AI Assited Tool For Child Language Learning

1st Disha Gayathri Umashankar
Applied Data Science and Analytics
SRH Univeristy Of Applied Sciences
Heidelberg, Germany

2nd Sandhya Tripathi
Applied Data Science and Analytics
SRH Univeristy Of Applied Sciences
Heidelberg, Germany

3rd Shruti Pardeshi
Applied Data Science and Analytics
SRH University of Applied Sciences
Heidelberg, Germany

***Abstract*—Traditional language learning methodologies, which are heavily based on textbooks, rote memorization, and instructor-led training, often lack the engagement and adaptability required for effective acquisition of second language in children. With advancements in Artificial Intelligence (AI), learning paradigms have shifted toward personalized, interactive, and adaptive frameworks that enhance learner engagement and comprehension. AI-driven solutions, particularly those utilizing Large Language Models (LLMs) and speech processing technologies, have demonstrated significant potential in facilitating dynamic and responsive language-learning experiences.**

This study introduces an AI-assisted language learning tool designed specifically for non-native English-speaking children aged 6 to 12. The proposed system employs a voicebot (Lingola) powered by LLMs and Google Speech-to-Text APIs to enable real-time conversational interaction. The tool facilitates context-aware dialogue generation, adaptive vocabulary reinforcement, and personalized feedback mechanisms, thereby promoting fluency, pronunciation accuracy, and language comprehension. Unlike traditional learning approaches, this system dynamically adjusts to the learner's proficiency level and provides an immersive learning environment that encourages active participation.

Experimental evaluation and qualitative analysis indicate that AI-driven voicebots can significantly improve engagement, retention, and spoken language confidence in children. The findings highlight the effectiveness of LLM-based conversational AI in second-language learning applications and demonstrate its potential to bridge the gap between structured language instruction and real-world language use.

By integrating LLMs with Google's speech processing capabilities, this study contributes to the ongoing advancements in AI-assisted educational technologies. Future work will focus on enhancing contextual understanding, refining speech interaction accuracy, and expanding multilingual support to further optimize personalized language learning experiences.

I. Introduction

Background and Motivation : The Role of Artificial Intelligence in Language Learning A crucial component of cognitive development, especially in early life, is language learning. Classroom instruction, rote memorization, and textbook-based activities are examples of traditional language learning techniques that frequently fall short of offering individualized and interesting learning opportunities. Innovative technologies to improve language acquisition have been made possible by the introduction of artificial intelligence (AI) into education, which has created new opportunities for interactive and adaptive learning. By enabling real-time engagement and tailored feedback, voicebots—AI-driven conversational agents—have in particular shown promise as a means of assisting second-language acquisition.

Challenges of Traditional Language Learning Approaches: Conventional language learning methodologies are often constrained by several limitations. Classroom-based instruction is inherently non-adaptive, failing to cater to the diverse learning speeds, proficiency levels, and engagement needs of individual learners. Furthermore, delayed feedback mechanisms in traditional educational settings hinder the development of real-time conversational skills, which are essential for language fluency. Additionally, passive learning strategies, such as text-based exercises and isolated vocabulary drills, may not effectively simulate naturalistic language use, limiting learners' ability to apply acquired knowledge in real-world communication scenarios.

Traditional approaches to language learning are frequently limited by a number of issues. Because it cannot accommodate the various learning styles, skill levels, and engagement requirements of individual students, classroom-based training is fundamentally non-adaptive. Furthermore, the development of real-time conversational skills—which are crucial for language fluency—is hampered by typical educational environments' delayed feedback processes. Furthermore, passive learning techniques like text-based exercises and isolated vocabulary drills might not accurately replicate naturalistic language use, which would restrict students' capacity to employ what they have learned in authentic communication situations.

Rationale for a Voicebot-Based Language Learning Approach: An innovative and successful strategy for raising kids' interest and language skills is the use of LLM-powered voicebots in language instruction. By facilitating genuine dialogues, offering real-time corrections, and adjusting to the learner's competency level, voicebots with LLMs can close the gap between conventional teaching techniques and contemporary AI-driven education. LLMs facilitate free-form, interactive discussions, which promote active learning, impromptu language output, and contextual comprehension, in contrast to pre-scripted AI models.

II. Objectives of the Study

This research aims to investigate the effectiveness of LLM-based Voicebots in enhancing children's language learning experiences, with a specific focus on:

1. Engagement metrics - Evaluating how interactive, AI-powered dialogues influence behavior, cognitive, and emotional engagement.

2.Fluency development - Evaluating how well LLM-driven interactions can enhance contextual usage, phrase form, and pronunciation.

3.Personalization and adaptability Examining how feedback systems driven by LLM might modify learning activities according to each learner's level of competency.

4.Gamification and Play-Based Learning: Investigating the impact of integrated game-based elements on motivation and educational outcomes.

By examining these aspects, this study seeks to contribute to the growing field of AI-assisted education, providing insights into the potential of voicebot-driven language learning frameworks as an alternative to traditional approaches.

