

**AI-enabled Intelligent Assistant to Improve Reading and  
Comprehension Skills in English Language  
(AI-Powered Gaming English Vocabulary Improvement Module)**

RP-24-25J-027

Project Proposal Report

M.D.A Sooriyaarachchi- IT21173790

B.Sc. (Hons) Degree in Information Technology Specialized  
in  
Software Engineering

Department of Computer Science and Software Engineering

Sri Lanka Institute of Information Technology  
Sri Lanka

July 2024

**AI-enabled Intelligent Assistant to Improve Reading and  
Comprehension Skills in English Language**  
**(AI-Powered Gaming English Vocabulary Improvement Module)**

RP-24-25J-027

Project Proposal Report

Supervisor: Prof. Dasuni Nawinna

B.Sc. (Hons) Degree in Information Technology Specialized in  
Software Engineering

Department of Computer Science and Software Engineering


Sri Lanka Institute of Information Technology  
Sri Lanka

July 2024

## DECLARATION

### DECLARATION

I declare that this is my own work, and this proposal does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Name	Student ID	Signature
M.D.A Sooriyaarachchi	IT21173790	

Signature of the supervisor  
(Prof. Dasuni Nawinna)

  
.....

Date

26 / 8 / 2024  
.....

## ABSTRACT

The development of AI-Powered Gaming English Vocabulary Improvement Module is to focus on the challenges of the Sri Lankan High school children in learning and memorizing the complex English words. As seen, many students face challenges in learning the vocabulary brought about by the fact that English is mostly a second language. This is impacting their general language comprehensiveness in addition to achievement at school. This research seeks to examine how these students can be better taught and how their interest in the practical process of vocabulary acquisition can be aroused. The suggested solution is a gaming platform heading advanced AI technologies as Retrieval Augmented Generation (RAG) and a Large Language Model (LLM) in specific because of the ability to create a dynamic Education setting to teaching.

In the current study, a technique of dynamic content generation is used and the vocabulary words inserted by the students are transformed into vector embeddings then deposited in the vector database. The LLM employs these embeddings to facilitate access to the desired educational content as synonyms, definitions and examples of the use of a term or word for the purpose of developing the games content to suit and assist Sri Lankan students. The following are; feedback, which may be real-time, contextual hints and adaptive levels which change with the performance of the students while they play the game.

It brings emergency capabilities that allow for the detection of difficulty and the provision of various levels of difficulty within the game, and gives feedback in real time to help every learner get the proper help through their education process. First of, second-by-second analysis shows that the proposed game style tremendously enhances the results in the preservation of newly studied vocabulary, making the learning process as close as possible to the requirements of a Sri Lankan high school student. Having reached its conclusion that an integration of RAG and LLM technologies makes for an effective subject matter in a flexible and feasible approach, the study solves Sri Lanka's language learning issues. This approach has a chance to significantly improve students' vocabulary skills in English.

.

## Table of Contents

DECLARATION .....	i
ABSTRACT .....	ii
LIST OF FIGURES .....	iv
LIST OF TABLES .....	v

## Contents

LIST OF ABBREVIATIONS .....	9
1. INTRODUCTION .....	1
1.1 Background & Literature survey .....	1
1.2 RESEARCH GAP .....	6
1.3 RESEARCH PROBLEM .....	9
2. OBJECTIVES .....	12
2.1 Main Objective .....	12
2.2 Specific Objectives .....	12
3. METHODOLOGY .....	14
3.1 System Architecture .....	14
3.1.1 Software Solution .....	17
3.1.2 Commercialization .....	20
4. PROJECT REQUIREMENTS .....	21
4.1 Functional requirements .....	21
4.2 Non-functional requirements .....	21
4.3 System requirements .....	21
4.4 User requirements .....	22
5. GANTT CHART .....	23
5.1 Work Breakdown Structure (WBS) .....	24
6. BUDGET AND BUDGET JUSTIFICATION .....	25
REFERENCES .....	26
APPENDICES .....	28
Appendix – A: Plagiarism report .....	28
Appendix – B : Sample Questionnaire .....	28

## LIST OF FIGURES

Figure 1. 1: Survey report on preference for a tool that generates.....	9
Figure 1. 2 : Survey report on the likelihood of using a tool .....	10
Figure 1. 3: Survey report on the importance of receiving .....	10
Figure 3. 1: System diagram for Module.....	14
Figure 3. 2: Use case diagram.....	18
Figure 5. 1: Gantt chart .....	23
Figure 5. 2: Work Breakdown Chart .....	24

## LIST OF TABLES

Table 1. 1: Comparison of former researches .....	7
Table 3. 1 : Technologies, techniques, architectures and algorithms used. ....	16
Table 6. 1 : Expenses for the proposed system .....	25

## **LIST OF APPENDICS**

Appendix - A : Plagiarism report .....	29
Appendix - B : Sample Questionnaire .....	29



## LIST OF ABBREVIATIONS

Abbreviation	Description
RAG	Retrieval-Augmented Generation
NLP	Natural Language Processing
ML	Machine Learning
AI	Artificial Intelligence
LLM	Large Language Model
SDLC	Software Development Life Cycle
DB	Database

# 1. INTRODUCTION

## 1.1 Background & Literature survey

The importance of English language in the different facets of life in many countries has surged and is now considered a factor that enhances students' performance. For people who are learning English as second language, especially high school learners in Sri Lanka, it is important to learn English vocabulary in order to perform well academically and as well to open up job opportunities [1]. However, there are difficulties of Sri Lankan students in the learning of English vocabulary which in turn hinder language development and academic achievement [2]. The conventional approach to vocabulary learning involving memorization and practice activities is relatively old fashioned and have not been effective in addressing students learning needs [3]. In light of these challenges, the following research recommends for the development of an AI-Powered gaming English Vocabulary Improvement Module. To this end, the present module developed employs state-of-the-art technologies like Retrieval-Augmented Generation (RAG) and Large Language Models (LLMs), vector databases, and can be made into a fun and more effective pedagogical tool that can be used to make high school students in Sri Lanka learn more vocabulary and perform better.

It worth to note that the process of applying the technology in education in general as well as in language learning in particular has undergone a radical revolution over the last several decades. Computer-assisted language learning has implemented this change since it focused on CALL hence changing from classroom to computerized learning mode. In the past, CALL systems' main role was to deliver to learners exercises and quizzes in digital format which most times was not interactive and could not be easily adjusted to meet the particular learning needs of a learner. Levy (1997) and Warschauer (1996) have been among the first scholars who tried to elucidate of how technology can be used in the course of language learning.

These early systems proved to have problems as the application and articles on the field of the educational technology grew; it was researched that these systems were not very flexible to suit the needs of learners. Thus, because early CALL systems were static in nature, these systems would not be able to deliver the experiences which are needed for learners to acquire language – this specifically targeted at vocabulary learning. This realization resulted in the emergence of procedures which involve artificial intelligence (AI) and machine learning (ML) in ALE [6]. They allow systems to monitor students' behaviors, comprehending their learning capabilities, and changing content presented to them in the process enhancing the learning process.

With a focus on AI-driven educational tools, several innovations in the sphere among which are NLP, LLMs have been noted, and mention of vector databases. AI can be a part of Natural Language Processing in which computers perform comprehension of text and generation of text. Tokenization, word embeddings, and sentence embeddings have been used to in creating smart learning solutions, which can transform and create text-based materials based on the learner.

### **Natural Language Processing (NLP):**

NLP serves as the base on which many educational assistances of the present are developed. It allows the system to interpret and produce human language which is very essential for the upbringing of educational content. Many methods, including tokenization that is the division of a textual material into distinctive words or phrases, are crucial for the processing of student responses [7]. The other important NLP technique is word embeddings, which enables the system to represent words within continuous-vector space and to capture semantic similarities between different words [8]. This is further generalized to whole sentences by what are known as sentence embeddings and allows the system to provide more contextually relevant responses [9]. A good understanding of these NLP ability on how to process student's input and how to give them meaningful content is crucial in the development of the system.

