

BioMentor - Personalized E-Learning Platform For English Medium A/L Biology Subject Students in Sri Lanka

LLM Based Provide answers for structured and essay type of questions and
evaluate answers based on approved resources.

24-25J-257

Project Proposal Report

Sajeevan S - IT21204302

B.Sc. (Hons) in Information Technology Specializing in
Software Engineering

Department of Computer Science & Software Engineering
Sri Lanka Institute of Information Technology
Sri Lanka

BioMentor - Personalized E-Learning Platform For English Medium A/L Biology Subject Students in Sri Lanka

LLM Based Provide answers for structured and essay type of questions and
evaluate answers based on approved resources.

24-25J-257

Project Proposal Report

B.Sc. (Hons) in Information Technology Specializing in
Software Engineering

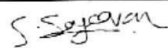
Department of Computer Science & Software Engineering

Sri Lanka Institute of Information Technology

Sri Lanka

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 our 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
Sajeevan S	IT21204302	

The above candidate is carrying out research for the undergraduate Dissertation under my supervision.



Signature of the supervisor:

(Dr. Sanvitha Kasthuriarachchi)

Date: ...22/8/24'



Signature of the Co-supervisor:

(Ms. Karthiga Rajendran)

Date: 22/08/2024

ABSTRACT

E-learning is an educational tool that uses technology to facilitate learning, providing students with easily accessible and adaptable resources to improve their education. Traditional learning methods often rely heavily on teacher support, which can limit students' ability to develop self-directed learning skills. To address this issue, we propose an innovative e-learning platform designed specifically for English-medium biology students. The proposed platform will help to enhance student engagement, understanding levels, and overall academic performance in Biology using the latest technological tools.

KEYWORDS

- Independent Learning
- E-learning Platform
- Advanced-Level Biology
- Self-Evaluation
- Structured Essay Questions
- Essay Writing Skills
- Large Language Models
- Self-Improvement Strategies

LIST OF ABBREVIATIONS

LLM	Large Language Model
A/L	Advanced-Level
DB	Database
API	Application Framework Interface
UI	User Interface
BERT	Bidirectional Encoder Representations from Transformers
GPT	Generative Pre-trained Transformer
VS CODE	Visual Studio Code
IDE	Integrated Development Environment

Table 1: List of Abbreviation

Table of Contents

Declaration.....	3
Abstract.....	4
Keywords.....	4
List of Abbreviations	5
1 Introduction	9
1.1 Background and Literature	9
1.2 Research Gap	10
1.3 Research Problem	11
2 Objectives	13
2.1 Main Objective	13
2.2 Specific Objectives	13
3 Methodology.....	14
3.1 Requirement Gathering.....	14
3.2 Past Research Analysis	14
3.3 Development Methodology	15
3.4 Project Management Methodology.....	15
3.5 Feasibility Study	15
3.5.1 Technical Feasibility.....	15
3.5.2 Schedule Feasibility	16
3.5.3 Economic Feasibility.....	16
3.5.4 Knowledge on Tools.....	17

3.6	Data collection Knowledge.....	18
1.	Data collection: Obtain the government-approved resource books, past papers, and answer schemes for Advanced Level Biology.	18
2.	Data-Preprocessing: data cleaning.....	18
3.7	System Analysis.....	18
3.7.1	Software Solution Approach	18
3.8	Tools & Technology.....	21
3.9	Project Requirements.....	22
3.9.1	Functional Requirements.....	22
3.9.2	Non-Functional Requirements	22
3.10	Project Scope	23
3.10.1	Project In Scope.....	23
3.10.2	Project Out of Scope.....	23
3.11	Testing.....	24
3.12	Timeline	25
3.13	Risk Management Plan	26
3.14	Communication Management Plan.....	27
4	Commercialization.....	28
5	Budget.....	29
6	Summary.....	30
7	References	31

Table 1: List of Abbreviation.....	5
Table 2: Risk Management Plan.....	26
Table 3: Cost Management Plan.....	29

Figure 1: Competitive Analysis	11
Figure 2: Jira Project Management.....	17
Figure 3: System Diagram.....	20
Figure 4: Work Breakdown Structure.....	20
Figure 5: Timeline	25

1 INTRODUCTION

1.1 Background and Literature

Students must perform well in A/L Biology to fulfill their professional goals, as the subject prepares them for further study in the biological sciences or medical area. But there are a lot of challenges for students with the A/L Biology syllabuses, which are very detailed and call for thorough examination of different concepts. The feedback and self-evaluation parts that are essential for mastering the structure and essay-style questions are not usually provided by texts, lectures, and assignments utilized for methods of preparation. Though many A/L Biology students struggle with self-evaluation in general and particularly in the personal domain, the literature emphasizes the importance of frequent and tailored feedback as a means of enhancing learning.

