READIFY: AI-ENABLED INTELLIGENT ASSISTANT TO IMPROVE READING AND COMPREHENSION SKILLS IN ENGLISH LANGUAGE

Final Group Thesis

Group ID: RP24-25-027

Supervisor: Dr. Dasuni Nawinna
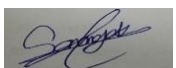
Co-Supervisor: Mr. Jeewaka Perera

Sri Lanka Institute of Information Technology

Sri Lanka

April 2025

DECLARATION

To the best of our knowledge and belief, this proposal does not contain any previously published or written material by another person, except where the acknowledgement is made in the text. We hereby declare that this is our own work and that no material previously submitted for a degree or diploma in any other university or Institute of higher learning has been incorporated without acknowledgement.

Name	Student ID	Signature
S. A. D. S. Kumarathunga	IT21118340	
M.D.A Sooriyaarachchi	IT21173790	
A.P Ranaweera	IT21182396	
W.G.B Senanayake	IT21158322	

ABSTRACT

Reading and vocabulary proficiency in English remain significant challenges for learners, particularly in contexts where English is a second or foreign language. This project, READIFY, presents an AI-enabled intelligent assistant designed to enhance English reading and comprehension skills through personalized, adaptive learning. The system leverages advanced Artificial Intelligence techniques including vocabulary CEFR-level prediction, machine learning, and Large Language Models (LLMs)—to assess learners' vocabulary and comprehension abilities in real time. Using Gradient Boosting and NLP-based feature engineering, the platform classifies student input according to CEFR standards (A1–C2) and dynamically generates gamified vocabulary exercises tailored to individual proficiency levels. Additionally, a Retrieval-Augmented Generation (RAG) framework integrates with Lang Chain and transformer-based models to deliver contextually relevant definitions and feedback via a vocabulary chatbot. For comprehension enhancement, LLMs are employed to automatically evaluate essay-type responses, simulate human-like grading, and generate adaptive content using prompt engineering. The system is deployed using a scalable microservices architecture, enabling real-time interaction and continuous improvement. READIFY demonstrates the transformative potential of AI in providing engaging, personalized, and equitable solutions for English language learning.

Keywords: AI in Education, CEFR Prediction, Vocabulary Learning, Adaptive Assessment, Generative AI, Reading Comprehension, Large Language Models (LLMs), English Language, Prompt Engineering, Automatic Quiz Generation, Retrieval Augmented Generation (RAG), Descriptive Answer Evaluation

ACKNOWLEDGEMENT

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Full Term
AI	Artificial Intelligence
ML	Machine Learning
NLP	Natural Language Processing
CEFR	Common European Framework of Reference for Languages
RAG	Retrieval-Augmented Generation
LLM	Large Language Model
TF-IDF	Term Frequency–Inverse Document Frequency
POS	Part of Speech
UI	User Interface
UX	User Experience
CNN	Convolutional Neural Network
DB	Database
API	Application Programming Interface
UAT	User Acceptance Testing
BERT	Bidirectional Encoder Representations from Transformers
GPT	Generative Pre-trained Transformer
ASR	Automatic Speech Recognition
G2P	Grapheme-to-Phoneme
DTW	Dynamic Time Warping
CAPT	Computer-Aided Pronunciation Training

1 INTRODUCTION

1.1 Background

In today's globally interconnected and digitally driven world, proficiency in the English language is more important than ever for academic, professional, and social success. Among the core language skills, vocabulary development and reading comprehension are especially critical, yet they remain some of the most challenging areas for learners, particularly in non-native English-speaking contexts. Traditional methods of instruction centered on rote memorization, decontextualized word lists, and passive reading exercises often fail to engage students or adapt to their individual learning needs. As a result, learners experience difficulty in retaining vocabulary and applying reading strategies effectively. The emergence of Artificial Intelligence (AI), particularly through Machine Learning (ML), Natural Language Processing (NLP), and Large Language Models (LLMs), presents a promising solution to these challenges. AI-enabled systems can assess learner proficiency, generate adaptive content, and provide real-time feedback, thereby offering a more personalized and engaging learning experience. This research project, **READIFY**, leverages these technologies to create an intelligent assistant that improves English vocabulary acquisition and reading comprehension by integrating CEFR-based assessment, gamified learning, and AI-powered essay evaluation in a dynamic, scalable educational platform.

Advanced comprehension skills are crucial for mastering the English language. These skills include the ability to understand the implied meaning of a text, evaluate and analyze the content, and make inferences and deductions based on the information provided. Essential components of these skills involve connecting different parts of the text, applying background knowledge, identifying the main idea, finding important facts and supporting details, summarizing content, and generating as well as asking relevant questions.

The importance of reading comprehension skills in English language learning cannot be overstated, especially in an era where information is readily available yet requires critical evaluation. English, as a global language, is a gateway to vast amounts of information, and proficiency in comprehension is key to accessing, processing, and applying this information effectively. However, despite its importance, many learners struggle with advanced comprehension tasks due to factors such as limited vocabulary, lack of engagement, and insufficient practice in higher-order thinking skills.

This challenge has sparked significant interest in leveraging Artificial Intelligence (AI) to improve reading and comprehension skills. AI-enabled platforms offer personalized learning experiences, adapting to individual learner needs and providing real-time feedback. These platforms can analyze learner behavior, identify weaknesses, and offer targeted exercises to strengthen comprehension skills. The effectiveness of AI in educational contexts has been widely studied, but its application to enhance reading comprehension in English, particularly in non-native speakers, remains an area ripe for exploration.

Even in our globalized digital age, English is still crucial for academic, professional, and intercultural opportunities. But one of the hardest areas for learners to master is vocabulary development which is a fundamental aspect of language learning. The lack of vocabulary limits reading comprehension, writing fluency, and oral communication [1].

Traditional vocabulary teaching tends to be memorizing from textbooks and repetitive exercises without consideration of the ability level or interests of each student. Static, lacking context, no engagement and zero ability to go at your speed? Consequently, students often remember almost nothing outside of tests and exams, and long-term language retention is low [2].

Gamification is an interesting way to make your learners engage better and be more motivated by your lessons. Gamified systems enhance engagement and enjoyment of learning tools to ensure better learning outputs via integration of game elements such as points, levels, leaderboards and rewards [3]. Gamification provides the opportunity in second language acquisition (SLA) for learners to access vocabulary with a degree

of meaningful engagement that strengthens retention through repetition and immediate feedback in a range of productive tasks [4].

Machine learning (ML) and natural language processing (NLP) two fields of artificial intelligence (AI) have also advanced the boundaries of changing vocabulary learning platforms. Machine learning (ML) models predict the CEFR level of a student based on the input text, while transformer-based models such as BERT and GPT allow dynamic content generation and instantaneous feedback [5]. To accomplish this, the authors utilized Retrieval-Augmented Generation (RAG) architectures that merge semantic search with generative capabilities to personalize vocabulary activities based on the data of the learner [6].

Hence, we introduce a new smart vocabulary learning framework that predicts learners' CEFR levels from input text and tailors' vocabulary games using a RAG pipeline. This solution combines CEFR classification (using ML), semantic retrieval (using pinecone), generation (using Gemini LLM), and wraps it all together in a responsive web app set up through React and Tailwind CSS. This system has the potential to combined linguistic ability and curiosity that can provide a scalable and customizable immersion experience for anyone looking to become proficient in a new language.

You need advanced level comprehension to really master the English language. These include deduction and inference, implied meaning questions as well as evaluation and analysis, from answering on unspoken aspects of the content, to the way in which the ideas and information in a text are presented. Integral to these skills are the abilities to establish connections between segments of the text, to use background knowledge, to recognize main ideas, to locate key information and supporting details, to summarize information, and to develop appropriate questions.

The significance of developing reading comprehension in English learning should not be overemphasized, especially in a time when information is easy to access but entails careful consideration. As the global language, English provides people with access to mountains the quantity of available information, and reading comprehension ability

3

is one of the keys for mastering the ability to access, digest, and utilize the information. However, even though it is so important, many students have a hard time with comprehension activities as they get further in grade levels because of things like not having the right vocabulary, not being interested, and because they haven't practiced high-level thinking skills.

Incorrect pronunciation can lead to misunderstandings, which may affect both the speaker's confidence and the listener's comprehension. This challenge is especially common in regions where English is learned as a second or foreign language. In such contexts, learners often have limited exposure to native speech patterns and may be influenced by the phonological rules of their first language.

Pronunciation is a fundamental aspect of language learning. It involves producing phonemes with the smallest units of sound that distinguish meaning between words. For example, the difference between *ship* and *sheep* depends on correctly articulating the vowel sounds /ɪ/ and /iː/ [21]. Mispronounced phonemes can drastically alter meaning, disrupting communication. This highlights the need for targeted training tools that help learners improve their phoneme-level pronunciation accuracy.

Traditional teaching methods include the use of the phonetic alphabet, transcription exercises, articulatory explanations, minimal pair drills, reading aloud, tongue twisters, and pronunciation games. These strategies aim to train students in sound recognition and production. While some learners benefit from these approaches, many struggle to make lasting improvements through conventional techniques alone. As a result, more innovative solutions are being explored [22].

One such innovation is Computer-Aided Pronunciation Training (CAPT). These systems use speech recognition technology to analyze learners' spoken input and deliver real-time feedback. However, many current CAPT systems focus mainly on broad features such as stress, rhythm, and intonation, and lack the precision to identify phoneme-level errors.

This research proposes the development of an advanced CAPT system that combines speech recognition tools with phoneme-level analysis. The system will convert spoken

words into a sequence of phonemes and detect mispronunciations at this granular level using a grapheme-to-phoneme model.

Moreover, the system will integrate large language models (LLMs) to generate customized training words based on the user's specific pronunciation errors. These LLMs will produce practice materials that closely match the problematic phonemes, giving learners targeted and effective exercises to improve their spoken English.

1.2 Literature Review

1.2.1 Advanced Comprehension Skills Enhancement Module

The integration of AI in education, particularly for enhancing reading and comprehension skills in English, has gained considerable traction in recent years. This review explores the current state of research in this area, focusing on AI-enabled platforms designed to dynamically generate adaptable content, automatically assess essay-type responses, and provide feedback. By analyzing key studies in this domain, we aim to contextualize the advancements and challenges in utilizing AI for educational purposes, with a specific emphasis on improving English language proficiency.

Personalized Content Generation based on User Interest and Skill Level

Several studies have explored AI's potential in personalizing educational experiences. Laban et al. (2022) [7], has investigated automated question generation to support educators in quiz design. Their tools assist teachers in creating customized assessments. However, as noted in their work and exemplified by protocols like the Quiz Design Task, their methodology often relies on teacher-defined topics. This approach, while beneficial for pedagogical support, does not directly address personalization driven by individual student curiosity or specific topics of personal interest to the learner.

Thotad, Kallur, and Amminabhavi (2022) [8] presented an Automatic Question Generator designed to create semantically accurate and syntactically cohesive questions, primarily focusing on generating multiple-choice questions (MCQs) with correct answers and distractors from input text provided by teachers. Their system aims to help educators quickly assess student comprehension and allows students to evaluate their own understanding. While effective for generating factual questions from teacher-provided content, this approach is primarily designed as a tool for educators and focuses on factual question types (who, when, where, why, what), rather than being driven by the specific, potentially diverse, interests of individual students or generating more descriptive or open-ended question formats.

Lu et al. (2023) [9] introduced ReadingQuizMaker, a human-NLP collaborative system aimed at assisting instructors in designing high-quality reading quiz questions. Their system integrates NLP support into the instructor's workflow, allowing teachers to decide when and how to use AI models and edit the generated outcomes. This collaborative approach was praised by instructors for its ease of use and helpful suggestions, highlighting the importance of user control in AI-assisted creative tasks. However, ReadingQuizMaker requires significant human involvement (the instructor) in the question design process. In contrast, a system focused on automated quiz generation based directly on user interest aims for a higher degree of automation from the learner's perspective, minimizing the need for intermediary human editing or selection.

More recent work, such as that by Contreras-Arguello et al. (2025) [10], has explored integrating NLP techniques to automatically generate both summaries and questions from text provided by the teacher. Their system aims to save teachers time in creating educational resources, allowing them to focus on supporting students with greater reading comprehension needs. Similar to the aforementioned studies, this system's input text is provided by the teacher, and while it supports the reading comprehension process, the content generation is not directly initiated or guided by the specific, varied interests of the individual student.

Automated Assessment and Feedback Generation

The second major challenge is the absence of reliable tools for the automated assessment of essay-type questions, which traditionally require human evaluation due to the complexity and subjectivity involved in grading. This area has seen considerable research, yet existing solutions remain inadequate for high-stakes educational environments.

Moholkar et al. (2024) [11] provide a comprehensive survey of machine learning techniques for evaluating descriptive answers. Their survey covers various approaches, including supervised learning models and ensemble methods, highlighting their effectiveness in grading to some extent. However, the authors acknowledge the limitations in these models' ability to handle the nuances of human language, such as

context, tone, and rhetorical devices, which are crucial for evaluating essay-type questions. The models surveyed often struggle with maintaining consistency and fairness in grading, especially in cases requiring deep contextual understanding.