### **Large Language Models (LLMs):**

Another expressive advancement in NLP is the creation of LLMs – large magnetic language models, as it is the case of OpenAI's GPT-3. These are trained on large corpora of text and can produce text which is indistinguishable from text written by a human being; this makes LLMs useful in education. For example, GPT-3 can provide contextually appropriate text, answers to questions and explanations, which can be useful for producing individualized educational content [6]. Nevertheless, the perspective of LLMs in the generation of the text is vast and, although their application in education contexts is relatively recent, the future is bright especially when tied with other Artificial Intelligence advancements, which can improve the learning model.

### **Vector Databases: A New Paradigm in Data Retrieval**

Being AI systems, like LLMs, data-heavy in generating and processing data, the issue of efficient data storage emerges. It is in this area that vector databases are most useful. Vector databases are expected to store and manage high-dimensional vector embeddings—essences that depict the semantic meanings in forms that are comprehensible and manipulable by machines. Unlike most common database that use key-value pair to store and query data, Vector databases search the contents based on similarity, and therefore, their application in contexts where meaning and context is important such as in generation of educational content [10].

Referring to the AI-Powered Gamified English Vocabulary Improvement Module, vector databases are utilized to store vocabulary embeddings, meaning and other related educational materials. These embeddings that models such as BERT [11] produce benefits the system in being able to search and bring out content that is related semantically to what the student typed in or said. For instance, if a student types “ecosystem”, the vector database allows the student to use other correlated terms such as “biodiversity” and “habitat” which enhances their learning process.

With the help of vector databases like FAISS [10] or Milvus, the system can perform real-time similarity search that constitutes the core of the formation of individual educational content that is individual and reacts on user's needs in real time. This capability is of optimum usage in educational context where it can play a central role in improving the learning experience by presenting the relevant content of right type.

AI, machine learning, and data management achievements can be clearly observed in language learning technologies, which nowadays are presented by such platforms as Duolingo and Memrise, which have already proven themselves in Personalized learning. Such platforms involve the learning machine algorithms that make the display of content personalized by responding to the progress of the learner. For instance, Duolingo applies the spaced repetition in combination with a game to review the learned vocabulary, and it has become one of the most effective and well-known applications for language learning [12].

However, these platforms are rather important development, they have rather general approach to learning and distinctive for providing more deep customization in order to teach specific vocabulary, which may be very important to persons from Sri Lanka, where major language is not English. In addition, these platforms are often based on pre-set content, something that restricts the individual approach that can be offered to the learners. On the other hand, the AI-Powered Gamified English Vocabulary Improvement Module seeks to transcend these limitations by using RAG, LLMs, and vector databases so as to create content material based on individual learner's needs and expertise within a student [13]. This approach is a major step forward in the development of educational technology, which is a more elastic and proactive system where study is tuned to the specifics of non-English speaking learners in Sri Lanka.

New trends in LLMs show that machine learning can cultivate individualized educational material and even GPT-3. For example, GPT-3 has demonstrated an uncanny learning effect in which is capable of writing like a human [6], and thus can be used in the education system. Nevertheless, LLMs are not very widely used in education, and there is a great deal of potential for their further development, especially in the combination of these models with other classes of AI, including vector databases, to develop more effective and sophisticated ones.

### **Retrieval-Augmented Generation (RAG):**

RAG is one of the latest state of arts that blend the advantages of both information retrieval and generative models. In educational contexts, RAG enables the access of appropriate educational content from a pool of knowledge and the educational content base that, in turn, is applied by the LLM to provide specific responses and exercises for the student. This approach has the potential of revolutionizing the way content in education is developed and presented to the learners in a way that is effective to every individual in the class in a real-time basis.

Salient conventional procedures in vocabulary learning have included the construction of applications for learning vocabularies, flash cards and the spacing repetition system (SRS). These strategies involve the rote use of words and phrases in academic contexts to reinforce the learning of vocabulary and although they can be useful, they are not very useful as they do not meet the needs of the learners [4]. In addition, a large number of approaches are not flexible and do not correlate to the learner's progress, which can negatively influence their efficiency enhancing the learner's long-term use of vocabulary.

On the other hand, the proposed AI-Powered Gamified English Vocabulary Improvement Module is a form of development on the existing techniques by incorporating RAG, LLMs, and vector databases for an improved learning environment. Traditional knowledge tools on the other hand can have prepared quasi-images, whereas this one produce information on the fly and based on the

student's input and his/her performance. This enables the provision of content that is more personalized meaning each child will be receiving content that will suit his or her capability level.

Besides, the utilized approach of gamification makes this module different from other approaches to teaching. The use of points, levels, and rewards are part and parcel of the module in order to adopt the game like features which cause enhancement in the interest level and motivational level of the students during learning process. It caters for one major drawback of most conventional pedagogy, that is the lack of learning interest as displayed by students.

The AI-Powered Gamified English Vocabulary Improvement Module also apply the concept of adaptive learning, where it is flexible enough to changing the difficulty level of the exercises of the student. This makes it hard for any of the students to be overchallenged while at the same time, none of the students is under-challenged so that the process of learning and improvement goes on. This adaptive approach is a major improvement from the normal approach where no consideration is given to the requirements and capacities of the learner.

The present study has the following major scholarly contributions to the field of educational technology and language learning. Firstly, it shows that RAG can work in educational tools illustrating its capability to reconsider the generation and delivery of personalized learning content. Although RAG has been discussed in other disciplines, its usage in learning especially in the learning of new vocabulary is relatively recent. Thus, this research may prove useful to RAG for developing strategies that enhance the provision of special learning activities for learners.

Second, this research is novel in incorporating gamification of education with the latest AI technologies. Although gamification has been implemented in learning for a while now and its benefits are known, the employ of AI technologies connecting RAG and LLMs are a new perspective that opens up new opportunities for creating tasks and increasing students' engagement. Through discussing this connection between gamification and AI, this study enhances current knowledge of integration between these two technologies in enhancing educational practice.

Third this research finds a particular niche in language learning tools required for the Sri Lankan student. As is true for many educational sectors, there is plethora of language learning resources, but there are no resources designed for the L2 learners in Sri Lanka. That is the kind of research that offers a scalable solution and that can be applied in other non-native English-speaking environments, thus providing a model of how technology, AI specifically and gamification in particular, can be used to address the needs of these learners.

The importance of this study is flagged in the proposition that it can facilitate raising achievements in education of Sri Lankan high school learners who experience a lot of difficulties when it comes to learning English vocabulary. In a way, this research has the ability to help students improve their performance as well as their future prospect; additionally, making their learning experience fun and enjoyable. If this module was implemented successfully, it may well also become good practice for other educational systems where technologies such as the AI or game-based learning are being sought as a way of providing solutions to questions that have been lingering within educational systems for a long time.

Moreover, the outcomes of this study will bring benefit in the aspects of students, educators, policymakers and the providers of fund for the implementation of quality education. In the case of educators, they shall get one more valuable tool when it comes to helping the learners to resolve their language difficulties and get them the deserved result. Thus, to policymakers, this research has captured the possibilities of how AI and gamification enhance education encouraging moves toward betterment of the current education systems. For fund providers this research presents feasibly solutions that can be adopted in different regions/community thus making a stride in solving issues to do with education for economic development.