III. Problem Statement

A. Deficiency of Interactive and Adaptive Learning Solutions

Conventional language learning approaches primarily depend on fixed and non-adaptive instructional techniques, such as textbooks, scripted exercises, and passive memorization activities. These traditional methods often fail to adapt to the learner's unique needs, leading to diminished engagement and suboptimal language acquisition.The lack of interactive and responsive tools restricts learners' ability to develop conversational proficiency, as most traditional systems do not provide real-time adjustments or contextual feedback.Additionally, these methods do not replicate natural speech interactions, which are crucial for improving fluency, pronunciation, and contextual understanding.

B. Requirement for Personalized and Engaging Educational Technologies

Tailored learning experiences play a vital role in maintaining engagement and improving learning outcomes. However, many existing digital and classroom language learning platforms employ uniform instructional models, overlooking individual differences in proficiency, learning pace, and specific areas that need improvement. Research highlights the benefits of adaptive and interactive learning environments, particularly those that incorporate conversational AI and dynamic feedback, in boosting motivation, engagement, and language retention. Despite this, there remains a gap in AI-powered, voice-based educational tools designed to deliver personalized, real-time language learning experiences.

To address these limitations, this study focuses on the development and evaluation of a voicebot powered by Large Language Models (LLMs). By utilizing LLM-driven contextual awareness and conversational adaptability, this approach aims to enhance interactivity, provide personalized feedback, and offer an engaging language learning experience, bridging the gap between traditional methodologies and AI-driven educational innovations.

C. Objectives and Contributions

Objectives The primary goal of this study is to design and implement an AI-driven, voice-interactive tool that enhances language learning for children through an engaging and adaptive conversational experience. The specific objectives of this research include:

- Development of an AI-Enabled Voice Interaction System – This study aims to create an intelligent voicebot (Lingola) capable of facilitating natural conversations to support language learning in children. Utilizing Large Language Models (LLMs), the system will be designed to comprehend user inputs, generate meaningful dialogue, and adjust responses based on the learner's proficiency level.
- Implementation of Speech-to-Text and Text-to-Speech Features – To enable seamless interaction, the system will integrate speech-to-text (STT) and text-to-speech (TTS) functionalities, allowing children to communicate verbally with the AI. These features will help analyze pronunciation, assess fluency, and provide instant feedback, ensuring an interactive and effective learning process.
- Creation of a Child-Centric User Interface (UI) with Adaptive Responses – Since young learners require an intuitive and engaging platform, this study will develop a user-friendly interface tailored for children, incorporating visual enhancements, gamification elements, and interactive features. Additionally, the AI will generate personalized responses based on the learner's progress, individual needs, and skill level, offering a customized and immersive learning experience.

Contributions

This research contributes to AI-assisted language education in several key ways:

- Introducing an LLM-driven voicebot specifically designed to create an interactive and adaptive second-language learning environment.
- Enhancing real-time engagement in language learning by integrating speech-to-text and text-to-speech capabilities, allowing children to practice spoken communication effectively.
- Providing a personalized and visually engaging learning platform, incorporating child-friendly UI components and interactive elements to cater to learners at different proficiency levels.
- Bridging the gap between traditional language learning methods and AI-driven conversational learning models, demonstrating the potential of voice-interactive AI in improving engagement, fluency, and pronunciation among children.

IV. Literature Review

AI-powered language-learning programs have drawn a lot of interest lately due to their capacity to deliver individualised and captivating educational experiences. Numerous research have looked into various facets of AI-assisted language acquisition, especially as it relates to young students. This section examines pertinent research on automatic speech recognition (ASR) for non-native child speech, AI language assistants for early language learners, game-based learning approaches, generative AI in child language acquisition, and intelligent tutoring systems.

A. Automatic Speech Recognition of Non-Native Child Speech for Language Learning Applications

Researchers faced a tricky challenge—how to make AI understand the unpredictable pronunciation of non-native children. With limited training data and varying speech patterns, traditional ASR systems struggled. But by improving phonetic modeling and adaptive learning, AI tutors started recognizing young learners more accurately. This meant better feedback, personalized learning, and a confidence boost for kids speaking a new language.

B. Exploring AI Language Assistants with Primary EFL Students

In classrooms, AI-powered language assistants became the new learning companions. Unlike textbooks, these AI tutors engaged students in real conversations, adjusted to their pace, and provided instant corrections. Teachers noticed that students felt more motivated and less afraid to make mistakes, making language learning more natural and interactive.

C. Game-Based Learning in Early Childhood Education: A Systematic Review and Meta-Analysis

Learning a language shouldn't feel like a chore—it should feel like play. Researchers found that gamification boosts vocabulary, pronunciation, and retention. AI-powered games transformed lessons into exciting challenges, where kids earned rewards, solved puzzles, and unknowingly absorbed language skills in the process.

D. Generative AI Pioneers the Future of Child Language Learning

Imagine a tutor who doesn't just teach but tells stories, holds conversations, and adapts on the fly. That's what generative AI brought to language learning. By creating personalized dialogues and immersive experiences, AI helped children practice real-world conversations in a fun and meaningful way, making language learning feel like second nature.