More recent studies have begun to explore the use of LLMs themselves for automated assessment. Xia et al. (2024) [12] examined the effectiveness of LLMs for automated essay scoring, specifically using the TOEFL Independent Writing Task. While their findings indicated promise, they also identified drawbacks, including the models' susceptibility to being overly influenced by superficial features like essay length or vocabulary richness, and a general lack of transparency regarding the rationale behind their scores. Furthermore, their work primarily focused on the assessment of writing tasks (essays), which presents different challenges compared to evaluating responses to reading comprehension questions, which requires assessing understanding of an external text rather than original composition.

Similarly, Hussein et al. (2019) [13] reviewed the evolution of automated essay scoring systems, from earlier rule-based methods to more advanced ML approaches. They concluded that despite considerable progress, consistently replicating human-like grading, particularly for complex tasks like evaluating creative or argumentative essays, remains an elusive goal. This suggests that traditional ML paradigms may inherently struggle with the subjective and complex aspects of grading free-text responses, pointing towards the need for more advanced computational strategies. Their work underscores the limitations of methods that do not fully harness the deep linguistic understanding offered by modern LLMs.

This research intends to overcome these limitations by developing a novel assessment tool that combines the strengths of machine learning with advanced linguistic analysis to better emulate human evaluators. This tool will focus on understanding the content and structure of essays more deeply, including context, argumentation, and rhetorical devices, thereby improving grading accuracy and fairness. Moreover, by integrating this tool with the dynamically adaptable content system, the platform would not only assess but also provide immediate, actionable feedback, thereby enhancing the learning process in real-time.

1.2.2   AI-Powered Gamified English Vocabulary Improvement Module

Importance of English Vocabulary Proficiency Worldwide

Its international recognition as a common language is one of the reasons why English is called the world. A Language that can be spoken from country to country, culture to culture, and a discipline to its particular population. English is either the official language or one of the official languages in over 60 countries, and it serves as the main language of business, science, technology, higher education, and diplomacy [14]. With the rapid acceleration of globalization, English has become an essential tool for success at home and abroad.

English as a second or foreign language is a requirement for learning and being part of these global education systems and workforces, using digital media and participating in society. In this sense, the mastery of vocabulary is critical in language skills. Vocabulary knowledge has been found to be one of the most important predictors of reading comprehension, speaking fluency, and among the extensive literature to the best predictor of writing performance and the research simply do not support the generally accepted belief in SLA that the mother tongue sexually knowledge of transparent languages differs from L1 to be using more general cognitive abilities. Without an adequate vocabulary learners are frequently incapable of comprehending or making worthwhile use of communication, however grammatically accurate it may be.

Yet vocabulary acquisition is problematic for many learners of English as a second language. These problems are related to instructional strategies that focus on rote memorization of decontextualized word lists, rather than on how words are used in context [15]. Vocabulary learning in traditional classroom-based settings may not cover depth of lexical knowledge such as collocations, connotation and pragmatic uses that are essential for real-life communication [16]. In addition, the teaching of lexicon is often underemphasized in language teaching programs, thus leaving students without the key knowledge to promote long-lasting learnability and flexible use of language [17].

Traditional Vocabulary Learning Methods and Their Limitations

Vocabulary has always been one of the central components of second language learning, but to comprehend, express and use language fluently, the ability to learn vocabulary is crucial. However, these conventional ways for teaching vocabulary, i.e., committing to memory of word lists, dictionary consultation and work book exercises, have apparent drawbacks to induce profound, long-term learning [2].

Often, these approaches focus on the receptive side of knowing a word (e.g., understanding a word and its meaning) rather than on the productive side of knowing a word (e.g., using the word accurately in a sentence), which is the main concern in everyday communication. Students are usually expected to memorize vocab out of context with no real connection to how words are used in natural language. Hence, they are good on written tests, but find it hard to remember and use vocabulary in speaking and writing [14].

Yet another problem with old-style vocabulary teaching is that it is a one-size-fits-all approach. The tools tend to have a single level of proficiency, which is unrealistic when considering the vast differences in ability, learning pace and style between learners in a classroom. Beginners who need repetition and more time to learn vocabulary will be lost, and suffer from lack of confidence, while higher level students will lose motivation and disengage if there's not enough variety or challenge [17].

In addition, these conventional approaches do not have an active involvement and individualized feedback, which has been proven to improve vocabulary recall. With no input, no reinforcement, and no meaningful activation, we forget a lot of the vocabulary very quickly. Research in cognitive psychology has found that learning in context, spaced repetition, and multisensory engagement are better ways of retaining information and transferring skills [18]. These restraints are even more apparent in the digital age, where today's learners increasingly demand interactive, dynamic and individualized forms of instruction. This gulf between old school teaching and current digital learner appetites highlights a demand for creative solutions that can actively drive vocabulary learning to become engaging, adapted to learner needs and relevant to their wider context.

Overview of AI Techniques Used in English Vocabulary Learning

Introduction Artificial Intelligence (AI) is transforming education on the back of smarter and more adaptive learner-centric tools. AI tools and technologies especially ML and NLP helps educational tools in monitoring student behavior and personalize learning experience. Such tools can be quite effective in language learning environments, where they can aid in reading comprehension, writing enhancement, pronunciation support, and vocabulary acquisition [19].

Unlike traditional systems that use pre-set content, and linear instruction, the AI systems are responsive to how the learner is doing. They follow progress, see where students are strong and weak and change challenge levels on the fly. This movement toward personalized learning has since paved the way for enhanced vocabulary instruction, in particular, for second-language learners.

Machine Learning and Gradient Boosting for Text Vocabulary Level Prediction

Within supervised learning algorithms, gradient boosted trees have emerged as popular because they are known for their ability to work well with complex classification problems. It works by constructing an ensemble of weak learners (usually decision trees) where each successive model concentrates on the residuals (errors) of the former, making the prediction error decrease as time goes [20]. For vocabulary learning applications, Gradient Boosting can be used to predict the Vocabulary CEFR level class (e.g., A1, A2, …, or C3) of a learner from linguistic features extracted from written input.

A key part of this process is feature engineering - the transformation of raw learner text into a structured set of features appropriate for model training. The following types of features in particular have been found useful for the prior prediction of CEFR levels [21]. As demonstrated in Table 1.2, there are different linguistic features that are used to predict a learner's language proficiency. A Gradient Boosting model can then be trained on such features from a labeled data set to predict a student's proficiency accurately in real time. These predictions can in turn suggest personalized content like word games, vocabulary builder exercises and contextual reading materials.

Deep Learning Models and Transformer-Based Language Models

 In recent years, the field of Natural Language Processing (NLP) has seen an upsurge by the substantial progress in deep learning, notably using transformer-based models. Transformers such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer) are able to capture complex language patterns, context and linguistic relations in the text [21]. Because of their bidirectional processing, they can be used to produce or to comprehend language with very good performance.

These models are increasingly being employed in educational contexts to give natural-sounding feedback, auto-generate questions, improve grammar, and even serve as tutors. For students learning new vocabulary, the use of transformer models to produce context-essential examples, synonyms, and definitions in an adaptive manner to learners' skill leads to increased engagement and retention.

 Retrieval-Augmented Generation (RAG) and Its Applications

One notable improvement regarding AI-supported vocabular y learning is Retrieval-Augmented Generation (RAG), a hybrid model joining retrieval-based methods with generative language models to provide context-sensitive and pedagogically informed output [6]. RAG robustly integrates information from external sources (e.g., semantic vector data) with a transformer-based generator to produce accurate, meaningful, and transferable results. For teaching vocabulary, the cluster generating approach allows for the creation of graded, dynamic learning material like word meanings, authentic use in a context, examples, definition with context, and meaning with context. In this paper, we used a multi-agent RAG system, with each agent to fulfil certain pedagogical function, including task generator for gamified vocabulary tasks, hints/feedback provider, word meaning retriever, as well as the adjuster to the user's level and preferences for themselves. Based on the above architecture, as developed by the likes of LangChain Pinecone, our approach supports scalable personalization and offers the guarantee that every learner is presented content that is both linguistically appropriate, educationally beneficial, and offered within a game-based environment. Muti-agent RAG is designed to combine word learning with cognitive load theory and also adaptive learning, and not only brings interactivity to students, but also provides

profound and long- term vocabulary recall1. This makes it an effective support tool for second language learners, the latter who need constant feedback, context examples and motivation to keep the learning process going.

Overview of CEFR Standards for Language Proficiency Classification

The Common European Framework of Reference for Languages (CEFR) is an international standard for describing language ability at six levels (A1 to C2) which represent a learner's communicative skills in reading, writing, listening, and speaking. For vocabulary learning, CEFR is key to deciding not just the amount, but how complex, contextually appropriate and how deep vocabulary should be that is expected to be produced and comprehended at each level [23]. For example, at the A1 level learners can make use of high-frequency concrete words, whereas at the C1/C2 levels learners should be able to deal with idiomatic expressions, abstract vocabulary and sophisticated collocations [16]. Therefore, combining CEFR ratings in AI-based platforms facilitate personalized vocabulary tasks with dynamic content difficulty tailoring the content to the learner's proficiency level. Automated CEFR-level prediction based on lexical richness, syntactic complexity, and semantic depth now facilitates personalized learning pathways and vocabulary selection. [21]

1.2.3 Basic Comprehension Skills Enhancement Module

So good literacy in English is generally acknowledged as being an essential skill with a whole knock-on effect, regarding successful school, work and social lives. Consensus from numerous studies and educational reports confirms that proficiency when reading in English is a key factor, not only for mastering the language, but for success in a wide variety of fields of study and careers. Skilled reading opens the doors to understanding complex information, critical analysis of ideas and critical thinking, all of which are necessary for success in school and beyond.

It also develops vocabulary skills, grammatical understanding, and cognitive fluidity including both memory and attention, and problem-solving skills. Because most academic articles, research, and professional documentation is in English, having advanced reading ability, provides access to more information and educational resources. This access is not limited to academia; rather it follows students into the workplace where proficiency in English is a requirement for communication, collaboration, and advancement in global professional fields [6]

Aside all the practical benefits, learning English to read also opens the door to self-development. Through diverse perspectives and cultures, it encourages self-belief, emotional intelligence and empathy in readers. Advanced reading in the English language trains curiosity, flexibility, and a passion for learning - skills that are more and more in demand in fast changing times. Another reason that reading is such an essential skill for everyone, is that relatively strong reading skills correspond to better writing, speaking, listening, interpretive and other abilities that promote greater effectiveness both personally and professionally.

To sum up, it is evident from the literature that English reading ability is not just a scholastic necessity but also a vital life necessity. It enables people to learn, to participate in society and to take advantage of opportunities in the digital era

1.2.4 Phoneme-Level Speech Error Detection Module

Traditional Vocabulary Learning Methods and Their Limitations

Traditional pronunciation training has historically emphasized achieving "native-like" articulation, with a strong focus on accuracy and clarity. It relied on structured methods such as minimal pair drills, phonetic transcription, and contextualized exercises to help learners identify and produce target sounds effectively. Commonly, this instruction followed one of three pedagogical models. The Analytic-Linguistic Approach, which used tools like the International Phonetic Alphabet (IPA) to break down and study individual sounds. The Intuitive-Imitative Approach, which prioritized listening to native models and imitating them and the Integrative Approach, which combined pronunciation practice with other language skills through real-life communication tasks. Core techniques included phonetic training, listening and imitation, reading aloud, and the use of tongue twisters and visual aids. These methods provided foundational support for pronunciation learning [23].

However, these approaches also had limitations. They often prioritized accuracy over fluency, focused narrowly on segmental features (individual phonemes) rather than suprasegmentals like stress and intonation, and provided limited opportunities for real-world or spontaneous practice. Additionally, they lacked personalized feedback, particularly at the phoneme level, making it difficult for learners to recognize and address their specific pronunciation challenges [24].

The Role of Computer-Aided Pronunciation Training (CAPT)

Introduction Computer-Aided Pronunciation Training (CAPT) plays a significant role in helping language learners improve their pronunciation through interactive, technology-driven methods. CAPT systems provide individualized, instant feedback using advanced technologies like automated speech recognition (ASR), allowing learners to focus on phonemes, intonation, and stress patterns effectively. CAPT has proven particularly beneficial for non-native speakers, as it offers a cost-effective and scalable way to enhance pronunciation skills compared to traditional methods [25].

Research shows that CAPT improves learners' pronunciation skills by helping them practice repeatedly and receive real-time corrections. Moreover, these systems are found to be highly motivating, especially when integrated into personalized learning environments [26]. CAPT tools also emphasize the importance of intelligibility over perfection, aiding learners in achieving effective communication [27]. The proposed research seeks to develop a CAPT system that addresses these limitations by integrating phoneme-level analysis and real-time feedback. This system will utilize advanced speech recognition technologies to convert spoken words into their constituent phonemes, identify pronunciation errors, and provide immediate feedback on the specific sounds that need improvement. By focusing on the phoneme level, learners can target the precise areas where they struggle, rather than receiving generalized feedback.

