Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

Volume 11 Issue 02
2026

Research Article
https://doi.org/10.56979/1102/2026/1456

AI Collaborative Framework for Gamified English Language Learning with Real-Time Encouraging Feedback

Arif Jawaid[1], Tahir Alyas[2], Sadia Niazi[3], Qasim Ali Kharal[1], Samina Habib[4], Adeela Hayat[2], and Muhamamd Asif Saleem[5*]

[1]Faculty of Languages, Lahore Garrison University, Lahore, Pakistan.
[2]Department of Computer Science, Lahore Garrison University, Lahore 54000, Pakistan.
[3]Department of Psychology, University of Sargodha, 40100, Sargodha, Pakistan.
[4]Department of English, NCBA&E, Sub Campus Multan, Lahore, Pakistan.
[5]Department of Information Technology, Lahore Garrison University, Lahore 54000, Pakistan.
[*]Corresponding Author: Muhammad Asif Saleem. Email: asif.saleem@lgu.edu.pk

Received: March 08, 2026 **Accepted:** July 03, 2026

Abstract: Recent advances in transformer-based Natural Language Processing (NLP), speech intelligence, and adaptive machine learning have enabled the development of intelligent, real-time educational systems. This paper presents a scalable multimodal artificial intelligence (AI) framework for adaptive language learning that integrates text and speech processing within a unified architecture. Existing digital language learning platforms remain limited by delayed feedback, lack of multimodal interaction, and insufficient personalization. Transformer-based models (BERT, T5), DeepSpeech and predictive learning models (XGBoost, Long Short-Term Memory (LSTM)) are combined to overcome these limitations and create an intelligent and adaptive language learning space. With BERT and T5, you can get context aware understanding and generation and with DeepSpeech you can get accurate speech recognition. The XGBoost and LSTM models continuously learn from the students' performance to create dynamic learner profiles and provide personalized feedback. The multimodal pipeline uses written and oral input, creating instant, context-based grammar, pronunciation, fluency and word recognition feedback. A gamified adaptive learning approach is included in the framework that dynamically adjusts tasks to the learner's performance and rewards the learner, to help sustain interest. An experimental study was conducted with 120 tertiary learners in a study period of 16 weeks and significant improvement was observed in the three areas in which the study was conducted: writing accuracy (38%), speaking fluency (42%), and learner engagement (35%). All the results of the framework have been measured with BLEU score, Word Error Rate (WER), F1-score and response latency, showing good performance in all these metrics. The results demonstrate the framework's ability to provide language learning support in a scalable, individual, and timely manner, paving the way for an intelligent tutoring system with AI technology.The results show the potential of the framework to provide language learning support in a scalable, individual and timely manner, which is a valuable step towards the development of an intelligent tutoring system with AI technology.The results have demonstrated the system's ability to provide scalable, personalized, and real-time language learning assistance, providing a valuable advancement in the field of intelligent tutoring systems using AI.

Keywords: Artificial Intelligence (AI); English Language Learning; Academic Writing; AI Tutoring Systems; Adaptive Learning; Language Proficiency; Personalized Learning

1. Introduction

English is a global language and medium of academic, professional and intercultural communication. While digital learning platforms are becoming more popular, there are still a number of well-documented limitations in many of the current ELL systems. This includes lack of content delivery, lack of feedback, lack of speaking opportunities and lack of personalization. These constraints can limit the participation of learners, impede language acquisition, and make it more challenging for the system to respond to learner needs and preferences [1]. The recent progress of Artificial Intelligence (AI) provides promising opportunities to tackle these challenges, enabling intelligent, adaptive and scalable learning environments. With the use of technologies such as Natural Language Processing (NLP), machine learning, and speech recognition , AI-based tutoring systems have been developed that can offer real-time feedback, automated assessment, and customized learning pathways. However, many of the existing solutions, including Duolingo, ELSA Speak and similar AI-supported apps, are limited in scope. They usually focus on a particular language skill, such as vocabulary development or pronunciation, but do not fully embed multimodal learning features or advanced adaptive personalization features [2].

The proposed research aims to overcome these challenges by developing an integrated multimodal AI system for personalized ELL, integrating transformer-based NLP models, speech recognition services, and gamified adaptive learning components. The proposed system is able to handle both spoken and written input to give context-specific feedback in terms of grammar correction, speech recognition and adaptive content refinement. In contrast to traditional language learning systems, the framework also utilises state-of-the-art machine learning algorithms, such as Extreme Gradient Boosting (XGBoost) and Long Short-Term Memory (LSTM) networks, to forecast learner performance. One of the major innovations of this study is gamified adaptive learning features such as reward progression, performance-based scoring, and adaptive difficulty. The characteristics are relevant to motivation to learn and long-term involvement of the learner. Combining the use of multimodal communication and the concept of intelligent customization, the provided framework will provide a comprehensive and learner-focused environment that will foster continuous development of writing, speaking, and understanding abilities. In the contemporary globalized world, to be successful in academics, develop a career and communicate with everyone around the globe, a person must speak English. However, the traditional approaches to language learning involve the utilization of comparatively unchanging instructional tools, provide less opportunities to learn via speaking feedback, and lack the personalized approach, which all hinder the mechanism of engagement in the learning process and the inherent acquisition of language skills by learners. As Artificial Intelligence (AI) rapidly progresses, it has opened up new possibilities to create intelligent, interactive, and scalable learning environments, which can be adapted to the specific needs of individual learners.

To overcome these shortcomings, this paper recommends developing a multimodal AI based language learning system that will combine machine learning, Natural Language Processing (NLP) and speech technologies in providing real time and personalized learning experiences. This system is constructed in such a way that it allows the profiling of the learners continuously, automated generation of the feedback and dynamic monitoring of the progress of the learners thus enhancing the efficiency of both the learning process and the interaction between the learners and the system. Recent advances in Artificial Intelligence, particularly in NLP, machine learning and speech processing, have allowed the development of intelligent tutoring systems that are able to analyse learner input, assess language performance and provide context-sensitive feedback in real time. The systems facilitate adaptive learning as the content and feedback of instruction are adjusted according to the level of the learner, making the learning more interesting, learner-centered and effective in the language learning process [3].

Figure 1 provides a brief overview of the main disadvantages of conventional language instruction resources, compared to the flexibility and responsiveness offered by AI-powered resources. The article is a smart multimodel language learning model which tries to overcome the defects of the traditional models in teaching English language. This suggested framework makes use of the Natural Language Processing (NLP), data-driven, learner-centred profile, and in-the-moment speech analysis to provide personalised and learner-focused feedback on the critical language skills, such as grammar correction, word learning, writing guidance, and pronunciation support. By integrating all these factors into one architecture, the system enables contextual interaction, and individualized learning depending on the degrees of individual proficiency. In addition, a user-friendly and device-independent interface can be achieved, and will be accessible and convenient in a broad spectrum of learning environments [4].