E. Intelligent Tutoring Systems for Language Learning

Every child learns differently, and AI-powered Intelligent Tutoring Systems (ITS) understood that. By analyzing strengths and weaknesses, these systems created personalized learning paths, adjusting difficulty levels and offering targeted feedback. The result? Self-paced learning that made mastering a new language more effective and enjoyable.

F. The Evolution of Language Learning: Exploring AI's Impact on Teaching English as a Second Language

The impact of AI on English as a Second Language (ESL) learning was undeniable. From personalized coaching and automated feedback to immersive experiences, AI transformed traditional teaching methods. It made language learning more accessible, adaptive, and engaging, catering to learners of all backgrounds and abilities.

V. Model Comparison and Evaluation

Both BLEU score and semantic similarity metrics are chosen because they capture different yet complementary aspects of text quality:

A. BLEU (Bilingual Evaluation Understudy) Score:

Lexical Accuracy: Measures ngram overlap between generated text and reference text, highlighting how closely the word choices and order match expected outputs.

Established Metric: Widely used in machine translation and text generation tasks, offering a standardized way to compare models.

B. Semantic Similarity:

Meaning Preservation: Evaluates whether the generated text conveys the same meaning as the reference, even if the wording differs.

Flexibility: Captures cases where paraphrasing or alternative phrasing is acceptable, ensuring that models like Llama2, Llama3, and DeepSeek are not penalized for valid linguistic variation.

By combining these metrics, we get more comprehensive evaluation of a model's performance. While BLEU focuses on surface-level similarity, semantic similarity ensures that the underlying message is intact. This dual approach is particularly useful for advanced language models where both precise wording and overall meaning are important. This section evaluates LLaMA2, LLaMA3, and DeepSeek on a children's audio dataset using BLEU and semantic similarity metrics. LLaMA3, which performed best, was selected for final training

C. Model Evaluation

1) Dataset and Metrics: The dataset consists of transcribed children's audio files. Models were evaluated using:

BLEU Score (Bilingual Evaluation Understudy): Measures n-gram precision.

Semantic Similarity: Assesses contextual accuracy via cosine similarity.

2) Results: Below are the evaluation results for each model:

LLaMA3 outperformed LLaMA2 and DeepSeek, making it the optimal choice.

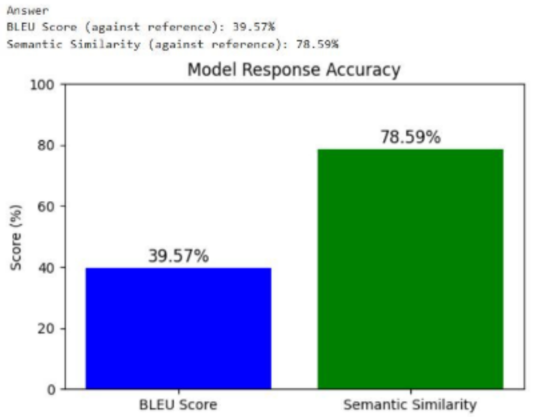


Fig. 1. BLEU Score and semantic similarity for LLaMA2.

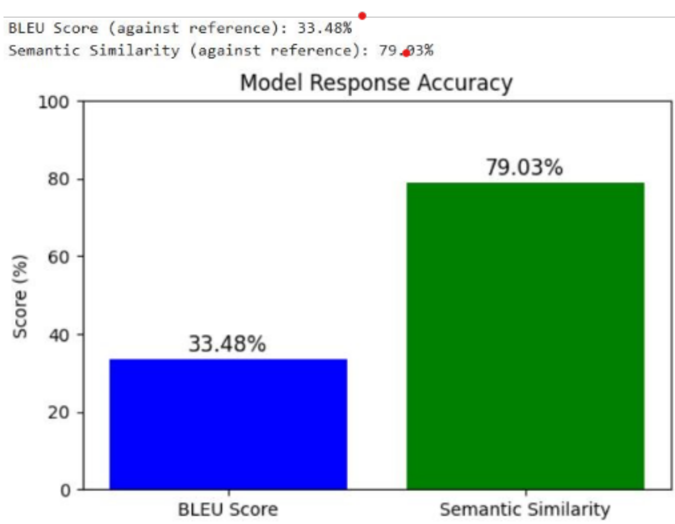


Fig. 2. BLEU Score and semantic similarity for Deepseek.

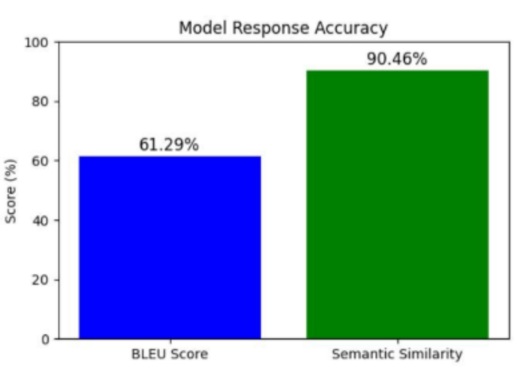


Fig. 3. BLEU Score and semantic similarity for LLaMA3.