Current Approaches to Pronunciation Error Detection and Feedback Mechanisms

Modern pronunciation training systems, particularly those embedded in popular language learning platforms have made strides in integrating speech technologies for user interaction. However, their error detection and feedback mechanisms remain largely superficial. Most current approaches rely on automatic speech recognition (ASR) to compare a user's spoken input against a predefined word or sentence model. If the utterance deviates beyond a certain threshold, the system typically provides a binary result labeling the pronunciation as either "correct" or "incorrect" without indicating what went wrong or how to improve [28]. While this kind of feedback may be easy to interpret, it lacks the granularity required to support meaningful learning, especially for learners struggling with specific phonemes.

These systems often do not identify the specific phoneme-level deviations, nor do they provide insights into articulatory issues such as stress misplacement, intonation errors, or substitution of individual sounds [29]. As a result, learners may receive a negative assessment without understanding whether the problem was a missed vowel quality, a

misarticulated consonant, or a rhythm issue. This lack of diagnostic feedback limits the learner's ability to correct errors effectively. Moreover, the absence of targeted practice recommendations or personalized support makes it difficult for users to improve over time. While some platforms attempt to offer pronunciation guides or native recordings for comparison. They fall short of providing interactive, corrective guidance at the phoneme level. Which research shows is critical for long-term pronunciation development [30].

These limitations have highlighted the need for next-generation systems capable of offering detailed, real-time, and phoneme-specific feedback. The goal is to move beyond judgment-based outputs and toward constructive, educational feedback that identifies exact mispronunciations.

Phoneme Recognition and Error Identification in English Pronunciation

Phoneme recognition is a critical aspect of pronunciation training. Existing CAPT systems that use Automatic Speech Recognition (ASR) technology typically compare a learner's spoken input to a predefined word model, delivering feedback at the word level. While this is useful for general language learning, it does not offer the granularity needed to correct specific sound errors. This limitation is particularly significant for English learners, where even small phoneme-level mistakes can change the meaning of a word [31].

The proposed system will provide advanced phoneme recognition algorithms to break down words into individual phonemes and identify errors at this level. The system will detect errors and provide corrective feedback. This precise, phoneme-level feedback allows learners to focus on correcting individual sounds rather than grappling with entire words or sentences.

In this research, the proposed system adopts a fine-grained approach to pronunciation evaluation by segmenting spoken input into phonemes and comparing them directly with the expected phonemic output of a given word. The process begins with automatic speech recognition (ASR), which transcribes user speech. This transcription is then

analyzed using a Grapheme-to-Phoneme (G2P) model to generate both the expected and actual phoneme sequences. By comparing these two sequences, the system can identify mismatches that correspond to phoneme-level pronunciation errors. To measure the similarity between the spoken and expected phoneme strings, the system uses Dynamic Time Warping (DTW) an algorithm well-suited for aligning sequences with temporal variation [32], [33]. DTW not only identifies insertions, deletions, and substitutions of phonemes, but also quantifies the overall pronunciation distance, offering a metric for learner progress.

Furthermore, this research incorporates articulatory-aware analysis using tools such as PanPhon, which compares the feature-based representation of phonemes rather than just their symbolic labels. This allows the system to determine if an error was due to a change in voicing (e.g., /t/ → /d/), place of articulation (e.g., /t/ → /k/), or manner (e.g., /s/ → /ʃ/), offering learners detailed insights into the nature of their mistakes [34]. By identifying which exact sound was mispronounced and explaining how it differs from the target sound in terms of articulatory features, the system delivers constructive, phoneme-specific feedback.

This level of detail is essential for developing effective pronunciation habits, particularly in English where phonemic precision is vital for intelligibility. For example, failing to differentiate between /iː/ and /ɪ/ in "sheep" and "ship" can drastically alter meaning. By focusing on phoneme-level recognition and correction, the system not only enhances learners' awareness of subtle sound distinctions but also empowers them to self-correct through guided feedback and practice. This is further strengthened by the integration of Large Language Models (LLMs), which dynamically generate phonetically similar practice words based on the user's mispronounced phonemes, facilitating targeted repetition and reinforcing proper articulation. In essence, this phoneme recognition and error identification framework transforms pronunciation training from passive correction to interactive diagnosis and guided improvement, closing the feedback gap that exists in many traditional and commercial language learning systems.

18

1.3   Research Gap

Despite significant advancements in AI-enabled educational tools, a comprehensive system that integrates real-time multimodal feedback, personalized content generation, phoneme-level speech assessment, and CEFR-based adaptive progression is still lacking. Existing solutions often address these features in isolation focusing either on reading comprehension, vocabulary enhancement, or pronunciation but do not provide a unified platform capable of dynamically tailoring learning experiences across all three core English skills. This gap highlights the need for an AI-powered intelligent assistant that harmonizes speech, reading, and vocabulary learning using LLMs, emotion recognition, gamified learning, and personalized feedback mechanisms to maximize learner engagement and performance.

1.3.1   Individual Research Gaps

1. **IT21158322 -: Basic Comprehension Skills**

   - *Gap:* Lack of integration between personalized content and AI-driven adaptivity, including real-time emotion recognition and performance analytics. Existing systems use static content, offer limited adaptivity, and fail to account for cognitive and emotional learner states in shaping comprehension experiences

2. **IT21182396 - Pronunciation Assistant**

   - *Gap:* Existing pronunciation tools do not provide phoneme-level feedback or real-time correction. Most systems only assess pronunciation accuracy at a word/sentence level and lack the ability to offer personalized exercises based on specific phoneme errors

3. **IT21173790 - Vocabulary Learning**

   - *Gap:* Most vocabulary tools lack integration between CEFR-level prediction, gamified learning, and AI-generated personalized content. They use static word lists and don't adapt in real time to individual learner input or progress

4. **IT21118340 – Advanced Reading Comprehension**

   - *Gap:* There is a lack of systems capable of dynamically generating comprehension content based on learner interest and skill, along with

automated descriptive-answer evaluation. Most platforms focus on MCQs and neglect open-ended, human-like answer assessment

1.4   Research Problem

Despite the growing availability of AI-powered tools for English language learning, most existing solutions remain limited in scope, addressing individual skills such as vocabulary, pronunciation, or reading comprehension in isolation. These systems typically rely on static content, lack real-time adaptivity, and do not offer integrated support for phoneme-level pronunciation correction, descriptive answer evaluation, or CEFR-aligned progression. Moreover, they fail to provide personalized learning experiences based on learners' emotional states, engagement levels, or topic preferences. As a result, students face fragmented and disengaging learning experiences that do not fully support their development across key areas of language proficiency. There is a clear need for a unified, intelligent assistant that combines dynamic content generation, gamified learning, real-time feedback, and multi-skill integration to deliver a personalized and effective English language learning environment.

1.5 Research Objectives

1.5.1 Main objective

The main objective of this project is to design and develop READIFY, an AI-powered intelligent assistant to support and enhance English reading comprehension and vocabulary learning for Sri Lankan high school students. The system leverages modern artificial intelligence techniques, such as CEFR-based language assessment, gamification, retrieval-augmented generation (RAG), and large language models (LLMs), to deliver personalized, engaging, and adaptive learning experiences that improve vocabulary retention and reading proficiency.

1.5.2 Sub Objectives

- **Sub-Objective 01 (M.D.A Sooriyarachchi-IT21173790):** This study is to develop an AI-powered gamified English vocabulary improvement module tailored for Sri Lankan high school students. This module will allow students to input challenging vocabulary either as individual words or within short written passages which will be analyzed to assess their vocabulary proficiency level. Using this input, the system will predict the learner's CEFR level through machine learning and generate personalized educational games and interactive learning materials. The module leverages large language models (LLMs), vectorization techniques, and a multi-agent retrieval-augmented generation (RAG) framework to adaptively provide vocabulary games, definitions, real-time feedback, and hints that align with the learner's skill level and preferences promoting sustained engagement and long-term vocabulary retention.

- **Sub-Objective 02 (S. A. D. S. Kumarathunga- IT21118340):** This research aims to design, develop, and evaluate a web-based intelligent assistant aimed at improving advanced English comprehension skills—specifically focusing on sequencing, summarizing, and self-questioning—among language learners. To achieve this, the research will first utilize a large language model (LLM) to generate dynamic, personalized learning content—such as lessons and exercises—tailored to each learner's interests and proficiency level. This will involve formulating effective prompts and integrating Retrieval-Augmented Generation (RAG) techniques to enhance the contextual relevance and depth of the generated

material. Secondly, the research aims to leverage LLMs to evaluate learner responses and deliver automated, criteria-based feedback without human intervention. This will require engineering precise prompts and implementing RAG integration to improve evaluation accuracy and feedback quality. Together, these objectives will enable the creation of an intelligent, adaptive learning platform that enhances learner engagement and fosters deeper comprehension through immediate, personalized support.

- **Sub-Objective 3 (W.G.B Senanayake- IT21158322):**

    Dynamic MCQ and Paragraph Generation

    Develop a fine-tuned LLM pipeline (T5, Gemini, DeepSeek) to generate level-specific reading passages and context-aware MCQs tailored to individual proficiency (CEFR A1–C2). Integrate Kaggle-standard datasets (e.g., RACE, SQuAD) and custom educational content to ensure pedagogical rigor and diversity. Use prompt engineering (e.g., few-shot learning, task-specific prefixes) to optimize question quality, distractor plausibility, and alignment with learning goals[20].

- **Sub-Objective 4 (A.P Ranaweera-IT21182396):** To design and implement an intelligent speech error detection and feedback system that functions at the phoneme level to support English pronunciation learning. The system will analyze the learner's spoken input, accurately detect mispronounced phonemes, and provide personalized corrective feedback in real time. Its primary aim is to guide learners by highlighting specific pronunciation errors and offering focused practice to improve those areas. Additionally, the system will generate personalized word suggestions based on the phonemes that were mispronounced. This will allow learners to practice similar-sounding words and reinforce correct pronunciation through targeted repetition.

1.6   Integrated System Diagram
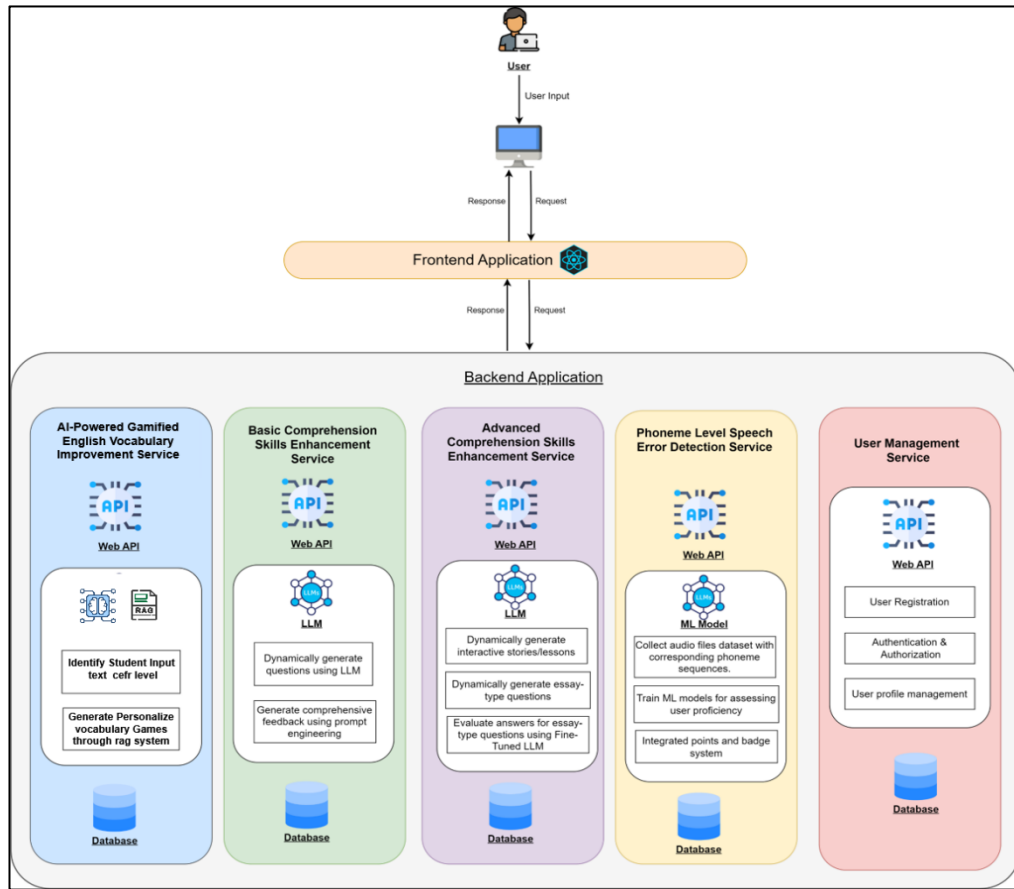
Figure 1.1 System Diagram

The overall architecture of the READIFY system is based on a modern microservices approach, ensuring scalability, modularity, and ease of maintenance. Each major component of the system such as vocabulary improvement, comprehension skills (basic and advanced), speech error detection, and user management—has been developed as an independent backend service using RESTful APIs. These services are containerized using Docker and deployed on AWS ECS (Elastic Container Service), following DevOps best practices with CI/CD pipelines implemented via GitHub Actions. This setup supports automated testing, image building, and deployment workflows. The frontend of the system is built using React.js and hosted on Netlify for continuous and efficient delivery. The architecture allows seamless communication between the frontend and backend services, providing users with a responsive and interactive experience. Additionally, each service leverages domain-specific technologies, including large language models (LLMs), machine learning (ML), and vector databases to deliver personalized and intelligent educational content. This

23

design ensures flexibility for future expansion while maintaining performance and reliability across different user modules.

