IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AI TUTOR IN A 3D ENVIRONMENT LIKE METAVERSE

[1]R Arun, [2]S V Akash, [3]P Augustin, [4] M Charan, [5]K Vanitha Sri
[1]Student, [2]Student, [3]Student , [4]Student, [5]Asst Professor
[1]CSE,
[1]Aalim Muhammed Salegh College of Engineering, Chennai, India

Abstract: AI Tutor is a virtual 3D intelligent tutoring system developed using Three.js, designed to revolutionize the way students interact with digital learning platforms. This system features a lifelike male or female avatar that communicates with users through real-time voice input and provides immediate, personalized feedback for grammar correction and language improvement. The tutor is capable of expressing human-like emotions through visual elements such as blinking eyes, moving lips, eyebrow gestures, and animated hair, which enhances the user's sense of engagement and realism. The primary goal of this project is to offer an immersive, interactive, and responsive learning experience, especially for students who struggle with grammar and communication skills.

***Index Terms* -** .AI Tutor ,3D Avatar, Speech Recognition, Natural Language Processing immersive learning experience, grammar correction, Real-Time interaction, AI Learning Assistant, Avatar's emotion rendering, Three.js powers the 3D visuals, while React.js enables a dynamic interface. Powered by Gemini API Key

I. INTRODUCTION

In the modern world, English communication has become an essential skill across various domains, including education, business, and technology. Despite the widespread availability of online learning platforms and language apps, many learners—especially non-native speakers—still face challenges in improving their spoken English fluency, pronunciation, and grammar in a personalized and interactive way. Traditional classroom- based methods often lack individual attention, while existing mobile applications may not offer real-time conversational feedback or personalized correction. With the growing adoption of Artificial Intelligence (AI), Natural Language Processing (NLP), and virtual simulation technologies, intelligent language tutoring systems are emerging as powerful tools for language acquisition. This project introduces AI Tutor 3D, an innovative web-based virtual assistant that combines Google's Gemini API for language understanding and real-time grammar correction, with a 3D animated avatar to deliver an immersive learning experience.

II. RELATED WORKS

Various intelligent tutoring systems have been developed in recent years to aid language learning, ranging from rule based applications to AI-driven conversational platforms. Early systems focused on predefined question-answer modules and lacked flexibility in natural communication. Recent advances in AI have enabled the use of large language models (LLMs) such as OpenAI's GPT and Google's Gemini, which significantly enhance real-time grammar correction and conversational depth. Projects like Duolingo use gamification, but lack real-time speaking interaction. Similarly, other platforms provide only text based assistance or delayed feedback. Moreover, most systems fail to incorporate a human-like interface that engages learners visually and emotionally. In contrast, this project integrates Gemini API for intelligent conversation with a 3D avatar, creating a unique combination of real time speech processing, NLP, and visual interaction, which together enhance the user's learning experience in a more immersive andinteractive manner. Numerous studies have explored the use of Artificial Intelligence (AI) in language education. Early intelligent tutoring systems (ITS) primarily relied on rule-based or decision-tree logic, which lacked the flexibility and adaptability required for personalized language learning. These systems were limited in their ability to understand user input in natural language, making the learning experience rigid and less engaging. However, the integration of Natural Language Processing (NLP) has significantly improved language-based learning systems, allowing for more dynamic interactions between users and machines.

In recent years, large language models (LLMs) such as GPT 3, ChatGPT, and Google's Gemini have revolutionized AI driven education tools. These models provide advanced capabilities in understanding, correcting, and generating human-like responses in real time. Applications such as Grammarly offer powerful grammar correction, while platforms like ChatGPT provide conversational learning experiences. However, most of these tools are either text based or lack a structured approach for oral communication training, which is essential for improving spoken English. Recent research in human-computer interaction highlights the potential of using virtual avatars for educational purposes. Studies show that learners are more engaged when educational content is delivered through animated characters or 3D models. A few experimental platforms have combined speech recognition with 3D avatars, but they lack the sophistication of real-time AI processing for conversation and correction. The proposed AI Tutor 3D project builds upon this foundation by integrating real-time speech recognition, Gemini API-powered NLP capabilities, and an interactive 3D avatar created in Blender. This system provides a more immersive and personalized environment for language learners, addressing the limitations of previous works and offering a comprehensive solution for spoken English practice and improvement.