Finally, it can be affirmed that AI-Powered Gamified English Vocabulary Improvement Module is a groundbreaking innovation in the sphere of foreign language learning as being a sustainable solution for the problems of Sri Lankan high school students concerning the difficulties encountered during learning English vocabulary. Thus, this research spotlights the capacity to transform traditional learning methodologies by applying the best AI technologies in combination with gamification elements and creating highly personalized, effective, and entertaining educational tools that – when used in practice – may lead to a rapid enhancement of the students' language skills.

## 1.2 RESEARCH GAP

This research project aims to develop an AI-powered gamified English vocabulary improvement module that will facilitate improving the vocabulary and reading abilities of students in Sri Lanka. The study addresses the challenge of how to make traditional vocabulary learning more effective and engaging, by integrating advanced AI technologies. Students have the facility to insert lists or words, individually, into this module. It also has LLM along with RAG, which personalizes these very educational games and activities in a dynamically generated format. Games are enriched by definitions, meanings, and Sinhala translations for comprehensive learning.

There continues to be an important gap in meeting the specific demands of non-native English speakers of Sri Lanka, despite major developments in language learning technologies and the introduction of AI in educational tools. Through gaming and adaptive algorithms, language learning services like Duolingo and Memrise have achieved significant progress in individualized learning. These systems, however, might not offer the level of deep personalization needed for the unique vocabulary learning issues experienced by Sri Lankan high school students, as they frequently focus on general language development [2]. Additionally, the content offered by these platforms is typically pre-programmed, limiting their ability to deliver truly personalized and contextually relevant educational experiences in real time [12].

The ability of Large Language Models (LLMs) to produce human-like text has been effectively showed by recent developments such as GPT-3. However, there is still a lack of research on the use of LLMs in educational settings, particularly when combined with other AI technologies like vector databases and Retrieval-Augmented Generation (RAG) [13][6]. The educational tools now on the market do not fully fulfill the promise of these technologies to build a fully responsive and adaptive learning environment that meets the specific needs of students.

The specific gap this research addresses is the lack of an educational tool that integrates RAG, LLMs, and vector databases to offer a highly personalized, engaging, and effective vocabulary learning experience tailored to the needs of Sri Lankan high school students. While traditional vocabulary learning strategies and established educational platforms have attempted to address vocabulary development, they often fall short in terms of effectively inspiring students or accommodating their specific learning needs, particularly in non-native English-speaking environments [1][3].

By creating the AI-Powered Gamified English Vocabulary Improvement Module, which will offer real-time, personalized content generation depending on the student's input and performance, our research aims to fill this gap. In addition to using gaming to engage students, the module will adjust the level of difficulty and information depending to each student's progress, making for a more successful learning environment. This research aims to develop a tool that not only addresses the current shortcomings of current platforms but also establishes a new benchmark for personalized and adaptive learning in the context of vocabulary acquisition for non-native English speakers by utilizing modern AI technologies [10][11].

The purpose of this project is to gain more understanding of the methods by which AI technologies, such as vector databases, RAG, and LLMs, can be used to produce personalized and flexible learning materials. The results of this study will close an important gap in the research literature and provide a scalable solution to the challenges students face with vocabulary learning around the world by presenting a new approach to language learning that can be personalized for many different kinds of non-native English-speaking environments [8].

Criteria	Research Paper 1[14]	Research Paper 2[15]	Research Paper 3[12]	Proposed Module
OBJECTIVE	Improve vocabulary learning through educational games	Investigate the effectiveness of online vocabulary games	Use educational games to enhance vocabulary retention in young learners	Enhance vocabulary acquisition and retention through AI-powered gamification
CUSTOMIZATION	No	No	No	Yes (Dynamic content generation based on student input)
ENGAGEMENT	Yes (Games increase student engagement)	Yes (Students preferred online games over traditional methods)	Yes (Games were engaging for young learners)	Yes (Gamification and real-time feedback to engage students)
ASSESSMENT & FEEDBACK	Yes (Pre-test, post-test, delayed post-test)	Yes (Pre-test and post-test, teacher feedback)	Yes (Pre-test, post-test, delayed post-test)	Yes (Real-time feedback and adaptive difficulty)
PERSONALIZATION	No	No	Yes	Yes (Content tailored to individual student needs using LLM and RAG)
DYNAMIC GAMES	No (Static games, predefined content)	No (Static online games with fixed content)	No (Predefined games without dynamic content)	Yes (Games dynamically generated based on student inputs and needs)



SINHALA DEFINITIONS	No	No	No	Yes (Includes Sinhala translations for better comprehension for Sri Lankan students)
DYNAMIC IMAGE GENRATE	No	No	NO	Yes

*Table 1.1: Comparison of former researches*

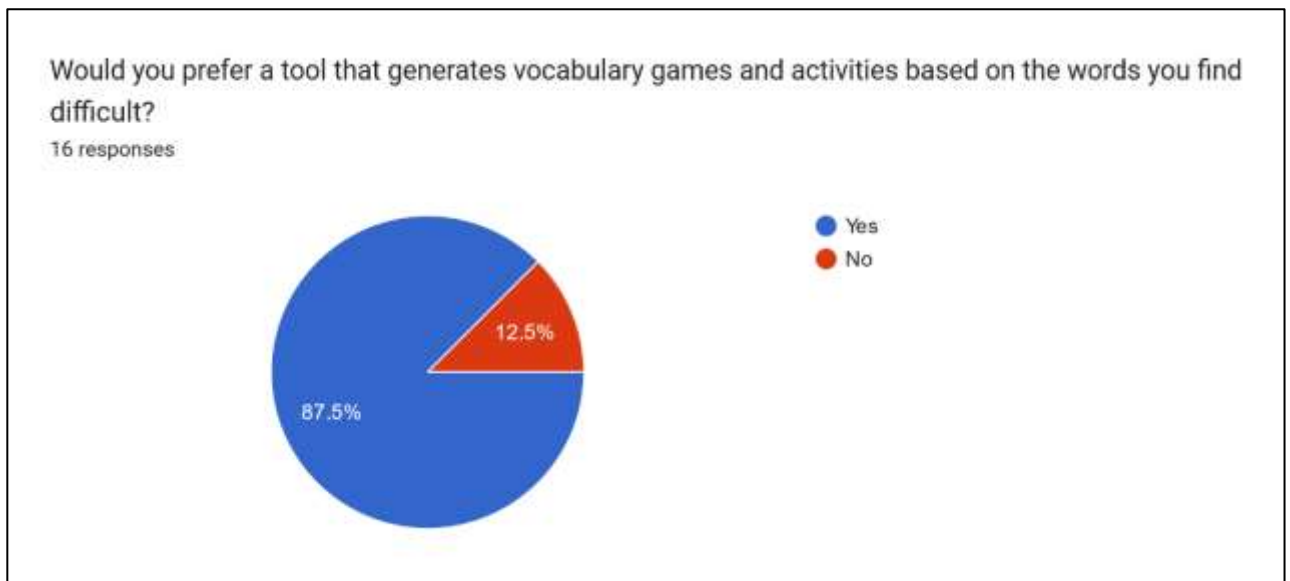
As in the above diagram we can see my research gap clearly.

### 1.3 RESEARCH PROBLEM

Teaching English vocabulary to non-native speakers is still a challenge, especially in countries like Sri Lanka. Traditional vocabulary learning methods often fail to engage students and adapt to their different requirements, resulting in poor retention and limited language development. Modern language learning platforms offer personalized and gaming features, but they typically rely on static information and fail to adapt in real-time to the specific vocabulary difficulties of individual students.