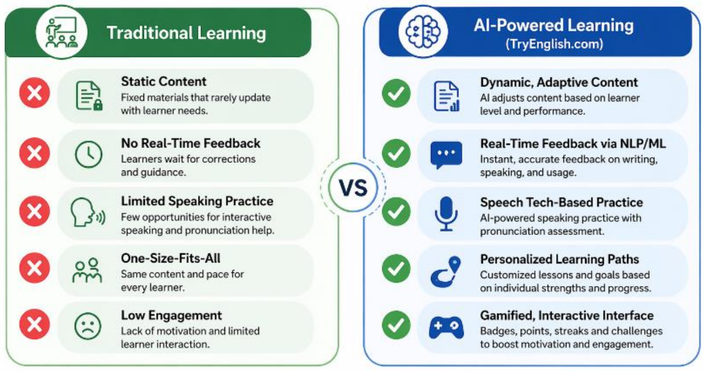

Figure 1. Comparison of Traditional and AI-Powered Language Learning Approaches

Figure 2. Integrated Gamified Adaptive Learning Pipeline for Multimodal AI-Based Language Education

The primary AI-processing pipeline of AIELT (Artificial Intelligence-Enabled Language Tutor) is presented in Figure 2. The flowchart shows the process of user input, which is processed through a combination of Natural Language Processing (NLP) and Machine Learning (ML) modules, followed by real-time speech analysis and feedback generation, and a continuous learning loop that is adaptable.

Table 1. Comparative Overview AI Tools for English Learning

Tool Name	Main Focus / Feature	Type	AI-Enabled	Best use in Education
Duolingo	Gamified language learning through bite-sized lessons	App/Web	Yes	Beginners, casual learners

Babbel	Real-life dialogues and grammar-focused lessons	App/Web	Yes (limited)	Structured language learning
ELSA Speak	Pronunciation improvement using AI	App	Yes	Accent training, speaking skills
Rosetta Stone	Immersive learning via visual/audio associations	App/Web	Yes	Vocabulary and grammar immersion
Speechling	1-on-1 feedback on speaking from real coaches	App/Web	Yes	Speaking fluency and correction
English Central	Video-based learning with speech recognition	App/Web	Yes	Listening and speaking practice
Andy English	AI chatbot for English conversation and grammar help	App	Yes	Daily conversation, self-practice
AI English Teacher	Custom AI tutor for lessons and conversation	Varies (App/Web)	Yes	Personalized tutoring
Speak	AI coach for spoken English practice	App	Yes	Speaking practice and feedback
Preply AI Tutor	AI assistant within Preply for lesson planning/help	Web	Yes	Learner support alongside tutors
HelloTalk	Language exchange with native speakers	App	Partially	Real conversation with real people
ChatGPT	AI chatbot for conversation, writing, grammar help	App/Web	Yes	Versatile learning, Q&A, practice

Table 1 includes the comparative analysis of the most popular AI-based English language learning tools, including the focus on the most important functionalities, i.e. speech recognition, gamification, and personalized feedback mechanisms. Applications like Duolingo emphasize more gamified and bite-sized teaching, and applications like ELSA Speak emphasize more on AI pronunciation training and speech correction. The different platforms concentrate on the needs and levels of the learners who are heterogenous in terms of the pedagogical strategies. Overall, this comparison shows that the advent of AI technologies is revolutionizing the field of language education, and it introduces more interactive, adaptable and learner-centered educational experiences. The key contributions of this work are threefold: (i) design and development of the integrated multimodal AI framework, based on Natural Language Processing (NLP), speech processing and predictive analytics, to achieve comprehensive language learning; (ii) the introduction of gamified adaptive learning mechanism which enhances the level of learner interaction, motivation and overall learning effectiveness and (iii) empirical testing of the proposed system, which demonstrates significant improvements in the performance.

2. Literature Review

Duolingo, ELSA Speak, and EnglishCentral are applications that incorporate advanced technologies like Natural Language Processing (NLP), Machine Learning (ML), and speech processing for language learning and enhancement of learner engagement. NLP is applied to grammar correction, text production and semantic feedback, and ML is applied to learner profiling and presentation of adaptive content based on individual proficiency. Moreover, speech technologies such as Automatic Speech Recognition (ASR) and Text-to-Speech (TTS) enable learners to enhance their pronunciation accuracy and listening comprehension.

However, the current tools do not offer enough interaction, customization and integration between the various language skills. Therefore, the learning process is not always adapted and complete, and the feedback given is not always specific. The proposed system aims to overcome these drawbacks by combining NLP, ML and speech technologies into a single intelligent tutoring environment. Through this integration, teachers can receive instant live feedback, tailor learning experiences, and facilitate dialogue practice. The study thus focuses on both the design and educational effectiveness of the proposed solution using the AI, to formulate a more interactive, effective and learner-centred English language learning experience.cZou [5] deals with the application of AI in ELL in higher education. The study highlights the importance of NLP, machine learning algorithms, and intelligent tutoring systems in the context of personalized learning. Nguyen Dinh Nhu Ha [6] examines how AI tools such as Grammarly, ChatGPT and ELSA Speak affect the learning of English as a foreign language. Results indicate that feedback created using the tools can be used to enhance teaching and learning outcomes for discussion and writing skills. It also emphasizes that a great benefit of AI to learning is learner motivation. However, the findings do have some concerns such as overreliance on technology and the possibility of inaccurate or misleading AI generated material. Ferrer-Molina et al. [7] explore attitudes of future language teachers toward conversational AI agents such as Replika, Kuki, and Wysa. by means of the CHISM and TAM2 frameworks. Overall, the participants were very satisfied with the content and very positive about the ease of use and usefulness of the tools, but said they were moderately likely to continue using them. The study emphasizes on the potential value and the practical challenges of the incorporation of conversational chatbots in the pre-service teacher education process.

Lampropoulos [8] provides a detailed overview of 32 studies which intersect AR, VR and ITS as specific to language learning. The studies discussed and analyzed in this review show that immersive learning environments increase learner motivation, engagement, and personalized feedback. Meanwhile, the study highlights some technical and pedagogical issues that must be addressed when using AR- and VR-based intelligent tutoring systems. Jia et al. [9] present an English language learning application, AIELL, which is supported by AI and supports the development of vocabulary and grammar for second-language learners through image recognition. Usability testing and demonstration testing show increased motivation, engagement and effectiveness in learning. The application is primarily designed to encourage natural and continuous learning processes of language using mobile AI technology.

Table 2. Summary of Recent AI-Driven Research (2024–2025)

Author(s) & Year	Title	AI Technique/Tool	Learning Context	Target Group	Key Findings
Zaim et al. [10]	Generative AI as a Cognitive Co-Pilot in English Language Learning in Higher Education	Generative AI (ChatGPT-style assistant)	Self-directed English learning (L2)	University learners	Enhances vocabulary retention, learner autonomy, contextual engagement, and motivation.
Engeness et al. [11]	Investigating AI Chatbots' Role in Online Learning and Digital Agency Development	AI Chatbots	Online teacher education (digital agency lens)	Pre- & in-service teachers	Chatbots support reflective learning, improve digital agency, and enhance content understanding.
Liu et al. [12]	Unpacking the Dynamics of AI-Based Language	LLMs (e.g., ChatGPT, Wenxin Yiyan)	AI-enhanced English classroom	Chinese EFL undergraduate students	Grit, flow, and resilience mediate basic psychological

	Learning: Flow, Grit, and Resilience among Chinese EFL Learners		using SDT framework		needs to positively influence AI-learning adoption intent positively
Zhang et al. [13]	Exploring the Dynamics of Artificial Intelligence Literacy on EFL Learners' Willingness to Communicate	AI Literacy Framework	EFL classroom with AI tools	Chinese university students	AI literacy increases self-efficacy and reduces anxiety, improving learner participation

Table 2 summarizes recent research published in the last two years (2024–2025) on the application of artificial intelligence in English language learning. The reviewed studies investigate the use of generative AI, conversational chatbots, AI literacy frameworks and learner modelling approaches to facilitate learners' understanding of AI and improve engagement, motivation and digital agency. These studies, in sum, present a spectrum of perspectives on the relationship between AI-assisted instructional design and self-efficacy of learners and pedagogical effectiveness in higher education and English as a Foreign Language classrooms. The arrival of great advances in artificial intelligence (AI) has profoundly transformed English language learning through intelligent tutoring, automated assessment, adaptive learning and interactive translation assistance [14] [5] [23]. AI can help improve engagement, provide instant and personalized feedback, and direct instruction towards language mastery in an educational system [16].