VI. Model Training

The AI model, Quantised LLaMA 3.1 is used for intent recognition and response generation.

The training dataset was sourced from Kaggle, containing large number of children's audio files for speech recognition and intent classification.

The model was trained using PyTorch on a CPU-based environment.

The Google Speech-to-Text API was used to transcribe the children's audio files, aiding in speech recognition accuracy.

Evaluation Metrics such as BLEU Score and Semantic Similarity were employed to assess the performance of the trained model.

Continuous refinements are made through user feedback and dataset expansion to improve performance.

VII. System Architecture and Methodology

A. System Overview

The Lingola system is an AI-powered, voice-interactive language learning assistant tailored for children between the ages of 6 and 12. It leverages Large Language Models (LLMs) to facilitate engaging, real-time dialogue, enabling children to practice language skills through spoken communication and receive instant feedback.

The system architecture is modular and consists of speech input capture, audio processing, AI-based response generation, and voice output synthesis. The process begins when a child speaks into a microphone, and their voice is recorded in WEBM format. The recorded audio is then converted into WAV format for compatibility with further processing. The Google Speech-to-Text (STT) API transcribes the spoken input into text, which is then processed by Lingola, the system's LLM-powered AI engine. The response generated by the model is converted into speech using the Google Text-to-Speech (TTS) API, ensuring an interactive learning experience. Additionally, a database is used to store user interactions, allowing for personalized and adaptive learning.

B. System Components

1) Speech-to-Text (STT) Module: The Speech-to-Text (STT) module is responsible for transcribing spoken language into text, serving as the first step in the processing pipeline. The system captures real-time speech input, initially saving it in WEBM format before converting it to WAV format using FFmpeg for improved processing accuracy.

The Google STT API is used to transcribe the speech, employing advanced deep learning techniques to ensure high accuracy. This module offers:

Reliable transcription, even when dealing with variations in pronunciation, accents, and tone.

Optimized audio conversion, ensuring the best input quality for speech recognition.

Seamless integration with AI response processing, enabling real-time interaction.

2) AI-Based Response Generation (Lingola Model): The transcribed text is processed by Lingola, a Large Language Model (LLM)-based AI assistant, which generates a contextually appropriate response tailored to the child's input. Unlike conventional rule-based chatbots, Lingola dynamically constructs responses, making interactions engaging and adaptive to the child's learning progress.

This AI module is structured to:

- Provide context-aware responses, adapting to different levels of language proficiency.
- Support interactive learning, offering explanations, corrections, and encouragement.
- Enable a dynamic, free-flowing conversational experience, without relying on predefined scripts.

3) Text-to-Speech (TTS) Module: Once a response is generated by Lingola, it is transformed into spoken language through the Text-to-Speech (TTS) module. The Google TTS API is employed to synthesize the response, producing natural-sounding speech that is suitable for young learners.

This module ensures:

- Clear and engaging speech output, promoting better listening comprehension.
- Dynamic voice modulation, making interactions more expressive and child-friendly.
- Immediate verbal feedback, helping children improve pronunciation and fluency.

4) Database and Adaptive Learning: A MySQL database is implemented to store and track user interactions, ensuring an adaptive and personalized learning experience. The system records:

- User details, such as name and age, to tailor responses accordingly.
- Conversational history, allowing the system to adapt responses based on prior interactions.
- Topics covered, helping structure a progressive learning journey.
- Engagement metrics, such as session duration and number of conversational turns, for tracking learner progress.
- By maintaining past interaction records, the system can:
- Modify learning content based on the child's progress.
- Adjust lesson complexity, making learning more effective.
- Enhance retention and fluency, reinforcing key language concepts over time.

5) User Interface (UI) Design: The system incorporates a user-friendly interface (UI) designed specifically for young learners. It features:

- Simplified navigation, making it easy for children to use.
- Visual enhancements, such as animations and icons, to improve understanding.
- Gamification elements, including progress tracking and rewards, to sustain motivation.

C. System Workflow

The interaction process follows a structured workflow, ensuring smooth real-time communication:

- Voice Input Capture – The child speaks into the microphone, and their voice is recorded in WEBM format.
- Audio Format Conversion – The recorded file is converted into WAV format for better processing.
- Speech-to-Text Processing – The Google STT API transcribes spoken words into text.
- AI-Based Response Generation – The transcribed text is processed by the Lingola LLM, which generates an appropriate response.
- Text-to-Speech Synthesis – The response is converted into speech output using the Google TTS API.
- Real-Time Interaction – The synthesized voice is played back to the child, maintaining an engaging conversation.
- Database Logging – The system records interactions and learning progress, enabling future customization.

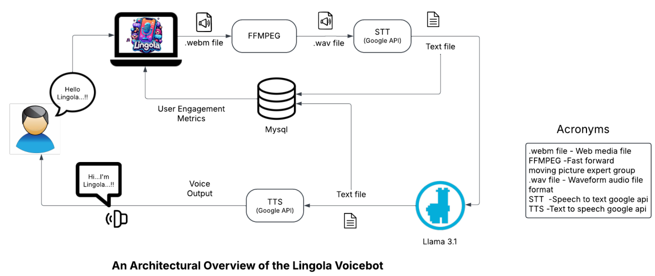


Fig. 4. An Architecture Overview of the Lingola Voicebot.