III. SYSTEM ARCHITECTURE

The proposed system integrates a 3D virtual environment with AI-powered language tutoring. The system architecture is divided into the following major components:

A. Virtual Environment Module Developed using Three.js and integrated within a web interface, this module forms the core immersive learning environment. Users interact with a 3D virtual tutor rendered as a humanoid figure. The environment mimics a digital classroom with avatars and voice interaction.

B. Speech-to-Text Engine This module captures user voice input through a microphone and converts it into text using developed as a modern web based application combining OpenAI's Whisper API.It supports multiple languages and is capable of recognizing accents and noisy backgrounds.

C. Natural Language Processing (NLP) The core of the system uses Gemini 1.5 Flash model via GenAI API to understand, correct, and respond to the user's input. It checks for grammatical errors, provides corrections, and generates human-like feedback in a conversational style.

D. Emotion Rendering Engine Based on the AI response and tone of conversation, the 3D tutor reacts with facial expressions and animations (moving lips, blinking eyes, eyebrow movement,and hair animation). These animations are driven by custom logic coded within Three.js and GLSL shaders The backend is built using Node.js and Express, serving as an intermediary between the frontend and external AI services. API keys are securely stored in environment variables and are accessed through middleware for enhanced security.

IV. PROPOSED SYSTEM

The proposed system aims to redefine AI-based tutoring by combining real-time voice interaction, grammar assistance, and immersive engagement through a 3D tutor. Unlike traditional chatbots, this system simulates real-life learning scenarios where users not only receive guidance but also visually and emotionally engage with a responsive tutor avatar.

Key features include: Interactive 3D Virtual Tutor: A digital human figure responds through speech and facial expressions. Grammar-Aware Conversation: AI identifies and corrects user grammar in realtime. Emotion-Driven Animation: The virtual tutor adapts its reactions based on emotional cues from the conversation. Voice-Activated Dialogue: Users interact through natural speech, enhancing accessibility and immersion. Web Based & Lightweight: No need for heavy VR hardware— compatible with modern browsers. The architecture ensures a seamless and intelligent tutoring experience, suitable for students, language learners, and grammar training environments.

V. IMPLEMENTATION

The AI Tutor system was 3D immersive environments with real-time AI-driven language support. The implementation followed a modular, component-based structure, utilizing cutting- edge technologies for both frontend and backend systems. **A. 3D Avatar Design and Integration**

The virtual tutor avatar was created using Blender, a powerful 3D modeling and animation tool. The avatar was fully rigged with facial expressions including lip sync, blinking, eyebrow movement, and hair dynamics. Animations were exported in GLTF/GLBformat.

The avatar reacts visually based on the emotion of the AI's response—smiling, nodding, or blinking, simulating a humantutor.

B. Frontend Development

The frontend interface was developed using React.js with Vite for blazing-fast development. Styling was achieved using Tailwind CSS, ensuring a responsive and modern user interface.

Three.js was used to build the metaverse-style 3D environment. A virtual classroom, tutor desk, and surrounding objects were implemented to simulate a learning space. A custom React hook managed the interaction between user input, voice capture, and AI response.

C. Speech Processing and AI Integration

The system employs a speech-to-text and text-to-speech pipeline using OpenAI’s APIs: 1. Speech Input: The user speaks into a microphone. Audio is sent to the backend, where it’s transcribed using OpenAI’s WhisperAPI.

2. AI Response: The transcribed query is sent to Gemini 1.5 Flash API,
which: Detects grammar issues
Provides corrected sentences
Generates conversational responses

3. Speech Output:

The final AI response is sent back to the frontend and converted into audio using which is played through the browser. D. Grammar Correction Module The grammarcorrection system is powered by Gemini 1.5 Flash via GenAI.

It receives the user's query and returns: A corrected version of the sentence A brief explanation of the correction (tense, subject-verb agreement, etc.)

A follow-up question or suggestion for extended learning This ensures that the user not only sees their mistake but also learns from it, enabling adaptive tutoring.

