

Verba-l: A Perfect Approach to Mastering Language

Afreen Aslam

Dept. of Computer Science and Engg.
ICCS College of Engg. and Management
Thrissur, Kerala, India

Anex Shaju

Dept. of Computer Science and Engg.
ICCS College of Engg. and Management
Thrissur, Kerala, India

Anjanakrishna P S

Dept. of Computer Science and Engg.
ICCS College of Engg. and Management
Thrissur, Kerala, India

Ms. Divya Jose

HoD, Dept. of CSE
ICCS College of Engg. and Management
Thrissur, Kerala, India

Abstract—The escalating global demand for English language proficiency, particularly for high-stakes evaluations such as the International English Language Testing System (IELTS), has necessitated the development of accessible, personalized learning tools that transcend traditional pedagogical methods. This paper presents Verba-L, an advanced multimodal AI platform engineered to bridge the gap between theoretical linguistic knowledge and functional communicative fluency within an adaptive digital environment. By integrating a sophisticated suite of Natural Language Processing (NLP) models—including a fine-tuned T5 Transformer for real-time, text-to-text grammar correction, Blenderbot for low-latency conversational engagement via the AI persona "Felix," and BERT for automated comprehension scoring—the system provides a holistic preparation ecosystem. The technical architecture follows a robust microservice framework, utilizing a Node.js backend for seamless user management and a Flask-based API to host computationally intensive AI tasks, supported by MongoDB Atlas for real-time progress tracking. Experimental results demonstrate significant model efficacy, characterized by a sharp decline in error rates during the training phase and the attainment of a grammatical accuracy of approximately 75 percent over 3,500 training steps. Ultimately, Verba-L offers a scalable, cost-effective, and interactive solution that empowers students to achieve the technical skills and confidence necessary for global academic and professional success.

Index Terms—Natural Language Processing, T5 Transformer, IELTS, Blenderbot, Adaptive Learning, Educational Technology.

I. INTRODUCTION

In the modern digital world, rapid technological growth has significantly changed the way education and learning take place. The easy availability of smartphones, high-speed internet, and online platforms has allowed students to access learning resources anytime and from anywhere, without depending only on traditional classrooms. Among all academic skills, proficiency in the English language plays a crucial role in shaping students' academic performance, career prospects, and ability to communicate at a global level. Even so, many

learners continue to struggle with issues such as correct grammar usage, clear pronunciation, listening comprehension, and confident communication, particularly in regions where English is not widely used in daily life.

Most traditional English learning approaches depend heavily on textbooks and limited classroom interaction, which may not fully meet the individual needs of learners. Many students feel hesitant to speak in English because they fear making mistakes, lack personalized guidance, or do not receive immediate feedback. In addition, conventional learning methods usually involve delayed assessments, making it difficult for students to regularly monitor their progress. These challenges indicate the necessity of intelligent, interactive, and adaptive learning systems that can effectively support learners in improving their English language skills.

With advances in Artificial Intelligence (AI), Natural Language Processing (NLP), and speech processing technologies, computer systems are now capable of understanding, analyzing, and responding to human language in a more meaningful and practical way. NLP techniques help machines process grammar, sentence structure, and context, while speech processing technologies convert spoken language into text that machines can interpret. Such technologies are already used in applications like virtual assistants, chatbots, and digital learning tools, proving their ability to make language learning more engaging and personalized.

Inspired by these technological developments, this paper presents a conceptual design for an AI-assisted English learning platform aimed at supporting learners in improving grammar, reading, listening, and conversational abilities. The proposed system is planned as a web-based platform that integrates grammar checking, conversational interaction, and IELTS-style assessment modules. The objective is to offer personalized feedback and adaptive learning paths based on individual performance. Since this work focuses on a proposed model, the discussion is limited to system design, method-

ological concepts, and expected outcomes rather than practical implementation results.

A. Background and Motivation

The growing use of AI-based applications in everyday life highlights users' preference for systems that provide ease of use, efficiency, and personalization. Intelligent tools such as chatbots and virtual assistants demonstrate how language-based interaction can simplify tasks and increase user involvement. In the field of education, these systems have the potential to go beyond simple information delivery and function as interactive learning companions.

Students preparing for English proficiency examinations often need regular practice, organized evaluation methods, and consistent feedback. Unfortunately, access to high-quality coaching and individualized instruction is limited for many learners. This issue becomes even more serious in remote areas or environments with fewer learning resources. An AI-driven learning platform, if developed, could help overcome these challenges by offering accessible, scalable, and exam-focused learning support.

The main motivation behind proposing this system is to reduce the gap between conventional teaching methods and modern intelligent learning environments. By combining text-based language analysis with speech-based interaction, the proposed platform aims to support overall language development. It is expected to promote active learner participation while minimizing the need for continuous human supervision.

B. Role of Artificial Intelligence in Language Learning

Artificial Intelligence plays an important role in improving language learning systems by enabling machines to understand and process human language with better contextual awareness. NLP techniques make it possible to detect grammatical errors, restructure sentences, analyze meaning, and interpret context. Similarly, speech recognition technologies allow spoken language to be converted into text, enabling effective speaking and listening practice.

In the proposed system, AI functions as the central element responsible for analyzing learner inputs, generating meaningful feedback, and adjusting difficulty levels based on performance. Conversational AI models are intended to simulate real-world communication scenarios, allowing learners to practice spoken English in a supportive and controlled environment. Over time, such interaction can help improve confidence, fluency, and accuracy.

Additionally, AI-based analytics can monitor learner performance patterns over time. These insights can be used to personalize learning activities and recommend specific exercises, thereby enhancing the effectiveness of the learning process.

C. Scope of the Proposed System

The English Learning Web Application is designed to create an IELTS-style learning environment that supports college students in improving their overall English proficiency through AI-assisted interactive learning. The scope of this project

covers all major language skills, including listening, reading, writing, and speaking, ensuring a well-rounded and organized approach to learning English effectively.

The application combines speech recognition, AI-based grammar correction, and conversational chatbot features to offer a personalized and adaptive learning experience. Users can speak freely and receive instant feedback on pronunciation and grammar, take part in real-time conversations, and practice structured exercises similar to those used in IELTS and other English proficiency examinations. The reading module provides automatically generated passages followed by comprehension-based questions, while the listening module includes audio-based exercises designed to strengthen listening abilities.