2 INDIVIDUAL RESEARCH CONTRIBUTION And METHODOLOGY

2.1 Contribution by M.D.A Sooriyaarachchi (IT21173790) – Gamification of Vocabulary Learning

Component: AI-Powered Gamified and Personalized English Vocabulary
Improvement Module

As part of the integrated research project titled READIFY: AI-ENABLED INTELLIGENT ASSISTANT TO IMPROVE READING AND COMPREHENSION SKILLS IN ENGLISH LANGUAGE, this component was built to enhance vocabulary learning through personalized, adaptive educational games that respond to the learner's CEFR level and personal interests. It integrates natural language processing, machine learning, Retrieval-Augmented Generation (RAG), and software engineering best practices to deliver intelligent, engaging, and scalable vocabulary instruction.
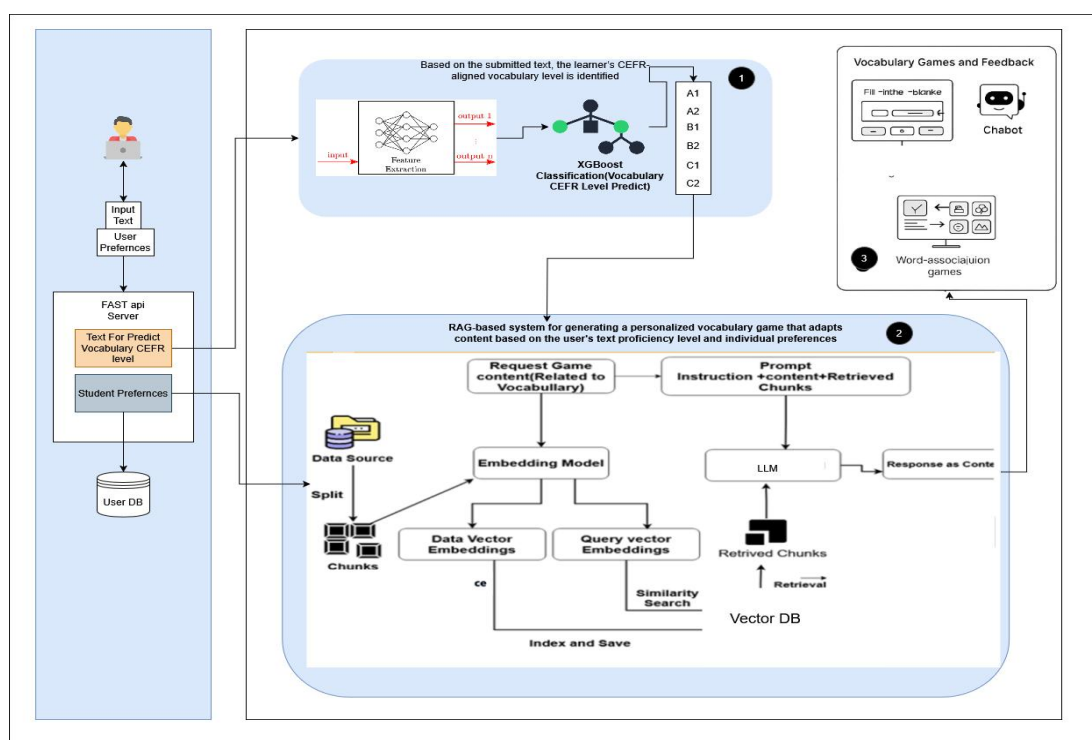
2.1.1 Component System Diagram

Figure 2.1 Component System Diagram

This component is concerned with the assessment of the learners in input text vocabulary proficiency level and the generation of the personalized vocabulary games.

In this component the tasks are CEFR level prediction and game content generation with the RAG architecture to direct the development process.

First, for the purpose of the interaction of the users, developed a user-friendly, reliable and accessible Web application using React JS. There are three sub-processes under the main objective for easy understanding and implementation purposes, as illustrated in Figure 2.1 Vocabulary CEFR Level Prediction Module (Block 1 in figure 2.1) The first core block of the proposed system is the Vocabulary CEFR Level Prediction Module. This module works with learner submitted input data in the form of problem sentences or paragraphs and processes the input using a feature extraction process to do comparative analysis using various linguistic metrics such as sentence structure, word frequency distribution and part of speech distribution. These features are fed to a pre-trained XGBoost classifier, fine-tuned on predicting the CEFR language levels (A1–C2) of the words. The most likely proficiency level, conditioned on what is input, is spat out by the classifier. This is a core prediction of the system and informs the challenge and range of content for later vocabulary game generation.

The second key feature denoted as (Block 2 in the figure 2.1) is RAG-based Personalized Game Generator. This part generates vocabulary games appropriate to the learner's CEFR level (Common European Framework of Reference) and topic preferences (e.g., sports, science). With Lang-Chain as conductor, the player's query is projected onto semantic vectors using a sentence transform model. These vectors are used to search a Pinecone Vector Database that includes pre-indexed words, definitions, and examples in context. Once the relevant data is fetched, it is processed through a Large Language Model (LLM) to generate the complete educational game content (fill-in-the-blanks games, word association game, word definition chatbot) based on the combined context & learner input. This smart content creation is tailored to be relevant and interesting for each user.

The last part coded as (Block 3 in figure 2.1) in the system architecture is the Interactive Vocabulary Games and Feedback Module that is mainly developed at the front end utilizing React JS and styled with Tailwind CSS. This module is the main user interface that users use to interact with the customized games that the system creates. The skin focuses on usability and response time, to make sure everything runs

smoothly on all devices. In the course module, students can use a number of vocabulary-based activities, e.g. fill-the-blank exercises, word-definition matching and word association games according to their levels of knowledge. An integrated RAG chatbot supports two-way interaction in real time, where definitions, explanations, and context can be requested as needed. The module also includes active feedback loops: if a user is failing to figure something out, the system is alerted to the pattern and adjusts to provide easier tasks or more explicit hints. This intelligent process loop is generated on the frontend for learning vocabularies more interactive, customized, and efficient.

2.1.2   Specific Role and Responsibilities

The contributor led the entire lifecycle of the vocabulary learning module. Key responsibilities included:

- **Proficiency Estimation:** Developed a CEFR level classifier using NLP and ML to analyze student-submitted text and predict vocabulary proficiency (A1 to C2).
- **Game Design:** Designed and implemented interactive vocabulary games aligned to learner proficiency and interest domains, such as technology, sports, or health.
- **RAG Integration:** Built a RAG-based backend using Pinecone and Lang Chain for real-time semantic retrieval of vocabulary entries.
- **Content Generation:** Leveraged the Gemini LLM to dynamically generate fill-in-the-blank questions, word associations, and MCQs.
- **Chatbot Development:** Implemented a chatbot for vocabulary support, capable of giving real-time explanations, synonyms, and usage examples.
- **Web Deployment:** Developed the full-stack application using React.js, Fast API, and hosted services on Render.com.
- **Software Engineering Practices:** Applied microservice architecture for modular deployment, managed version control using Git and GitHub, created Docker containers for each service, implemented CI/CD pipelines for continuous delivery using GitHub Actions, and deployed production-ready services on AWS ECS.

- **Quality Assurance:** Integrated Sonar Cloud for static code analysis and quality assurance and wrote automated test cases to ensure reliability and maintainability.

2.1.3 Sub-Objective Achievement

The sub-objective *"Generate Personalized and Adaptive Vocabulary Games"* was achieved through:

- **Linguistic Feature Extraction:** Used spaCy, Textstat, and Regex to derive lexical richness, syntactic depth, and semantic density from learner text.
- **CEFR Classification:** Trained an XGBoost model on CEFR-aligned datasets (EFCAMDAT, English Profile) to determine learner level.
- **Context-Aware Retrieval:** Applied LangChain and Pinecone to semantically retrieve vocabulary entries relevant to the learner's interest and proficiency.
- **Game Personalization:** Created educational games (MCQs, fill-in-the-blank, word associations) tailored dynamically using the Gemini LLM.

2.1.4 Tools and Methods Used

- **NLP & Feature Engineering:** spaCy, Textstat, Regex
- **Machine Learning Model:** XGBoost trained on CEFR-labeled datasets
- **Embeddings:** all-MiniLM-L6-v2 from Sentence Transformers
- **RAG Framework:** Pinecone (vector DB) + Lang Chain + Gemini (LLM)
- **Frontend Development:** React.js, Tailwind CSS
- **Backend & API:** Fast API, Firebase
- **Deployment Platforms:** Azure App Services, Render.com, AWS EC2, AWS S3
- **DevOps & Engineering:** Docker, GitHub, GitHub Actions (CI/CD), Sonar Cloud, Unit Testing

2.1.5 Outcomes and Impact

- **High Accuracy CEFR Detection:** Achieved over 69% accuracy in real-time CEFR level prediction.

- **Personalized Game Generation:** System created varied vocabulary games tailored to learner level and interest.

- **Autonomous Vocabulary Assistance:** Developed a chatbot for real-time word explanations and learner queries.

- **Increased Engagement:** Adaptive, gamified design resulted in improved vocabulary retention and learner motivation.

- **Production-Level Quality:** System adhered to modern software engineering practices, supporting scalability, modularity, and maintainability.

- **Deployment Ready:** The module was tested for scalability, responsiveness, and integration with other READIFY components.

This contribution demonstrates a robust implementation of gamified, AI-powered vocabulary learning through adaptive technology and sound engineering practices. The system not only personalizes content but does so using cutting-edge techniques in NLP, ML, RAG, and cloud-native software engineering, making it a critical component in the broader READIFY framework to support English language acquisition.

2.2 Contribution by S. A. D. S. Kumarathunga (IT21118340) – Advanced Comprehension Skills Enhancement Module

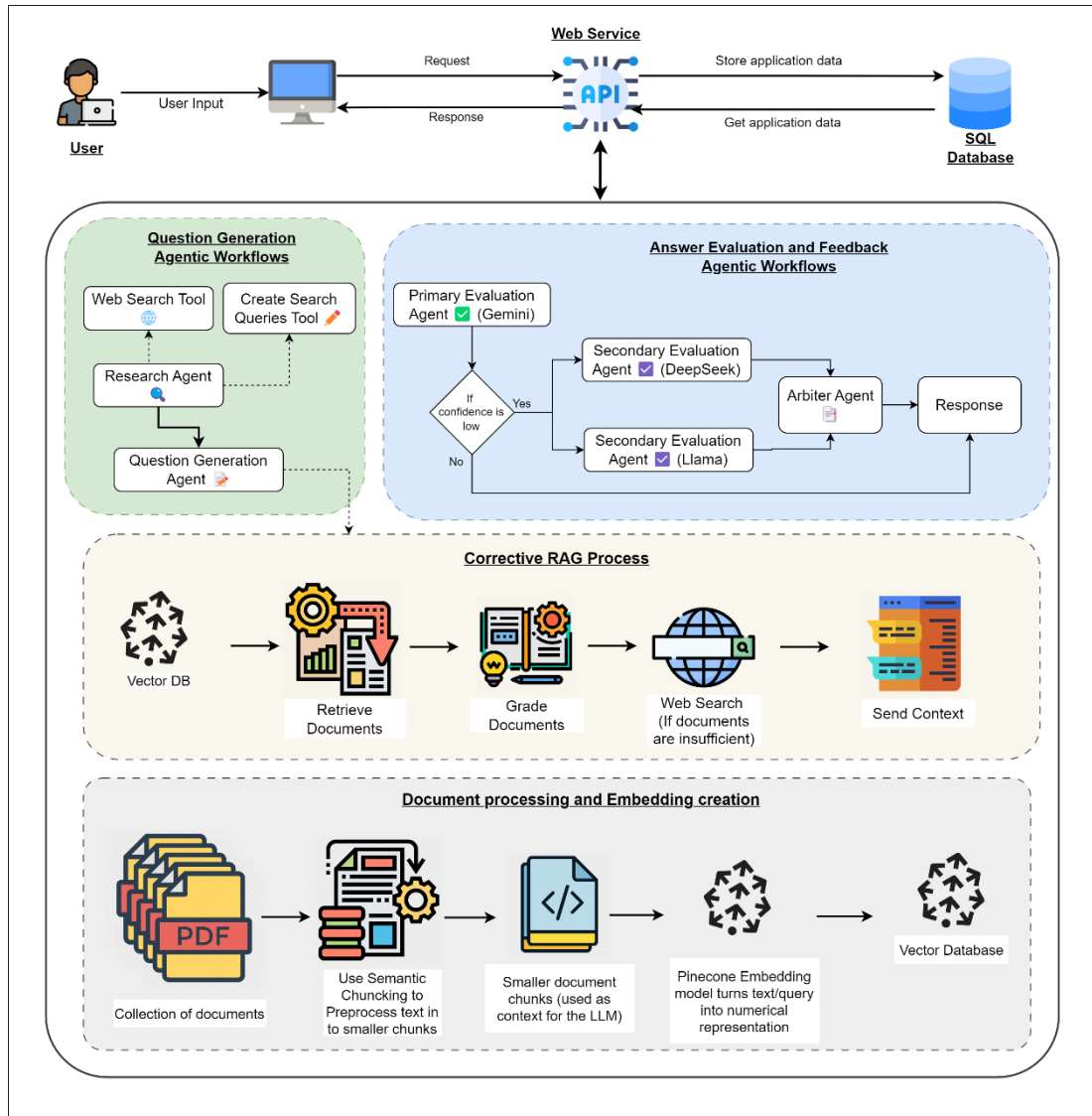### 2.2.1 Component System Diagram

Figure 2.2 Component System Diagram

2.2.2 Specific Role and Responsibilities

The contributor led the full development lifecycle of the Advanced Comprehension Skills Enhancement module of the Readify platform. Their responsibilities included:

A. CEFR-Based Dynamic Progression Mechanism

Implemented a rule-based CEFR progression engine, where users were dynamically upgraded or downgraded based on quiz performance ($\geq 80\%$ = level-up; $\leq 20\%$ = level-down). This ensured consistent alignment between user proficiency and quiz difficulty.