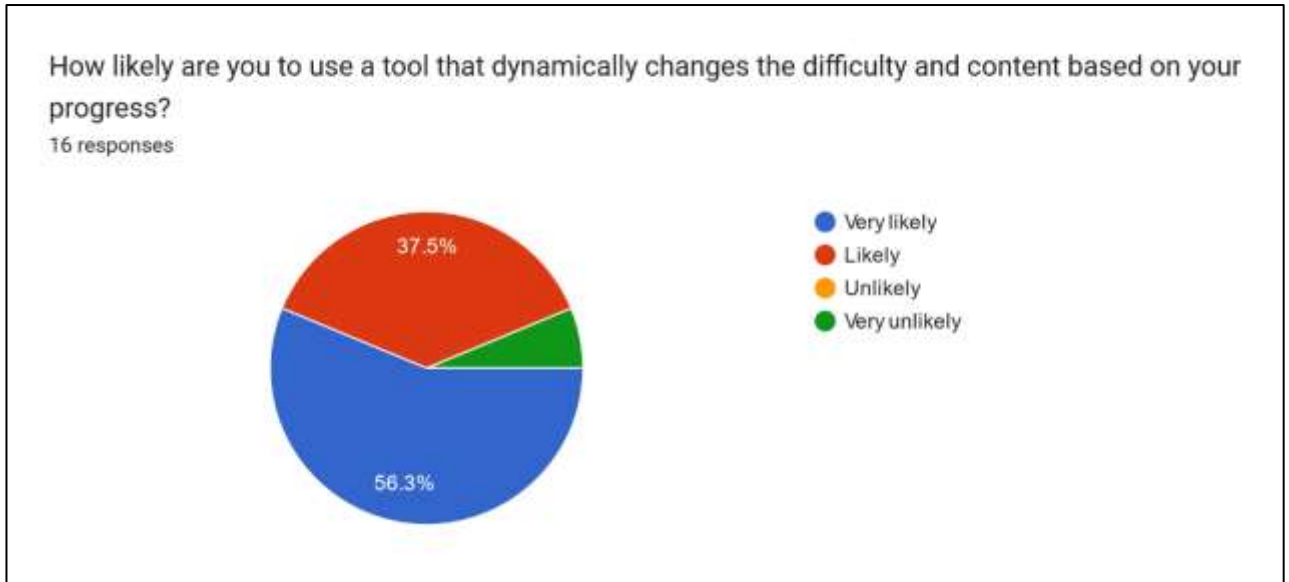
Furthermore, existing AI-driven educational tools have not fully utilized the potential of advanced technologies such as Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG) to create highly personalized and contextually relevant learning experiences. This gap in the application of the latest technologies has resulted in a lack of educational tools that can dynamically generate content and provide a tailored learning path that adapts to the student's changing requirements.

**The core research problem**, therefore, is the need to develop an AI-powered educational tool that can offer a fully adaptive, personalized, and engaging vocabulary learning experience for Sri Lankan high school students. In order to dynamically create content, such as educational games (such as word search puzzles, snake games, or matching puzzles, Image Matching Games) based on challenging words entered by students while reading paragraphs, this program should incorporate LLMs, RAGs, and vector databases. These activities have to be academically demanding and relevant to the context, responding to the unique vocabulary development needs of non-native English speakers.



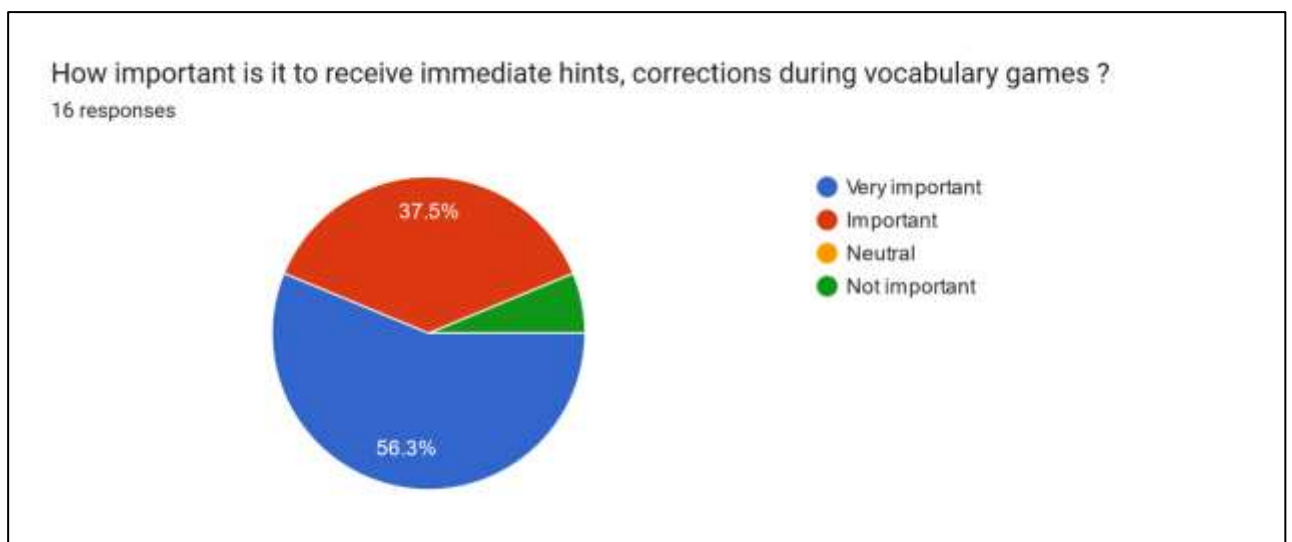
*Figure 1.1: Survey report on preference for a tool that generates vocabulary games based on difficult words.*

According to the above chart, a significant number of respondents (87.5%) would prefer a tool that creates vocabulary games and exercises based on the terms they find challenging. This indicates a strong interest in personalized vocabulary learning tools among the participants.



*Figure 1.2 : Survey report on the likelihood of using a tool that dynamically changes difficulty and content based on progress*

As shown in the second chart, 93.8% of respondents overall said they would be likely to use a tool that modifies the material and difficulty dynamically based on the user's progress. This highlights the importance it is for educational tools to have adaptive learning features.



*Figure 1.3 : Survey report on the importance of receiving immediate hints and corrections during vocabulary games.*

The final graph demonstrates that 93.8% of respondents value receiving quick hints and corrections during vocabulary games. Real-time feedback systems are essential for improving the efficacy of learning, as highlighted by this.

## **2. OBJECTIVES**

### **2.1 Main Objective**

The main objective of this study is to develop an AI-Powered Gamified English Vocabulary Improvement Module tailored for Sri Lankan high school students. This module will enable students to input challenging English vocabulary words, which will then be used to generate personalized and interactive educational games. The system will make use of modern AI technologies, such as vectorization techniques, large language models (LLMs), and retrieval-augmented generation (RAG), to create dynamic game experiences that adapt to the specific needs for learning of each student and provide immediate feedback, effectively improving vocabulary retention and learning.

### **2.2 Specific Objectives**

There are Three specific objectives that must be reached in order to achieve the overall objective described above.

#### **Sub Objective 01**

Identify and Analyze Student Vocabulary Needs

To collect and analyze vocabulary inputs from students, determining their specific learning requirements and difficulty levels. This will involve developing a data collection mechanism within the application where students can input or select vocabulary words, they find challenging. Utilizing natural language processing techniques and vectorization, these inputs will be converted into vector embeddings that represent the semantic relationships between words. The analysis of these vectors will inform the customization of game content, ensuring that it effectively addresses individual learning gaps.

#### **Sub Objective 02**

Generate Personalized and Adaptive Educational Games

To create interactive and engaging educational games that adapt to the student's proficiency level and learning pace by leveraging vectorized representations of vocabulary data. This will involve using advanced AI technologies such as LLMs and RAG, which employ vectorization to represent student-inputted vocabulary in a high-dimensional space. The system will dynamically generate games (such

as word search puzzles, snake games, or matching puzzles, Image Matching Games) and content that is semantically aligned with the student's learning needs. For example, games like a vocabulary-based Snake Game will be implemented where students collect letters to form words, matching definitions, or context usage challenges. The games will adjust in difficulty and complexity in real-time, responding to the student's performance to keep them engaged and appropriately challenged.