In addition, the web application features a progress-tracking dashboard that enables users to view their scores, observe improvement over time, and recognize areas that require more attention. A streak-based scoring mechanism encourages regular practice and consistent participation, while detailed analytical insights help users understand their learning patterns better. The system is built with scalability in mind, allowing future enhancements such as additional learning modules, support for more languages, and advanced AI updates to further improve conversational fluency.

From a technical standpoint, the web application uses Node.js with the Express framework for backend processing, Flask for handling AI-based text and speech analysis, and MongoDB to securely store user progress and performance data. The frontend is developed using standard web technologies such as HTML, CSS, and JavaScript to ensure a clear, engaging, and easy-to-use interface. The platform is designed to support multiple users at the same time, making it suitable for individual learners, college language laboratories, and online language learning environments.

In conclusion, the scope of this project goes beyond basic English learning by offering an AI-powered, interactive, and scalable platform that helps students build fluency and confidence in English communication. By using advanced speech recognition, chatbot-based interaction, and adaptive learning strategies, the system aims to provide a practical, real-world language learning experience that prepares students for academic challenges and professional communication needs.

II. OBJECTIVES

The main objective of this project is to develop an IELTS-style learning environment that helps college students improve their English language skills. Many students find it difficult to use English effectively, especially in academic and professional situations where strong communication abilities are required. This web application is designed to reduce this gap by providing an interactive, AI-supported platform that closely aligns with real-world English proficiency examinations such as IELTS.

The platform offers an immersive learning experience by replicating IELTS-based listening, reading, writing, and speaking activities. Structured reading and listening sections

help strengthen comprehension skills and vocabulary, while the chatbot-based conversation module allows users to practice speaking in real time and receive immediate feedback. Through the integration of speech recognition and AI-powered grammar correction, learners are guided on correct pronunciation, sentence formation, and proper grammatical usage.

In addition, the system includes progress-tracking features through personalized dashboards, enabling students to regularly monitor their performance, identify weaker areas, and improve their skills accordingly. This continuous evaluation process supports steady and measurable language improvement.

The AI assistant, functions as a virtual tutor by engaging users in interactive conversations, pointing out mistakes, and offering constructive feedback to gradually build confidence in English communication. The overall goal of the platform is to make English learning more accessible, engaging, and effective, helping students succeed in academic work, professional environments, and everyday communication.

By combining real-time speech processing, natural language processing models, and data-driven performance analysis, this web application delivers a complete AI-based approach to English language learning. Its interactive and personalized design encourages consistent practice and helps learners continuously improve their grammar, listening skills, and conversational fluency.

III. EASE OF USE

Ease of use plays a major role in how effectively any educational technology is adopted and used. For learners to regularly engage with an AI-assisted language learning platform, the system must be easy to understand, simple to access, and supportive of different user needs. The proposed system is conceptually designed with a strong focus on usability, ensuring that students with different levels of technical knowledge can use the platform comfortably. By emphasizing simplicity, clarity, and guided interaction, the platform aims to reduce entry barriers and encourage continuous learner participation.

A. User-Friendly Interface Design

The proposed platform is planned with a clean and well-organized user interface that presents learning modules in a clear and structured format. Each feature, including grammar practice, conversational interaction, reading, and listening, is intended to be clearly labeled and visually separated to prevent confusion. Simple design elements, consistent page layouts, and logical navigation paths are expected to help users move smoothly between different learning activities. In addition, the interface is designed to offer clear instructions and supportive prompts, guiding learners through each activity step by step. These design choices are expected to reduce mental effort and allow students to concentrate more on learning tasks rather than struggling with system complexity.

B. Accessibility and Device Compatibility

To reach a wider group of learners, the proposed system is intended to be accessible through standard web browsers without the need for specialized hardware or software installation. This approach ensures compatibility with desktops, laptops, tablets, and mobile phones. A responsive interface design is envisioned so that the platform can automatically adjust to different screen sizes and support learning across various devices. Accessibility considerations also focus on learners with different levels of English proficiency and learning speed. The use of simple language for instructions, clear feedback messages, and flexible learning pace options is intended to make the system usable for both beginners and advanced users.

C. Maintaining the Integrity of the Specifications

The system is conceptually designed to support learners through guided interactions rather than complicated setup processes. New users are expected to receive brief explanations on how to use the platform and begin their learning activities. During practice sessions, helpful hints and suggestions may be provided to clarify system responses and guide learners toward improvement. The conversational learning module is designed to encourage active participation by allowing users to make mistakes without fear of negative judgment. By maintaining a supportive and friendly interaction style, the system aims to improve learner confidence and sustain motivation. A key objective of the proposed system is to minimize the learning curve commonly associated with advanced AI-based applications. By simplifying user interactions and handling complex processes such as grammar evaluation and speech recognition in the background, learners can focus entirely on practicing and improving their language skills. Visual indicators and simple performance summaries are expected to present progress information in an easy-to-understand manner. These usability-focused features are anticipated to enhance user engagement, encourage consistent practice, and support long-term learning outcomes. A platform that is easy to use and navigate is more likely to be used regularly, thereby increasing its overall educational value.

IV. RELATED WORK

Gowthamy et al. proposed an improved AI-based voice assistance system that uses speech recognition, natural language processing, and machine learning to perform computer-related tasks through voice commands. The system is mainly designed to improve accessibility and efficiency by allowing users to interact hands-free through a graphical user interface. It also relies on external APIs to retrieve information and execute commands. Their study highlights the effectiveness of NLP-based command understanding and adaptive learning techniques in improving user interaction with computing systems. However, the system is focused mainly on utility-based operations rather than educational objectives. It does not include features such as structured language skill development, continuous learner evaluation, or real-time corrective feedback,

which limits its usefulness in English language learning applications that require pedagogical support.