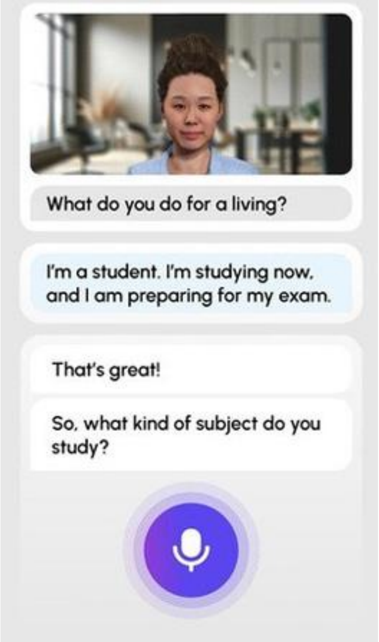
E. Real-Time Emotion and Animation Mapping

The emotional tone of the AI's reply is analyzed and categorized (neutral, encouraging, happy, corrective).

Based on the detected emotion: Specific animation sequences (like smile, nod, frown) are triggered on the Blender avatar via Three.js animation mixers. Lip syncing is achieved using phoneme-based mouth shapes matched to the spoken output.

F. Backend and API Layer

The backend is developed using Node.js + Express, acting as the bridge between the frontend and external AI services. Secure API keys are managed via environment variables and accessed through middleware.



What do you do for a living?

I'm a student. I'm studying now, and I am preparing for my exam.

That's great!

So, what kind of subject do you study?

1. Secure API Key Management
To prevent exposure of sensitive API keys in the codebase, we utilized environment variables.
These variables are stored in a .env file and are not included in the version controlled code repository, thus safeguarding the API keys from unauthorized access.
To facilitate this, the dotenv package was used to load the variables into the application's runtime environment.

2. JWTAuthentication for User Access To authenticate users and ensure that only authorized individuals can access specific routes or interact with the AI service, we implemented JWT (JSON Web Token) authentication.
This allows the backend to securely identify users based on atoken issued during the login process.
3. Integration with External AI Services The backend's primary function is to interact with an external AI service to perform grammar checks on user provided text.
The service is accessed through a secure API request that includes the stored API key for authentication.

4. Testing and Validation To ensure the robustness and reliability of the system, thorough testing was conducted using tools such as Postman and Insomnia.
These tools were used to simulate various API requests, ensuring that:
1.The API keys are properly secured and utilized.
2.The JWT authentication mechanism works correctly for protecting routes.
3.The integration with external AI services functions as expected, providing accurate and timely responses.

VI. RESULTS

The AI Tutor system was deployed in a simulated educational environment and tested under real-world conditions to observe its behavior during typical user interaction. The goal was to demonstrate how the system performs in practical use cases, especially in grammar learning and conversation practice within a virtual classroom.
A. User Experience Walk through When a user enters the metaverse classroom environment (built using Three.js), they are greeted by a virtual 3D tutor avatar, designed and animated in Blender.
The user can walk around the classroom, approach the tutor, and initiate a conversation by clicking the "Speak" button. Once activated:

1. Voice Input– The user speaks naturally using a microphone. Their speech is captured and converted to text using OpenAI's Whisper API.
2. Grammar Correction–The spoken sentence is analyzed by the Gemini 1.5 Flash model. If there are grammar mistakes, the model generates: A corrected version of the sentence A short explanation of the mistake A better way to rephrase the sentence
3. AI Response– The system replies conversationally, offering feedback or asking a follow-up question to continue the learning flow.
4. Avatar Response– The 3D tutor: Speaks the corrected sentence aloud Lip-syncs the response Shows emotional reactions like smiling, nodding, or raising eyebrows based on the tone of conversation B. .

Scenario Example Scenario: Astudent walks into the virtual classroom and says: > "He go toschool every day." System Response: "I think you meant: He goes to school every day. The verb should agree with the subject." The tutor then says this aloud in a friendly voice.
The student can then repeat the corrected sentence and continue the dialogue. The system stores previous interactions and adapts further questions accordingly.