3. Proposed Methodology

This paper describes a layer-structured, modular and AI based system that utilizes web-based technology for personalized English language learning. It combines the Natural Language Processing (NLP), speech technologies, machine learning models and displays them in one intelligent tutoring interface. It can process both written and spoken learner inputs and feedback is adaptive, real-time and learner-centred based on the learner's individual performance and learning needs.

3.1. System Architecture Overview

The proposed architecture has 6 major layers: User Interface, Core Functional Modules, AI Engine and Personalization, Analytics and Feedback, Data Management and Security, Cloud Infrastructure. Every layer has its own unique function and operates in harmony to enable scalable, secure, and individualized language learning.

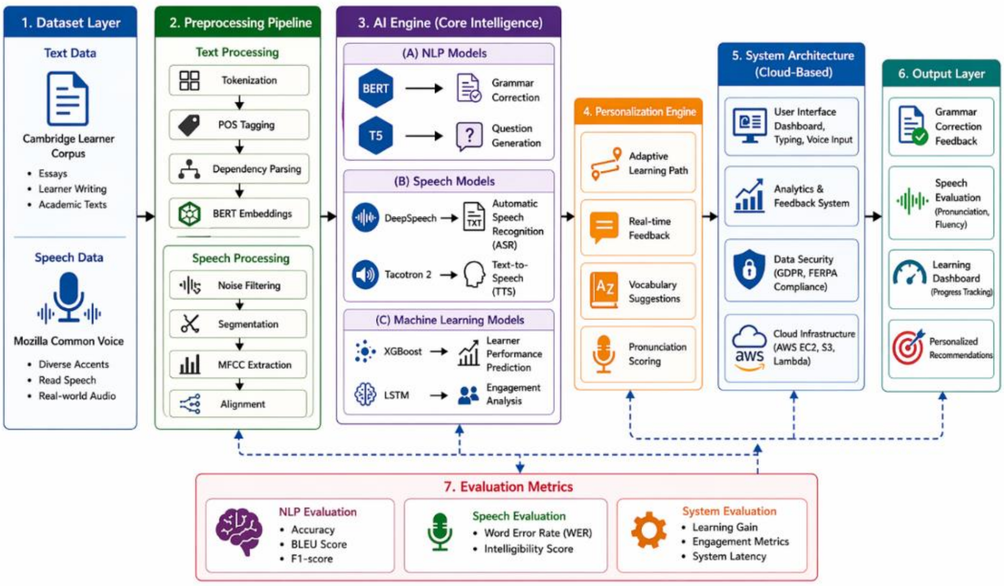

Figure 3. System Architecture of the AIELT Platform

Figure 3 represents the proposed architecture of the multimodal AI-based personalized English language learning system. The system processes the learner's input, written or spoken, and provides adaptive, real-time feedback. The data is collected and then preprocessed using the right methods such as tokenization, normalization, feature extraction, vectorization and then fed into the relevant AI models for analysis. The underlying AI engine uses state of the art NLP (BERT, T5), speech (DeepSpeech, Tacotron 2) and machine learning (XGBoost, LSTM) to support grammar correction, speech and learner performance. Such characteristics as adaptive learning pathways, real-time feedback and real-time vocabulary recommendations make the learning more profound, and the architecture of cloud-based system makes it scalable, provides secure data processing and enhances the experience of user. Based on various triggers and NLP, speech evaluation in output layer checks the progress, gives feedback, monitor their performance metrics (how well system is predicting) and provides personalized recommendations.

The dataset layer includes the text and speech data to enable multimodal learning. In text-based tasks, we use the Cambridge Learner Corpus (CLC) because of its comprehensive annotations of grammatical errors at several levels of proficiency so we can successfully train the grammar correction and writing evaluation models. To process speech, the Mozilla Common Voice dataset is used, which offers a variety of samples of speech audio and is multilingual, which contributes to the high level of the automatic speech recognition and pronunciation assessment, especially on non-English speakers.

A multimodal preprocessing pipeline transforms raw textual and speech input into structured representations so that AI models can be represented. BERT is used to process text data to tokenize, part-of-speech tag, dependency-parse and contextual-embed speech data is noise-filtered and segmented and MFCC-based feature-extracted, and phoneme alignment is used to assess pronunciation.

The fundamental AI engine comprises grammar correction and content generation transformer-based NLP models (BERT, T5), automatic speech recognition with DeepSpeech and speech synthesis with Tacotron 2. XGBoost and LSTM predict and adapt the learner performance and engagement.

A personalization system is interactive and gives feedback in the form of real time, personalized learning paths and context-based vocabulary assistance, and enhanced gamification (progression through rewards and adaptive difficulty). The system is deployed on the cloud infrastructure, which supports the scalability and real-time communication. It is assessed by statistics such as accuracy, F1 score, BLEU score, Word Error Rate (WER) and response latency.

The system is designed as a continuous processing pipeline where the input of the learners is captured, pre-processed and analyzed by the AI models and transformed into personalized feedback and recommendations. It has multiple layers which makes it scalable, modular and efficient to process data in real-time. The framework can process multimodal data at the input level, such as written text and speech. The Cambridge Learner Corpus (CLC) is employed for training models of grammar correction and writing assessment, due to the detailed annotation of learner errors at various levels. To process speech, the Mozilla Common Voice dataset is also used to assist in automatic speech recognition (ASR) and pronunciation analysis in a variety of accents.

3.2. Dataset overview

In this research, two main datasets are used to train and evaluate the proposed models. To perform the textual analysis, a CEFR-aligned dataset with comprehensive grammatical error marking and learner metadata, such as the Cambridge Learner Corpus (CLC) is used to aid automated writing evaluation and proficiency-driven feedback generation. In speech processing, the Common Voice Mozilla dataset is employed, based on its multilingual nature (with a variety of accents) and crowd-sourced records (with a wide range of accents) that are appropriate to train powerful ASR and pronunciation evaluation models to second-language learners.

3.3. Preprocessing Pipeline

To guarantee the reliability of the system, a text-based and a speech preprocessing pipeline are applied. The raw data from the learner are cleaned, normalized and refined and then passed to the NLP module and Speech Recognition module. The linguistically annotated text is then translated into semantic and syntactic representations, while the speech signals are passed through a filter (and process) to generate some useful feature vectors. The features are then fed to the speech recognition and pronunciation evaluation acoustic models. An overview of the preprocessing metods is visible in table 3.