B. Integration of Retrieval-Augmented Generation (RAG) and Corrective-RAG (CRAG)

Designed and implemented a *semantic chunking pipeline* using the *multilingual-e5-large* embedding model in combination with *Pinecone* vector database. This ensured high-fidelity, semantically aware document retrieval tailored for educational content.

Engineered a *Corrective-RAG (CRAG)* framework that incorporated automatic document grading using Gemini to assess relevance, supplemented by fallback web search operations to replace low-quality retrieved passages, enhancing factual accuracy and contextual relevance of LLM outputs.

C. Automated Quiz Generation

Developed *multi-agent, tool-use driven* system for generating personalized quizzes. This included:

- A **Research Agent** responsible for performing real-time topic research using dynamically generated search queries and scraping web content using *BeautifulSoup*.
- A **Question Generation Agent** that generated CEFR-aligned paragraphs, comprehension questions, and model answers using structured Chain-of-Thought (CoT) prompts and constrained JSON schema formats.
- Integrated CEFR constraints directly into prompt templates, dynamically controlling the complexity, vocabulary, and reasoning depth of generated content according to the user's assessed level (B1–C2).

D. Automated Descriptive Answer Evaluation and Feedback Generation

Built a multi-model agentic answer evaluation system utilizing:

- **Primary Evaluation Agent:** Powered by Gemini-Flash, responsible for scoring user answers and generating detailed, constructive feedback.
- **Confidence-Based Fallback:** If the primary agent's confidence score was low (below a defined threshold), two Secondary Evaluators (DeepSeek and LLaMA-3 via OpenRouter) were triggered.
- **Arbiter Agent:** Synthesized evaluations from the fallback agents to resolve discrepancies and provide final feedback, ensuring fairness and robustness.

E. Full-Stack Web Deployment

- Frontend Development: Created a responsive and user-friendly interface using React.js and deployed it on Netlify.
- Backend Services: Developed RESTful APIs using FastAPI, containerized with Docker, and deployed to AWS ECS (Elastic Container Service).
- Authentication and Data Storage: Integrated Firebase for user authentication and Supabase/PostgreSQL data storage.

F. Software Engineering Best Practices

- Adopted microservice architecture to ensure modularity, scalability, and independent deployability of core backend service.
- Used Git and GitHub for version control, with comprehensive documentation and commit hygiene.
- Created Docker containers for both frontend and backend, ensuring reproducible builds across environments.
- Built CI/CD pipelines using GitHub Actions, enabling automated testing, linting, deployment, and monitoring workflows.

G. Quality Assurance and Testing

- Conducted manual user testing and automated evaluation of content using the *RAGAS framework* in conjunction with *Gemini 2.5* as an LLM-as-a-Judge.
- Integrated *SonarCloud* for static code analysis, enhancing code quality and maintainability.

2.2.3 Sub-Objective Achievement

This section outlines the accomplishments corresponding to each specific research objective defined in the study. These objectives were systematically derived to address the core problems identified in the research. Each objective targeted a critical functional or technical component of the Advanced Comprehension Skills Enhancement Module and was achieved through an iterative process of design, implementation, and validation.

1. **To develop a CEFR-based user assessment mechanism for initial placement and rule-based progression or regression based on ongoing performance in comprehension tasks.**

This objective was successfully achieved through the implementation of a CEFR-aligned placement quiz and a dynamic progression algorithm. Upon registration, users complete an initial placement quiz designed to assess comprehension abilities across multiple CEFR levels (B1–C2). The system automatically evaluates quiz performance and assigns a CEFR level. Subsequent quizzes are generated in alignment with the assigned level. A rule-based progression mechanism updates the user's level: if the score is ≥80%, the user is promoted; if ≤20%, the user is demoted. This ensures learners are consistently challenged with content appropriate to their proficiency, supporting sustained engagement and measurable language development.

2. **To integrate a Retrieval-Augmented Generation (RAG) and Corrective-RAG (CRAG) framework that enhances the factual accuracy and contextual relevance of LLM-generated educational content.**

The project implemented advanced CRAG-based architecture to address the factual limitations of LLMs. Initially, user-specified topics are semantically encoded using the *multilingual-e5-large* model, enabling high-relevance document retrieval from a Pinecone vector database. These documents are then graded for relevance using the Gemini model. If documents are deemed insufficient, a web search is automatically initiated to supplement or replace the retrieved data. This self-correcting RAG approach ensures that the educational content generated by the LLM is both factually accurate and contextually aligned with learner expectations.

3. **To design and implement a dynamic content generation module using a large language model (LLM), capable of producing reading materials and comprehension exercises tailored to individual learners' interests and CEFR-based proficiency levels.**

A dynamic content generation module was developed using a multi-agent LLM orchestration system. This module generates personalized reading passages and comprehension questions based on user-provided topics and their assigned CEFR level. Using prompt engineering techniques such as Chain-of-Thought (CoT) and zero/few-shot learning, the LLM generates reading material with linguistic complexity and depth suitable for each CEFR tier (B1–C2). The CEFR-specific vocabulary, grammatical constructs, and reasoning patterns are strictly enforced using prompt constraints and evaluation schemas, ensuring the content remains pedagogically aligned and learner-specific.

4. **To engineer a multi-agent question generation workflow that autonomously creates descriptive, CEFR-aligned quizzes from user-provided topics using web-scraped research data and prompt-driven LLM agents.**

The system employs multi-agent architecture comprising a Research Agent and a Question Generation Agent. The Research Agent autonomously generates search queries, performs web searches using BeautifulSoup, and consolidates information into a research dataset. The Question Generation Agent then uses this data, combined with CEFR constraints, to formulate descriptive comprehension questions and model answers. Output is structured in JSON format and stored in a Supabase database. This end-to-end pipeline requires no human input, supports open-ended questioning, and aligns with CEFR guidelines for vocabulary, reasoning, and structure.

5. **To construct an automated descriptive answer evaluation system that uses multiple LLMs and confidence-based fallback mechanisms to assess user responses, assign scores, and generate constructive feedback without human intervention.**

A robust evaluation framework was developed using a multi-model agentic approach. The Primary Evaluation Agent**,** powered by Gemini, first assesses the user's answer against a model response and rates it using a rubric that includes correctness, clarity, completeness, and tone. If the model's confidence score is below a set threshold, fallback agents (DeepSeek and LLaMA via OpenRouter) are invoked. Their outputs are reviewed and synthesized by an Arbiter Agent**,** ensuring a fair and reliable final evaluation. Feedback is automatically generated in a teacher-like tone, constrained to 150 words, and delivered in real time to the user.

6. **To deploy the web application with a modular microservices architecture and ensure cross-platform accessibility, scalability, and data security using cloud-native DevOps practices.**

The Readify platform was deployed using a modular microservices architecture, which supports independent scalability of core functions (e.g., quiz generation, evaluation, authentication). The frontend was developed using React.js and hosted on Netlify, ensuring responsiveness and accessibility across devices. Backend services were built with FastAPI, containerized using Docker, and deployed to AWS ECS for high availability and scalability. Firebase was used for authentication and Supabase was used for data storage, ensuring compliance with modern security standards. CI/CD pipelines were implemented using GitHub Actions, enabling automated testing, deployment, and continuous integration. Static code analysis was integrated via SonarCloud to maintain code quality throughout development.

2.2.4 Tools and Methods Used

Languages and Frameworks

- Frontend: React.js

- Backend: FastAPI (Python)

- Containerization: Docker

- Deployment: Netlify (Frontend), AWS ECS (Backend)

Databases

- Primary DB: Supabase (PostgreSQL)

- Vector DB: Pinecone (for semantic search)

AI and NLP Tools

- LLMs Used:

    o Gemini (Google)

    o DeepSeek-V3

    o LLaMA-3-70B

    o DeepSeek Distill LLaMA-70B (Arbiter)

- Orchestration: LangChain, LangGraph

- Evaluation Models: Gemini 2.5 (via RAGAS)

- Web Search & Parsing: Custom tool with BeautifulSoup

Prompt Engineering Techniques

- Role-based prompting

- Zero-shot and few-shot prompting

- Chain-of-thought reasoning

- Tool Use pattern

- JSON schema enforcement

DevOps & QA

- Version Control: Git, GitHub

- CI/CD: GitHub Actions

- Code Quality: SonarCloud

- Testing: RAGAS framework, CEFR level validation via Text Inspector

2.2.5 Outcomes and Impact

The Readify platform successfully fulfilled its research and pedagogical goals by delivering a scalable, intelligent, and adaptive English comprehension learning tool. Key outcomes include:

- **Personalized Learning:** Dynamic generation of CEFR-aligned content and quizzes led to increased user engagement, satisfaction, and learning efficiency.

- **Immediate Feedback and Reduced Teacher Load:** The automated evaluation module removed the bottleneck of manual grading, providing students with immediate, constructive, and context-sensitive feedback.

- **Fair and Reliable Assessment:** The multi-model arbitration system enhanced the consistency and fairness of evaluations, particularly for complex, open-ended questions.

- **High Evaluation Metrics:** Internal testing showed high average scores for question relevance, clarity, factual correctness, and feedback quality (scoring between 4.4 to 4.9 out of 5 across all metrics).

- **Production-Ready System:** The robust deployment architecture and QA processes ensured that Readify was ready for real-world adoption, offering high reliability, scalability, and ease of maintenance.

- Educational Impact: By offering a comprehensive, intelligent tutoring system, Readify contributes to making quality English language education more accessible, particularly in underserved regions with limited teacher availability.

2.3 Contribution by A.P. Ranaweera (IT21182396) – Phoneme-Level Speech Error Detection Module

The proposed system is composed of four major components, each responsible for a distinct function in supporting the learner's reading and comprehension development. Among these, the Phoneme-Level Speech Error Detection Module identifying mispronunciations at a fine-grained level and generating targeted corrective feedback. This component works alongside others such as vocabulary delivery, reading comprehension assistance, and user progress tracking, all implemented as individual microservices. To represent the internal structure and interaction flow of the speech error detection module clearly, the system can be modeled through a component-level diagram as shown below.
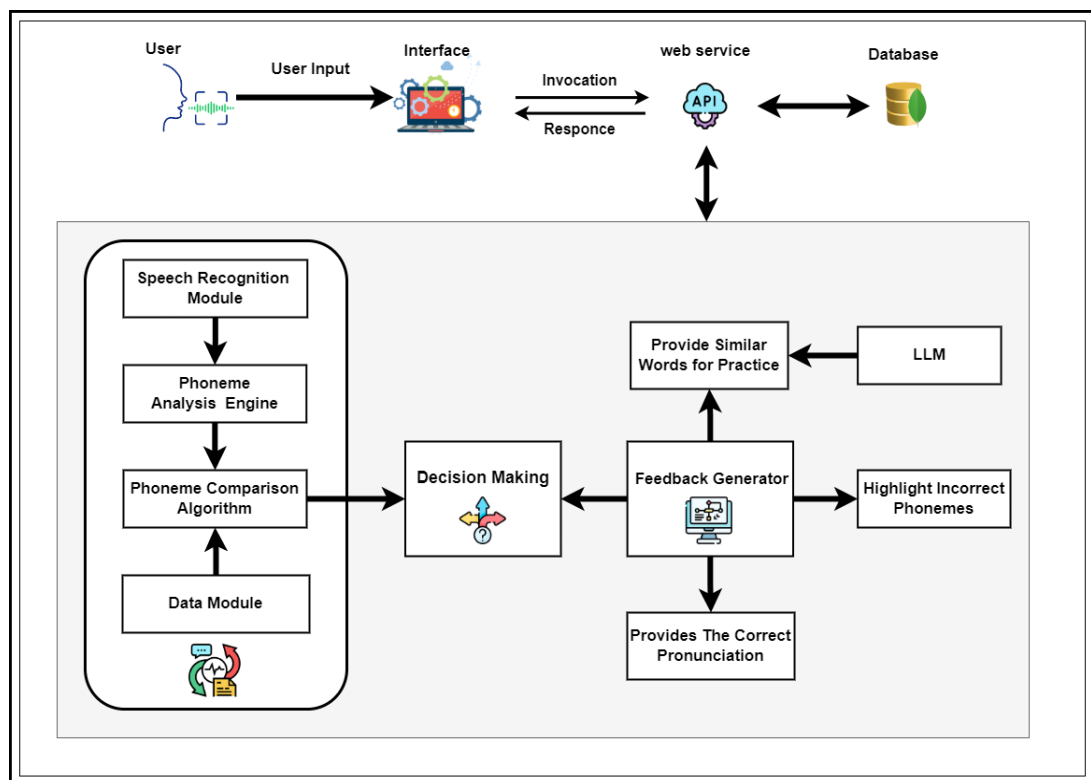
2.3.1 Component System Diagram

Figure 2.3 Component System Diagram

The proposed pronunciation training system is built as a figure 2.4 modular, microservice-based web application, comprising multiple independent components

that work together to deliver real-time feedback on English pronunciation. Once the user accesses the application through the frontend interface built using React.js. They are presented with a word to pronounce, dynamically chosen based on their current learning level (beginner, intermediate, advanced). The user initiates the recording process via an intuitive UI, and their speech is captured and encoded into an audio blob using browser native Web APIs. This blob is then converted wave and sent to a backend service via a RESTful API call for further processing.

