EnFluent: Personalized Language Learning through Adaptive Learning Management Systems

N.V. Kannangara, Y.U.K.K. Chandrasiri, P.T.B. Gunatilake, A.K.C.S.Madusan, Dr. Subodha Charles
Department of Electronic and Telecommunication Engineering
University of Moratuwa, Sri Lanka

Abstract—Learning a language can be challenging, especially when traditional methods fail to meet the unique needs of each learner. Enfluent is an adaptive LMS that aims to revolutionize personalized language learning for IELTS students by offering a range of innovative features. It includes a course recommendation system that suggests courses based on the individual band scores for reading, writing, speaking, and listening. The platform provides a learning profile that clearly shows the current band score for each skill. An automated grading and feedback system evaluates all four skills according to official IELTS criteria. Enfluent also introduces an interactive chat avatar, allowing students to practice speaking in a way that simulates real-life conversations. The system tracks engagement and stress levels using heart rate data collected from a smartwatch through a mobile application. Additionally, it uses webcam data to detect engagement, offering valuable insights for both students and course creators. A dedicated chatbot is available to answer course-related questions, while a content-aware chatbot provides answers linked to video materials by using video processing techniques. To keep learners actively involved, the platform also features activities such as generating pop quizzes during video playback, making the learning experience more interactive and engaging.

Keywords—Language learning, IoT, Engagement, Speech recognition, Avatar

I. Introduction

Learning a language is an intricate process that integrates our emotions, physical sensations, and mental abilities. Traditional methods often adopt a one-size-fits-all approach, ignoring the unique physical and emotional experiences of each individual, particularly when practicing speaking [1], [2]. This can result in a lack of engagement and less effective learning outcomes. Consequently, there is a clear need for more adaptive and personalized learning environments that address each learner's unique needs and challenges [3], [4].

Existing Language learning platforms like Duolingo, Babbel and official IELTS preparation tools mostly use cognitive data to personalize course contents.These platforms use surveys, quizzes, and automated assessments to measure learners' skills and adjust lessons accordingly.While some include basic AI features like chatbots and interactive avatars, most of them ignore real-time physiological data analysis. The majority of existing systems do not keep track of or adjust to learners' physical states, such as stress levels and heart rate, which are vital for effective learning [5]. Moreover, the feedback systems of these platforms often provide only pre-defined responses, lacking the complexity and personalization necessary to address individual learner needs. [6], [7].

A significant gap in the existing methods is the lack of systems that can simultaneously monitor and adjust to learners' physiological and cognitive states [8]. Traditional platforms overlook the emotional and physical aspects that impact language learning, such as fatigue from prolonged hours of study or nervousness during speaking activities [9].These solutions are unable to provide a fully personalized learning experience because they are unable to dynamically modify the learning environment in response to real-time data, [6]. This could cause learners to become disengaged or receive insufficient assistance that isn't personalized to their unique needs, limiting their overall language proficiency development. [2], [10].

In order to address these problems, the EnFluent project introduce a novel form of adaptive learning environment that combines IoT data with AI-driven avatars. By using smartwatches, Our platform monitors important physical and emotional signals like heart rate and stress levels, providing a comprehensive picture of each learner's state. This real-time information is combined with interactive 3D avatars that offer personalized feedback and support. This approach makes learning more immersive and responsive. In contrast to other platforms, EnFluent adapts the learning materials and paths based on the individual needs of each learner, guaranteeing that the content remains interesting and efficient.

We compared EnFluent against several best language learning apps, such as Duolingo, Babbel, and official IELTS preparation tools, in order to evaluate its effectiveness. We looked at features like personalized learning, automatic grading, integration of IoT data, AI chat capabilities, interactive avatars, and automatic updates to learner profiles. The findings demonstrated that Our platform is superior in a number of crucial areas. It stands out for its ability to seamlessly integrate physiological data and offer real-time, personalized feedback through AI-driven avatars. EnFluent can dynamically modify the learning process, improving engagement and language proficiency through the application of reinforcement processes and machine learning models. These findings demonstrate how our solution can advance language acquisition by filling significant gaps in the available systems.

Enfluent advances personalized learning by addressing the shortcomings of existing language learning tools and introducing a comprehensive solution that combines IoT and AI technologies to enhance language skills effectively and provide necessary support.This helps learners overcome the mental and physical barriers that come with learning a new language. [1], [2].

II. Literature Review

This literature review explores existing language learning platforms and technologies to identify key features, limitations, and gaps in the current landscape. The aim is to understand how tools like speech recognition, AI-driven personalization, learner engagement tracking, and interactive avatars are being utilized and how they can be improved to create a more adaptive and effective learning experience. By analyzing these systems, we lay the foundation for understanding the unique contributions of our proposed solution, Enfluent.

A. Related Work

In this section, we review existing work related to language learning platforms, discuss their features, and limitations.

Talkpal [11] AI is an AI-powered platform that consists of personalized conversations and real-time feedback. It includes role-play and debate modes that engage users in real-life scenarios, helping to improve language skills and critical thinking. In addition, there is a photo mode for vocabulary enhancement and advanced speech recognition technology for pronunciation feedback. These features position Talkpal AI as a comprehensive tool for language learners aiming to enhance their fluency and confidence in communication. Babbel [12] is a language learning platform introduced for beginners and intermediate learners. It consists of short and interactive lessons to improve practical conversational skills and vocabulary. Babbel uses surveys to determine user proficiency for content personalization.This customized approach guarantees an enjoyable and effective learning journey that caters to each learner's individual needs.

DP Education IELTS [13] provides free resources for IELTS preparation, which consist of video lessons and tests created by subject matter experts. There is a feature called T-AI to resolve queries and provide a degree of personalization, making it a helpful tool for test-takers. With its free and accessible resources, DP Education IELTS enables learners to confidently achieve their desired band scores.