D. Summary

The Lingola system introduces an interactive AI-driven language learning tool that integrates speech recognition, AI-generated responses, and speech synthesis to create a dynamic and adaptive learning experience. By utilizing LLMs for real-time conversation generation, along with Google's STT and TTS APIs, the system offers an innovative approach to enhancing language fluency, pronunciation, and engagement in children.

Through personalized and interactive dialogue, Lingola helps young learners practice spoken language skills effectively, making it a powerful tool for early-stage language acquisition.

VIII. Implementation Details

A. Technologies Used

AI frameworks and contemporary web technologies were used in the creation of the AI-assisted language learning tool. A summary of the technologies utilised is provided below:

- Programming Language: Python
- Platform: Visual Studio Code (VS Code)
- Backend Framework: Flask

- Frontend Technologies: JavaScript, HTML, and CSS
- Database: MySQL
- AI Model: Quantised LLaMA 3.1
- APIs: Google Speech-to-Text and Google Text-to-Speech
- Libraries:
- Torch: A deep learning framework used for AI model implementation.
- SacreBLEU: A tool for evaluating machine translation performance.
- NumPy: A library for numerical computing and handling data arrays.

Matplotlib: A visualization library for generating graphs and analysis reports.

- Sentence Transformers: A framework for sentence embeddings and semantic similarity calculations.
- Util: A module from Sentence Transformers used for utility functions.

B. Frontend Development

- HyperText Markup Language (HTML): used for defining features like buttons, forms, and layout and to organise web pages.
- Cascading Style Sheets (CSS): In charge of enhancing the user experience by designing the web pages using colour, fonts, and layouts.
- JavaScript (JS): Enables real-time responses to user input by enabling dynamic content and interactivity on web pages.

The frontend is organized into two primary sections. The first section is designed for children, featuring a dedicated login portal, interactive voicebot (Lingola) engagement, educational games, and a display of game results.

The second section is tailored for teachers, providing a login interface and detailed reports on children's engagement with the voicebot and their overall learning outcomes.

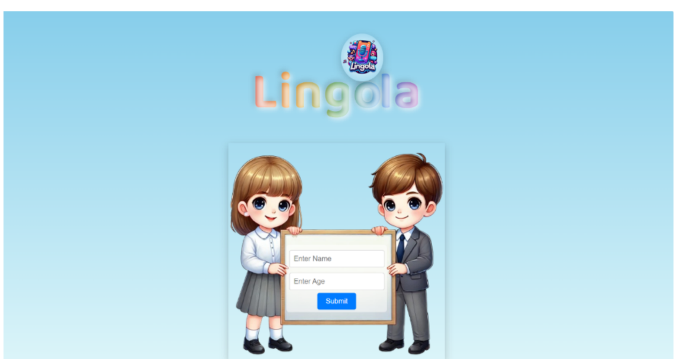


Fig. 5. Login Page: Allows users to authenticate and access the system.

C. Backend Development

Flask: A lightweight Python web framework that makes it easier to manage databases, integrate APIs, and run AI models. It facilitates server-side functions and effectively handles user requests.

The backend is responsible for:

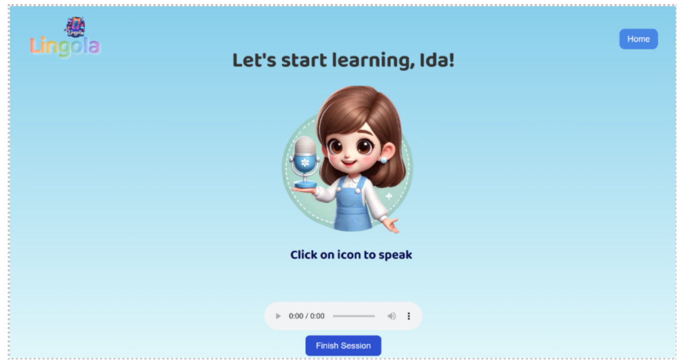


Fig. 6. Voicebot Page: The main interactive interface where users engage with the AI-driven language learning assistant.

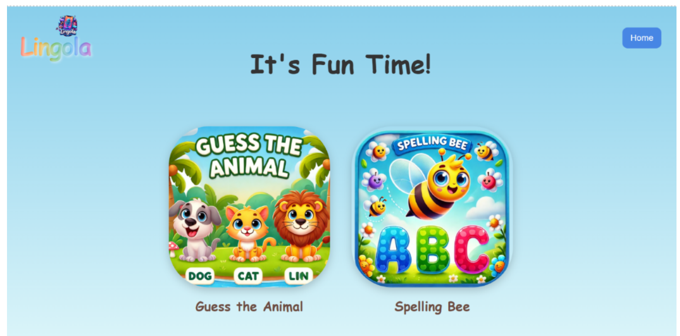


Fig. 7. Game Page: Following the knowledge-based interaction between the child and Lingola, a brief game is presented to assess the child's performance.