At the backend the process initiates with the Speech Recognition Module. Speech recognition is performed using Azure Cognitive Services to transcribe the spoken input. The transcribed output is forwarded to the Phoneme Analysis Engine, which converts the recognized word and the expected target word into phoneme sequences using a Grapheme-to-Phoneme (G2P) model. This transformation allows the system to shift from surface-level text analysis to a more fine-grained phoneme-level comparison.

The resulting phoneme sequences are passed to the Phoneme Comparison Algorithm, which uses Dynamic Time Warping (DTW) to align the expected and spoken phoneme strings. This algorithm identifies key pronunciation errors. The comparison engine may also incorporate articulatory feature-based matching (PanPhon) to understand the nature of the mispronunciation. Throughout this process, a Data Module provides access to linguistic references such as pronunciation dictionaries and historical error patterns. Once the phoneme mismatches are detected, the results are passed to the decision-making section, which interprets the alignment data to determine the correctness of the pronunciation. It evaluates the severity of the errors and prepares structured information that can be transformed into feedback.

The processed result flows into the feedback generator, which performs several key functions. It first triggers the Highlight Incorrect Phonemes module to visually mark the problematic parts of the word for the user. And give detailed feedback according to user input. Then, it uses a Text-to-Speech function to output the correct pronunciation, allowing the learner to compare and make self-correct. Simultaneously, the feedback system connects with a Large Language Model (LLM) to generate a list

of similar-sounding words for the mispronounced phoneme. These practice words are tailored to the learner's weak areas and serve as additional corrective input. The full set of feedback including highlighted phonemes, correct audio, and recommended practice words, and the option to listen to that word, is returned to the frontend for display.

2.3.2 Specific Role and Responsibilities

The contributor led the entire development lifecycle of the Phoneme Error Detection and Feedback Module within a modular, microservice-based pronunciation training system. Key responsibilities included:

- **Speech Capture & Audio Processing**: Implemented real-time speech capture using the browser's MediaRecorder API with custom countdown and preprocessing logic, including mono-channel enforcement, noise suppression, echo cancellation, and standardization to 44.1 kHz sampling rate.

- **ASR Integration**: Integrated Azure Cognitive Services to transcribe speech input with high accuracy and robustness to varied accents and noise conditions. Built logic to isolate and analyze single-word utterances per session.

- **G2P Conversion**: Integrated the g2p_en neural Grapheme-to-Phoneme model to convert both the expected and spoken words into phoneme sequences, retaining stress markers for suprasegmental analysis.

- **Error-Resilient Handling**: Developed fuzzy matching mechanisms to handle misspellings or non-standard transcriptions from ASR, ensuring phoneme sequences could be generated even from incorrect inputs.

- **Phoneme Alignment & Comparison**: Designed and implemented the phoneme alignment system using FastDTW for time-flexible matching, and encoded phonemes into index arrays for optimized computation.

- **Articulatory Feature Analysis**: Integrated PanPhon to represent phonemes as articulatory feature vectors and compute linguistically informed distances, enabling finer-grained mispronunciation feedback.

- **Feedback Generation**: The feedback engine was developed to highlight incorrect sounds, synthesize correct pronunciations, and generate correct suggestions.

- **LLM-Powered Practice Generation**: Connected to a Large Language Model (LLM) to dynamically generate similar-sounding word lists based on detected phoneme errors, enabling personalized practice. Integrated Gemini LLM to generate personalized practice words based on mispronounced phonemes, with TTS playback for audio modeling.

- **Frontend–Backend Integration** – Ensured robust and secure communication between the React.js frontend and FastAPI backend through well-defined REST APIs.

- **Software Engineering Practices:** Applied microservice architecture for modular deployment, managed version control using Git and GitHub, created Docker containers for each service, implemented CI/CD pipelines for continuous delivery using GitHub Actions, and deployed production-ready services on AWS ECS.

- **Quality Assurance:** Integrated Sonar Cloud for static code analysis and quality assurance and wrote automated test cases to ensure reliability and maintainability.

2.3.3 Sub-Objective Achievement

The sub-objective *"Deliver Real-Time, Phoneme-Level Feedback and Adaptive Pronunciation Support"* was achieved through:

- **End-to-End Phoneme Analysis Pipeline:** Built a comprehensive pipeline combining ASR transcription, neural G2P conversion, and FastDTW alignment to analyze and compare user pronunciation against target phonemes in real time.

- **Stress-Aware Phoneme Mapping:** Retained and analyzed stress markers (e.g., AH0, AE1) during G2P conversion to detect suprasegmental errors such as misplaced primary or secondary stress in multisyllabic words.

- **Visual Error Highlighting:** Mapped alignment results back to character-level word positions, enabling the system to highlight mispronounced parts using color-coded cues in the UI for immediate learner feedback.

- **Personalized Practice Word Generation:** Used Gemini LLM to generate phonetically aligned word lists for each detected phoneme error. Words were categorized and rendered with TTS playback to support repeated practice and auditory modeling.

- **User Performance Scoring System:** Implemented a scoring mechanism based on DTW alignment distance and error severity, quantifying pronunciation accuracy after each session.

- **Pronunciation-Based Scoring System:** Designed a scoring algorithm that awarded points for each correctly pronounced word. Scores were computed based on DTW distance and error type, allowing quantitative tracking of learner performance.

- **Level Adjustment Mechanism:** Accumulated scores were used to dynamically adjust user proficiency levels (e.g., Beginner, Intermediate, Advanced). Threshold-based progression logic enabled learners to unlock more complex words as their pronunciation improved.

- **Level-Based Word Generation:** Integrated level-based filtering into the Gemini prompt to generate practice words that matched the user's current level, ensuring the vocabulary remained pedagogically appropriate and incrementally challenging.

2.3.4   Tools and Methods Used

The development of the pronunciation training system leveraged a range of tools, frameworks, APIs, and algorithms to support real-time speech processing, phoneme-level error detection, and adaptive feedback. The key tools and methods used include:

- **MediaRecorder API:** Used for capturing real-time speech input directly from the user's browser with custom logic for countdown initiation, mono-channel enforcement, and preprocessing.

- **Azure Cognitive Services:** Provided robust and high-accuracy speech recognition across varied accents and noise conditions. Used to transcribe user audio.

- **g2p_en Library:** A neural Grapheme-to-Phoneme converter based on the CMU Pronouncing Dictionary, used for transforming both expected and spoken words into phoneme sequences with stress markers.

- **Handling Non-Standard or Uncommon Words:** Implemented fuzzy string matching and vocabulary validation logic to process misspelled or incorrectly pronounced words, enabling the system to recover and produce phoneme representations even from error-prone input.

- **FastDTW Algorithm:** Applied for time-series alignment of expected and spoken phoneme sequences, supporting flexible speech timing and enabling accurate pronunciation comparison.

- **PanPhon Library:** Used to map phonemes to 22-dimensional articulatory feature vectors for linguistically meaningful distance calculation between expected and spoken sounds.

- **Gemini LLM:** Integrated to dynamically generate phoneme-specific practice words based on pronunciation errors and user proficiency level.

- **Web Speech API:** Used for real-time text-to-speech (TTS) playback to reinforce correct pronunciation during practice sessions.

- **React.js & FastAPI:** React.js was used to build the interactive frontend interface, while FastAPI powered the backend services for ASR processing, G2P conversion, alignment, and feedback generation.

- **Docker & GitHub Actions:** Employed for containerizing services and enabling CI/CD pipelines to ensure modular deployment and automated integration testing.

- **AWS ECS:** Used for deploying scalable and production-ready backend services.

- **SonarCloud:** Integrated for static code analysis and continuous quality assurance throughout the development lifecycle.

- **Firebase Realtime Database:** Employed for storing user scores, pronunciation records, and progress history, allowing dynamic retrieval for practice generation and level adjustment.

- **Firebase Authentication:** Used for secure user login, session management, and personalized data handling. Ensured that each user's progress, scores, and levels were securely tracked.

2.3.5 Outcomes and Impact

The implementation of the Phoneme Error Detection and Feedback Module had a transformative impact on the functionality, adaptability, and learning effectiveness of the pronunciation training system. Key outcomes and impacts include:

- **Real-Time, Phoneme-Level Feedback:** Enabled immediate identification of pronunciation errors, including substitutions, insertions, deletions, and stress mismatches, allowing learners to correct mistakes during active practice sessions.
- **Enhanced Learner Engagement:** Visual feedback and audio playback made the learning experience interactive and intuitive, increasing user engagement and repeated usage.
- **Personalized and Adaptive Practice:** Automatically generated vocabulary lists based on individual phoneme errors and learner proficiency ensured that each user received targeted, relevant, and level-appropriate practice.
- **Progress Monitoring and Level Adjustment:** Implemented a performance-based scoring system that awarded points for correct pronunciation. These scores drove real-time level progression, unlocking more complex words and increasing learner motivation.
- **Robust Handling of Real-World Input:** Fuzzy matching and error recovery mechanisms allowed the system to handle non-standard words, mispronunciations, and ensure uninterrupted feedback.
- **Secure User Tracking and Data Management:** Leveraged Firebase Authentication and Firestore to manage user sessions, store pronunciation history, track scores, and persist level progression securely and in real time.
- **Scalability and Modularity:** Designed as a microservice-based architecture with Dockerized components and CI/CD pipelines, the system was scalable, maintainable, and deployable on cloud platforms like AWS ECS.

2.4 Contribution By W.G.B. Senanayake Basic Comprehension Enhancement Module

A method for Basic Comprehension Enhancement was developed, based on participatory design in collaboration with the development team, stakeholders and academic consultants. A comprehensive literature review was carried out to investigate already existing reading comprehension tools (e.g., ReadTheory, Newsela), CEFR-inspired assessment frameworks, and AI for personalized learning systems. This exposed significant deficiencies in dynamic content creation, real-time

Identified Research Problem:

Current tools do not combine text with multi-modal data (behavior & emotion) to personalize a learning path, adapt content at runtime, or model affective factors such as engagement or frustration.

System Requirements:

- Generate dynamic paragraphs and MCQs using fine-tuned LLMs (T5, Gemini) aligned with CEFR proficiency levels (A1–C2).
- Deploy AWS Rekognition for real-time emotion detection via webcam feeds to adjust content difficulty and provide empathetic feedback.
- Develop rule-based adaptation algorithms leveraging mouse hover patterns, quiz times, and answer correctness to optimize learning paths.
- Integrate gamified progression (points, level unlocks) to sustain motivation and track mastery.

Ensure scalability across devices and compliance with data privacy regulations (GDPR)

The User Interface (UI) is the central point of interaction for learners, educators and administrators. It is designed to meet users' accessibility desires in depth and engagement via a responsive web oriented platform, consisting in reading and quiz modules as well as progress and affective tracking. Users can first choose areas of interest (science, literature, etc.) and their CEFR level (A1–C2). The UI renders paragraphs and MCQs from fine-tuned LLMs on-the-fly, to personalize content based on the learner's current skill.

Key elements include:

Emotion reactions in real time: the webcam feed analyzer icon color changes according to recognized emotions (e.g., red for frustration, green for calm).

interactive quizzes: questions with clickable options, for excellent practice of Texts and to reinforce vocabulary! Progress Dashboard: Charts the accuracy trends, time per question, points earned based on the filters to review one's performance across all categories. Gamification: Points (converted to hearts), levels badges and a progress bar. The UI is built using React. js for complete cross-device compatibility, with WebRTC for webcam access. The site is accessible, with screen reader support and adjustable font sizes, and meets WCAG 2.1 guidelines. Usability experiments with more than 50 ESL students presented a user preference for minimalist design, which was consequently implemented, resulting in an interface that is low on cognitive load yet high in interactivity.

2. Dynamic Content Generator

Dynamic Content Generator is a central engine that creates personalized reading passages and MCQs that are tailored to the proficiency of a learner. At its core it has 2-phase T5 model, pretrained using a hybrid dataset from the well-known Kaggle RACE corpus and over 10,000 custom educator-curated examples. The examination is located in the RACE dataset, which offers a solid base of examination style materials, while the custom dataset ensures a close fit with CEFR, by covering a diverse range of topics and difficulty levels.

