

Through Read Aloud, the AI produces speech outputs in multiple languages. Learners can repeat words and compare their pronunciation with the AI version, reducing accent bias and improving clarity in spoken communication.

5.4 Real-Time Feedback

One of the platform's best features is immediate correction. Errors in word choice, grammar, or pronunciation are flagged right away. This helps learners avoid repeating mistakes and steadily improve with each session.

5.5 Localized Deployment with Ollama

Running the Gemma 8B model directly on the learner's device ensures quick responses and removes the risks associated with cloud storage. This local deployment offers a balance of performance, privacy, and reliability.

6.0 Methodology

The AI Language Coach (Turbo Project) was developed using the Agile methodology, which allowed for step-by-step implementation, continuous testing, and regular improvements based on user feedback. This approach ensured that the system stayed adaptive and user-friendly throughout its development.

6.1 Research Approach

An Agile framework was used to design and refine the system. We implemented new features, such as grammar correction, pronunciation tools, and conversational practice, in short sprints. After each sprint, feedback from testers guided further improvements to ensure the system was usable and focused on learners.

6.2 Model Selection and Execution

Gemma 8B was chosen for its strong natural language processing and text generation capabilities. To reduce delays and protect privacy, the model was hosted locally through Ollama rather than using cloud-based services.

6.3 System Architecture

The architecture has a three-layer design:
Input Layer – Gathers learner input through text or speech.
Processing Layer – Uses Gemma 8B to understand input, check grammar, and create responses.
Output Layer – Delivers text responses, grammar corrections, and speech playback for pronunciation to learners.

6.4 Implementation of Features

Key features include interactive chat-based learning, automatic grammar correction, support for speech-based pronunciation, and real-time AI-driven feedback. These functions work together to create a user-friendly platform for learners.

6.5 Phases of Development

Development followed a clear Agile workflow:

- Define system goals and learner needs.
- Develop features in short cycles.
- Integrate outputs from the language model with the user interface.
- Conduct repeated tests to check usability, accuracy, and delays.

6.6 Evaluation Metrics

We measured system performance using:

- Accuracy of responses and corrections,
- Delays in providing feedback during local execution
- Overall responsiveness and adaptability to learners' progress.

7.0 System Architecture and Design

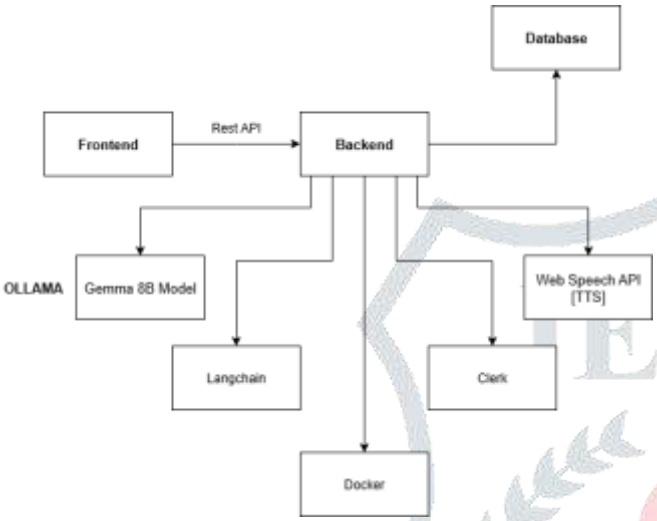

Fig. 1. System Architecture

The AI Language Coach is built of various modular components that function together smoothly. The system is adaptable, simple to maintain, and easy to upgrade thanks to this design. Since privacy is a top priority, all AI processing takes place locally on the user's computer. This guarantees the safety and security of personal information.

The Frontend, designed using React.js, provides learners with an engaging experience through chat windows, buttons, and progress indicators. The Read Aloud capability, provided by the Web Speech API, allows learners to practice pronunciation by converting AI text answers into speech. The Node.js-based Backend handles the basic functionality and processes user queries, while the Gemma 8B model provides grammatical correction and tutor-style responses. LangChain manages conversation flow and context, and everything is routed locally through Ollama for faster and more private responses.