Duolingo [14] is a language learning app with its gamified approach to engage users. This platform consists of reading, writing, listening, and speaking exercises while tailoring lessons to user progress. Its intuitive and engaging format increases popularity among learners.Duolingo's intuitive and game-like format makes it a popular choice for learners across all ages and proficiency levels.

Busuu [15] combines adaptive learning with social interactions. It provides a comprehensive and interactive learning environment through AI-powred reviews and live lessons, and feedback from native speakers. This makes Busuu a versatile tool for learners aiming to improve multiple language skills simultaneously. By merging technology with community support, Busuu provides a tailored and immersive experience for mastering new languages.

The following Table 1 summarizes the features of existing language learning platforms [11]–[15].

Feature	**Talkpal AI**	**Babbel**	**DP Education IELTS**	**Duolingo**	**Busuu**
Personalized Learning	AI-driven conversations	Survey-based	T-AI for query resolution	Adaptive lesson plans	AI-powered personalized reviews
Automatic Grading	Not available	Basic feedback	Not available	Automated for basic tasks	Feedback from native speakers
IoT Integration	Not available	Not available	Not available	Not available	Not available
Interactive Avatar	Not available	Not available	Not available	Basic avatar	Not available
Speech Recognition	Advanced	Basic	Not available	Limited	Limited
Learner Profile Updates	Limited	Task-based	None	Goal-driven	Performance-based

TABLE I
Comparative Analysis of Existing Language Learning Platforms

Current language learning platforms have several shortcomings that make it difficult for them to provide complete adaptive learning experience. Most of them do not have automatic grading systems that can grade all key skills, including writing, speaking, listening and reading. They also lack integration with IoT devices like smartwatches and cameras, which enable more personalized and interactive learning. Furthermore, most platforms do not offer advanced features like interactive chat avatars and AI-driven chat systems which enable learners more engaged. Lack of real-time feedback system prevents learners from quickly adjusting their learning paths, reducing the overall effectiveness of the process. Addressing these shortcomings is crucial for fostering a more engaging, personalized, and effective language learning experience for all.

Enfluent aims to solve these problems with advanced features that fill those gaps. Enfluent's innovative approach sets a new standard for adaptive, interactive, and comprehensive language learning systems.

B. Learner Engagement Detection

The EnFluent platform emphasizes maximizing learner engagement. To achieve this, we integrated a Learner Engagement Detection system that uses heart rate data from the user's smartwatch as a time series. Existing research supports heart rate as an objective measure of engagement, showing strong correlations with other studies and datasets. [16] [17] [18]

We argue that heart rate correlates with engagement during lectures, a connection grounded in physiological principles. Although the brain constitutes only 2 percent of body weight, it consumes 20 percent of oxygen and 25 percent of glucose. When cognitive load increases—interpreted here as engagement—oxygen and glucose usage also rises. During sedentary activities like sitting in a classroom, the body compensates for this increased demand by elevating heart rate and, potentially, respiration. Our approach focuses on measuring heart rate as a simple and accessible physiological marker of engagement. While we acknowledge that correlation does not imply causation, this method provides a straightforward mechanism that aligns with physiological reasoning [16].

This understanding informs techniques to enhance user engagement during course lessons on the platform. Literature indicates that incorporating active learning activities is the most effective strategy to boost engagement. Studies reveal that during activities like discussions, learners experience an increase in mean heart rate. These activities require greater cognitive effort, resulting in heightened physiological responses. Researchers using ANOVA analysis confirmed that all active learning activities led to statistically significant increases in heart rate compared to each learner's baseline [17].

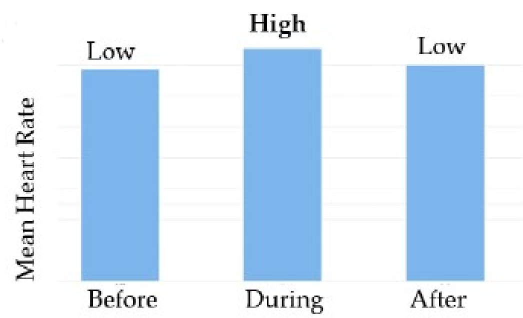

Fig. 1. Mean Heart Rate during an Active Learning Activity

Active Learning Activity	**F Value**	**Critical Value for F**	**Had Heart Rates Statistically Significant Differences?**
Quiz	0.181	3.890	No
Discussion in group/pair	37.492	3.865	Yes
Share responses	11.834	3.910	Yes
Individual activity	24.341	3.857	Yes
Interaction with professor	6.801	3.860	Yes

TABLE II
ANOVA-ACTIVE LEARNING ACTIVITIES.

Table II highlights the statistically significant differences in heart rate when learners engage in ANOVA active learning activities. If the F-value exceeds the critical F-value, it indicates that the means of the two variables differ significantly. [17]

By leveraging these insights, EnFluent aligns its approach with evidence-based strategies to create a more engaging learning experience.

C. Speech recognition and pronunciation assessment

In order to perform automatic speech recognition (ASR), first it is essential to perform feature extraction of the speech signal. Existing audio feature extraction methods mainly fall into two categories [19], which are spectral feature analysis where the spectral representation of the speech signal is analyzed and the temporal feature analysis of the audio signal.

All the feature extraction methods come after the preprocessing steps including denoising and normalizing of the audio signal. To analyze spectral features researchers commonly use methodologies of Mel-Frequency Cepstral Coefficients (MFCC), Principal Component Analysis (PCA), Linear Predictive Coding (LPC), Linear Predictive Cepstral Coefficient (LPCC), Perceptual Linear Prediction (PLP) [19].In these methods MFCC represents the audio data as how they are perceived by human ear but results in reduced accuracy when the audio signals are noisy. Researchers commonly use Discrete Wavelet Transforms (DWT), Wavelet Packet Transforms (WPT), Relative Spectra–Perceptual Linear Prediction (RASTA-PLP) for the analysis of temporal features of audio data. For increased accuracy in speech recognition and to enforce robustness to noise, practitioners pair DWT with MFCC which allows to capture both temporal and frequency data of the audio signal [19].