### **Sub Objective 03**

#### **Implement Real-Time Feedback**

For instance, to give an instant response while students are playing a game, thus availing information to the students as whether their answers are right or wrong and at same time giving hints as well as encouraging remarks. This will include incorporation of feedback mechanisms within the games in order to facilitate learning of the students and real time correction of their wrong doings. The feedbacking will be done in a way that will encourage learning as it will keep on reminding and encouraging the learners as they go through the educational material.

### 3. METHODOLOGY

#### 3.1 System Architecture

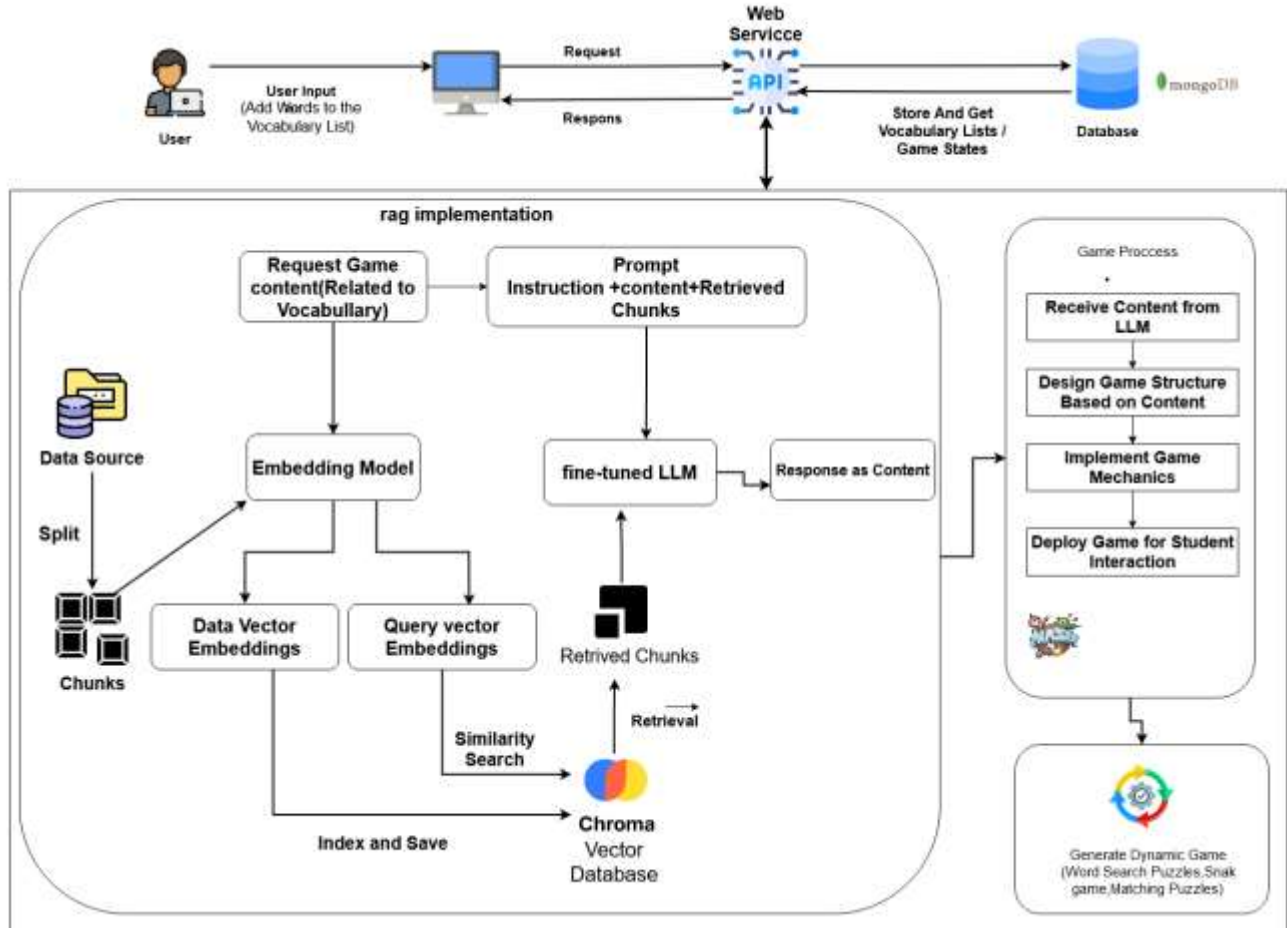


Figure 3. 1 : System diagram for overall AI-Powered Gamified English Vocabulary Improvement Module

In the above diagram, the users interact with the system through a web interface, enabling them to access various services related to vocabulary enhancement and game generation. Here's how the system works.

1. Users interact with the system via a web interface where they input vocabulary words, they find challenging. This inputted triggers the backend processes that generate personalized educational content.

2. The system uses Retrieval-Augmented Generation (RAG) to convert the inputted words into vector embeddings, which are used to search a vector database for relevant content. The retrieved information is then processed by a fine-tuned Large Language Model (LLM) to create tailored educational content for the games.
3. The content generated by the LLM is structured into a game format, such as word search puzzles, snake games, or matching puzzles, Image Matching Games. The system implements game mechanics and deploys the games, making them ready for the user to play.
4. The games are dynamic, adapting to the user's progress and providing real-time feedback. This ensures a personalized and engaging learning experience, enhancing vocabulary retention through interactive gameplay.



*Table 3. 1 : Technologies, techniques, architectures and algorithms used*

<b>Technologies</b>	ReactJS, Python, MongoDB, Chroma, Large Language Model (LLM), FAISS/Elasticsearch, Unity, Phaser.js, Django, Node JS, Hugging Face
<b>Techniques</b>	Vectorization, Gamification, Embedding Models, Similarity Search
<b>Algorithms</b>	Self-Attention Mechanism, Transformer
<b>Architectures</b>	Microservices Architecture, RAG (Retrieval-Augmented Generation)

### 3.1.1 Software Solution

- **Requirements Gathering**

- **Data gathering**

We have found data sets of English Vocabulary online websites like Kaggle, High School English book. We are going to use that data sets in our project. And we will use large language model and fine-tune it.

- **Conducting a survey**

To gain insights into the effectiveness of current vocabulary learning methods and identify the challenges faced by students, we conducted a survey. The survey targeted Sri Lankan high school students and educators, including close and open-ended questions designed to gather information about their experiences with vocabulary learning, preferences for game-based learning, and familiarity with English vocabulary. The results have guided the design and development of our module, ensuring it meets the needs of the target audience.

- **Feasibility Study (Planning)**

- **Economic feasibility**

The project's development costs and benefits are covered in the economic feasibility report. The approach would be unsuccessful without a suitable economic feasibility plan in place. As a result, the suggested approach ought to be more economical and effective.

- **Scheduled feasibility**

A "scheduled feasibility study" is a feasibility study that is often carried out according to a defined timeline or schedule, frequently as part of a wider project management process. This kind of feasibility study is organized and prepared in advance to evaluate the viability of a project at particular checkpoints or stages of development.

- **Technical feasibility**

Before starting our project, a technical feasibility analysis is a crucial part of the overall feasibility assessment. It focuses on determining whether a suggested project can be technically performed successfully. In other words, it looks to see if the project can be carried out with the resources and technologies at hand.

- **Design (system and software design documents)**

The high-level requirements obtained during the earlier phases are translated into an exact blueprint for how the software system will be created during the Design Phase of the Software Development Life Cycle (SDLC).