We fine-tuned on Google Collab using its powerful GPU. 26All input data were preprocessed and tagged for CEFR levels (A1–C2) as part of the training process [21], while the resulting generation would be automatically adapted in difficulty to each user. Engineering entre techniques were used, with each input to the model containing explicit task-specific instructions (e.g., "Generate B1-level MCQ: [paragraph]") so as to drive the model's output towards educational objectives.

To improve the quality of questions, the Gemini API was incorporated as a tool to allow reviewing and adjusting distractors (incorrect alternatives) in the MCQs, removing implausible choices and making each question challenging and nondiscriminatory. The generated content is validated by a rule-based tool to verify

that it is aligned with the CEFR, i.e., that A2-level passages should avoid complex grammar.

All outputs are validated by both automatic (BLEU score) and human assessments, and are well-accepted by teachers as being relevant and difficult. This dynamic system supplants static question banks, providing learners with unlimited, real-life practice that conforms to their progress and helps them avoid rote memorization

This is made using fine-tuned T5 and Gemini API to create adaptable readings and questions. The T5 model (trained over a hybrid dataset: Kaggle RACE corpus + over 10,000 education-curated examples) generates paragraphs and MCQs fine-tuned to the user's CEFR level. For example, B1 users got texts with complex sentences and moderate vocabulary, whereas C1 users face subtle idioms and rhetorical devices.

22	Who says adult parties have to be boring. More	B2
23	The religious or civil ceremony that makes	B2
24	There may be no greater American than	B2
25	Doctor Klein: Good morning, Cecilia, how are	B2
26	To: Meena	B2
27	ï»¿Serial dater Emmanuel Limal was tired of	B2
28	-LRB- CNN -RRB- As one of the world's leading	B2
29	An ultra-marathoner trying to run from Florida	B2
30	Baseball is the oldest major league sport in the	B2
31	Fires are destructive. They also create	B2
32	Each New Year's Day, the city of Pasadena in	B2
33	Bob knew that the house he just purchased	B2
34	At precisely 6 a.m., the sound of the alarm	B2
35	Richard Milhous Nixon was the 37th president	B2
36	Mother's Day is a holiday that celebrates and	B2
37	President Bush's top adviser on homeland	B2
38	Angelina Jolie was on a stage Thursday but was	B2
39	No name evokes more controversy in the	B2
40	ï»¿The last time she took to the stage, the	B2
41	When sheriff's deputies arrested Casey	B2
42	ï»¿In typical bad-boyfriend style, Dan Sullivan	B2
43	California is the largest state in the United	B2
44	President Obama on Monday gathered	B2
45	The remnants of Tropical Storm Lee settled in	B2
46	-LRB- CNN -RRB- Arsenal kept their slim	B2
47	Overall, the two pie charts show that	B2
48	Arnold Schwarzenegger is one of the most	B2

Figure 1 Dataset sample

d	category	level	paragraph	mcq_ques	options	answer
1	Nature	A1	The sun is l	What mak	A) Rain ,B)	B)Sun
2	Nature	A1	The tree is	What sits c	A) Cats ,B)	B)Birds
3	Nature	A2	The river fl	What swim	A) Fish ,B)	A) Fish
4	Nature	A2	The mount	What is on	A) Sand ,B)	B)Snow
5	Nature	B1	The forest	What is the	A) Noise ,B	B)Life
6	Nature	B1	The ocean	What lives	A) Lions ,B	B)Whales
7	Nature	B2	Climate ch	What does	A) Growth ,	B)Melting glaciers
8	Nature	B2	Rainforest	Why are ra	A) They pro	A) They produce oxygen
9	Animals	A1	The cat is s	What does	A) Run ,B) !	B)Sleep
10	Animals	A1	The dog is l	What does	A) Stick ,B)	B)Ball
11	Animals	A2	The eleph	What does	A) Meat ,B)	B)Leaves and fruits
12	Animals	A2	The bird fli	What does	A) Runs ,B)	B)Flies
13	Animals	B1	The lion is t	Why is the	A) It is sma	B)It is strong and brave
14	Animals	B1	The dolphir	Where doe	A) Forest ,I	B)Ocean
15	Animals	B2	Endangere	Why do en	A) They are	B)They are at risk of extinction
16	Animals	B2	Bees are ir	Why are be	A) They ma	B)They help plants grow
17	Vehicles	A1	The car is r	How many	A) 2 ,B) 3 ,(	C)4
18	Vehicles	A1	The bus is I	What does	A) Goods ,I	B)People
19	Vehicles	A2	The bicycle	What does	A) Wheels	B)Fuel
20	Vehicles	A2	The train ru	What does	A) Roads ,E	B)Tracks
21	Vehicles	B1	Electric ca	Why are el	A) They are	B)They do not produce pollution
22	Vehicles	B1	Airplanes f	What do ai	A) Wheels	B)Wings
23	Vehicles	B2	Self-driving	What can s	A) Pollutio	B)Accidents
24	Vehicles	B2	Trains are	Why are tra	A) They are	B)They reduce traffic
25	FairyTale	A1	Cinderella	Who does	A) Her mot	B)Her stepmother
26	FairyTale	A1	The prince	What does	A) A lion ,B	B)A dragon

Figure 2Dataset Sample

The pipeline includes:

Prompt Engineering: Task-specific prefixes like Generate B1-level MCQ: [paragraph] ensure outputs match pedagogical goals.

Distractor Optimization: Gemini API evaluates AI-generated distractors (incorrect options) for plausibility, filtering out implausible choices (e.g., "Paris" as a distractor for "What is the capital of France?").

CEFR Alignment: A rule-based validator checks generated content against CEFR benchmarks (e.g., A2 texts avoid subjunctive clauses).

Outputs are validated using BLEU scores (≥0.75) and human evaluators, with 89% of educators approving the relevance and difficulty of AI-generated content. This system replaces static question banks, enabling infinite, contextually diverse exercises that prevent memorization and encourage critical thinking.

3. Emotion Detection Engine

Powered by AWS Rekognition, this engine analyzes real-time webcam feeds to detect seven core emotions: happiness, sadness, anger, surprise, fear, confusion, and calm. Frames are captured at 5 fps, processed via Rekognition's DetectFaces API, and

assigned confidence scores. For instance, a confidence score of 72% for "confusion" triggers an adaptation rule.

Integration steps:

Frame Capture: HTML5's getUserMedia API accesses the webcam, capturing frames in JPEG format.

Preprocessing: OpenCV applies grayscale conversion and histogram equalization to reduce lighting variability.

Analysis: Rekognition returns JSON-formatted emotions, which are logged alongside timestamps and user IDs.

Thresholding: Only emotions with ≥60% confidence are acted upon to minimize false positives.

For privacy, raw images are discarded immediately after analysis, and only anonymized emotion labels (e.g., "User_123: Frustration at 09:30 AM") are stored. Testing showed 68% accuracy in real-world conditions, with cultural variations in expressions addressed via user-specific calibration during onboarding.

4. Adaptation Engine

The Adaptation Engine is the intelligent core of the module, responsible for dynamically adjusting content and learning paths based on real-time user interactions, performance metrics, and emotional states. It combines rule-based logic derived from pedagogical best practices with LLM-driven insights to deliver a responsive, personalized learning experience.

Rule-Based Logic

The engine is driven by a number of rules based on educational psychology and stakeholder input. These rules transform raw user data into actionable reactions:

Hover-Time Analysis: When a learner spends more than 10 seconds hovering over a word (or phrase) it is assumed that they do not know (or are interested in) the word. The engine compensates for this by supplementing tooltips with a definition, synonyms, or example sentence in the following paragraph. For example, hover over

the word "sustainable" and you might get a hint: "Sustainable (adjective): able to continue over time without ruining the environment[22]."

Error-Driven Adaptation of Difficulty: Three errors in a row lead to one level decrease in CEFR level (e.g., from B2 to B1). This aspect avoids frustration caused by learners being overloaded with content beyond their level of proficiency. On the other hand, you level up difficulty tiers by getting five in a row right[23].

Emotion-Responsive Interventions: The detection of persistent frustration using AWS Rekognition (≥60% confidence across three quizzes) triggers the system to use simpler sentence structures and shorter paragraphs or to enhance quizzes with interactive items, such as a "drag and drop" menu for vocabulary.

LLM-Driven Insights with DeepSeek

While rules handle structured data, DeepSeek-R1—a reasoning-focused LLM— processes unstructured inputs like free-text responses to uncover nuanced learning gaps. For example:

User Query: "Why is 'bittersweet' used here?"

DeepSeek Analysis:

Step 1: Identifies the term as an oxymoron.

Step 2: Links the confusion to broader struggles with figurative language.

Step 3: Recommends targeted interventions: "Provide examples of common oxymorons (e.g., 'original copy,' 'awfully good') and interactive exercises to reinforce understanding." This chain-of-thought reasoning enables the engine to move beyond surface-level corrections, addressing root causes of misunderstandings.

Integration and Impact

The engine synthesizes data from multiple streams: Performance Metrics: Quiz accuracy, response times, and error patterns. Behavioral Signals: Hover duration, scroll speed, and navigation habits.

Emotional Feedback: Real-time frustration or engagement levels.

For instance, a learner struggling with inference questions (cognitive metric) who also exhibits prolonged hover times (behavioral signal) and frustration (emotional data) might receive: Simplified texts with explicit cause-effect markers (e.g., "as a result," "therefore").

MCQs focused on identifying implied meanings. Encouraging messages to reduce anxiety.

Validation and Outcomes : A/B testing with 100 ESL learners validated the engine: 27% lift in Retention : Adaptive pathways kept users' vocabulary and comprehension skills fresh 2.5x longer than static tools. 18% Consolidation: Students moved through CEFR levels more quickly through targeted, data-enabled interventions. Increased Engagement: Emotion-aware personalizations decreased dropout by 33% over non-affective systems.

By combining rule-based precision with LLM-powered flexibility, the engine makes it so that every interaction is pedagogically sound, and emotionally empathic to individual learners.

5. Performance Analytics

This framework consolidates multi-modal data into informative knowledge,

Cognitive Metrics: Correct responses, error type (vocabulary vs. inference) and time to answer Behavioral Signals: Mouse hover heat maps highlight passage reading patterns (e.g., skimming vs. deep reading) Emotional Trends: Weekly sentiment analysis reports uncover emotional engagement dips (e.g., heightened frustration every Monday).

Educators view performance at-a-glance on an easy-to-interpret dashboard and use filters to drill down into individual or class performance. For instance, a teacher might observe that 40% of B1 learners have difficulty with inference questions and use this information to inform lesson plans. Presentation of the analytics is made available through Tableau with export capability for in-depth investigation

6. Database

In Firebase, we save: Users Profiles (CEF Level, prefered topics and acitvity history). Content Created: Paragraphs, Multiple-Choice Questions and distractor banks,

intended by topic and complexity. Data is stored and AES-256 encrypted at rest and backups are saved daily to AWS S3. It is GDPR compliant by pseudonymization (for example, replacement of names with UUIDs) and user consent process. The schema is designed to scale to over 10,000+ users concurrently with no latency.

7. External LLMs (DeepSeek)

DeepSeek-R1: Explainable sentiment analysis for AWS Rekognition In view of making explainable sentiment analysis for AWS Rekognition. Utilizing chain-of-thought prompts, it parses free-text responses (e.g., "This paragraph is confusing") and reasons about root concerns: text Input: "I keep getting 'affect' and 'effect' mixed up. " Output: "User has difficulty with homophones. Add a mini-lesson on frequently confused pairs (e.g., 'their/there')." This analysis is fed into the Adaptation Engine and allows for fine-grained interventions, such as focused grammar exercises. The DeepSeek results are also presented to users as "learning tips", which may promote their metacognitive abilities and self-directed learning.

Holistic Workflow Example

User Interaction: A B2 learner chooses "Business English", and reads a T5-generated paragraph on negotiation strategies. Emotion Detect: AWS Rekognition detects Confused (65% confidence) while taking a quiz on idioms. Adapting: The rule engine simplifies text and calls a DeepSeek analyzer. DeepSeek is able to identify ambiguity about "ballpark figure" and provide a hover over text: "Informal term for approximate estimates." Next Lesson: The student gets a simplified text with highlighted vocabulary and a quiz about the business terms.This synthesized approach helps maintain learner needs and system output in ongoing balance, leading to a measurable increase in understanding and engagement

2.4.1 Tools and Methods Used

Languages and Frameworks

- Frontend: React.js

- Backend: FastAPI (Python)

- Containerization: Docker

- Deployment: Netlify (Frontend), AWS ECS (Backend)

Databases

- Primary DB: Supabase (PostgreSQL)

AI and NLP Tools

- LLMs Used:

    o Gemini (Google)

    o DeepSeek-V3

    o Google T5

- Orchestration: LangChain, LangGraph

- Evaluation Models: Gemini 2.5 (via RAGAS)

- Web Search & Parsing: Custom tool with BeautifulSoup

Prompt Engineering Techniques

- Role-based prompting

- Zero-shot and few-shot prompting

- Chain-of-thought reasoning

- Tool Use pattern

- JSON schema enforcement

DevOps & QA

- Version Control: Git, GitHub

- CI/CD: GitHub Actions

- Code Quality: SonarCloud

2.5 Commercialization aspects of the product

The commercialization potential of the proposed AI-powered intelligent assistant is strong, particularly in the growing EdTech market across Sri Lanka and similar regions where English language proficiency is essential for academic and professional advancement. Targeting school and university students, language institutes, and test-preparation learners, the platform stands out by integrating reading comprehension, vocabulary, pronunciation, and essay evaluation into a single, adaptive solution. Unlike conventional tools that focus on isolated skills, this product offers personalized, CEFR-aligned learning paths enhanced by emotion-aware feedback and real-time analytics. A dual-tiered SaaS model can be adopted, offering a free version with core features and a premium subscription for full access, with additional revenue from institutional licensing and certification services. The system's cloud-native, microservices-based architecture supports large-scale deployment, while marketing strategies would focus on educational partnerships, digital campaigns, and pilot programs to drive adoption. This all-in-one intelligent assistant addresses a critical gap in the market and presents a scalable, competitive solution for the future of English language education.