There has been increasing evidence that technology-enhanced learning can close these gaps by using creative evaluation strategies as the usage of digital tools and online platforms in education and training keeps growing. However, the analysis of the literature reveals that further research is necessary to determine how these technological solutions can help A/L Biology students overcome their challenges with self-assessment and feedback. This is where the study fits in, focusing on the specific difficulties that students have and exploring Paraphrase the ways in which different technologies could improve students' exam performance and preparation.

1.2 Research Gap

While there are currently AI-based learning tools such as Phind AI and Claude AI that can generate suggestions and evaluate answers, they do not fully meet the needs of Advanced Level Biology students in Sri Lanka. These platforms do not offer feedback and evaluation based on the Sri Lankan A/L marking scheme and are not customized to meet the specific curriculum and standards of the Sri Lankan A/L Biology syllabus.

For students looking to enhance their understanding and exam performance, the lack of resources tailored to the Sri Lankan A/L Biology syllabus is a problem. Independently evaluating their answers and making sure they meet the requirements of the Sri Lankan A/L marking scheme is a task that many students find difficult.

To address this gap, our project, Bio Mentor, focuses on providing a solution that aligns with the Sri Lankan A/L Biology curriculum and marking criteria. Unlike existing tools, Bio Mentor provides an extensive evaluation and feedback system tailored to the specific needs of Sri Lankan English-medium biology students. By doing so, Bio Mentor aims to close the technological gap in education by offering a more efficient platform for generating answers to structured and essay-type questions, as well as evaluating answers in alignment with the A/L Biology requirements in Sri Lanka.













	For Srilankan A/L Bio syllabus	Answer based on the Srilankan A/L system	Answer Evaluation and suggestion
			
			
			

Figure 1: Competitive Analysis

1.3 Research Problem

When it comes to answering structured and essay-style questions, biology students frequently ability to grow into independent learning and self-assessment skills. Furthermore, due to financial limitations and other issues, not all Sri Lankan students have access to qualified biology teachers, particularly in the English-medium Advanced Level Biology stream. This lack of access makes the gaps in education greater and prevents students from getting the individualized feedback they depend primarily on their biology teachers for feedback and evaluation, which may limit their need to get better grades.

Apart from these difficulties, many students find it challenging to identify and address their personal weaknesses without professional support. The traditional approaches to evaluation and feedback frequently don't scale effectively, which leaves students without the necessary support to meet their goals in education. The challenge of achieving academically is further increased by the simple fact that, without support, it can be extremely challenging to locate proper answers to structured and essay-style questions.

There is a significant need to explore how technology-based tools can be effectively developed and utilized to support students in finding answers to structured and essay questions, as well as in independently improving and evaluating their answers. This strategy could greatly enhance learning results by developing a more knowledgeable and independent group of A/L Biology students.

2 OBJECTIVES

2.1 Main Objective

Provide a platform for English-medium biology students to find answers to structured and essay-type questions, evaluate their answer and offer suggestions and feedback to improve the answer.

2.2 Specific Objectives

The following are the sub-objectives of conducting this research.

- **Answer Generation for a Question:** The objective of the research is to create an integrated system that would be able to provide answers to the questions of both the structural and essay type questions.
- **Provide Suggestions for Improvement:** In this case, this research will specifically involve trying to look at the given answers so as to notice the improvements that need to be made.
- **Provide Feedback for Answers:** The research will also develop a set of guidelines for the provision of constructive feedback on the given answers.

3 METHODOLOGY

3.1 Requirement Gathering

Requirement Gathering involves engaging with both former Advanced Level Biology students and those currently studying to understand the challenges they face in preparing for A/L exams and term-wise Biology exams. We seek to discover the specific issues that students face through surveys and conversations, such as the complexity of self-evaluation without guidance from professionals, having no way to obtain individualized feedback, and the difficulty in accurately answering structured and essay-type questions.