- **Use case diagram**

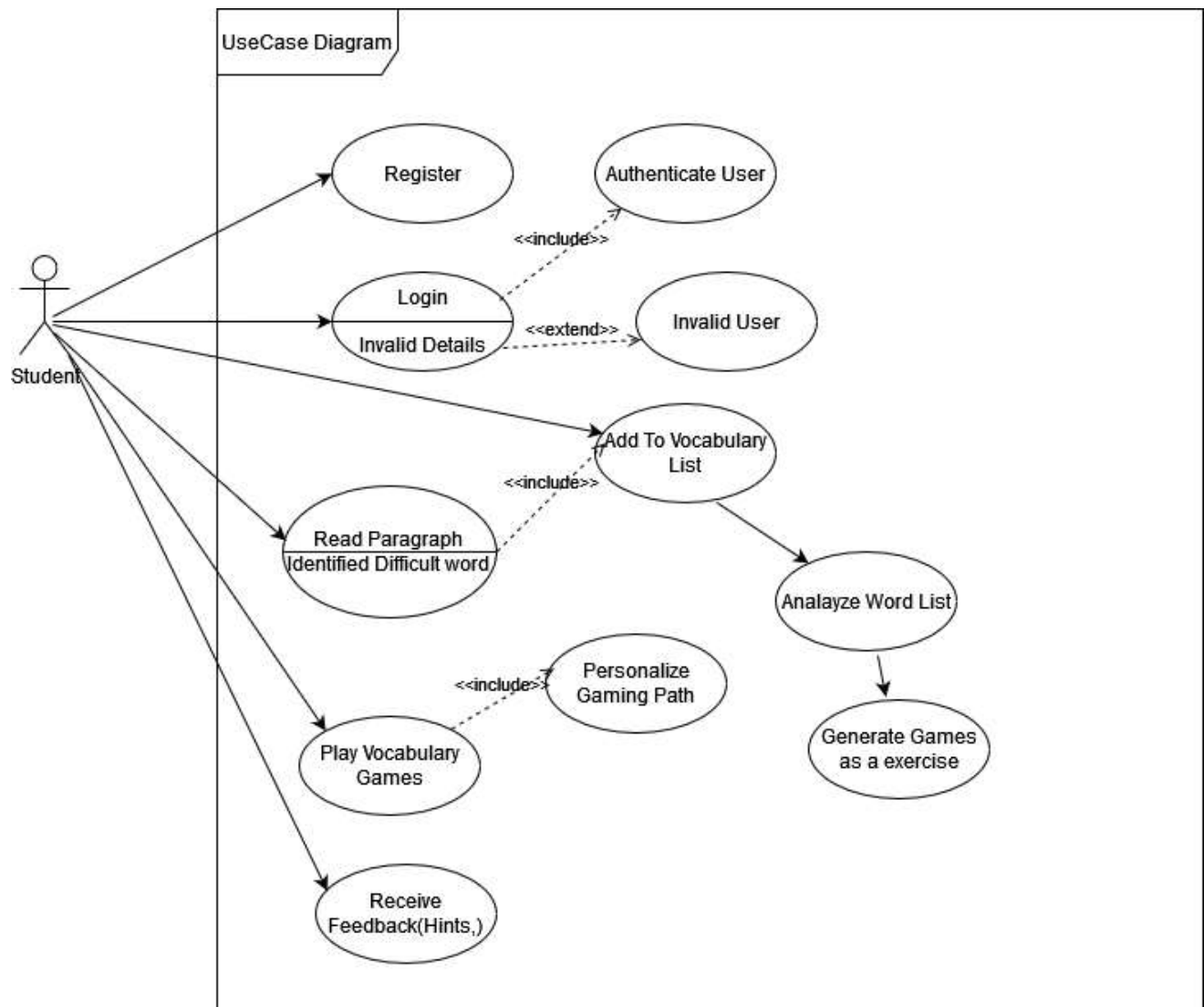


Figure 3.3 :Use Case Diagram of AI-Powered Gamified English Vocabulary Improvement Module

## • **Implementation (Development)**

The implementation process is being carried out as shown below.

### Development of Core Features:

- Develop the user interface to allow seamless input of vocabulary words and interaction with dynamically generated games.
- Implement AI-powered content generation using LLMs and RAG, ensuring real-time adaptability and personalized learning experiences.

### Backend Development:

- Set up microservices architecture to manage different functionalities such as game generation, vocabulary tracking, and user progress.
- Integrate vector databases for efficient storage and retrieval of learning materials and user data.

### Integration and Deployment:

- Integrate all system components and test for compatibility and performance.
- Deploy the solution on cloud infrastructure to ensure scalability and accessibility.

## • **Testing**

- Develop unit tests for each module, including the LLM, game mechanics, and feedback system. These tests will verify that individual components work as expected.
- Test the integration between the frontend, backend, and game components, ensuring that data flows smoothly and the system responds correctly to user input.
- Conduct end-to-end testing of the entire system, simulating real-world use cases to identify and fix any issues.
- Test the system with a group of students and educators to make sure it meets their needs and expectations. Collect feedback and make necessary adjustments before the final deployment.

### 3.1.2 Commercialization

- Target Audience and Market
  - High school students
  - English language learners
  - English Educators
- Business Strategy

This application's business model is based on a Software as a Service (SaaS) model, with two different tiers designed to meet different user requirements.

1. Advertisements, especially from educational institutions, ensure that Free Tier is supported. The widespread adoption among learners is encouraged by broad accessibility. Users are introduced to the app's features and benefits through this tier as their entry point.
2. The Premium Version is available through a subscription-based model and offers an ad-free experience. An environment that has extra features and resources.

By introducing these two tiers, the app can attract a wide range of users, from casual users who use the free version to dedicated learners and educators who opt for the premium services. By utilizing a dual-tiered approach, revenue streams can be diversified and user engagement and retention can be improved, positioning the app as a valuable tool in the educational landscape.

#### 3.1.2.1 Future Scope

Future updates may include features that foster collaborative learning, like multiplayer games or peer-to-peer challenges. By encouraging students to interact with their peers, learning could become more interactive and socially engaging

## 4. PROJECT REQUIREMENTS

### 4.1 Functional requirements

- The system must allow students to manually input a list of challenging vocabulary words or upload them all at once
- The system must validate the input to ensure only valid words are entered.
- The system must use a Large Language Model (LLM) to generate game content based on the inputted vocabulary
- The system should dynamically generate game content based on the vocabulary words provided by the student.
- Immediate feedback should be provided to students by the system during gameplay.

### 4.2 Non-functional requirements

1. User friendliness – The interface should be easy to use, with an emphasis on visual appeal, simple controls, and clear navigation.
2. Reliability – The system should perform without any Problem in given time
3. Performance – The system must provide real-time feedback as a hints, Motivations and game content generation with minimal latency.

### 4.3 System requirements

As per the system requirements, the AI-Powered Gamified English Vocabulary Improvement Module has the following system requirements. To ensure the proposed system functions effectively, these requirements specify the necessary software resources and tools.

- **ReactJS and Node.js:** Designed a web application interface that is user-friendly, responsive and interactive, with the ability to input difficult vocabulary words, interact with dynamic educational games, and track learning progress.
- **Python and Django:** These serve as the primary backend programming languages and framework, respectively, for managing backend operations, integrating AI models, and handling user data and game logic.
- **Unity and Phaser.js:** Essential for developing and rendering the educational games. Unity and Phaser.js provide the necessary environments for implementing game mechanics, visual interactions, and animations.
- **MongoDB and Chroma:** MongoDB is utilized as the primary database to store user data, vocabulary lists, and game progress, while Chroma is used to manage vector embeddings and retrieve educational content quickly and efficiently.
- **Microservices Architecture:** The system is built using a microservices architecture, where each component (e.g., game engine, user management, content generation) operates

independently but communicates with others through APIs, ensuring scalability and maintainability

- **RAG (Retrieval-Augmented Generation):** The architecture combines retrieval-based and generative AI models to generate personalized educational content dynamically, improving the relevance and effectiveness of learning materials.
- **Docker:** Used to containerize the various components of the system, ensuring consistent deployment across different environments and simplifying the integration of different services.