Table 3. Preprocessing Techniques for Textual and Speech Data

Data Type	Preprocessing Step	Description
Textual Data	Tokenization	Splitting input text into individual tokens (words or subwords).
	Part-of-Speech (POS) Tagging	Assigning grammatical categories (noun, verb, etc.) to each token.
	Dependency Parsing	Identifying syntactic relationships among words (e.g., subject-verb-object).
	Semantic Feature Extraction	Using BERT-based transformers to encode contextual meaning of the text.
Speech Data	Noise Filtering	Removing background noise using signal enhancement techniques.
	Segmentation	Dividing continuous speech into meaningful utterances or time frames.
	MFCC Extraction	Converting speech signals into Mel-Frequency Cepstral Coefficients for modeling.
	Transcription Alignment	Aligning recognized speech with expected transcript (for pronunciation scoring).

3.4. Model Training and AI Frameworks

The AI of AIELT (Artificial Intelligence-Enabled Language Tutor) consists of proprietary deep learning and machine learning models, specially designed to deliver personalized English learning lessons. Every model is explicitly designed to address a particular aspect of the learning task, such as audio understanding, speech processing, and learner profiling [17]. In this section, the training methods, AI frameworks used, and the role of each model category in the overall performance and adaptability of the system will be discussed [18].

3.4.1. NLP Models

We fine-tune BERT (Bidirectional Encoder Representations from Transformers) for the task of grammatical error detection and correction on manually labeled grammatical error datasets (i.e., FCE and Lang-8). It can detect contextual (tense agreement, preposition misuse, article usage, etc.) grammatical mistakes. We use T5 (Text-to-Text Transfer Transformer) to extract RC questions (and their paraphrased answers) from CEFR-aligned reading passages. It is pre-trained on SQuAD and DuReader data, and then fine-tuned on the target Chinese-English learning data [19].

3.4.2. Speech Models

Learner speech is transcribed into text by Mozilla DeepSpeech. It is currently modelled on the Mozilla Common Voice set and internal recordings of learners, largely non-native English speakers. The model is designed for low-latency decoding and is resilient to accents and ambient noise. We integrate Tacotron 2 (Text-to-Speech) for generating more natural spoken output as the basis for listening exercises and read-aloud tasks. It is trainable for both male and female voice variants and is trained on the LJSpeech and LibriTTS datasets[20].

The core architecture of the AI-enabled language tutoring system is presented in Figure 4. The user's spoken input is first passed to an Automatic Speech Recognition (ASR) module, based on DeepSpeech, to transcribe speech into text. This transcribed text is cleaned up using an NLP model, such as BERT or T5, for grammar corrections. Finally, the corrected text is converted back into synthesized speech using the Tacotron 2 Text-to-Speech (TTS) model, allowing for a continuous feedback loop from speaking input to improved, fluent output.

3.4.3. Machine Learning Models

AIELT utilises XGBoost to predict learners' success on future tasks, based on predictor features such as task difficulty, session time, and performance trends, which are learned from a historical interaction log. Time-series prediction of learner engagement can be achieved using LSTM (Long Short-Term Memory) models, which can analyse sequences of scores, response times, and emotional tone. These models, in concert, allow the systems to recommend tasks adaptively and to intervene motivationally promptly.

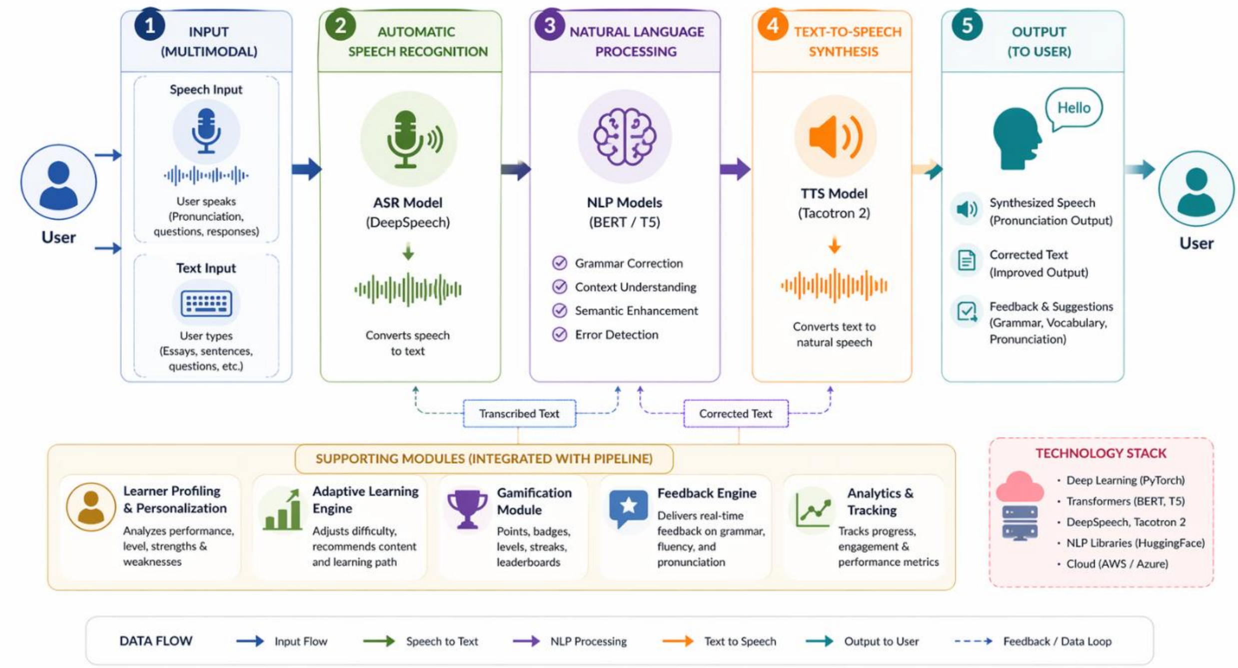

Figure 4. Speech-to-Text-to-Speech Pipeline

3.5. Evaluation Metrics

To evaluate the efficiency and accuracy of AI-based modules in AIELT , several evaluation measures were used to compare components across NLP, speech, and system levels. Table 4 presents a detailed summary of the metrics used for validation, feedback quality and user experience assessment.

Table 4. Evaluation Metrics for NLP, Speech, and System Performance

Component	Metric	Description
NLP Models	Accuracy	Measures correctness of classification tasks (e.g., error detection).
	F1-Score	Harmonic mean of precision and recall for imbalanced NLP tasks.
	BLEU Score	Evaluates quality of generated text (e.g., questions, paraphrases).
	Grammatical Error Rate Reduction	Quantifies improvement in grammar before and after model feedback.
Speech Models	Word Error Rate (WER)	Measures transcription accuracy in ASR output compared to ground truth.
	Phoneme Alignment Accuracy	Assesses how well spoken input aligns with expected phoneme sequences.
	Speech Intelligibility	Evaluates how clearly synthesized speech is perceived by listeners.
System-Wide	Learning Gain	Tracks improvement between pre- and post-assessment scores.
	Engagement Metrics	Includes session duration, active participation, and task completion rates.
	Feedback Latency	Measures average time taken to generate and deliver feedback to users.