Yves and Chan introduced a machine learning-based approach to predict IELTS speaking proficiency levels using linguistic features extracted from transcribed speech. Their system converts speech into text using automatic speech recognition and then applies natural language processing techniques to extract lexical and syntactic features. This approach allows for more objective and consistent scoring aligned with IELTS assessment standards, thereby reducing dependence on human examiners. Although the model effectively emphasizes factors such as lexical richness, it relies completely on text-based analysis. As a result, important spoken language aspects like pronunciation accuracy, fluency, and intonation are not considered. Moreover, the absence of an interactive learning environment or real-time feedback limits the system's role to score prediction rather than active speaking skill improvement.

Jing examined the use of speech recognition sensors combined with artificial intelligence to enhance learners' oral pronunciation skills. The proposed method captures spoken input through speech recognition and provides model pronunciation using speech synthesis, allowing learners to compare their speech with standard references. Real-time feedback helps users identify pronunciation errors and improve accuracy. While the study demonstrates the strength of AI-based pronunciation training, its focus is limited only to pronunciation. Other language skills such as grammar, listening comprehension, and conversational ability are not addressed. In addition, the reliance on specialized sensor hardware and the variations in AI-generated feedback present challenges for large-scale and web-based learning systems.

Ericsson and Johansson studied the long-term impact of conversational AI tools on English speaking practice among lower secondary school students. Their system integrates conversational agents with automatic speech recognition to give learners more opportunities to practice speaking outside regular classroom settings. The results show improvements in learner confidence, engagement, and willingness to speak English over extended use. However, the system is more suitable for beginners, as advanced users experience repetitive interactions due to predefined dialogue structures. Speech recognition errors and technical limitations also affect the learning experience, indicating the need for more adaptive conversational models, detailed feedback mechanisms, and coverage of additional language skills.

Barnwal et al. developed ARIVA, an AI-powered voice assistant that uses speech recognition, NLP, and text-to-speech technologies to perform everyday tasks such as scheduling, web searches, and information retrieval. The assistant employs intent recognition and entity extraction techniques to respond efficiently through both audio and visual outputs while maintaining low computational complexity. Although ARIVA performs well as a utility-oriented assistant, it is not designed as a language learning system. The absence of structured learning activities, linguistic error correction, and performance tracking limits its application in exam-focused or skill-based

English learning environments.

Subhash et al. proposed a general-purpose AI-based voice assistant capable of performing a range of functions including system control, language translation, and information access through voice interaction. The system integrates automatic speech recognition, text-to-speech synthesis, and language modeling to enable natural communication between humans and machines. While the assistant offers flexibility and customization across multiple domains, it prioritizes task completion rather than language learning. Features such as adaptive feedback, individualized assessment, and structured learning progression are not included, highlighting the clear gap between general-purpose voice assistants and specialized AI-driven English learning platforms.

V. PROPOSED SYSTEM DESIGN

The proposed system is an AI-assisted English learning platform designed to help learners improve their grammar, speaking, listening, and reading skills in a single, adaptive environment. It brings together multiple AI components, including natural language processing, text evaluation, speech recognition, and comprehension analysis, to create an interactive learning experience that closely mirrors standardized English proficiency tests.

The platform uses a modular design approach, where each learning component functions independently while remaining connected through a central controller. This makes the system scalable and allows for future expansion, such as adding more advanced AI models or supporting additional languages. The platform aims to solve common learning challenges, including limited access to human tutors, delays in receiving feedback, and the lack of a structured, personalized learning environment.

Conceptually, the system acts as a virtual tutor, providing real-time corrections, scoring user responses, analyzing performance patterns, and offering tailored recommendations. The ultimate goal is to build a fully automated, efficient, and engaging platform that supports learners in academic, professional, and certification-focused English language preparation.

A. System Architecture Design

The architecture of the proposed system is designed as a multi-layered framework consisting of an integrated user interface, a centralized AI processing core, and a secure data management backend, all working together to create an adaptive English learning environment. At the user interface layer, learners interact with the platform through modules for grammar correction, speaking practice, reading comprehension, and listening exercises, each designed to capture text or audio inputs and display AI-generated feedback in a simple and intuitive manner. The captured inputs are transferred to the AI processing layer, which forms the core intelligence of the system and comprises several specialized modules, including a grammar analysis engine that detects and corrects sentence-level errors, a speech recognition and evaluation module that converts spoken responses into text

while assessing pronunciation and fluency, a conversational dialogue generator that produces context-aware responses for interactive speaking sessions, and comprehension evaluators that analyze user answers for reading and listening tasks. These AI modules operate independently but communicate through internal APIs, ensuring modularity, scalability, and parallel processing. All processed results, user scores, activity logs, and improvement patterns are stored within the data management layer, which maintains structured user profiles, historical performance records, and adaptive difficulty settings that allow the system to personalize the learning experience over time. This layered architecture not only enables efficient data flow and real-time feedback but also ensures that future upgrades—such as more advanced AI models, new learning modules, or expanded language support—can be seamlessly integrated without restructuring the entire system.

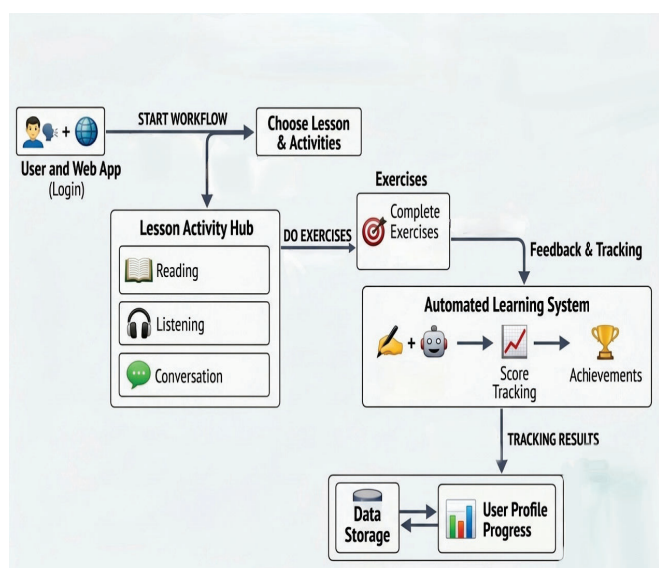


Fig. 1. SYSTEM ARCHITECTURE

B. System Flow Chart

The data flow begins when the user interacts with the frontend, whether by logging in, signing up, submitting written input, or speaking through a microphone. These interactions are captured by the interface and sent to the Node.js/Express.js backend, where authentication and session validation are performed to ensure secure access. Once the user is verified, inputs related to learning activities are forwarded to the Flask API for AI processing, while the backend simultaneously communicates with MongoDB Atlas to store user profiles, update progress logs, and maintain session activity. After the AI service completes its processing, the results are returned to the backend, where they are integrated with the user's stored data to generate a comprehensive response. Finally, the processed outputs are displayed on the frontend dashboard, providing the user with real-time feedback, performance metrics, and clear visual insights into their learning progress.