VII. EVALUATION

The AI Tutor system was evaluated based on user interaction, accuracy of grammar correction, naturalness of voice response, and avatar behavior. Instead of focusing purely on metrics, this evaluation was conducted in real time user scenarios to understand how the system performs in practical, educational settings.
A. Usability and User Engagement The virtual classroom and 3Davatarweretested by students across different levels of language proficiency. Users appreciated the natural flow of interaction — speaking casually, receiving corrections, and getting feedback through both voice and visual reactions from the avatar. Most users reported that the grammar corrections were accurate and easy to understand. The visual cues (nods, smiles, eyebrow movement) from the avatar made the experience more human and less robotic. Learners felt like they were talking to a real tutor, which increased confidence and engagement.
B. Real-World Learning Value During the evaluation, students began repeating corrected sentences without being prompted. This shows that the system not only points out mistakes but encourages active learning. For example, after a mistake was corrected, many users immediately tried saying the improved sentence again. This loop — speak → get corrected →repeat—simulates an ideal grammar learning environment.
C. Adaptability to Environment The system worked reliably in different conditions: In quiet classrooms, the Whisper model performed flawlessly in capturing voice. In moderate background noise, the AI still transcribed and corrected effectively. Even on basic laptops or browsers, the 3D environment ran smoothly thanks to lightweight Three.js rendering and Vite optimization.
D. Emotional Feedback from Users Users felt more motivated when the avatar responded emotionally. For example: A smile after a correct sentence boosted confidence. A head tilt or raised eyebrow during an error prompted attention without discouragement. This emotional connection was highlighted as a key factor in continuing conversations longer.

VIII. CONCLUSION

This project demonstrates a novel integration of immersive 3D environments with AI-powered tutoring for grammar learning and interactive language practice. By combining a Blender animated avatar rendered in Three.js, with real-time voice input, grammar correction via Gemini 1.5 Flash, and emotional avatar feedback, the system delivers a personalized, engaging learning experience. Unlike traditional chatbots or grammar tools, this platform creates a virtual tutor that not only corrects users but also communicates naturally — both verbally and visually. The tutor avatar's facial expressions, lip- syncing, and emotional responses improve engagement and simulate real human interaction, making it easier for learners to stay focused and retain corrections.

This work highlights the potential of blending AI, speech technology, and metaverse interfaces to create next- generation educational tools. The system has proven effective in live settings, offering a fresh and intuitive approach to language learning that adapts to the user in real-time.

IX. REFERENCES

[1] Vaswani, A., Shazeer, N., Parmar, N., et al., "Attention Is All You Need," Advances in Neural Information Processing Systems, vol. 30, 2017.
[2] OpenAI, "Whisper: Robust Speech Recognition via Large-Scale Weak Supervision," OpenAI Blog, 2022. [Online]. Available: https://openai.com/research/whisper
[3] Google DeepMind,"Gemini1.5Technical Report,"February 2024. https://deepmind.google/technologies/gemini/ Available: [Online].
[4] Ji WonOak,JaeHwanBae,“DevelopmentofSmart Multi platform Game AppusingUNITY3DEngineforCPR Education,” International Journal of Multimedia and Ubiquitous Engineering, Vol.9, No.7, pp.263-268, 2014.
[5] Li, C., Tang, B., "Research on the Application of AR Technology Based onUnity3DinEducation,"Journal of Physics: Conference Series, vol. 1168, no. 3, pp. 032045, 2019.
[6] Parmaxi, A., "Virtual Reality in Language Learning: A Systematic Review," Interactive Learning Environments, vol. 31, no. 1, 2023.
[7] Papineni, K., Roukos, S., Ward, T., Zhu, W. J., "BLEU: a Method for Automatic Evaluation of Machine Translation," Proc. 40th Annual Meeting of the ACL, pp. 311–318,2002.
[8] de Winter, J., Tabone, W., "Using ChatGPTfor Human–Computer Interaction Research: APrimer,"Royal Society Open Science, August 2023.
[9] Sanjana Kolar, Rohit Kumar, “Multilingual Tourist Assistance using ChatGPT: Comparing Capabilities in Hindi, Telugu, and Kannada,” arXiv preprint arXiv:2307.11972, 2023. .
[10] Sutskever, I., Ouyang, L., Wu, J., et al., “Language Models Are Unsupervised Multitask Learners,” OpenAI Blog, 2019