3.2 Past Research Analysis

Past Research Analysis reveals that various studies have explored the application of large language models across different fields. Research has utilized models such as BERT, GPT-2, and GPT-4 to address diverse challenges in natural language processing. For instance, studies have demonstrated the effectiveness of BERT-based models in reading comprehension question answering [1] [2][4], as well as in developing conversational chatbots using BERT and platforms like Google Dialog flow [3].

During the past research analysis, the focus was to identify the methodologies, and the tools used to build all existing tools and platforms. Moreover, it helped to identify the problems that the past researchers faced.

3.3 Development Methodology

In the development of the Bio Mentor platform, we will utilize the Agile development methodology to ensure a dynamic and adaptive approach. Agile uses short development cycles called sprints to make steady progress, allowing for ongoing improvements based on supervisor and co-supervisor feedback and changing needs.

3.4 Project Management Methodology

For project management, we will use Jira to effectively plan, track, and manage the development of the Bio Mentor platform. Jira's capabilities will help us organize tasks, assign responsibilities, and monitor progress throughout the project lifecycle.

3.5 Feasibility Study

3.5.1 Technical Feasibility

The LLAMA 3 and LLAMA Index models will be utilized by the Bio Mentor project to enhance the platform's functionality. LLAMA 3 will be assessed for its efficiency in generating answers to essay-style and structured questions, as well as in evaluating uploaded answers and providing suggestions and feedback to improve responses for Advanced Level Biology. Additionally, the LLAMA Index's ability to quickly locate and index study materials. If the resources are text-based, the system will extract all the text, including tables, and convert it into embeddings using an embedding model. For image-based resources, the system will use a multimodal approach to extract details from each page.

3.5.2 Schedule Feasibility

The project schedule will be built up with phases for model setup, API development, testing, and deployment to allow for the integration of LLAMA 3 and LLAMA Index. Each step will have important benchmarks set to ensure that progress is monitored, and corrections are made as necessary. To fulfill deadlines, resource allocation will be closely monitored, with a focus on the effective completion of development and testing tasks.

3.5.3 Economic Feasibility

As a way of minimizing costs, the Bio Mentor project will first be built using the quantized LLAMA 3 and LLAMA Index models on Google Colab and other similar platforms. This method allows for early development and testing to be done without having to put a lot of money down. The project shall later shift to commercial services such as Google Colab Pro or Azure Virtual Machines in case of need for more powerful computational resources.

3.5.4 Knowledge on Tools

Every member should be aware of the project management, other supporting tools, and development tools to build the proposed system. The current Jira board regarding the project management plans.