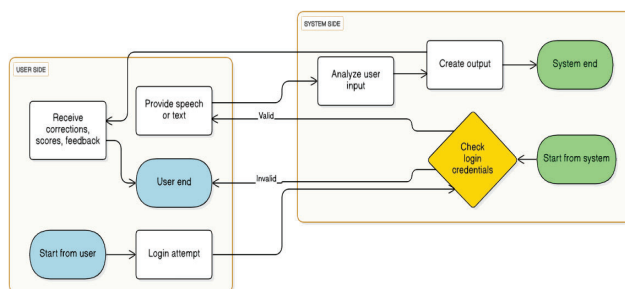


Fig. 2. MODULE 1

When the backend receives AI-related requests from the frontend, it forwards them to the Flask API, which manages all model-level processing. The API handles each request using specialized AI components, including DeepSpeech for speech-to-text conversion, transformer-based models for grammar correction and text refinement, BERT models for reading and listening comprehension scoring, and conversational models such as BlenderBot for interactive dialogue. Once the processing is completed, the AI-generated results are returned to the backend, which updates the user's performance metrics, learning history, and progress records in MongoDB Atlas. The system also incorporates a continuous improvement loop, whereby datasets from open-source repositories are periodically used to retrain and enhance models such as T5, BlenderBot, and BERT. This iterative training process ensures that the platform becomes increasingly accurate, responsive, and capable of addressing a broader range of language learning tasks over time.

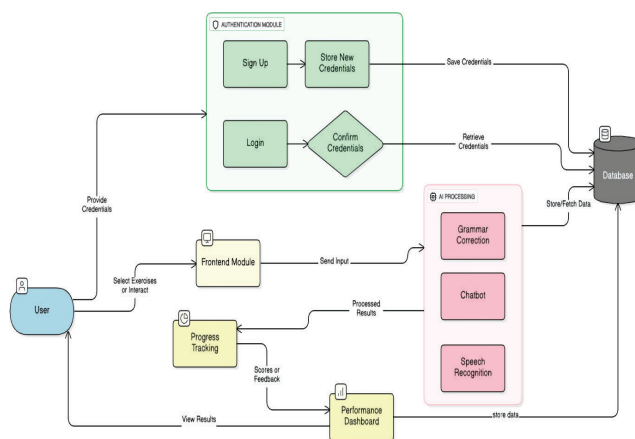


Fig. 3. MODULE 2

A closer examination of the Flask microservice reveals how multiple AI models collaboratively process user inputs. Upon receiving raw text or speech, a routing mechanism determines the appropriate model for the specific task. For grammar correction, the input is directed to the T5 Transformer, which

identifies errors and generates refined output. For conversational interactions, the input along with contextual information is processed by BlenderBot to produce coherent and contextually relevant responses. In reading or listening comprehension tasks, the user's submitted answers are evaluated by the BERT model, which analyzes content and assigns corresponding scores. Once each model completes its respective operation, the Flask API consolidates the outputs—including corrected text, chatbot responses, comprehension scores, or speech transcriptions—into a unified result. This aggregated output is then sent back to the Node.js backend, ensuring a seamless, organized, and efficient flow of intelligent feedback to the user.

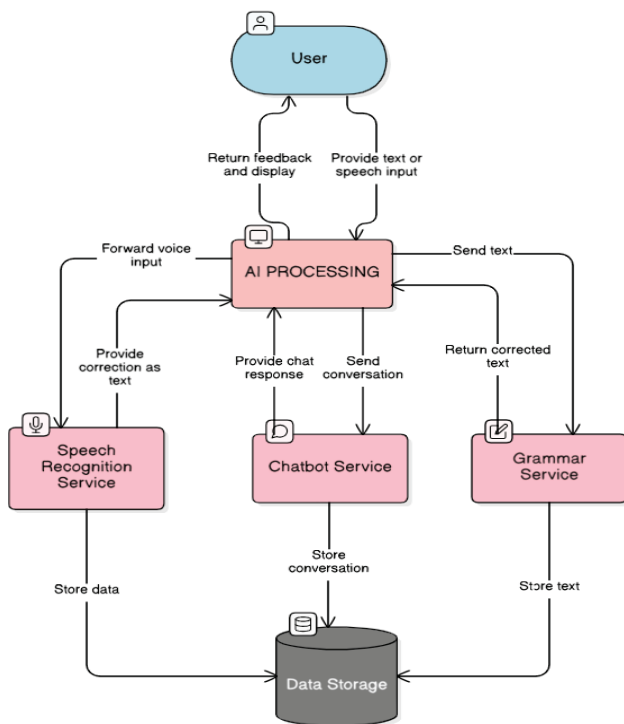


Fig. 4. MODULE 2.1

VI. METHODOLOGY

The development of the NLP-based English Language Assistant follows a clearly defined and systematic methodology to ensure reliable performance, accuracy, and future scalability. The overall approach is divided into several well-organized stages that guide the progress of the project from initial planning to final deployment. These stages include data collection, appropriate model selection, system implementation, training, performance evaluation, and deployment.

A. Data Collection

To develop an AI-driven English learning platform, a diverse and well-balanced dataset is required to ensure accurate model training and overall system effectiveness. The dataset

is designed to cover multiple aspects of language learning, including speech data, grammar correction data, conversational data, and listening and reading materials. Speech data is collected from publicly available sources such as Mozilla Common Voice, which offers a wide range of spoken English samples. These recordings include different accents and speaking styles, helping improve the performance of speech recognition models.