Researchers have developed several methodologies to detect mispronunciations and to evaluate the pronunciation of speech [20]. Classification based on acoustic features with techniques of Linear Discriminant Analysis (LDA) with using MFCC feature extraction methods [21]. Also, researchers have worked on likelihood based Hidden Markov Model (HMM), Gaussian mixture models (GMM) and they evaluated based on the HMM based log posterior scores as a goodness of pronunciation measurement for the speech signal [22]. Apart from those parametric modeling, researchers have proposed on deep learning based models to assess pronunciation without using an alignment component with the help of CNN-RNN-CTC models [23] [20]. In addition to that researchers have used transformer encoders to detect mispronunciations and automatic assessment of the prosodic and segmental scores of the speech signal [24].

Enfluent builds upon the existing ASR and speech evaluation methodologies to provide more robust and accurate pronunciation assessment system.

D. Avatar Creation in Language Learning

Avatars are widely used in modern education systems, especially in digital learning environments, due to their ability to create engaging and interactive experiences. In language learning, avatars play a significant role by acting as virtual instructors or companions that improve the learning process.

Research shows that realistic avatars can enhance learning by providing visual and emotional support. According to Bailenson [25], avatars with human-like traits create a stronger sense of connection between users and the system. Elements like synchronized lip movements and natural gestures Make interactions feel more realistic. Also Lee [26] found that avatars with changing responses improve engagement and understanding by mimicking real conversations.

Despite these promising findings, some gaps remain. Many studies focus on the visual and emotional aspects of avatars but overlook how they can adapt to individual learners in real time. Additionally, most research emphasizes short-term engagement, leaving long-term impacts and scalability unexplored. Filling these gaps is key to maximizing avatars' potential in education.

Our study builds on this by adding advanced features to avatar systems. We include adaptive feedback and real-time responses to learners' emotions and needs. These improvements make learning more personal and effective, addressing current research gaps and advancing avatar-based language learning.

Although current language learning platforms offer useful tools and features, many lack the depth needed for a truly personalized and engaging experience. Key challenges include limited real-time feedback, minimal IoT integration, and weak support for adaptive learning. These limitations highlight the need for a more comprehensive solution. Enfluent addresses these issues by integrating advanced technologies such as engagement detection, robust speech analysis, and emotionally responsive avatars—offering a smarter and more effective path to language learning.

III. METHODOLOGY

This section outlines the methods and technologies used in the design and implementation of the EnFluent language learning platform. It describes the system's core components, including learner profiling and grading, IoT-based engagement tracking, speech evaluation, intelligent chatbot assistance, and 3D avatar integration. Each component is developed using modern AI techniques and cloud-based tools to deliver a personalized and adaptive learning experience.

A. Learning Profile and Automatic Grading

When a user signs up for EnFluent for the first time, the platform initiates an onboarding process. After completing basic authentication and authorization, the platform provides an onboarding test. This test must be completed before the user can enroll in any course. The onboarding test includes four separate components: Reading, Writing, Listening, and Speaking. The platform evaluates the user's performance in each component and assigns a band score (ranging from 1 to 9) for each category. Based on these scores, the platform creates a personalized learning profile for the user. This profile contains individual band values for Reading, Writing, Listening, and Speaking.

Once the onboarding process is complete, users can enroll in the courses offered on the platform. The platform also suggests courses based on the user's band scores and the recommended band level for each course. At the end of each course, the platform re-evaluates the user's performance. It then updates the learning profile based on the new scores obtained by the user.

To conduct evaluations during the onboarding process and at the end of courses, the platform uses automated grading systems. These tests are modeled after IELTS assessments. The Reading test requires users to read a paragraph and answer multiple-choice questions (MCQs). The Listening test involves listening to audio clips and responding to MCQs. These tests are graded automatically by checking the correctness of the answers, providing a straightforward method to calculate band scores.

For the Writing test, the platform presents two tasks: describing a given graph and writing an essay on a specific topic. Automated grading for this component utilizes GPT-4, accessed through Azure AI. The backend system sends the test details, user responses, and a fine-tuned prompt to GPT-4. This prompt is specifically designed to evaluate the writing based on key criteria: Task Achievement, Coherence and Cohesion, Lexical Resources, and Grammatical Range and Accuracy. The system calculates a final band score based on these aspects and updates the platform accordingly.

The Speaking test requires users to record their speech, which the platform evaluates using Azure's Speech-to-Text (STT) service. This service employs multi-aspect feature extraction, providing separate scores for accuracy, prosody, and content. The accuracy score assesses elements like Task Achievement, Coherence and Cohesion, Lexical Resources, and Grammatical Range and Accuracy. Based on these individual scores, the platform calculates a final band score for the Speaking test.

This structured approach ensures personalized learning and effective evaluation throughout the user's journey on the platform.

B. IOT integration

The EnFluent platform integrates Internet of Things (IoT) technologies to monitor learner engagement and stress levels through physiological and behavioral data. This hybrid engagement tracking system utilizes both smartwatch-based heart rate data and webcam-based facial expression recognition to provide adaptive and responsive learning experiences.

The heart rate of learners is continuously measured using smartwatches equipped with photoplethysmography (PPG) sensors. The collected data is transmitted to an Android device via Bluetooth and stored in Health Connect App. To facilitate cloud synchronization, a custom mobile app named *HC-Gateway*, developed using the Expo framework, reads heart rate data from Health Connect and transmits it to a cloud database through a Python FastAPI backend. The data is visualized on a React-based dashboard with a Spring Boot API.