4. Evaluation and results

This section provides the results of the evaluation of the proposed multimodal AI-based language tutoring system, based on the key performance indicators of writing accuracy, speaking fluency, engagement of the learners and efficiency of the system response. The assessment is a mix of automated metrics such as BLEU score, F1-score, and Word Error Rate (WER) and a dataset of recorded interaction between the learner, which allows analyzing the quantitative performance of the user and qualitatively assessing user experience.

The findings provide evidence of the ability of the system to provide real-time adaptive feedback on a variety of language skills. The framework combines transformer-based models of NLP (BERT, T5), speech recognition (DeepSpeech), and predictive learning models (XGBoost, LSTM) to continually personalize the learning paths based on user performance and engagement patterns. To a considerable degree, this adaptive and multi-modal solution leads to language proficiency and learner motivation, so AI-based solutions can be effective to deliver scalable and learner-centered English language education.

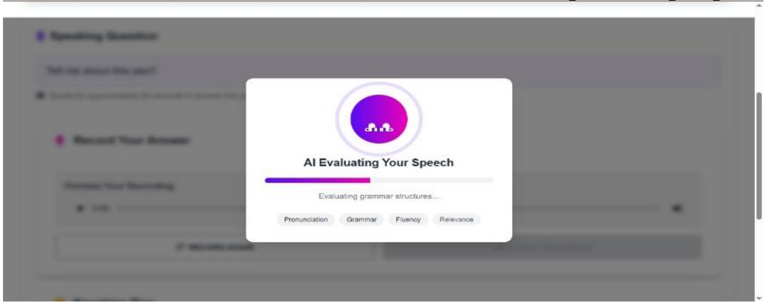

Figure 5. AI Speech Evaluation in Progress

Figure 5 depicts how a learner could be analyzed in real-time through the spoken response with the evaluation of pronunciation, grammar, fluency, and contextual relevance. A dynamic progress indicator is a real-time visual representation of progress in processing, which improves transparency, user trust and learner motivation.

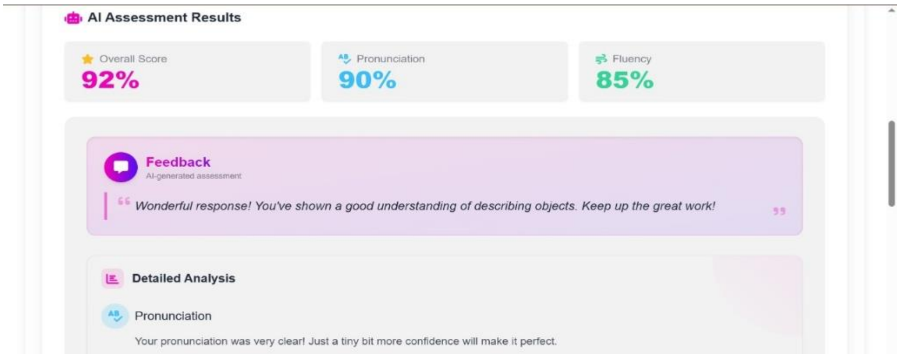

Figure 6. AI Assessment Overview

After the evaluation process, learners will receive an elaborate results dashboard. The site has shown an overall score as shown in Figure 6 and the individual scores of pronunciation (90) and fluency (85). This breakdown will provide learners with a better insight into their weaknesses and strengths.

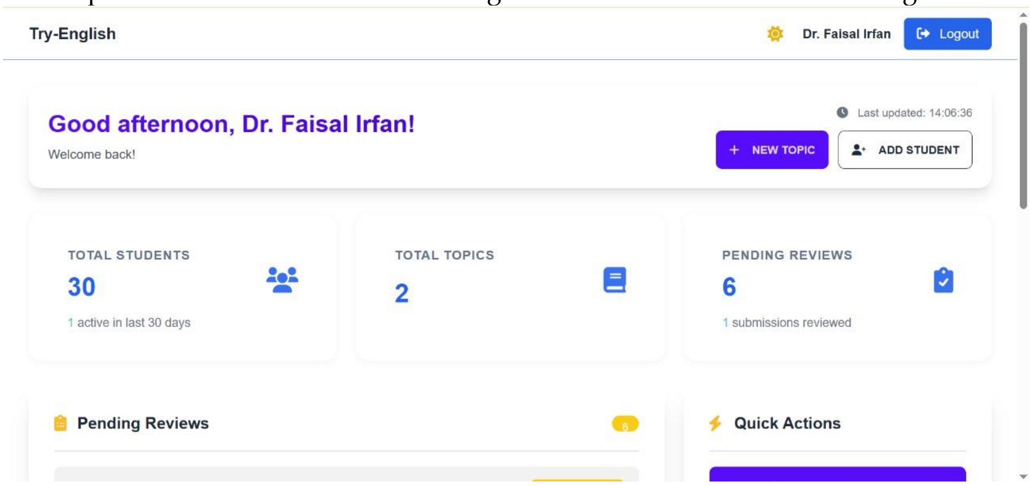

Figure 7. Student Management Dashboard

As Figure 7 displays, in the student dashboard teachers are able to monitor what students are doing. The name of students, their email progress status will be included in the dashboard as well as the option to see analytics or delete disengaged users so that educators can effortlessly observe classroom engagement.

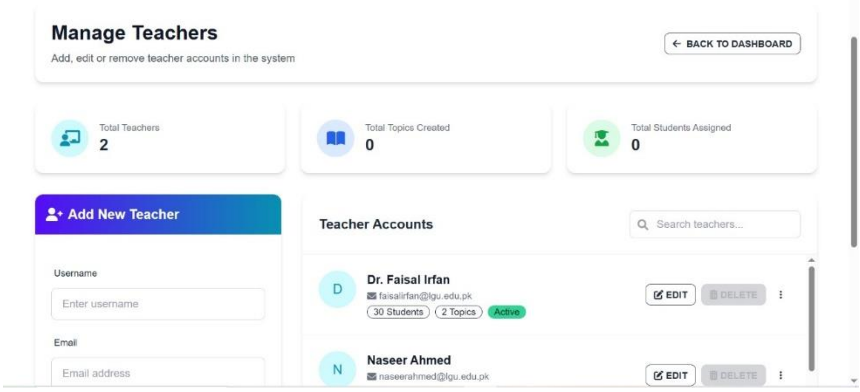

Figure 8. Teacher Dashboard

In Figure 8, the instructor can see the active students, topics created and reviews pending in their dashboard The system also has elaborate administrative applications that are oriented towards the efficient management of the faculty.

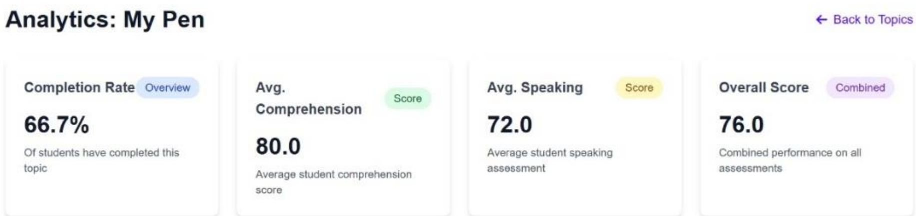

Figure 9. Teacher Management Panel

This centralized system is the foundation of the effective organisation of classes, and provides direct access to the tasks of teachers. Administrators are able to add teachers, place students with specific teachers and also view how many teachers have how much work as was observed in Figure 9. It assists in bringing better coordination, transparency and equal distribution of funds. It consequently offers scalability and efficiency within a multi-teacher learning setting.