#### 4.4 User requirements

This web service will be developed for

- **High School Students**
  - The system is primarily designed for Sri Lankan high school students to help them improve their English vocabulary through interactive and personalized games.
- **English Language Teachers**
  - Teachers can use the system to support their classroom teaching by assigning
- **Parents**
  - Parents can use the system to support their children's vocabulary learning at home.

## 5. GANTT CHART

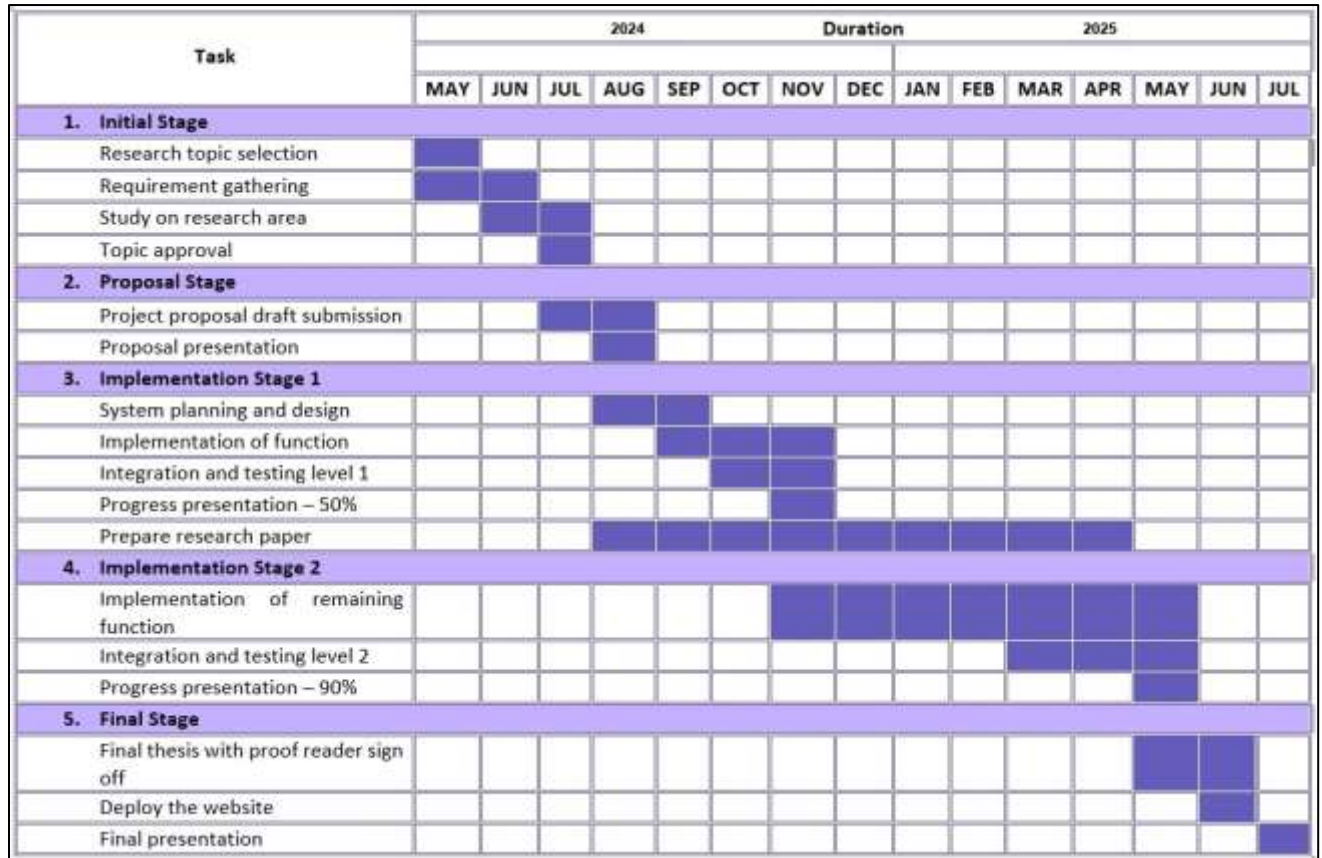


Figure 5. 1 : Gantt chart



## 5.1 Work Breakdown Structure (WBS)

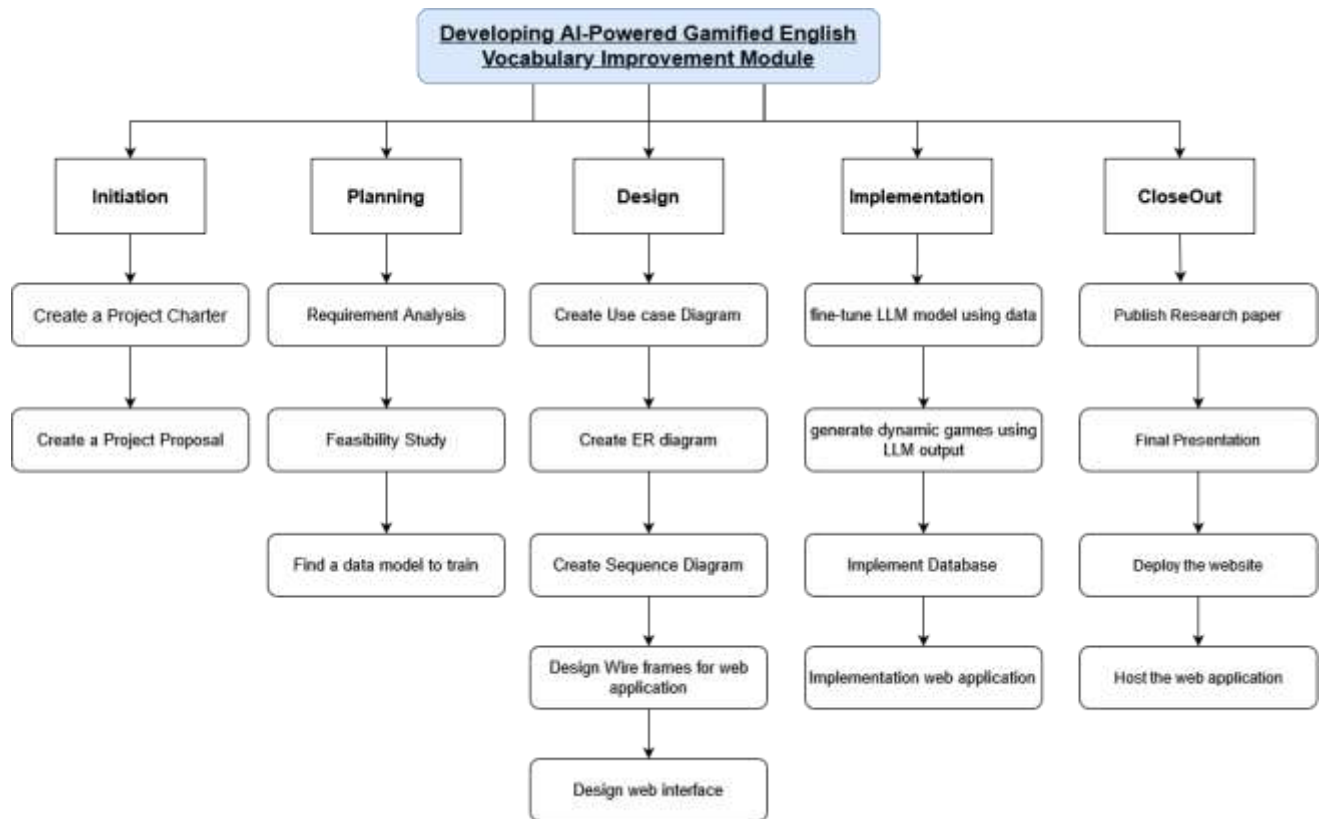


Figure 5. 2 : Work Breakdown Chart

## 6. BUDGET AND BUDGET JUSTIFICATION

Ex penses	
Requirement	Cost (LKR.)
Deployment cost	8,000.00
Domain	4,845.00
Wi-fi / Mobile data	3,000.00
<b>Total Cost</b>	<b>15,845.00</b>