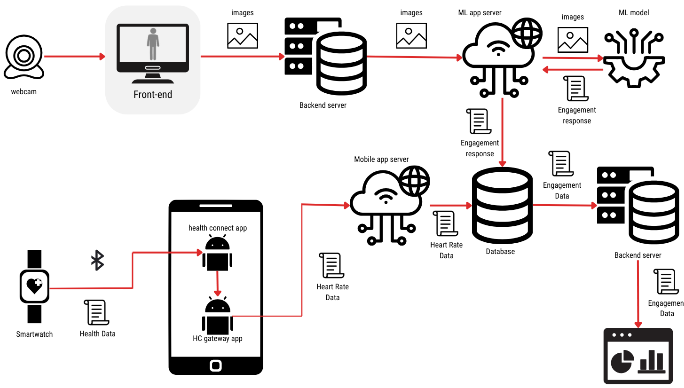

Fig. 2. IoT Integration Architecture

Heart Rate–Based Engagement and Stress Detection

Three methods were evaluated to quantify stress levels from heart rate variability (HRV) data. The first used machine learning classifiers, where stress levels were labeled as high, moderate, or low. Algorithms such as Logistic Regression, Random Forest, K-Nearest Neighbors (KNN), K-Means, and BIRCH were compared, with KNN achieving the highest accuracy of 99%. The second method used the Triangular Index, which showed lower accuracy. The third and most reliable method was the Baevsky Stress Index [27], computed as:

$$SI = \frac{AMo \times 100}{2 \times Mo \times MxDMn} \tag{1}$$

where SI is the stress index, AMo is the amplitude of the mode, Mo is the mode of Normal-to-Normal Intervals (NNI), and $MxDMn$ is the range of NNIs. This clinically validated method provides a robust numerical assessment of stress. Engagement is calculated as directly proportional to heart rate levels [16], under the assumption that cognitive and emotional involvement leads to elevated cardiovascular responses.

Webcam-Based Engagement Detection

In addition to physiological data, EnFluent employs a webcam-based facial expression recognition module to analyze real-time affective states. This system detects engagement-relevant emotions such as focus, confusion, boredom, or frustration by analyzing visual cues—eye gaze, mouth movement, and eyebrow position.

A deep learning model based on YOLOv11 was selected for emotion classification due to its high inference speed and accuracy. The model was trained using the Affective States Computer Vision dataset, which includes a large volume of labeled emotion images. Training was conducted on the Kaggle platform using an NVIDIA Tesla T4 GPU.

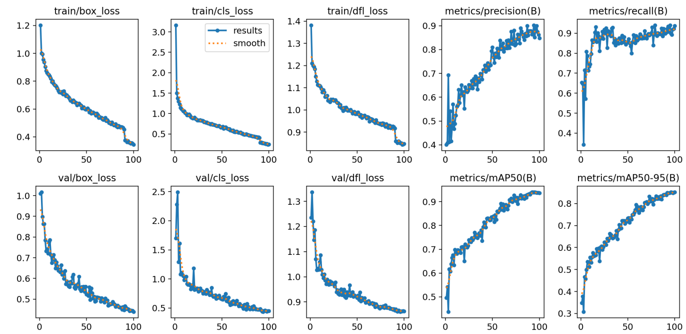

Fig. 3. Training Loss Plots

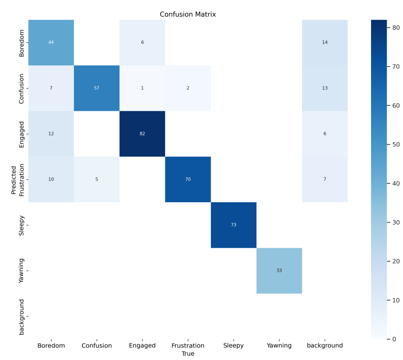

Fig. 4. Model Performance Matrix

The final model performance achieved 93.7% accuracy, 0.8485 precision, and 0.9365 recall—well within the system requirements of >90% accuracy, <30s latency, and high recall.

Emotion-Based Engagement Scoring

To convert visual emotion detection into a quantitative engagement score, each detected emotion was mapped to a predefined score, as shown in Table III. Engagement is calculated as the weighted average of emotion scores over time, enabling continuous tracking.

TABLE III
EMOTION TO ENGAGEMENT SCORE MAPPING

Emotion	Engagement Score
Engaged	5
Boredom	3
Confusion	2
Frustration	2
Sleepy	0
Yawning	0

This scoring framework allows the system to:

- Monitor moment-by-moment changes in engagement
- Trigger context-sensitive feedback through the AI avatar
- Adjust course difficulty based on learner affect
- Record longitudinal engagement trends in learner profiles

System Integration and Impact

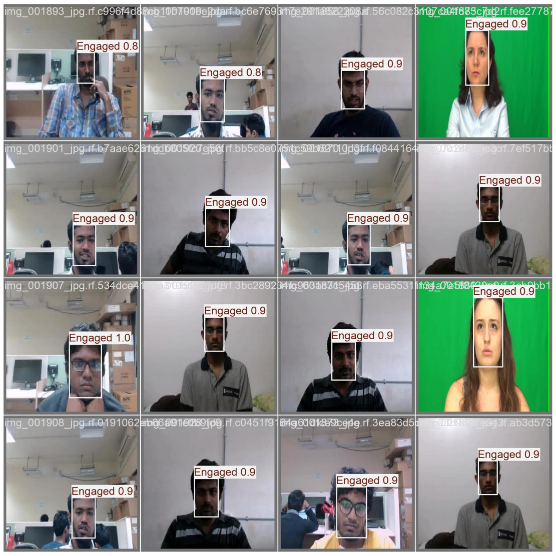

Fig. 5. Webcam Engagement Model Evaluation

The webcam module is seamlessly embedded in the platform's interface and operates during interactive learning sessions. By combining real-time physiological (heart rate) and behavioral (facial expression) data, EnFluent delivers adaptive and emotionally aware feedback to learners.

C. Speech evaluation module

The speech evaluation module we developed for EnFluent language learning platform provides automatic grading ability of speech tests which we use to create the learner profile of the users. Additionally, the module provides detailed description of user's performance of the speech in word and phoneme level with the accuracy, fluency, completeness and overall pronunciation scores. Following diagram represents the overall architecture of the speech evaluation module according to IELTS speaking evaluation criteria.