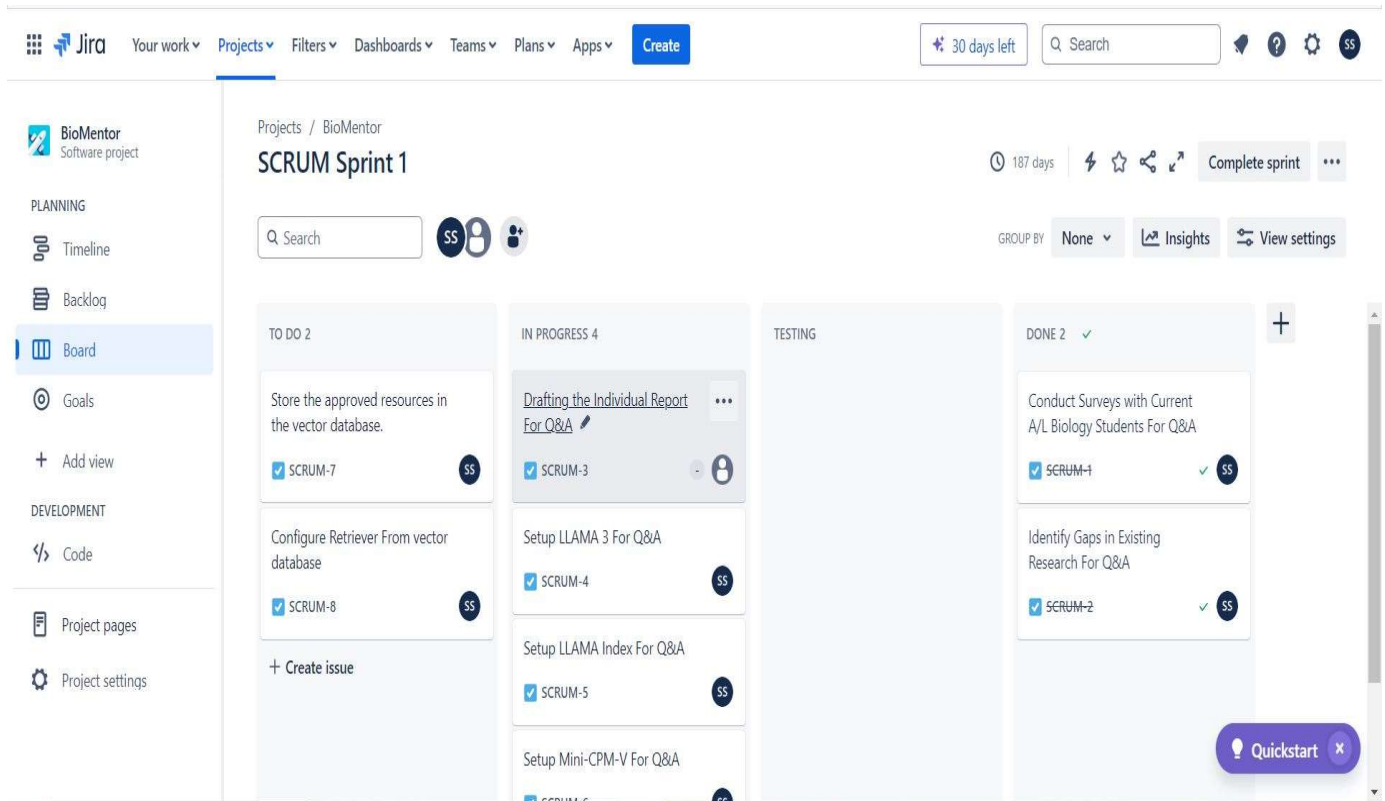


Figure 2: Jira Project Management

3.6 Data collection Knowledge

Data collection and preprocessing are essential tasks.

1. Data collection: Obtain the government-approved resource books, past papers, and answer schemes for Advanced Level Biology.
2. Data-Preprocessing: data cleaning

3.7 System Analysis

3.7.1 Software Solution Approach

The Bio Mentor platform will follow a structured approach to efficiently handle and process educational resources and user queries. Here's a detailed approach to the solution:

1. Resource Upload and Processing
 - a. Text-Based Resources: The system will extract and convert text, including tabular data, into embeddings for resources in text format, including those containing tables. It will do this by using an embedding model.
 - b. Image-Based Resources: For image-based resources, such as scanned PDFs, the platform will employ a multimodal approach. The extracted information will be processed and stored in the vector database.
2. Fine-Tuning the Model
 - a. Data Collection for Fine-Tuning: To refine the LLM, the model will be fine-tuned using a curated dataset of past exam papers, including both structured and essay-type questions.

- b. Training Process: The fine-tuning will involve adjusting the model's parameters to enhance its performance on biology-related questions.
- 3. Query Handling and Response Generation
 - a. Query Processing: When a user submits a question, the system will break it down into manageable chunks to understand the query context and specifics.
 - b. Retrieval from Vector Database: The parsed query will be used to retrieve relevant information from the vector database.
 - c. Response Generation: Using a fine-tuned LLM, specifically tailored with past papers and both structured and essay-type questions, the platform will generate comprehensive responses.
- 4. Feedback and Improvement
 - a. User-Submitted Answers: When users submit their answers to structured or essay-type questions, the system will evaluate the answer with generated answer. Feedback and suggestions for improvement will be provided using carefully crafted prompts.