- User authentication and session management.
- Processing user input and sending it to the AI model for intent recognition and response generation.
- Integration with Google Speech-to-Text API for recognizing spoken language.
- Integration with Google Text-to-Speech API to provide verbal feedback to users.
- Storing and retrieving conversations from the database.

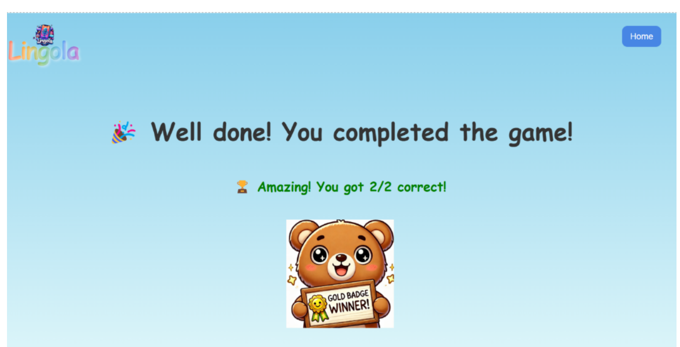


Fig. 8. Reward Page:Upon completing the game, the child is awarded either a Gold or Silver badge based on their performance assessment.

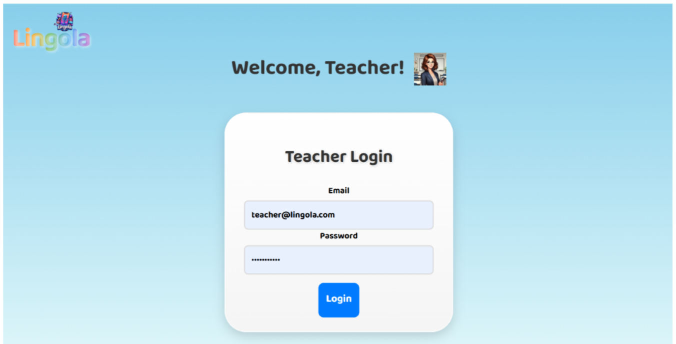
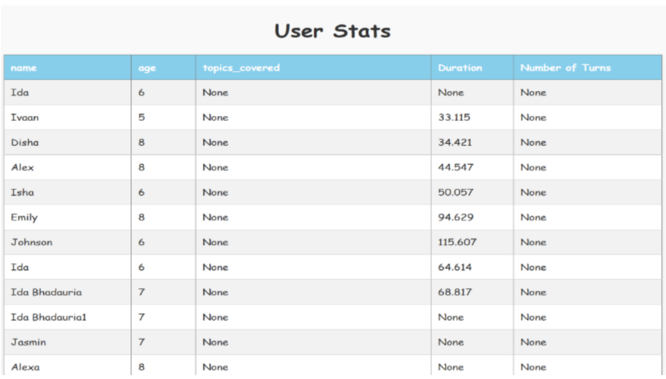


Fig. 9. Teacher Login Page: A separate login page has been given to the Teacher to monitor the performance of the children interacting with the Lingola Voicebot.



User Stats

name	age	topics_covered	Duration	Number of Turns
Ida	6	None	None	None
Ivaan	5	None	33.115	None
Disha	8	None	34.421	None
Alex	8	None	44.547	None
Isha	6	None	50.057	None
Emily	8	None	94.629	None
Johnson	6	None	115.607	None
Ida	6	None	64.614	None
Ida Bhadauria	7	None	68.817	None
Ida Bhadauria1	7	None	None	None
Jasmin	7	None	None	None
Alexa	8	None	None	None

Fig. 10. Results/Statistics Page: This page provides a comprehensive overview of the session, displaying key conversation metrics such as duration, number of turns, timestamps, and topics covered, along with the game outcomes.

D. Database Design

MySQL: An effective way to store structured data is with a relational database management system. It enables the easy execution of queries while keeping user credentials and conversation log data.

The database consists of two tables:

- User Table: Stores user credentials like Id, user name, age and session data like duration, number of turns, start-end timestamps and topics covered of individual user.

- Conversation Table: Maintains a record of user interactions with the voicebot including queries and model responses from individual users.

- Prompts Testing: Lingola was rigorously evaluated using a diverse set of interactive prompts to ensure its effectiveness across multiple learning domains. The tests covered:

- Alphabet Instruction: Delivering engaging activities to help children recognize and recite letters.
- Numerical Concepts: Teaching numbers and basic counting skills through interactive exercises.
- Vocabulary Building: Introducing and reinforcing new words to expand the child's lexicon.
- Animal Identification: Familiarizing children with animal names through contextual conversation.
- Topics Covered: Tracks the subjects the child has engaged with during interactions, providing insights into the specific areas of learning.
- Age-Appropriate Content: Curate and deliver learning materials specifically designed for the child's developmental stage, ensuring that all content remains strictly appropriate by excluding any sexual or obscene references.
- Error Correction: Offer prompt, constructive feedback when mistakes occur during interactions, and motivate the child to improve through positive reinforcement and guidance.