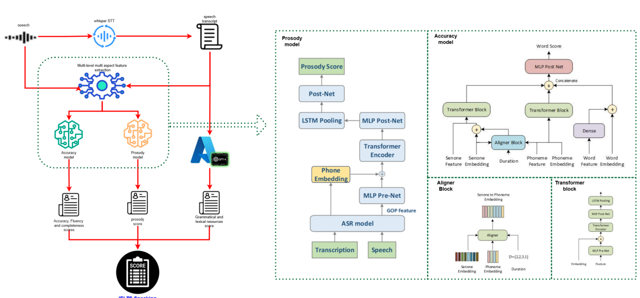

Fig. 6. speech evaluation architecture

In the implementation, we used a hierarchical Transformer model for pronunciation accuracy scoring and a Transformer encoder combined with LSTM pooling for evaluating prosodic aspects. The platform was developed to assess pronunciation and fluency using these azure pronunciation assessment pre-trained models. We adopted a scripted assessment method, where the transcript was first extracted using the Whisper medium Speech-to-Text (STT) model. The input audio was pre-processed and converted to single-channel PCM format with 16kHz sampling rate and 16-bit sample width (.wav) to ensure compatibility with the models. For improved efficiency, the backend developed using Python FastAPI processed the audio input as segmented chunks before feeding them into the evaluation pipeline. The output includes detailed pronunciation scores at both the word and phoneme levels. Additionally, completeness scores were calculated to indicate how many transcript words were pronounced, and fluency scores were estimated based on the similarity of the user's speech to native speaker patterns. The speech evaluation pipeline outputs scores for fluency, pronunciation accuracy, and prosody, along with a combined pronunciation score.

To evaluate grammatical accuracy and lexical resource usage, the transcript was analyzed using GPT-4-based large language models guided by IELTS evaluation criteria. Finally all the assessment criteria (pronunciation, fluency, grammar, and vocabulary) were integrated to estimate the user's overall IELTS speaking band score.

D. Chatbots Creation

The EnFluent platform integrates two intelligent chatbot systems to enhance the user experience through contextual and personalized conversation: a Course Content-Based Chatbot and a Video Content-Based Chatbot. Both systems leverage AI technologies including OpenAI's embedding models, Whisper for transcription, and Azure GPT-4 for generating natural language responses.

Course Content-Based Chatbot

This chatbot allows users to interact conversationally with their uploaded learning materials. Upon uploading documents (e.g., PDF or DOCX), the backend splits the content into smaller text chunks and converts them into embeddings using OpenAI's embedding model. These embeddings are stored in a vector database alongside metadata. When a user poses a question, it is also embedded and matched against the database to retrieve relevant chunks. These chunks and the question are then sent to Azure GPT-4, which generates a context-aware answer. The final output includes the response and links to the relevant document segments.

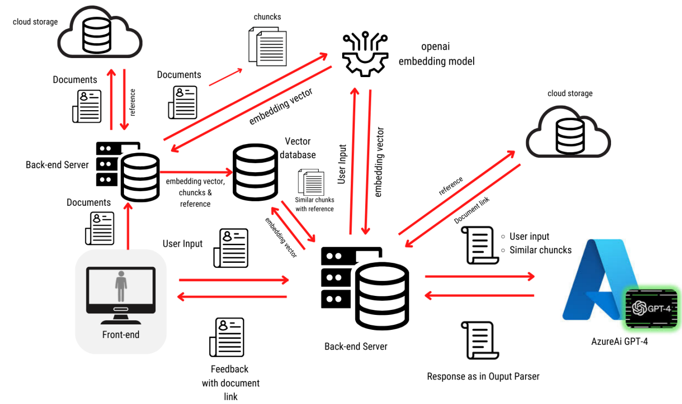

Fig. 7. Course-Based Chatbot Architecture in EnFluent

Video Content-Based Chatbot

This module enables learners to ask questions about video lessons. Uploaded videos are processed by extracting their audio and transcribing it using OpenAI's Whisper model. The transcript is stored and linked to the video. When a question is asked, the system searches the transcript for relevant segments. These are passed, along with the user's query, to GPT-4 to generate an appropriate answer. The system returns the answer to the user.

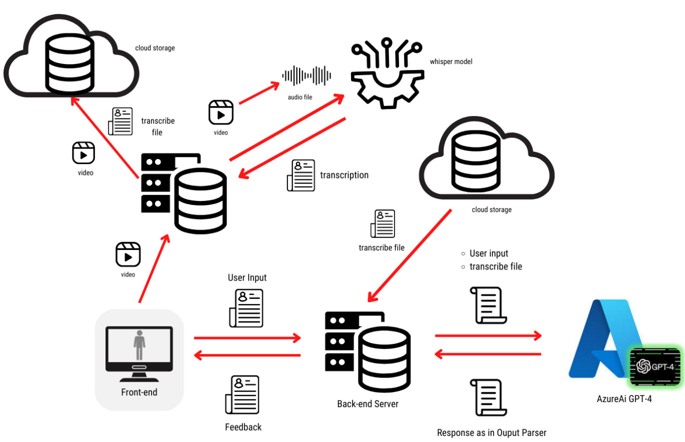

Fig. 8. Video-Based Chatbot Architecture in EnFluent

By combining document segmentation, vector similarity matching, speech-to-text transcription, and GPT-4-based generation, these chatbots offer personalized, meaningful dialogue. This interactive methodology improves learner engagement and provides efficient access to learning resources in a natural, user-friendly format.

E. Avatar Creation

The interactive 3D avatar for EnFluent was developed in multiple stages. The goal was to make it realistic, responsive, and efficient. This section explains the process of building, improving, and smoothly integrating the avatar into the adaptive learning platform.