The grammar correction dataset is created using data extracted from the C4 200M dataset, along with manually prepared sentences that contain grammatical errors and their corrected forms. This combination allows the model to learn how to identify common mistakes and apply appropriate corrections, supporting better sentence structure and grammatical accuracy. Conversational data is obtained from open-source chatbot datasets as well as real-world English conversations, enabling the chatbot module to generate meaningful and context-appropriate responses. This data helps the system understand conversational flow, intent, and language context.

Listening and reading materials are carefully selected to reflect an IELTS-style learning environment. These materials include passages and audio recordings from educational sources, which are structured to improve comprehension skills through repeated practice. By aligning the content with standard proficiency tests, the system allows learners to train under realistic and exam-oriented conditions.

Before using the collected data for model training, preprocessing steps are applied to improve quality and consistency. Tokenization is used to divide sentences into smaller textual units, making it easier for language models to analyze and process the data. For speech data, noise reduction techniques are applied to remove background disturbances and improve transcription accuracy. Text normalization is performed to standardize sentence forms, correct inconsistencies, and ensure balance across the dataset.

Through these preprocessing techniques, the dataset becomes structured, reliable, and suitable for training AI models effectively. The quality and diversity of the data play a key role in the performance of the English learning assistant, making data collection and preparation a critical stage in the development of the English Learning Web Application.

B. Model Selection

The selection of appropriate AI models plays an important role in delivering an effective and efficient learning experience in the proposed web application. Different models are integrated to support grammar correction, speech recognition, chatbot-based interaction, and listening and reading evaluation. Each model is chosen based on its suitability for language learning tasks and its ability to provide accurate and real-time feedback.

The grammar correction module is built using a fine-tuned T5 Transformer model. This model is trained on large datasets consisting of incorrect sentences along with their corrected versions, enabling it to recognize grammatical patterns and apply accurate corrections based on context. By using

deep learning techniques and natural language processing, the model is designed to adapt and improve over time as it processes more user-generated input.

For speech recognition, the system integrates DeepSpeech along with the Web Speech API. DeepSpeech is an open-source speech recognition model trained on diverse speech datasets, which helps achieve reliable transcription accuracy across different accents and speaking styles. The Web Speech API supports real-time speech-to-text conversion, allowing users to practice speaking exercises and receive instant feedback. Together, these tools help learners improve pronunciation, clarity, and overall fluency through interactive voice-based practice.

The chatbot module is powered by Blenderbot, developed by Facebook AI, and fine-tuned using educational and conversational dialogue datasets. This model enables the chatbot to engage users in meaningful conversations while providing feedback related to grammar usage, sentence formation, and language clarity. Blenderbot is designed to generate natural, human-like responses, making it suitable for realistic conversational practice. The chatbot can also adjust its responses based on the learner's proficiency level, offering more personalized learning support as users progress.

For listening and reading evaluation, BERT-based NLP models are used to assess user comprehension. These models analyze learner responses to reading passages and listening exercises to evaluate understanding and provide corrective suggestions. Due to BERT's strong ability to capture contextual relationships within text, it can offer accurate and meaningful feedback. This evaluation component plays a key role in simulating IELTS-style assessments, where strong comprehension skills are essential for achieving high scores.

Overall, the selected models are chosen for their accuracy, scalability, real-time performance, and suitability for educational applications. By integrating these AI-powered components, the proposed web application creates an adaptive learning environment that provides personalized feedback, supports gradual skill development, and helps learners prepare effectively for English proficiency examinations such as IELTS.

C. Implementation

The English Learning Web App is built with several interconnected components, including the frontend, backend, AI processing, and database management. The frontend uses JavaScript, HTML, and CSS to create an intuitive, easy-to-use interface. It handles interactive features like real-time speech-to-text conversion, chatbot conversations, and user progress dashboards. By integrating the Web Speech API, learners can interact using their voice and receive immediate feedback. The responsive design ensures the platform works smoothly across desktops, laptops, tablets, and mobile devices, providing a seamless learning experience.

On the backend, Node.js with Express.js manages tasks such as user authentication, session handling, and API requests. It processes inputs from users, manages their sessions, and

communicates with the database. The backend also connects the frontend with the AI models, ensuring smooth data flow and accurate responses. To protect users' personal data, secure authentication methods like JWT-based login are implemented.

AI functions are handled through a Flask-based API, which powers grammar correction, chatbot interactions, and speech-to-text processing. The AI models analyze user inputs, generate real-time feedback, and return corrections or suggestions. Flask allows the frontend and AI components to communicate efficiently, keeping interactions fast and responsive. API endpoints are set up to manage grammar checks, conversational practice, and listening comprehension exercises, providing an adaptive and personalized learning experience.

User progress, conversation history, and learning analytics are stored in MongoDB, a flexible NoSQL database. MongoDB can handle both structured and unstructured data, making it ideal for tracking performance metrics. The database keeps records of chatbot interactions, listening exercises, and reading scores, allowing the system to give personalized recommendations based on past performance. Real-time updates ensure learners can track their progress efficiently.

All these components work together to create a smooth, cohesive system. Users can practice, interact, and improve their English skills while the web app adapts to their performance. Its modular structure also allows for future enhancements, such as adding new learning modules, improving AI features, or supporting additional languages. Overall, this implementation provides a robust, interactive, and adaptive platform for students preparing for English proficiency tests like IELTS, making language learning engaging and effective.

D. Training and Evaluation

The training and evaluation process is designed to make sure each AI model performs well and adapts effectively to real-world learning situations. The Grammar Model is fine-tuned using a dataset containing thousands of sentence pairs, each with incorrect and corrected versions. Its performance is evaluated with BLEU (Bilingual Evaluation Understudy) and ROUGE (Recall-Oriented Understudy for Gisting Evaluation) scores, which measure how accurate and fluent the corrected sentences are. The model is trained iteratively to improve its ability to identify and correct a wide variety of grammatical errors, helping learners write more accurately.

The Speech Recognition Model is trained using diverse English speech samples from datasets like Mozilla Common Voice. This allows the system to understand different accents, speech speeds, and pronunciations. The model's accuracy is measured using the Word Error Rate (WER), which calculates the percentage of errors in transcriptions compared to the original speech. Lower WER indicates better performance. The model is refined through repeated training until it can provide reliable real-time speech recognition for learners.