Clerk manages user accounts and authentication, while Docker ensures uniform performance across multiple platforms, which aids the system. User profiles, preferences, and progress are saved with style by SQLite on the local machine. Its modular architecture allows for the addition of new languages or capabilities without interfering with existing ones. Furthermore, the system is secured by running artificial intelligence locally and using Docker. is adaptable and prepared for the future, allowing it to grow in response to the changing demands of instructors and students.

7.1 Flowchart

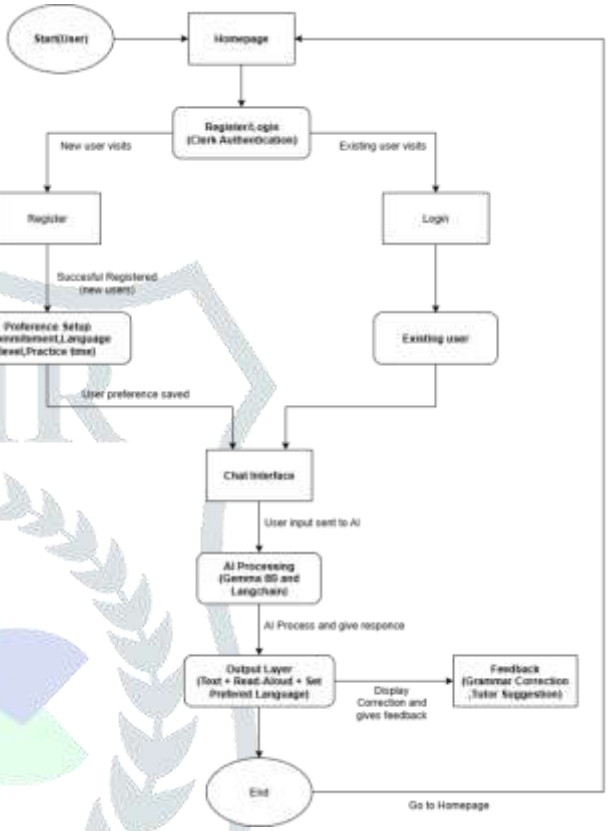

Fig. 1. Flowchart

The AI Language Coach aims to give students a tailored, seamless experience. The system finds out whether a visitor is a returning one or a new one when they interact with the site first. New users can quickly sign up and select their preferences—including language level, objectives, and amount of practice time.

Most of the user-AI tutor interaction is via the Chat Interface, where more instruction is given. Learner messages or inquiries are managed by the system (which runs locally with Ollama) using the Gemma 8B model. The AI offers constructive ideas including vocabulary alternatives, better sentence structure, and grammatical corrections. It also includes a Read Aloud option that lets readers listen to the AI's replies to practice their pronunciation. With this interactive loop, pupils may practice, get comments, and develop at every stage.

The system also offers customized learning paths.

It highlights areas that need more attention, recommends fresh subjects, and adjusts the difficulty depending on each student's progress. The AI Language Coach will include more features in the future, like sophisticated artificial intelligence capable of recognizing feelings or providing context sensitive answers. This will guarantee a dynamic and interesting learning environment as the platform evolves and grows with its users.

8.0 Results and Discussion

In this study, we created the AI Language Coach, a platform that helps learners practice languages in a personalized way, using the Gemma 8B model. The system adjusts to each learner's progress, provides instant feedback, and offers chat-based exercises for sentence formation, vocabulary, and grammar. We also added a read-aloud feature using the Web Speech API, so users can improve their pronunciation and listening skills.

The results show that the system works well, providing quick and accurate responses thanks to the integration of Ollama for hosting. Learners get real-time corrections and can practice conversations interactively, making it more engaging. Overall, the AI Language Coach has proven to be an effective, flexible, and user-friendly tool for learning languages.