*Table 6. 1 : Expenses for the proposed system*

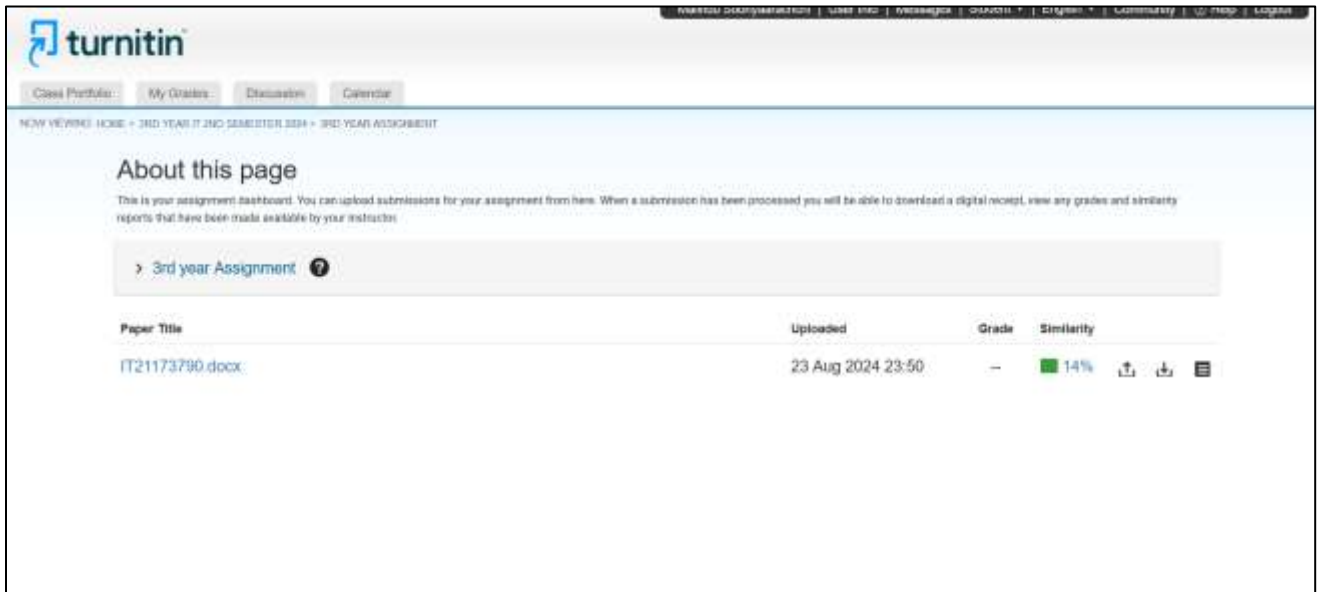
## REFERENCES

- [1] E. Ridge, "Crystal, David. 2003. English as a Global Language. Second edition. Cambridge University Press," *Per Linguam*, vol. 20, no. 1, 2011. [Online]. Available: <https://doi.org/10.5785/20-1-80>.
- [2] I. S. P. Nation, *Learning vocabulary in another language*. Cambridge University Press, 2001. [Online]. Available: <https://doi.org/10.1017/cbo9781139524759>.
- [3] N. Schmitt, "Review article: Instructed second language vocabulary learning," *Language Teaching Research*, vol. 12, no. 3, pp. 329–363, 2008. [Online]. Available: <https://doi.org/10.1177/1362168808089921>.
- [4] S. Alzahrani, "Linking multimedia vocabulary CALL research to SLA cognitive theories," *American Journal of Educational Research*, vol. 5, no. 7, pp. 821–827, 2017. [Online]. Available: <https://doi.org/10.12691/education-5-7-19>.
- [5] M. Warschauer, "Computer-assisted language learning: An introduction," in *Multimedia language teaching*, S. Fotos, Ed. Tokyo: Logos International, 1996, pp. 3-20.
- [6] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, and D. Amodei, "Language Models are Few-Shot Learners," *arXiv*, 2020. [Online]. Available: <https://doi.org/10.48550/arxiv.2005.14165>.
- [7] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to information retrieval*. Cambridge University Press, 2008.
- [8] T. Mikolov, K. Chen, G. S. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv*, 2013. [Online]. Available: <https://doi.org/10.48550/arxiv.1301.3781>.
- [9] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using Siamese BERT-networks," in *Proc. 2019 Conf. Empirical Methods in Natural Language Processing and 9th Int. Joint Conf. Natural Language Processing (EMNLP-IJCNLP)*, 2019, pp. 3982-3992.
- [10] J. Johnson, M. Douze, and H. Jegou, "Billion-Scale Similarity Search with GPUs," *IEEE Trans. Big Data*, vol. 7, no. 3, pp. 535–547, 2021. [Online]. Available: <https://doi.org/10.1109/tbdata.2019.2921572>.

- [11] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *arXiv*, 2018. [Online]. Available: <https://doi.org/10.48550/arxiv.1810.04805>.
- [12] L. von Ahn, "Duolingo: Learn a language for free while helping to translate the web," in *Proc. 2013 Int. Conf. Intelligent User Interfaces*, 2013, pp. 1-2.
- [13] P. Lewis, E. Perez, D. Kiela, M.-F. Moens, and W.-T. Yih, "Retrieval-augmented generation for knowledge-intensive NLP tasks," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 9459-9474.
- [14] A. M. Saleh and A. S. Ahmed Althaqafi, "The Effect of Using Educational Games as a Tool in Teaching English Vocabulary to Arab Young Children: A Quasi-Experimental Study in a Kindergarten School in Saudi Arabia," *Sage Open*, vol. 12, no. 1, 2022. [Online]. Available: <https://doi.org/10.1177/21582440221079806>.
- [15] F. W. M. Yip and A. C. M. Kwan, "Online vocabulary games as a tool for teaching and learning English vocabulary," *Educational Media International*, vol. 43, no. 3, pp. 233–249, 2006. [Online]. Available: <https://doi.org/10.1080/09523980600641445>.

## APPENDICES

### Appendix – A: Plagiarism report



The screenshot shows the Turnitin user interface. At the top, there's a navigation bar with links like 'NEW! TURNITIN', 'USER INFO', 'MESSAGE', 'DASHBOARD', 'COURSE', 'COMMUNITY', 'HELP', and 'LOGOUT'. Below this, there are tabs for 'Course Portfolio', 'My Grades', 'Discussion', and 'Calendar'. The main content area is titled 'About this page' and explains that this is the assignment dashboard where users can upload submissions and view grades and similarity reports. A button labeled '> 3rd year Assignment' with a help icon is visible. Below this, a table lists the submission details:

Paper Title	Uploaded	Grade	Similarity
IT21173790.docx	23 Aug 2024 23:50	—	14%

The similarity score of 14% is represented by a green bar. There are also icons for downloading the report and other actions.

### Appendix – B : Sample Questionnaire

[Form Link](#)