The avatar was designed in Character Creator 4, creating a realistic 3D character with accurate features. The avatar was styled to resemble a professional language instructor, building trust and engagement with learners.

Next, the avatar model was optimized using Blender. Unnecessary details were removed to simplify its structure and reduce file size without losing visual quality. The high-poly model was optimized into a low-poly version to boost performance on the web platform. Materials were also simplified for better compatibility with web platforms. Finally, the optimized model was exported in GLTF/GLB format, a lightweight file type perfect for interactive web applications. [28]

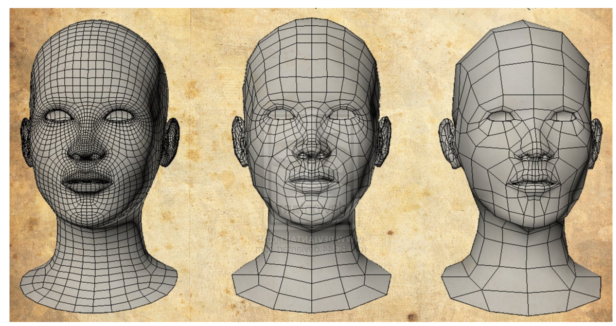

Fig. 9. Web Optimizing Avatar Model High Polly to Low Polly

Animations, including facial expressions, gestures, and idle behaviors, were created in Blender and exported alongside the model. These animations were essential for mimicking human-like interactions, providing the avatar with the capability to respond naturally to user inputs.

To integrate the avatar into the web-based environment, React Three Fiber, a React-based library for rendering 3D content with Three.js, was utilized. This framework allowed for seamless loading and manipulation of the GLB files within the web application. Scripts were developed to dynamically control the avatar's animations, enabling context-sensitive responses and actions based on system feedback.

Lip synchronization, a critical feature for realistic interaction, was implemented using Rhubarb Lip Sync. This tool generated lip-sync data from audio files, enabling precise synchronization of the avatar's lip movements with speech. The pipeline for this process included capturing user speech through Azure Speech Services, which converted the audio input into text. The text was processed using OpenAI GPT to generate a contextually relevant response, which was then converted into audio using the ElevenLabs. [29] The resulting audio file was analyzed to produce a waveform, from which Rhubarb [30] generated the corresponding lip-sync data in JSON format.

The final step involved synchronizing the lip-sync animation with audio playback in real-time. The avatar's lip movements were seamlessly coordinated with the generated speech, pro-

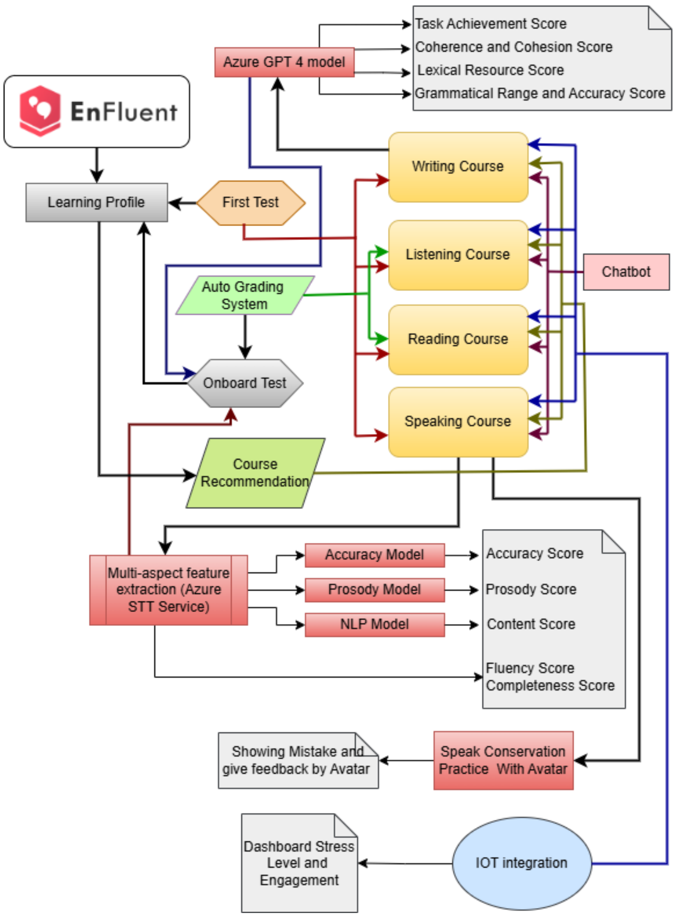

Fig. 10. Enfluent Project Block Diagram

viding users with an immersive and realistic conversational experience. In addition to lip-syncing, the avatar's gestures and facial expressions were dynamically adjusted based on the context of the conversation, further enhancing the interactive experience.

The EnFluent platform brings together multiple advanced technologies to support data-driven, user-centered language learning. By combining automatic assessment tools, engagement monitoring, conversational AI, and interactive avatars, the system creates an immersive environment tailored to individual learners. The methodology described ensures scalability, real-time responsiveness, and educational effectiveness across diverse user needs.

IV. Results

Our platform delivers a comprehensive and personalized language learning experience through an Adaptive Learning Management System. By integrating all the features described earlier, we have developed and deployed a unified platform. This platform enables users to have a personalized language learning experience while preparing for IELTS and other similar exams.

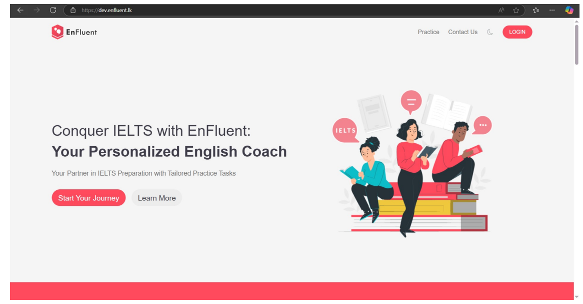

Fig. 11. EnFluent Platform deployed in enfluent.lk

A. Writing Task Evaluation

To validate the effectiveness of EnFluent's automated grading system for writing tasks, a comparison was conducted between scores assigned by human IELTS examiners and scores generated by the platform. Five writing samples (A1 to A5) were reviewed by trained IELTS examiner, and the same responses were evaluated by the platform's auto-grading module using GPT-4. This section presents a statistical comparison and visual analysis of the results.