The Chatbot Model is fine-tuned on datasets containing conversational dialogues, especially those aimed at English learning and correction. Its responses are evaluated using

perplexity scores, which show how well the model can generate human-like replies. User feedback is also critical in improving the chatbot, ensuring that it provides meaningful, context-aware answers. The chatbot is continuously updated based on interactions with learners, making conversations more engaging and effective for practicing spoken English.

For Listening and Reading Evaluation, BERT-based NLP models are used to analyze responses to comprehension exercises. These models are trained on annotated IELTS-style passages and validated by comparing user answers to expert-verified solutions. This evaluation ensures that the system can accurately gauge comprehension skills and provide helpful feedback for improvement.

Overall, the training and evaluation methodology ensures that each AI component meets high standards of performance. By continuously refining the models based on evaluation metrics and learner interactions, the system adapts to different learning needs. This makes the English Learning Web App a dynamic, intelligent, and effective tool for improving grammar, pronunciation, comprehension, and conversational skills.

E. Model Deployment

Deploying the AI models is a crucial step to make sure they are accessible, scalable, and efficient for real-time use. The Flask API serves as the core of this deployment, enabling smooth communication between the AI models and the frontend. It handles requests for grammar correction, chatbot interactions, and listening comprehension feedback, providing fast responses with minimal delay. Flask's lightweight and efficient design makes it an excellent choice for serving machine learning predictions in real time.

To improve scalability and reliability, Docker and cloud services are used in deployment. Docker containers package the AI models along with their dependencies and API services, ensuring that the system can be deployed consistently across different environments. This containerized approach minimizes compatibility issues and makes maintenance easier. Cloud platforms like AWS, Google Cloud, or Azure host the models, allowing multiple users to use the system simultaneously without any drop in performance.

Continuous updates are an important part of the deployment strategy. The AI models are periodically retrained using new data from user interactions and feedback, keeping them accurate and adaptive to evolving learning patterns. Automated retraining pipelines process new data, update model parameters, and deploy improved versions seamlessly, without interrupting the user experience. These updates gradually enhance grammar correction, chatbot conversational ability, and speech recognition accuracy over time.

By combining Flask, Docker, and cloud-based deployment, the English Learning Web App provides a scalable, real-time, and continuously improving AI-powered learning platform. This robust deployment strategy ensures that learners receive high-quality feedback, adaptive interactions, and an effective environment for mastering English skills.

VII. RESULTS

The results of the Verba-L platform show that integrating AI technologies like speech recognition, grammar correction, and chatbot interaction creates an effective and interactive English learning environment. The system provides real-time feedback and supports all key language skills, helping users improve continuously. Overall, it demonstrates a scalable and efficient solution for enhancing language proficiency.

The proposed system, upon implementation, is expected to yield significant improvements across grammar, speaking, reading, and listening proficiency. Leveraging an advanced grammar-correction model, learners are anticipated to receive immediate feedback on sentence structure, tense consistency, and vocabulary usage, enabling more efficient identification and rectification of errors.

The speech-recognition and conversational AI components are designed to enhance speaking fluency by allowing learners to engage in real-time spoken interactions while receiving automated feedback on pronunciation, pacing, and coherence. For reading and listening tasks, the platform provides IELTS-style comprehension exercises, offering exam-like scenarios in which difficulty levels dynamically adjust to the learner's performance.

Furthermore, the system is expected to generate personalized learning pathways through continuous tracking and analysis of individual strengths and weaknesses, resulting in targeted recommendations and structured practice routines. The integration of visual dashboards, streak indicators, and progress graphs promotes sustained engagement and motivation by presenting clear, actionable insights into performance trends over time.

Collectively, these anticipated outcomes suggest that the system could substantially strengthen English language skills, facilitate exam-oriented preparation, and provide a scalable solution suitable for classrooms, training centers, and remote learning contexts once fully deployed.

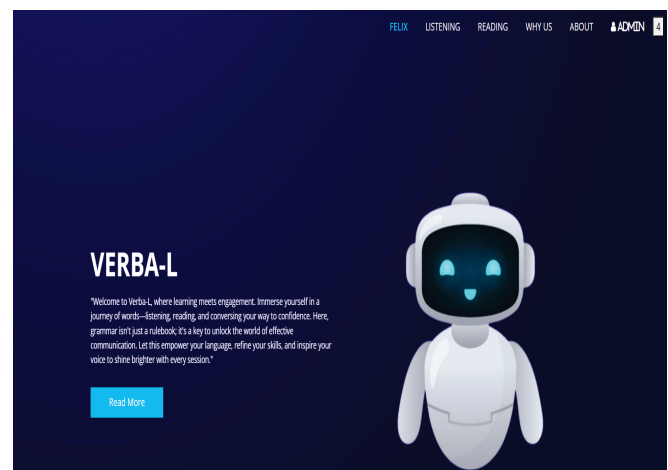


Fig. 5. Home Page

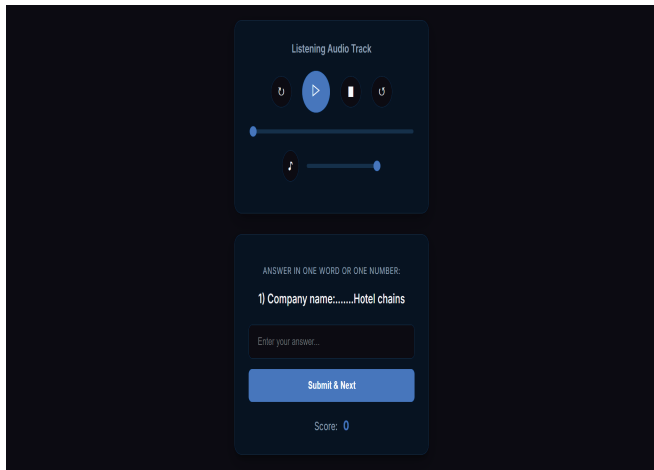


Fig. 6. Listening Module

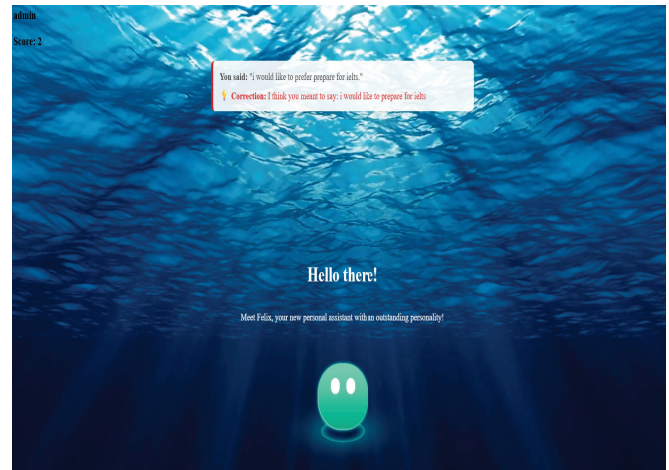


Fig. 8. Chatbot Interaction (Felix)