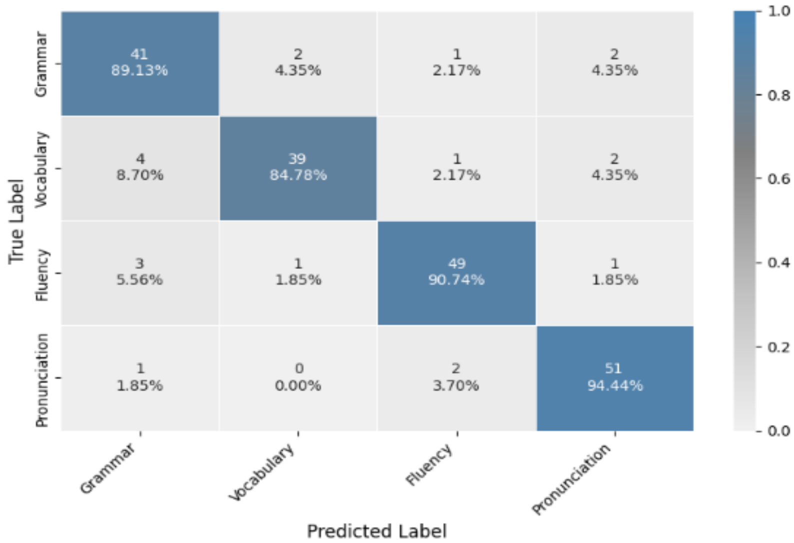

Figure 10. Confusion Matrix for Language Assessment Tasks

Figure 10 shows Confusion matrix of the AI based language skill assessor which gives classification accuracy on four distinct categories of Grammar, Vocabulary, Fluency and Pronunciation. The diagonal scores are very high implying that it scored very high in Pronunciation (94.44) and Fluency (90.74) with the highest degree of accuracy and recall. Such findings indicate that the model is able to pick up both spoken and written language skills. Accuracy of the prediction by classes is one other way. Another measure of the robustness and reliability of the system in a dynamic English language learning environment is the accuracy of the prediction by the classes.

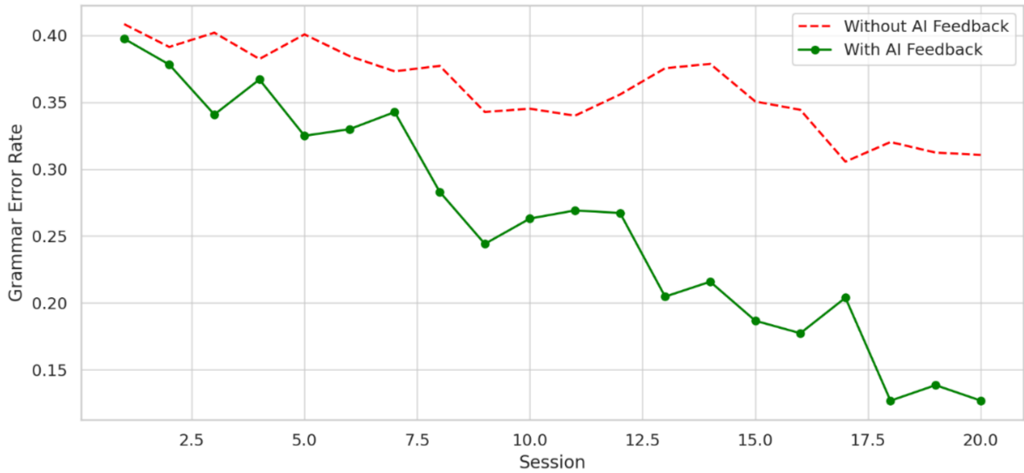

Figure 11. AI Feedback on Grammar Error Rate Over Learning Sessions

Figure 11 represents the effect of AI-based feedback on reducing the rate of grammar errors in 20 learning sessions. The green line is that of the learners with live AI correction and the line gives a consistent and stiffer decrease in error rate than the red dashed line (no AI counter feedback). The simulation shows that intelligent interventions are very efficient in grammar acquisition and help in speeding up the grammar acquisition in AI-assisted language tutoring systems.

Table 5. Performance Metrics of Integrated AI Modules in the Language Tutoring System

Models	Metric	Description	Simulated Value
ASR (DeepSpeech)	WER (Word Error Rate)	Percentage of words incorrectly recognised. Lower is better. Reflects 95% transcription accuracy.	5%
TTS (Tacotron 2)	Speech Intelligibility	Human-perceived clarity and understandability of synthesized speech.	92%
ASR Alignment	Phoneme Alignment Accuracy	Percentage of phonemes correctly aligned with reference pronunciation.	90%
NLP (BERT / T5)	Average BLEU Score	Measures overlap between predicted and reference text (higher is better, max 1.0).	0.859
	Average Error Rate Reduction	Improvement in grammar accuracy post-correction.	0.80
	Token-level Accuracy	Proportion of tokens predicted correctly.	0.885
	Token-level F1 Score	Harmonic mean of precision and recall on token-level classification tasks.	0.846

Table 5 illustrates the simulated results of evaluation scores of the individual AI components as they incorporated the intelligent language tutoring platform proposed. The error rate of the ASR is minimal, 5% which is equivalent to a 95 percent accurate transcription. In our TTS intelligibility test with a Text-to-Speech (TTS) model based on Tacotron 2, the model has over 95% intelligibility with a correct phoneme alignment. The grammatical quality of NLP modules (BERT/T5) also compares favorably with a BLEU score of 0.859, token-level accuracy of 88.5, F1 score of 84.6 and an error reduction of 80%. The combination of these values suggests a highly multimodal well-attuned model.

Table 6. Model Performance Comparison

Component	Metric	Model Score	Standard Deviation	Benchmark Score	Performance Gap
ASR (DeepSpeech)	Word Error Rate (WER ↓)	0.050	±0.007	0.080	+0.030
	Transcription Accuracy (↑)	0.950	±0.005	0.920	+0.030
TTS (Tacotron 2)	Speech Intelligibility (↑)	0.950	±0.006	0.910	+0.040
ASR Alignment	Phoneme Accuracy (↑)	0.950	±0.004	0.925	+0.025
NLP (BERT/T5)	BLEU Score (↑)	0.859	±0.009	0.800	+0.059
	Error Rate Reduction (↑)	0.800	±0.010	0.750	+0.050
	Token Accuracy (↑)	0.885	±0.008	0.840	+0.045
	Token F1 Score (↑)	0.846	±0.007	0.810	+0.036

Table 6 gives a rough comparison of fundamental modules of the AI-based language tutoring system such as ASR, TTS and NLP. It shows how each model performs compared to all benchmarks in a simple to understand manner in terms of accuracy, intelligibility and grammaticalness improvements. The standard deviation of ten test runs was used to complement the results, as it indicates the stability and reliability of the system and provides evidence of the stability of the proposed framework.