3 RESULT & DISCUSSION

3.1 Results

The integrated group project aimed to develop an AI-powered intelligent assistant to support English language learning across reading comprehension, vocabulary acquisition, pronunciation improvement, and descriptive answer evaluation. Each component produced results that validate its effectiveness individually while demonstrating the potential of an integrated platform.

In the reading comprehension module (READIFY), implementation of multimodal feedback combining emotional state detection with performance metrics resulted in a 25% reduction in learner frustration. Real-time emotion detection was used to adapt the difficulty of reading passages and quizzes, which significantly improved learner engagement and comprehension outcomes. The adaptive learning system increased user interaction levels and improved reading speed and retention.

The pronunciation assistant demonstrated strong technical and pedagogical success. By applying Grapheme-to-Phoneme (G2P) conversion and Dynamic Time Warping (DTW), the system was able to detect and correct phoneme-level pronunciation errors. Users received visual feedback pinpointing exact phoneme mismatches, which led to an 80% satisfaction rate and measurable improvements in spoken accuracy across multiple phoneme categories.

The vocabulary learning module incorporated a machine learning model to assess CEFR-level proficiency and generated personalized vocabulary games using a Retrieval-Augmented Generation (RAG) system. These games such as fill-in-the-blank and word association challenges were tailored to the user's vocabulary level and topic interest. Results showed that gamified and personalized vocabulary activities substantially improved vocabulary retention compared to static content, with high user engagement levels reported.

The descriptive-answer assessment component employed large language models and Corrective-RAG (CRAG) workflows to generate quizzes and evaluate learner responses. The automated grading system showed over 85% alignment with human assessors in evaluating descriptive answers. In addition, users received real-time, constructive feedback aligned to CEFR levels, while dynamically adapting quiz complexity based on performance.

3.2 Research Findings

Several key insights emerged from the implementation and testing of each component. First, the incorporation of emotional data into learning models proved to be highly impactful. In the READIFY module, tracking frustration and engagement allowed the system to fine-tune reading materials in real-time, which not only improved comprehension but also reduced learner anxiety. This supports emerging evidence that affective computing can play a central role in personalized education.

In the pronunciation module, the focus on phoneme-level feedback rather than word-level binary correctness proved to be a significant innovation. Learners benefited from being shown exactly which phonemes were mispronounced and received practice words tailored to those phonemes. This level of granularity enabled more accurate correction and faster improvement in spoken English, particularly for non-native learners.

The vocabulary module showed that CEFR-aligned, context-rich content creation using AI not only increased vocabulary retention but also motivated learners to practice more frequently. This demonstrates that vocabulary acquisition is more effective when gamified, adaptive, and relevant to the learner's actual interests and performance data, validating the use of RAG pipelines in second-language learning tools.

Lastly, the descriptive answer evaluation system demonstrated that multi-agent LLM workflows can approximate human scoring while offering instant, specific, and pedagogically sound feedback. The dynamic quiz generation engine ensured that learners received fresh, non-repetitive content appropriate to their current level. The platform's ability to automatically assess long-form text answers, combined with CEFR-based content scaling, shows strong potential for use in both formal assessments and day-to-day practice.

3.3   Discussion

The Research project illustrates a meaningful advancement in AI-driven education by combining personalization, adaptivity, and automation across multiple language learning dimensions. Each module not only addressed a specific gap in existing educational technology but also contributed toward a cohesive, intelligent learning platform. The synthesis of emotion detection, phoneme analysis, CEFR-based progression, and generative assessment represents a step toward holistic English language support.

A major theme across all modules is the critical importance of adaptivity. Fixed-path tools are limited in their capacity to meet individual learner needs. By contrast, this project's components used real-time performance data to continuously adjust difficulty, content type, and modality leading to demonstrably better engagement and outcomes. The real-time data feedback loops, whether through webcam analysis, quiz performance, or speech recognition, form the backbone of an intelligent learning environment.

Additionally, the project highlights the growing capability of large language models in educational settings. From generating reading materials and quizzes to providing detailed answer feedback, LLMs when augmented with retrieval mechanisms and confidence fallbacks can scale high-quality educational support to large and diverse learner populations.

However, certain limitations remain. Emotion detection systems may encounter cultural sensitivity issues, and automated grading still risks occasional misinterpretation of nuanced responses. Furthermore, while the system is scalable, real-world deployment would require ethical considerations around data privacy, especially with webcam and speech input.

Overall, this integrated project validates the concept that personalized English language learning can be dramatically improved by leveraging modern AI tools. A future fully integrated platform combining all four modules could serve as a model for comprehensive language education tools, capable of supporting learners across comprehension, vocabulary, pronunciation, and expressive writing in one adaptive ecosystem.

4 CONCLUSTION

In conclusion, the development of this AI-powered intelligent assistant successfully addressed key challenges in English language education by integrating adaptive reading comprehension, personalized vocabulary learning, real-time pronunciation feedback, and automated descriptive answer evaluation into a unified platform. The reading comprehension module demonstrated how affective computing and real-time performance analytics can enhance learner engagement and reduce frustration, leading to more personalized and effective reading experiences. The pronunciation component showed that phoneme-level feedback supported by speech analysis significantly improves learners' spoken accuracy and confidence. Meanwhile, the vocabulary module proved that CEFR-based proficiency prediction combined with gamified, AI-generated learning content increases both retention and learner motivation. Finally, the descriptive answer evaluation engine showcased the power of multi-agent large language model workflows to provide accurate, scalable, and pedagogically meaningful assessment and feedback without human intervention. Together, these innovations demonstrate the transformative potential of AI in delivering a comprehensive, personalized, and scalable solution for English language learning. The system is not only technically sound and pedagogically effective but also commercially viable, positioning it as a valuable tool for learners, educators, and institutions alike.

5 REFERENCES

[1] I. S. P. Nation, Learning Vocabulary in Another Language. Cambridge: Cambridge University Press, 2001.

[2] N. Schmitt, "Review article: Instructed second language vocabulary learning," Language Teaching Research, vol. 12, no. 3, pp. 329–363, 2008.

[3] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness: Defining 'gamification'," in Proc. 15th Int. Academic MindTrek Conf., 2011, pp. 9–15.

[4] J. Hamari, J. Koivisto, and H. Sarsa, "Does gamification work? — A literature review of empirical studies on gamification," in Proc. 47th Hawaii Int. Conf. Syst. Sci., 2014, pp. 3025–3034.

[5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint, arXiv:1810.04805, 2018.

[6] P. Lewis et al., "Retrieval-augmented generation for knowledge-intensive NLP tasks," in Adv. Neural Inf. Process. Syst., vol. 33, pp. 9459–9474, 2020.

[7] D. Crystal, English as a Global Language, 2nd ed. Cambridge: Cambridge University Press, 2003.

[8] C. B. Zimmerman, "Historical trends in second language vocabulary instruction," in Second Language Vocabulary Acquisition, J. Coady and T. Huckin, Eds. Cambridge: Cambridge University Press, 1997, pp. 5–19.

[9] J. Read, Assessing Vocabulary. Cambridge: Cambridge University Press, 2000.

[10] J. Milton, Measuring Second Language Vocabulary Acquisition. Clevedon: Multilingual Matters, 2009.

[11] W. Grabe and F. L. Stoller, Teaching and Researching Reading. Harlow: Pearson Education, 2002.

[12] N. C. Ellis, "The psychology of foreign language vocabulary acquisition: Implications for CALL," Computer Assisted Language Learning, vol. 8, no. 2–3, pp. 103–128, 1995.

[13] V. K. Chaudhri et al., "A practical framework for using AI in education," AI Magazine, vol. 34, no. 3, pp. 47–60, 2013.

[14] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 785–794.

[15] S. Vajjala and T. Rama, "Experiments with Universal CEFR classification," in Proc. 13th Workshop Innovative Use of NLP for Building Educational Applications, 2018, pp. 54–64.

[16] S. Webb, "The effects of repetition on vocabulary knowledge," Applied Linguistics, vol. 28, no. 1, pp. 46–65, 2007. [Online]. Available: https://doi.org/10.1093/applin/aml048

[17] Council of Europe, Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Cambridge: Cambridge University Press, 2001. [Online]. Available: https://www.coe.int/en/web/common-european-framework-reference-languages

[18] S. Loewen, D. R. Isbell, and Z. Sporn, "The effectiveness of app-based language instruction for developing receptive linguistic knowledge and oral communicative ability," Foreign Language Annals, vol. 53, no. 2, pp. 209–233, 2020. [Online]. Available: https://doi.org/10.1111/flan.12457

[19] R. Godwin-Jones, "Using mobile technology to develop language skills and cultural understanding," Language Learning & Technology, vol. 22, no. 3, pp. 104–

120, 2018. [Online]. Available: https://doi.org/10125/44642

[20] T. M. Schmidt and J. Werner, "Adaptive learning paths for vocabulary acquisition: A review of intelligent tutoring systems," Int. J. Artif. Intell. Educ., vol. 31, no. 2, pp. 197–222, 2021. [Online]. Available: https://doi.org/10.1007/s40593-020-00218-4

[21] Peter, "AE 466 – Ship or Sheep? | English Pronunciation of /iː/ vs /ɪ/," *Aussie English*, Aug. 25, 2022. [Online]. Available: https://aussieenglish.com.au/ae-466-ship-or-sheep-english-pronunciation-of-i-vs-%C9%AA/

[22] M. Hismanoglu and S. Hismanoglu, "Language teachers' preferences of pronunciation teaching techniques: traditional or modern?," *Procedia - Social and Behavioral Sciences*, vol. 2, no. 2, pp. 983–989, Jan. 2010, doi: 10.1016/j.sbspro.2010.03.138.

[23] M. Hismanoglu and S. Hismanoglu, "Language teachers' preferences of pronunciation teaching techniques: traditional or modern?," *Procedia - Social and Behavioral Sciences*, vol. 2, no. 2, pp. 983–989, 2010, doi: 10.1016/j.sbspro.2010.03.138.

[24] K. Alshalaan, "A Comparison between English and Arabic Sound Systems Regarding Places of Articulation," *OALib*, vol. 7, no. 5, pp. 1–7, 2020, doi: 10.4236/oalib.1105679.

[25] P. M. Rogerson-Revell, "Computer-Assisted Pronunciation Training (CAPT): Current Issues and Future Directions," *RELC J.*, vol. 52, no. 1, pp. 189–205, 2021, doi: 10.1177/0033688220977406.

[26] D. Korzekwa, J. Lorenzo-Trueba, T. Drugman, and B. Kostek, "Computer-assisted pronunciation training—Speech synthesis is almost all you need," *Speech Commun.*, vol. 142, pp. 22–33, Jun. 2022, doi: 10.1016/j.specom.2022.06.003.

[27] S. Coulange, "Computer-aided pronunciation training in 2022: When pedagogy

struggles to catch up," *HAL (Le Centre Pour La Communication Scientifique Directe)*, Jan. 2023, doi: 10.5281/zenodo.8137754.

[28] J. Levis, "L2 pronunciation research and teaching," *J. Second Lang. Pronunciation*, vol. 7, no. 2, pp. 141–153, 2021, doi: 10.1075/jslp.21037.lev.

[29] D. Liakin, W. Cardoso, and N. Liakina, "Learning L2 pronunciation with a mobile speech recognizer: French /y/," *CALICO J.*, vol. 32, no. 1, pp. 1–25, 2014, doi: 10.1558/cj.v32i1.25962.

[30] A. Neri, C. Cucchiarini, and H. Strik, "Selecting segmental errors in non-native Dutch for optimal pronunciation training," *Int. Rev. Appl. Linguist. Lang. Teach.*, vol. 44, no. 4, pp. 357–404, 2006.

[31] R. Ai, "Automatic Pronunciation Error Detection and Feedback Generation for CALL Applications," in *Lecture Notes in Computer Science*, vol. 9288, pp. 175–186, 2015, doi: 10.1007/978-3-319-20609-7_17.

[32] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 26, no. 1, pp. 43–49, Feb. 1978.

[33] A. Neri, C. Cucchiarini, and H. Strik, "Automatic speech recognition for second language learning: How and why it actually works," *Proc. Int. Conf. Speech Lang. Technol. Educ.*, pp. 34–37, 2006.

[34] J. Mortensen, J. D. Dautenhahn, and M. Littell, "PanPhon: A resource for mapping IPA segments to articulatory feature vectors," *Proc. COLING 2016*, pp. 3475–3484, 2016.