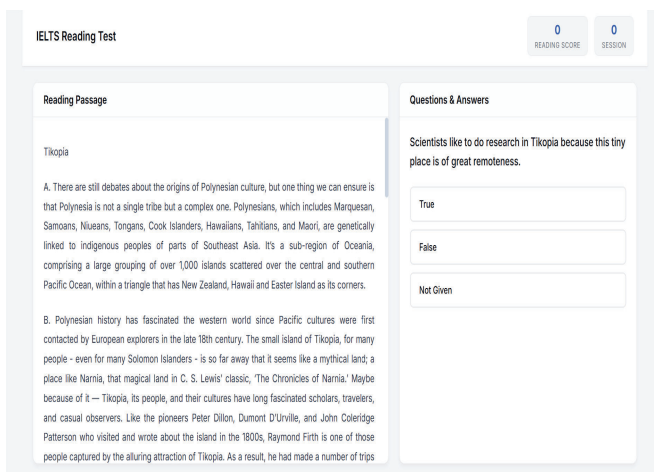


Fig. 7. Reading Module

VIII. DISCUSSION

The proposed system highlights the potential of integrating artificial intelligence into language learning, offering a practical approach for learners to improve grammar, speaking, reading, and listening skills. By combining multiple AI components—such as grammar correction models, speech recognition, conversational dialogue, and comprehension scoring—the system provides a comprehensive framework that goes beyond traditional learning tools.

One of the key advantages of this conceptual platform is its adaptive and personalized nature. The system is designed to analyze individual performance and adjust the difficulty level of exercises accordingly, ensuring that learners of varying proficiency levels receive an appropriate challenge. This adaptiveness encourages consistent engagement and helps learners focus on areas that require improvement, making the learning process more efficient and tailored.

In addition, the modular architecture ensures that each AI component can operate independently while contributing to the

overall learning experience. This design supports scalability and allows future enhancements, such as the integration of more advanced transformer models, additional language modules, or immersive learning techniques.

However, several conceptual challenges must be considered. Accuracy in speech recognition can vary due to accents, pronunciation differences, and background noise. Furthermore, ethical concerns, including data privacy, secure handling of user information, and unbiased model responses, are essential considerations for real-world deployment. Despite these challenges, the system's conceptual framework demonstrates a promising path toward creating an interactive, AI-assisted learning environment suitable for academic institutions, remote learners, and professional training programs.

IX. CONCLUSION

This paper presents a conceptual framework for an AI-driven English learning platform aimed at enhancing grammar, speaking, reading, and listening skills through interactive and adaptive modules. The proposed system combines natural language processing, speech recognition, and personalized feedback mechanisms to offer a structured and engaging learning experience.

Although the platform is presented at a conceptual level, the analysis indicates significant potential for improving learners' language proficiency in both academic and professional contexts. By providing real-time feedback, tailored learning paths, and comprehensive performance tracking, the system addresses common challenges faced by English learners, such as limited access to tutors, delayed feedback, and unstructured learning approaches.

Future work could focus on integrating more advanced AI models, expanding support for multiple languages, and incorporating immersive technologies such as augmented reality or virtual classrooms. Overall, this conceptual platform lays the foundation for a scalable, intelligent, and adaptive learning tool

that can be used by educational institutions, training centers, and individual learners seeking to improve their English skills.

ACKNOWLEDGMENT

Above all, we attribute the successful completion of this project to the grace and generous blessings of the Almighty. We express our sincere gratitude to all those who have extended their valuable guidance, encouragement, and support throughout the course of this work.

The authors would like to thank the Chairman, Dr. Sojan V. Avirachan, the Executive Director, Dr. Jai M. Paul, and the Principal, Dr. Vibin Antony P., of ICCS College of Engineering and Management, for providing a supportive academic environment and continuous encouragement. We extend our special and heartfelt gratitude to Ms. Divya Jose, Head of the Department of Computer Science and Engineering, and Project Guide, for her invaluable guidance, insightful suggestions, and constant support during the conceptualization and development of this research. We also extend our heartfelt gratitude to Ms. Simi Cheriyan, Project Coordinator, for her invaluable guidance.

We also acknowledge our classmates, peers, and department staff for their cooperation, constructive feedback, and valuable discussions that helped refine our ideas. Finally, we thank ICCS College of Engineering and Management for providing the necessary resources, motivation, and collaborative environment that made it possible to develop and present this project concept successfully.