Table IV shows the scores assigned by human examiners and platform for the five sample answers.

TABLE IV
Raw Examiner Scores for Writing Tasks

	A1 (4.0)	A2 (6.5)	A3 (8.0)	A4 (5.5)	A5 (7.5)
T1	4.0	6.0	6.0	5.5	6.5
T2	4.0	6.0	7.0	5.5	6.5
T3	4.0	5.5	6.5	5.5	6.5
T4	4.0	6.0	6.5	5.5	7.0
T5	4.0	6.0	7.0	5.5	6.5
T6	4.0	6.0	6.5	5.5	6.5
T7	4.0	6.0	6.5	5.5	6.5
T8	4.0	6.0	6.5	5.5	6.5
T9	4.0	5.5	6.5	5.5	6.5
T10	4.0	6.0	6.5	5.5	6.5

The summary of these comparisons is shown in Table V, including average examiner scores, platform scores, mean differences, and standard deviations.

TABLE V
Auto-Grading vs Examiner Evaluation Summary Table

Answer ID	Examiner Score	Platform Mean Score	Mean Difference	Std. Dev. of Difference
A1	4.0	4.0	0.0	0.0
A2	6.5	5.9	-0.6	0.21
A3	8.0	6.6	-1.4	0.21
A4	5.5	5.5	0.0	0.0
A5	7.5	6.55	-0.95	0.16

Figure 12 visualizes how closely the platform's grading aligns with human examiners. Responses A1 and A4 show exact agreement, while A3 demonstrates the largest discrepancy.

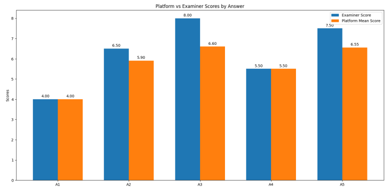

Fig. 12. Platform vs Examiner Scores by Answer

Figure 13 shows the standard deviation of scoring differences. A1 and A4 show no variance, indicating perfect alignment. A2 and A3 show minor deviations (around 0.21), which are acceptable in the context of subjective evaluation.

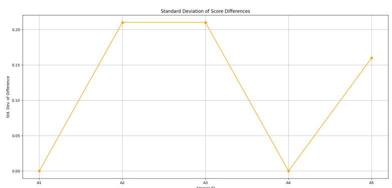

Fig. 13. Standard Deviation of Score Differences

Overall, the results confirm that EnFluent's automated grading system produces consistent and reasonably accurate writing band scores. While minor underestimations are observed in higher-band responses, the platform maintains acceptable alignment with human evaluation, especially in mid- and low-band responses. These findings support the reliability of the auto-grading module for formative assessment in language learning.

B. Speech Evaluation

The speaking assessment module was evaluated using both topic-based scripted prompts and avatar-based interactive conversations. The pronunciation evaluation pipeline outputs fluency, accuracy, and prosodic scores, which are combined to compute an overall pronunciation score. Additionally, the model provides a phoneme-level breakdown for each word, with results formatted in JSON. This includes detailed segment-level pronunciation accuracy, enabling granular analysis of learners' spoken output.

For the assessment of pronunciation and the fluency and coherence aspect we use transformer based multi-task, multi-view, multi-modal MDD with Pearson Correlation Coefficient (PCC) tested with the speechocean762 dataset results are as follows. [31]

Accuracy	0.70
Fluency	0.72
Prosody	0.84

Fig. 14. Speech model PCC scores

For the evaluation of the Lexical resources, Grammatical range and accuracy we used GPT-4 based LLMs. As GPT based evaluations on IELTS evaluation criteria report Intraclass Correlation Coefficient (ICC) of 0.81 (ICC $\approx$ 0.81, 95% CI 0.702–0.887) with human graders [32].

V. Conclusion

Our EnFluent platform provides personalized language learning experience to each of its users by evaluating their language skills in the aspects of reading, writing, listening and speaking. The integration of learner's vitals to the platform using IoT to recognize the learner's engagement and learning state, along with the automatic grading aspects of the language learning differentiates our platform from the traditional language learning methods. In addition to that, the availability of the interactive 3D avatar for speaking practices makes the learning experience much interactive and enhances the user engagement.

The EnFluent platform provides support for multiple groups related to the language education field. It helps language learners by providing personalized feedback on their performance and provides an interactive learning platform compared to the existing static online platforms. In addition to that it provides detailed reports about user engagement levels for each course material and assessment activities to the course provider. In conclusion, the EnFluent platform addresses the issues of traditional language learning methods by integrating interactive avatars and the technologies of IoT, generative AI to provide personalized, interactive language learning experience.

REFERENCES

[1] T. Ingkavara, P. Panjaburee, N. Srisawasdi, and S. Sajjapanroj, "The use of a personalized learning approach to implementing self-regulated online learning," *Computers and Education: Artificial Intelligence*, vol. 3, p. 100086, 2022.

[2] M. A. Hernández-Mustieles *et al.*, "Wearable biosensor technology in education: A systematic review," *Sensors*, vol. 24, no. 8, 2024.

[3] S. Khosravi, S. G. Bailey, H. Parvizi, and R. Ghannam, "Wearable sensors for learning enhancement in higher education," *Sensors*, vol. 22, no. 19, p. 7633, 2022.

[4] S. Chu, B. Garcia, and N. Rani, "Research on wearable technologies for learning: A systematic review," *Journal of Educational Technology*, vol. 50, no. 1, pp. 12–30, 2022.