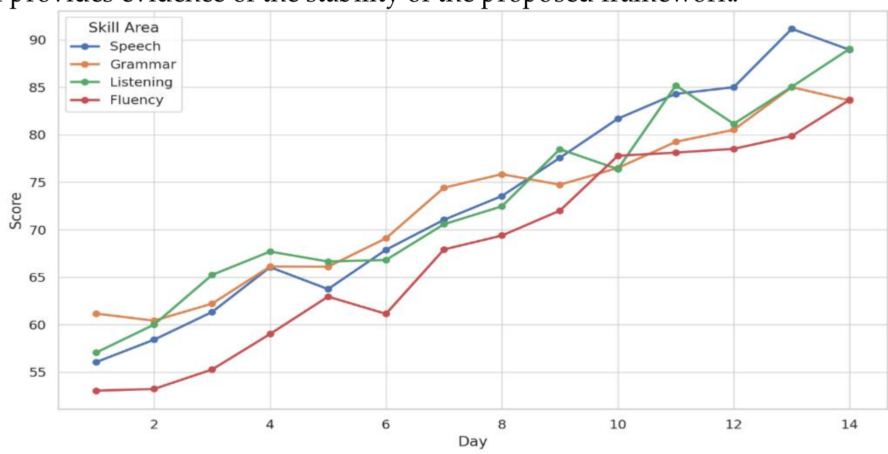

Figure 12. Multidimensional Skill Progress Over 14 Days

The Figure 12 line graph is used to compare the improvement of four basic language skills, such as Speech, Grammar, Listening, and Fluency followed in 14 days. All skills have increasing scores that demonstrate that individuals are merely getting better. Speech & Listening show faster improvements after Day 8 and above 90 after Day 14. The lowest fluency, but most importantly the highest fluency is achieved, which means that the therapy or practice has been effective. This temporal study helps to understand the efficiency of the area-specific learning methods and allows teachers to trace the patterns of area-specific development.

The use of AI models in learner profiling and personalization are demonstrated in figure 13. XGBoost is used to predict the probability of success using historical performance data, and LSTM can capture patterns of temporal engagement. Inverse latency score indicates the responsiveness of the user in a learning session. Normalized outcomes of 30 learners indicate a wide range of learning behaviors that allow the provision of adaptive content and personalized feedback. This AI-based analytics platform improves decision-making and maximizes personalized learning journeys by creating real-time insights.

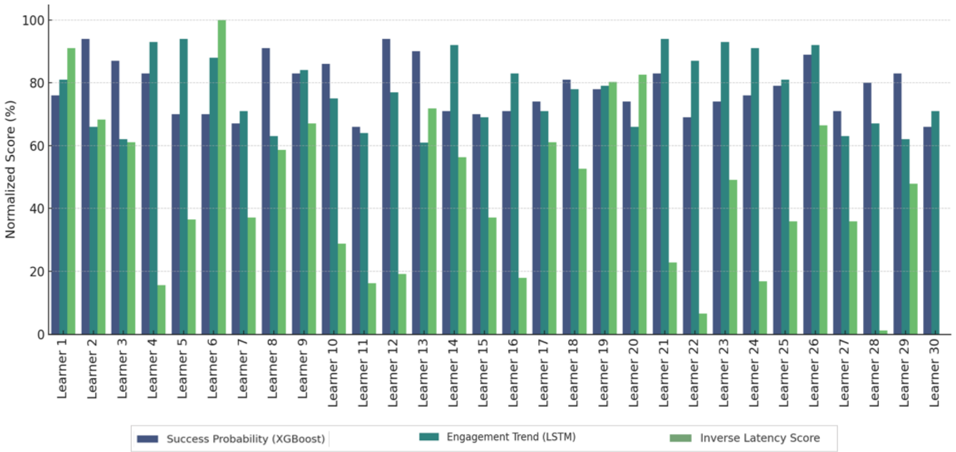

Figure 13. AI-Driven Personalization Metrics for Learner Insights

5. Discussion

The findings of this research indicate that the suggested multimodal AI-based language learning system proves to be effective in enhancing the main language skills, such as writing precision, speaking fluency and general learner engagement. The combination of transformer-based NLP models, speech recognition and predictive machine learning methods provide the system with real-time, context-driven feedback which greatly improves the learning experience as opposed to the traditional, non-living e-learning environments. The improvements in performance, which include pronunciation and fluency, reflect the effectiveness of the incorporation of different learning modalities in text and speech. The results align with recent studies in NLP exploring AI optimization, such as multi-objective reinforcement learning that aims to optimize the model for different evaluation metrics to enhance flexibility and performance. In general, these results support the importance of optimization methods to be integrated in the intelligent and adaptive design of educational technologies [21].

One of the key strengths of the proposed framework is that it offers personalized learning trajectories. The predicted learner performance is achieved through XGBoost and the changes of the learner engagement is achieved by LSTM. The system dynamically adjusts content difficulty and gives feedback depends on the learners' behavior and progress based on these predictions. This type of personalisation may help learners learn efficiently, and can help to maintain their interest and motivation for extended periods. These experiment results also prove the reliability and the effectiveness of the AI models used. It demonstrates a good precision when it comes to the confusion matrix and good evaluation indicators like BLEU score, F1-score and low Word Error rate (WER) indicating that it can perform well both in NLP and speech fronts. This suggests the model can also be leveraged to handle variations in learner inputs, for instance differences in writing styles, speech accents etc., thus enabling it to deployed in the real world to heterogeneous learning settings.

However there are some limitations with these promising results. The system relies heavily on surface language features represented through grammar, pronunciation and fluency level, and is unable to assess higher order skills such as critical thinking, organizing discourse or argumentative writing across contexts. Moreover, feedback generated by AI can lead to dependency and thus reduce the capacity of learners to independently solve problems themselves. Variations in level of proficiency and engagement pattern of learners also presents a problem of providing a consistently high level of improvement in all users.

The recent researches underline the effectiveness of AI-driven collaborative learning systems that bring together students, teachers, and intelligent systems that facilitate interaction and contemplative learning. These plans integrate human teaching and AI learning to aid in developing analytical skills and interaction. The results of this study provides evidence of the impact hybrid and interactive designs can have in the design of new AI-driven education systems [22].

6. Conclusion

This paper suggested a multimodal AI platform of personalized English language learning as an integration of NLP, speech technologies and machine learning in a single adaptive system. The framework overcomes the major shortcomings of the traditional and currently existing digital platforms through its ability to provide real-time feedback, personalized learning trajectories, and interactive speech-based interaction. The experimental findings show significant gains in writing accuracy, speaking fluency and engagement of the learner, which evidences the effectiveness of the multimodal inputs combined with the adaptive and gamified learning mechanisms. The predictive learner modeling also adds to the capability of the system to provide context-dependent and individualized instruction. The framework is now focused more on linguistic accuracy rather than on higher-order cognitive ability and is subject to the problem of AI bias as well as over-dependence on machine feedback. The next step of the work will be to add explainable AI, increase multilingual support, and test the system on a large scale, using real-world applications.

Data Availability Statement:
The datasets used and/or analyzed during the current study are available from the corresponding author on request.

Funding Statement:
This research received no external funding.

Competing Interests: The authors declare no competing interests.
Ethics, Consent to Participate, and Consent to Publish declarations: not applicable.