9.0 Conclusion

In this project, we created the AI Language Coach, a platform that makes learning languages more personalized and fun. Using the Gemma 8B model, the system provides real-time feedback, lets learners practice through conversations, and even reads aloud to hear and help with pronunciation. This makes learning feel more interactive, helping users improve at their own pace while keeping them engaged.

So far, the AI Language Coach has been working great, offering fast and accurate responses with the help of Ollama for hosting. It's easy to use and really helps learners get better at their language skills. Looking ahead, we plan to add cool new features like certification for learners, and speech recognition that will give feedback on pronunciation and help improve it. We'll also let the system teach via speech, making the platform even more engaging and interactive for learners everywhere.

10.0 References

1. Guntamukkala Gopi Krishna et al. [1] (2023) Multilingual NLP. International Journal of Advanced Engineering and Nano Technology (IJAENT) Volume-10 Issue-6.
2. Viet Dac Lai et al. [2] (2023) ChatGPT Beyond English: Towards a Comprehensive Evaluation of Large Language Models in Multilingual Learning. Adobe Research. Computation and Languages. arXiv:2406.11409v2.
3. Libo Qin et al. [3] (2025) A survey of multilingual large language models. Patterns: A Cell Press jounal.Volume 6 Issue 1.
4. Sachin Goyal et al. [4] (2024) Advancements in Natural Language Processing: Leveraging Transformer Models for Multilingual Text Generation. Pacific Journal of Advanced Engineering Innovations.
5. Kalin Kopanav et al. [5] (2024) Comparative Performance of Advanced NLP Models and LLMs in Multilingual Geo-Entity Detection. Proceedings of the Cognitive Models and Artificial Intelligence Conference.
6. Gaurav Kashyap et al. [6] (2024) Multilingual NLP: Techniques for Creating Models that Understand and Generate Multiple Languages with Minimal Resources. International Journal of Scientific Research in Engineering and Management
7. Mohammed Mohsen et al. [7] (2024) Artificial Intelligence in Academic Translation: A Comparative Study of Large Language Models and Google Translate. Article. Psycholinguistics Vol 35 No 2.
8. Matt Zandstra et al. [8] (2020) Docker. Technical book. Docker containerization.
9. Thomas Mesnard et al. [10] (2024) Gemma: Open Models Based on Gemini Research and Technology. Google DeepMind. Open model series.
10. Morgane Riviere et al. [11] (2024) Gemma 2: Improving Open Language Models at a Practical Size. Google DeepMind. Open model series.
11. Aishwarya Kamath et al. [12] (2025) Gemma 3 Technical Report. Google DeepMind. Open model series.
12. Nam Nguyen et al. [13] (2024) CodeGemma: Open Code Models Based on Gemma.

13. Francisco Marcondes et al. [15] (2025) Using Ollama. Natural Language Analytics with Generative Large-Language Models**.** Natural Language Analytics with Generative Large-Language Models**.**
14. Hause Lin et al. [16] (2025) ollamar: An R package for running large language models**.** The Journal of Open-Source Software.
15. Hangseo Choi et al. [17] (2025) Domain-Specific Manufacturing Analytics Framework: An Integrated Architecture with Retrieval-Augmented Generation and Ollama-Based Models for Manufacturing Execution Systems Environments. Application of Artificial Intelligence in Industrial Process Modelling and Optimization
16. Javraisinh Gohil et al. [18] (2025) Developing a User-Friendly Conversational AI Assistant for University Using Ollama and LLama3. Data Science, Agents & Artificial Intelligence (ICDSAAI), International Conference.
17. Salini Suresh et al. [19] (2025) Emotional Intelligence in Chatbots: A Study on Enhancing User Experience with Llama3 and Ollama.
Data Science, Agents & Artificial Intelligence (ICDSAAI), International Conference.