REFERENCES

- [1] A. Rahman and P. Tomy, "Voice Assistant as a Modern Contrivance to Acquire Oral Fluency: An Acoustical and Computational Analysis," *World J. English Lang.*, vol. 13, no. 1, 2023, doi: 10.5430/wjel.v13n1p92
- [2] T. Tiratatri, K. Sukittivarapunt, and A. Pyae, "Designing an LLM-Based IELTS Question Generator, Assessment, and Personalized Training System: Architecture and Research Agenda," *Proc. 2025 22nd International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTICON)*, 2025, doi: 10.1109/ECTI-CON64996.2025.11101665.
- [3] S. Ayothi, A. Kumar, J. Gowthamy, A. S., and G. Sreedhar, "Enhanced AI Voice Assistance using Machine Learning and NLP," *Conference Paper*, 2023, doi:10.1109/STCRS9005.2073.10396893.
- [4] J. Wang, "Speech recognition sensors and artificial intelligence automatic evaluation application in English oral correction system," *Meas.: Sensors*, vol. 32, 2024, Art. no.101070.
- [5] E. Ericsson and S. Johansson, "English speaking practice with conversational AI: Lower secondary students' educational experiences over time," *Comput. Educ. Artif. Intell.*, vol. 5, 2023, Art. no. 100164, doi: 10.1016/j.cacal.2023.100164.
- [6] E. Ericsson and S. Johansson, "English speaking practice with conversational AI: Lower secondary students' educational experiences over time," *Comput. Educ. Artif. Intell.*, vol. 5, 2023, Art. no. 100164, doi: 10.1016/j.cacal.2023.100164.
- [7] D. M. Yves and J. H. Chan, "Measuring Spoken English Proficiency Level Based on IELTS Speaking Test Using Machine Learning Models," *Proc. 2024 IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology (WIAT)*, 2024, doi: 10.1109/WIAT62293.2024.00069.
- [8] Malodia, Suresh, Nazrul Islam, Puneet Kaur, and Amandeep Dhir. "Why do people use Artificial Intelligence (AI)-enabled voice assistants?," *IEEE Transactions on Engineering Management* (2021).
- [9] Raina, Vineet, and Srinath Krishnamurthy. "Natural language processing." In *Building an Effective Data Science Practice*, pp. 63-73. Apress, Berkeley, CA, 2022.
- [10] Balakrishnan, Janarthanan, Yogesh K. Dwivedi, Laurie Hughes, and Frederic Boy. "Enablers and inhibitors of AI-powered voice assistants: a dual-factor approach by integrating the status quo bias and technology acceptance model." *Information Systems Frontiers* (2021):1-22
- [11] R. Sangpal, T. Gawand, S. Vaykar and N. Madhavi, "JARVIS: An interpretation of AIML with integration of gTTS and Python," 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), Kannur, India, 2019, pp. 486-489, doi: 10.1109/ICICICT46008.2019.8993344.
- [12] J. Buddhiraja, Himank, and Nikhil Sharma. "IntelliAssistant-AI based Personal Assistant." In *Proceedings of the International Conference on Innovative Computing and Communication (ICICC)*. 2021.
- [13] Buddhiraja, Himank, and Nikhil Sharma. "IntelliAssistant-AI based Personal Assistant." In *Proceedings of the International Conference on Innovative Computing and Communication (ICICC)*. 2021
- [14] . Nasirian, Farzaneh, Mohsen Ahmadian, and One-Ki Daniel Lee. "AI-based voice assistant systems: Evaluating from the interaction and trust perspectives." (2017).
- [15] A. Kumar, D. Kaur and A. K. Pathak, "VOICE ASSISTANT USING PYTHON," 2022 International Conference on Cyber Resilience (ICCR), Dubai, United Arab Emirates, 2022, pp. 1-4, doi:10.1109/ICCR56254.2022.9995997.
- [16] . M. Soto and S. Allongue, "A semantic approach of virtual worlds interoperability," *Proceedings of IEEE 6th Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises*, Cambridge, MA, USA, 1997, pp. 173-178, doi: 10.1109/ENABL.1997.630810.
- [17] Patil, Jaydeep, Atharva Shewale, Ekta Bhushan, Alister Fernandes, and Rucha Khartadkar. "A Voice Based Assistant Using Google Dialogflow and Machine Learning." *NEW ARCH-INTERNATIONAL JOURNAL OF CONTEMPORARY ARCHITECTURE* 8, no. 2 (2021):1103-1111
- [18] J. Nair, S. Pillai, G. S. Nair and A. T, "Emotion Based Music Playlist Recommendation System using Interactive Chatbot," 2021 6th International Conference on Communication and Electronics Systems (ICES), Coimbatre, India, 2021, pp. 1767-1772, doi:10.1109/ICES51350.2021.9489138.
- [19] J. Nair, S. Pillai, G. S. Nair and A. T, "Emotion Based Music Playlist Recommendation System using Interactive Chatbot," 2021 6th International Conference on Communication and Electronics Systems (ICES), Coimbatre, India, 2021, pp. 1767-1772, doi:10.1109/ICES51350.2021.9489138.
- [20] Devi, N. Vasunthira, and R. Ponnusamy. "A Systematic Survey of Natural Language Processing (NLP) Approaches in Different Systems." (2016): 192-198.
- [21] Bansal, Mohit, and Dr TK Thivakaran. "Analysis of Speech Recognition using Convolutional Neural Network." *Journal of Engineering Sciences* 11, no. 1 (2020): 285-291.
- [22] . Khurana, Diksha, Aditya Koli, Kiran Khatter, and Sukhdev Singh. "Natural language processing: State of the art, current trends and challenges." *Multimedia tools and applications* (2022): 1-32
- [23] . Kalyanathaya, Krishna Prakash, D. Akila, and P. Rajesh. "Advances in natural language processing—a survey of current research trends, development tools and industry applications." *International Journal of Recent Technology and Engineering* 7, no. 5C (2019): 199-20
- [24] GAUR, VISHVAKETAN. "DESKTOP ASSISTANT PROJECT BASED ON VOICE RECOGNITION AND FACE DETECTION." (2021).
- [25] Sharif, Khairunisa, and Bastian Tenbergen. "Smart home voice assistants: a literature survey of user privacy and security vulnerabilities." *Complex Systems Informatics and Modeling Quarterly* 24 (2020): 15-30.
- [26] Bartle, Vince, Janice Lyu, Freesoul El Shabazz-Thompson, Yunmin Oh, Angela Anqi Chen, Yu-Jan Chang, Kenneth Holstein, and Nicola Dell. "'A Second Voice': Investigating Opportunities and Challenges for Interactive Voice Assistants to Support Home Health Aides." In *CHI Conference on Human Factors in Computing Systems*, pp. 1-17. 2022.
- [27] Terzopoulos, George, and Maya Satratzemi. "Voice assistants and smart speakers in everyday life and in education." *Informatics in Education* 19, no. 3 (2020): 473-490
- [28] Pal, Debajyoti, Chonlameth Arpikanondt, Suree Funilkul, and Vijayakumar Varadarajan. "User experience with smart voice assistants: the accent perspective." In *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pp. 1-6. IEEE, 2019.