These comprehensive tests confirmed that Lingola can adapt its responses to support effective, context-sensitive learning while actively correcting errors and reinforcing key concepts.

- Engagement Metrics Calculation:

- Duration (D):
  Calculate by subtracting the session's start date time from the session's end date time.
  For example, if a session starts at 11:45:56 and ends at 12:15:56, the Duration is 30 minutes (or 30.0 in decimal hours).
- Number of Turns (T):
  Count each interactive exchange between the child and the voicebot.
  For instance, if the child speaks once and the bot replies once, that counts as two turns in total.
- Engagement Score (E):

$$E = w_1 \times D + w_2 \times T + w_3 \times GE \tag{1}$$

where:

- D is the calculated Duration,
- T is the total number of turns, and
- GE is the game result (e.g., 2 for Gold, 1 for Silver).

The weights w_1, w_2, and w_3 determine the relative importance of each metric. For instance, if conversational depth (T) is more critical than time spent (D), we can set $w_2 > w_1$.

IX. Model Challenges

The development of Lingola, an AI-assisted language learning tool, presented several challenges related to model selection, computational efficiency, and system integration. These limitations are discussed below:

A. Model Selection and Evaluation

Selecting an appropriate Large Language Model (LLM) was a crucial challenge. Multiple models were evaluated, and a comparative analysis based on key performance metrics was conducted. The results indicated that LLaMA 3.1 was the most suitable model for real-time conversational interactions within Lingola, offering a balance between response accuracy and computational efficiency.

B. Computational Resource Constraints

LLaMA 3.1 requires high computational power, particularly GPU acceleration, for efficient real-time response generation. Running the model solely on a CPU-based system led to increased processing time and latency, impacting the system's ability to deliver smooth, interactive conversations.

C. Implementation of a Quantized Model

To mitigate high computational costs, a quantized version of LLaMA 3.1 was employed, reducing the model size and improving processing speed. However, quantization resulted in minor variations in response generation, as the reduced model had lower precision compared to the full version. This occasionally affected the accuracy and contextual relevance of generated responses. We also tried with google collab and college server but due to not having pro option we are unable to use it,

D. Integration with Google Speech-to-Text API

Incorporating Google's Speech-to-Text API into the system posed challenges in ensuring seamless interaction between speech input and LLM-driven response generation. Synchronizing speech recognition outputs with the AI model's response pipeline required additional tuning and optimization to maintain accuracy and real-time responsiveness.

E. Audio Format Compatibility

The system processes spoken input from users, which is initially captured in WebM format. However, Google's Speech API supports only WAV format, requiring an additional step of format conversion. Ensuring efficient and lossless conversion was a technical challenge that needed to be addressed to maintain audio quality and minimize processing delays.

X. Technical Challenges

The deployment of Lingola presented several technical challenges, particularly concerning computational efficiency, real-time processing, and system integration. The key technical difficulties encountered during implementation are as follows:

A. High Computational Requirements for Model Processing

Large Language Models (LLMs) such as LLaMA 3.1 require significant computational power, particularly GPUs, to efficiently process real-time responses. Running the full model on a CPU-only system resulted in slower response generation, affecting the interactivity and engagement of the voice-based learning experience. Ensuring low-latency, real-time interactions remains a key technical challenge.

XI. Furure Scope

1.Enhancing Speech Recognition for Children The Google Speech-to-Text API occasionally struggles with child-specific speech patterns, mispronunciations, and accents.

Future enhancements can involve fine-tuning speech recognition models specifically for children's speech variations to improve recognition accuracy and response quality.

2. Expanding Multilingual Support

Currently, Lingola focuses on English learning, but future versions can extend support for multiple languages such as French, Spanish, Mandarin, and German.

Expanding multilingual capabilities will make the tool accessible to a wider audience, including children from different linguistic backgrounds.

3. Optimizing Model Performance and Reducing Latency

The current deployment requires high computational resources, leading to latency in real-time response generation, especially when running on CPU-based systems.

Future improvements will focus on optimizing LLaMA 3.1, exploring lighter AI models, and utilizing cloud-based inference to reduce processing time and improve system responsiveness.

4. Personalization and Adaptive Learning

Future versions of Lingola can include adaptive learning algorithms that analyze user progress and adjust the difficulty level, vocabulary suggestions, and feedback accordingly.

Implementing a recommendation system will allow children to receive customized exercises and interactive lessons based on their learning patterns and areas of improvement.

5. Improved Conversational AI Capabilities

Currently, LLMs generate responses based on input text, but context retention and memory-based learning can be implemented to make the conversation more interactive and consistent.

Enhancing the AI's ability to recognize emotions and provide engaging responses can improve user satisfaction and learning motivation.

6. Mobile and Offline Support

Future iterations of Lingola could include a mobile application, making it more accessible to children who do not have access to desktops or laptops.

Offline functionality could also be developed, allowing children to access pre-trained AI responses, exercises, and interactive lessons even in low-connectivity environments.

7. Gamification for Better Engagement

Incorporating gamification features such as leaderboards, rewards, badges, and achievement tracking could increase motivation and engagement among children.