[5] S. Böttcher, S. Vieluf, E. Bruno, and B. Joseph, "Data quality evaluation in wearable monitoring," *Scientific Reports*, vol. 12, no. 1, p. 21412, 2022.

[6] S. Shahhosseini, Y. Ni, E. K. Naeini, and M. Imani, "Flexible and personalized learning for wearable health applications using hyperdimensional computing," in *Proceedings of the Great Lakes Symposium on VLSI 2022*, 2022, pp. 357–360.

[7] A. F. Abate, L. Cascone, M. Nappi, F. Narducci, and I. Passero, "Attention monitoring for synchronous distance learning," *Future Generation Computer Systems*, vol. 125, pp. 774–784, 2021.

[8] P. Andruloniw, K. Kowalik, and P. Zwierzykowski, "Unsupervised learning data-driven continuous qoe assessment in adaptive streaming-based television system," *Applied Sciences*, vol. 12, no. 16, p. 8288, 2022.

[9] A. Costa, J. A. Rincon, C. Carrascosa, V. Julian, and P. Novais, "Emotions detection on an ambient intelligent system using wearable devices," *Future Generation Computer Systems*, vol. 1794, pp. 479–489, 2018.

[10] D. F. T. Arciniegas, M. Amaya, A. P. Carvajal, P. A. Rodriguez-Marin, L. Duque-Muñoz, and J. D. Martinez-Vargas, "Students' attention monitoring system in learning environments based on artificial intelligence," *IEEE Latin America Transactions*, vol. 20, no. 1, pp. 126–132, 2021.

[11] Talkpal, "Ai language teacher," https://talkpal.ai/, [Accessed: 25-Dec-2024].

[12] Babbel, "Learn languages on android or ios," https://www.babbel.com/mobile, [Accessed: 25-Dec-2024].

[13] DP Education, "Best ielts online courses for free," https://dpielts.lk/, [Accessed: 25-Dec-2024].

[14] Duolingo, "The world's best way to learn a language," https://www.duolingo.com/, [Accessed: 25-Dec-2024].

[15] Busuu, "Learn languages online," https://www.busuu.com/, [Accessed: 25-Dec-2024].

[16] D. K. Darnell and P. A. Krieg, "Student engagement, assessed using heart rate, shows no reset following active learning sessions in lectures," *PLoS ONE*, vol. 14, no. 12, p. e0225709, 2019.

[17] A. C. L. Nocua, J. P. C. Gonzalez, I. A. C. Jimenez, J. S. G. Acevedo, F. Marcolin, and E. Vezzetti, "Assessment of cognitive student engagement using heart rate data in distance learning during covid-19," *Education Sciences*, vol. 11, no. 9, p. 540, 2021.

[18] S. Houzangbe, O. Christmann, G. Gorisse, and S. Richir, "Effects of voluntary heart rate control on user engagement and agency in a virtual reality game," *Virtual Reality*, vol. 24, no. 4, pp. 665–681, 2020.

[19] M. Labied and A. Belangour, "Automatic speech recognition features extraction techniques: A multi-criteria comparison," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 8, 2021.

[20] Y. E. Kheir, A. Ali, and S. A. Chowdhury, "Automatic pronunciation assessment - a review," in *Conference on Empirical Methods in Natural Language Processing*, 2023.

[21] H. Strik, K. Truong, F. de Wet, and C. Cucchiarini, "Comparing different approaches for automatic pronunciation error detection," *Speech Communication*, vol. 51, no. 10, pp. 845–852, 2009.

[22] Y. Kim, H. Franco, and L. Neumeyer, "Automatic pronunciation scoring of specific phone segments for language instruction," in *5th European Conference on Speech Communication and Technology (Eurospeech)*. ISCA, 1997, pp. 645–648.

[23] W.-K. Leung, X. Liu, and H. Meng, "Cnn-rnn-ctc based end-to-end mispronunciation detection and diagnosis," in *ICASSP 2019 - IEEE International Conference on Acoustics, Speech and Signal Processing*, 2019, pp. 8132–8136.

[24] Y. Gong, Z. Chen, I.-H. Chu, P. Chang, and J. Glass, "Transformer-based multi-aspect multi-granularity non-native english speaker pronunciation assessment," in *ICASSP 2022 - IEEE International Conference on Acoustics, Speech and Signal Processing*, 2022, pp. 7262–7266.

[25] J. Bailenson, K. Patel, A. Nielsen, R. Bajscy, S.-H. Jung, and G. Kurillo, "The effect of interactivity on learning physical actions in virtual reality," *Media Psychology*, vol. 11, no. 3, pp. 354–376, 2008.

[26] S. Lee, J. A. Mogle, C. L. Jackson, and O. M. Buxton, "What's not fair about work keeps me up: Perceived unfairness about work impairs sleep through negative work-to-family spillover," *Social Science Research*, vol. 81, pp. 23–31, 2019.

[27] S.-G. Lee, Y. D. Song, and E. C. Lee, "Experimental verification of the possibility of reducing photoplethysmography measurement time for stress index calculation," *Sensors*, vol. 23, no. 12, 2023.

[28] A. Aderhold, Y. Jung, K. Wilkosinska, and D. W. Fellner, "Distributed 3d model optimization for the web with the common implementation framework for online virtual museums," in *2013 Digital Heritage International Congress*, vol. 2, 2013.

[29] "Eleven labs introduction," https://elevenlabs.io/docs/product/introduction, [Accessed: 29-Dec-2024].

[30] "Mouth animation using rhubarb lip sync," https://en.esotericsoftware.com/forum/d/10769-mouth-animation-using-rhubarb-lip-sync, [Accessed: 29-Dec-2024].

[31] M. Ma. (2023, Nov.) Speech service ignite update – new enhancement features for pronunciation assessment. MSFT.

[32] O. Koraishi, "The intersection of ai and language assessment: A study on the reliability of chatgpt in grading ielts writing task 2," *Language Teaching Research Quarterly*, vol. 43, pp. 22–42, 08 2024.