References

1. C. Yu, “Artificial Intelligence Algorithm Assisted Automatic Detection and Feedback Mechanism for University English Translation Errors,” in *Proceedings - 2025 International Conference on Digital Analysis and Processing, Intelligent Computation, DAPIC 2025*, Institute of Electrical and Electronics Engineers Inc., 2025, pp. 541–545. doi: 10.1109/DAPIC66097.2025.00106.
2. A. S. Jothi, M. B. Alazzam, S. F. Sulaiman, K. Abirami, R. Saranya, and B. K. Bala, “Speech Recognition and Machine Learning: Revolutionizing English Language Oral Assessments,” in *Proceedings of 2025 3rd International Conference on Intelligent Systems, Advanced Computing, and Communication, ISACC 2025*, Institute of Electrical and Electronics Engineers Inc., 2025, pp. 576–581. doi: 10.1109/ISACC65211.2025.10969223.
3. X. J. Jiang, “Automatic Correction Method for English Composition Errors Assisted by Artificial Intelligence,” in *Proceedings - 2024 International Conference on Interactive Intelligent Systems and Techniques, IIST 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 264–268. doi: 10.1109/IIST62526.2024.00029.
4. K. Chandrasekhar, S. Sarwar, P. R. Alapati, K. S. Punithaasree, P. Pathak, and B. K. Bala, “Empowering Spoken English Assessment: Advanced Deep Learning Techniques for Automated Evaluation,” in *2024 International Conference on Intelligent Systems and Advanced Applications, ICISAA 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/ICISAA62385.2024.10829249.
5. D. Zou, "Leveraging Artificial Intelligence for Enhanced Language Teaching and Learning in Higher Education," in *Emerging Technologies for Education*, Singapore: Springer, 2025. [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-96-8430-4_4
6. N. D. N. Ha, "Harnessing Artificial Intelligence for English Language Learning: Benefits and Challenges," in *Proc. CIC 2024*, Atlantis Press, 2024. [Online]. Available: https://www.atlantis-press.com/proceedings/cic-24/126007495
7. E. Ferrer-Molina, M. Martínez-García, B. Gallego-Arrufat, et al., “Using Chatbots as AI Conversational Partners in Language Learning,” *Appl. Sci.*, vol. 12, no. 17, 2022. https://www.mdpi.com/2076-3417/12/17/8427 mdpi.com+15mdpi.com+15mdpi.com+15
8. G. Lampropoulos, “Augmented Reality, Virtual Reality, and Intelligent Tutoring Systems in Education and Training: A Systematic Literature Review,” *Appl. Sci.*, vol. 15, no. 6, June 2025. https://doi.org/10.3390/app15063223 mdpi.com
9. F. Jia, D. Sun, Q. Ma, and C.-K. Looi, “Developing an AI-Based Learning System for L2 Learners’ Authentic and Ubiquitous Learning in English Language,” *Sustainability*, vol. 14, no. 23, 2022. https://doi.org/10.3390/su142315527 mdpi.com+5mdpi.com+5arxiv.org+5
10. M. Zaim, S. Arsyad, B. Waluyo, H. Ardi, M. Al Hafizh, M. Zakiyah, W. Syafitri, A. Nusi, and M. Hardiah, “Generative AI as a Cognitive Co Pilot in English Language Learning in Higher Education,” Educ. Sci., vol. 15, no. 6, Art. no. 686, 2025. https://www.mdpi.com/2227-7102/15/6/686
11. I. Engeness, M. Nohr, and T. Fossland, “Investigating AI Chatbots’ Role in Online Learning and Digital Agency Development,” Educ. Sci., vol. 15, no. 6, Art. no. 674, 2025. https://www.mdpi.com/2227-7102/15/6/674
12. Y. Liu, F. Wang, and X. Chen, “Unpacking the Dynamics of AI Based Language Learning: Flow, Grit, and Resilience among Chinese EFL Learners,” Behav. Sci., vol. 14, no. 9, Art. no. 838, 2024. https://www.mdpi.com/2076-328X/14/9/838
13. Q. Zhang, H. Nie, J. Fan, and H. Liu, “Exploring the Dynamics of Artificial Intelligence Literacy on English as a Foreign Language Learners’ Willingness to Communicate: The Critical Mediating Roles of Artificial Intelligence Learning Self Efficacy and Classroom Anxiety,” Behav. Sci., vol. 15, no. 4, Art. no. 523, 2025. https://www.mdpi.com/2076-328X/15/4/523
14. W. Zhao, S. Huang, and L. Yan, “ChatGPT and the Future of Translators: Overview of the Application of Interactive AI in English Translation Teaching,” in *2024 4th International Conference on Computer Communication and Artificial Intelligence, CCAI 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 303–307. doi: 10.1109/CCAI61966.2024.10602989.
15. V. M. Jayakumar, R. Rajakumari, P. R. Alapati, S. Otero-Potosi, D. N. Malleswari, and M. Karthik, “Enhancing English Language Assessment in Educational Settings using Natural Language Processing Techniques,” in *Proceedings of 2025 3rd International Conference on Intelligent Systems, Advanced Computing, and Communication, ISACC 2025*, Institute of Electrical and Electronics Engineers Inc., 2025, pp. 438–443. doi: 10.1109/ISACC65211.2025.10969428.

16. J. Wu, Y. Wang, F. Chen, X. Yin, and Y. He, “Impact of AI-Powered Adaptive Learning Platforms on English Reading Proficiency: Evidence from Structural Equation Modeling,” *IEEE Access*, 2025, doi: 10.1109/ACCESS.2025.3571055.
17. A. Jawaid et al., "Explainable AI-Powered Writing Tutor for Academic English Learning and Improvement," 2025 International Conference on AI-Driven STEM Education and Learning Technologies (AISTEMEDU), Al Khobar, Saudi Arabia, 2025, pp. 1-10, doi: 10.1109/AISTEMEDU67077.2025.11403888.
18. Y. Jing and G. Jia, “Application of Artificial Intelligence (AI) Technology in Intelligent Recommendation of English Personalized Learning System,” in *Proceedings - 2024 3rd International Conference on Artificial Intelligence and Autonomous Robot Systems, AIARS 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 455–460. doi: 10.1109/AIARS63200.2024.00090.
19. H. Zhang, “A Corpus System Study of Oral English Ability Improvement Based on Big Data Analysis,” in *2024 IEEE 2nd International Conference on Image Processing and Computer Applications, ICIPCA 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 555–559. doi: 10.1109/ICIPCA61593.2024.10709317.
20. M. V. Kumar, N. Prakash, S. Prema, M. Manchanda, and R. Changala, “Harnessing AI for English Language Learning through Personalized Methods and Adaptive Technologies,” in *International Conference on Intelligent Systems and Computational Networks, ICISCN 2025*, Institute of Electrical and Electronics Engineers Inc., 2025. doi: 10.1109/ICISCN64258.2025.10934428.
21. M. J. Iqbal, M. A. Khan, T. Alyas, S. Abbas, A. Jawaid, and F. M. Zanzotto, “A multi-objective reinforcement learning approach to prompt optimization in NLP,” Procedia Computer Science, vol. 275, pp. 966–974, 2026, doi: 10.1016/j.procs.2026.01.110.
22. A. Jawaid, S. M. Usman, T. Alyas, S. Abbas, A. Fatima, and M. I. Ul Haq, “Collaborative methodology by integrating students, AI and teacher to teach English literature for active and reflective learning,” Procedia Computer Science, vol. 275, pp. 724–732, 2026, doi: 10.1016/j.procs.2026.01.084.
23. Al-Ezzi, R., & Al-Qudah, I. (2024). English translation of verbal humour in Egyptian comedy films. *World Journal of English Language*, *14*(1), 157-157.