AI-driven storytelling, role-playing exercises, and interactive challenges can further improve user involvement, making learning more enjoyable.

8. Integration with Educational Platforms and Schools

Lingola can be integrated into educational platforms and school curriculums to provide structured language learning assistance.

Collaborating with educators and language experts can help ensure alignment with standardized learning methodologies, making the tool suitable for formal education settings.

9. User Feedback and Continuous Improvement

User feedback plays a crucial role in improving the effectiveness of Lingola. Future updates will incorporate insights from children, parents, and educators to refine AI-generated responses, UI design, and learning modules.

Regular feedback collection through in-app surveys, performance assessments, and interactive testing will help identify areas of improvement and feature enhancements.

XII. RESULTS AND FINDINGS:

A. Summary of Findings

The evaluation of Lingola demonstrated its effectiveness in enhancing second-language acquisition for non-native English-speaking children aged 6 to 12. By integrating LLaMA 3.1 and Google Speech-to-Text API, the tool enabled real-time conversational interactions, improving pronunciation, vocabulary retention, and engagement. The system's adaptive learning mechanism utilized user data to personalize learning experiences, track progress, and adjust responses based on individual learning patterns.

B. Insights from User and Conversation Data

The database architecture, consisting of Users and Conversation tables, provided valuable insights into learning behavior, engagement levels, and AI-human interaction patterns.

1) User Table Analysis::

- The Users table recorded key metrics such as session duration, number of conversational turns, and topics covered during each learning session.
- Data analysis revealed that children who engaged in longer sessions with more conversational turns exhibited better fluency and vocabulary retention over time.
- The topics covered field allowed tracking of individual progress, helping to identify areas where additional reinforcement was needed.

2) Conversation Table Analysis: :

- The Conversation table stored the entire dialogue history between the child and the AI model, enabling analysis of interaction patterns, learning difficulties, and conversational fluency.
- By examining frequent errors and response trends, the system could identify common learning challenges and adjust AI-generated responses to better assist the child.
- Longitudinal data analysis of conversations allowed insights into how children adapted to AI-driven language learning over multiple sessions.
- We prove that higher engagement between the child and the model correlates with an accelerated learning rate and improved overall performance.

XIII. CONCLUSION

A. Summary

This study offers a thorough AI-assisted approach to learning language for kids ages 6 to 12 that makes use of audio processing tools and Large Language Models (LLMs). Through the integration of Google Speech-to-Text (STT), Text-to-Speech (TTS), and voicebot-driven conversational AI, the system facilitates personalized, adaptive, and real-time learning experiences. The Lingola platform effectively overcomes the drawbacks of conventional language learning techniques by improving fluency, pronunciation, and engagement.

B. Impact

Young learners are given individualized and interesting language-learning experiences through the use of Lingola's interactive AI-powered technology. While the adaptive database-driven method adapts learning to each learner's progress, the quantized LLaMA 3.1 model guarantees effective real-time dialogue creation. This technology promotes natural language acquisition through conversational AI, reducing reliance on scripted lectures and passive learning techniques.

C. Final Remarks

The successful application of AI-driven language learning in this study highlights the potential for expanding its capabilities to multilingual support, improved contextual understanding, and gamification features. Future enhancements in speech recognition for children's speech variations, response optimization, and mobile accessibility can further improve user experience. This work demonstrates AI's growing role in education technology, paving the way for smarter, more interactive, and personalized learning systems.

ACKNOWLEDGMENT

We sincerely thank Professor. Sina Mehraeen for his invaluable guidance and support throughout this project. His expertise and insights were instrumental in refining both the technical and functional aspects, significantly shaping the quality and direction of our work.

REFERENCES

[1] J. Underwood, "Exploring AI language assistants with primary EFL students," in CALL in a Climate of Change: Adapting to Turbulent Global Conditions – Short Papers from EUROCALL 2017, K. Borthwick, L. Bradley, and S. Thouësny, Eds. Research-publishing.net, 2017, pp. 317–321.

[2] Simone Wills, Yu Bai, Cristian Tejedor García, Catia Cucchiarini, and Helmer Strik,in Proceedings of the 12th Symposium on Languages, Applications and Technologies (SLATE 2023), Alberto Simões, Mario Marcelo Berón, and Filipe Portela, Eds. Licensed under Creative Commons License CC-BY 4.0, Article No. 11, pp. 11:1–11:8, 2023.

[3] M. S. Alotaibi, "Game-based learning in early childhood education: A systematic review and meta-analysis," Frontiers in Psychology, vol. 15, Art. no. 1307881, 2024. doi: 10.3389/fpsyg.2024.1307881.

[4] Pohang University of Science and Technology (POSTECH), "Generative AI pioneers the future of child language learning," ScienceDaily, Jul. 26, 2024.

[5] L. Konyrova, "The evolution of language learning: Exploring AI's impact on teaching English as a second language," Eurasian Science Review, vol. 2, no. 4, ISSN 3006-1164, 2024.

[6]https://github.com/dishagu/case_study_child_language_learning_2024.git