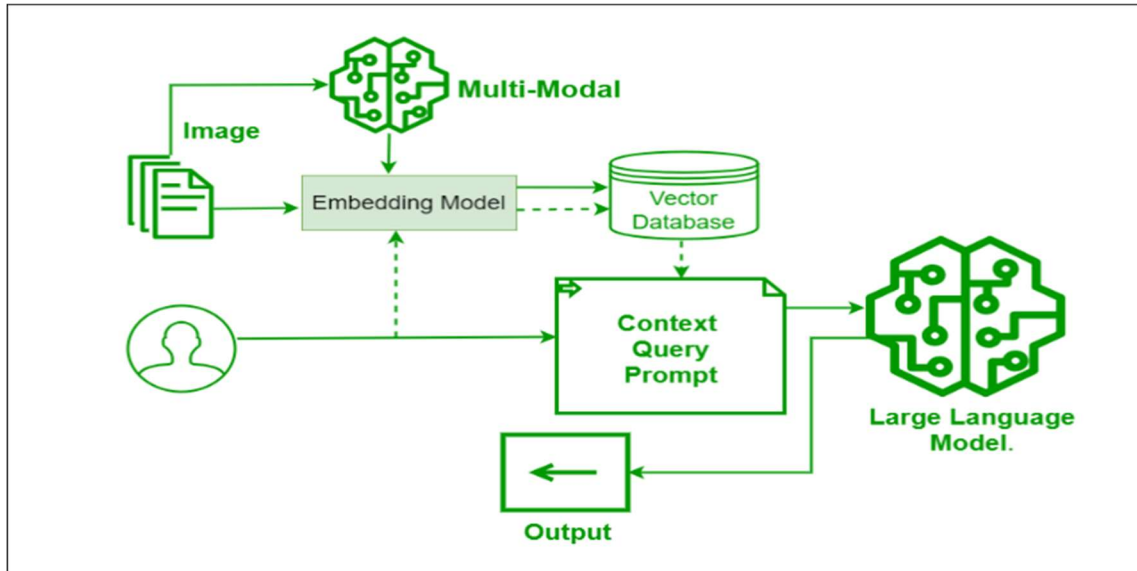


Figure 3: System Diagram

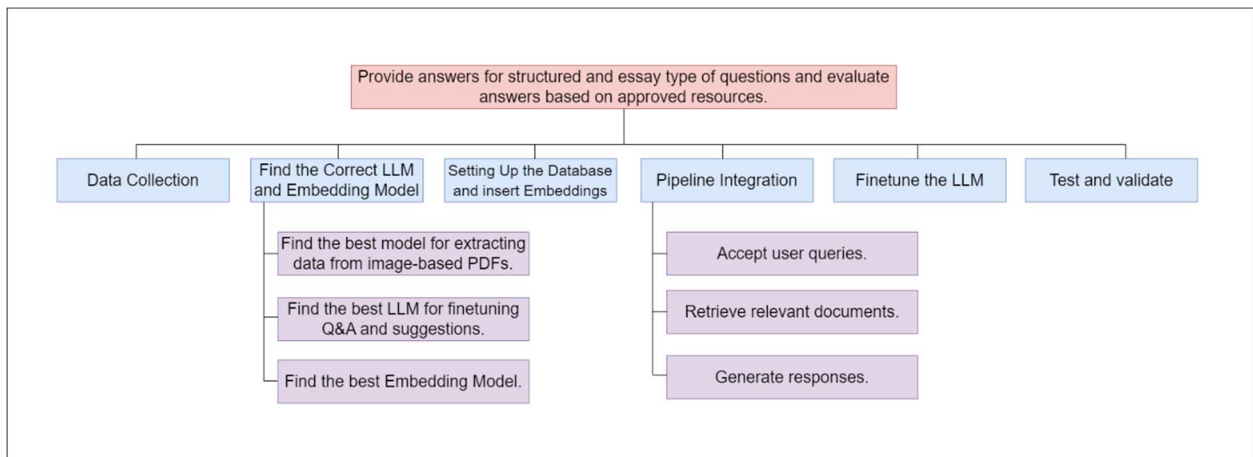


Figure 4: Work Breakdown Structure

3.8 Tools & Technology

- Programming Languages – Python
- Frameworks – LLAMA 3, LLAMA Index, Mini CPM-V, TensorFlow, OCR
- Database Systems – MongoDB, FAISS
- Web Development Frameworks – React JS (Front End Development), Python
- UI Design Implementation – Figma
- Diagramming Tool – Draw.io
- IDE – VS Code
- Version Control System– Git (GitHub)
- Collaboration Tools – Microsoft Teams or WhatsApp.
- Testing Tools
 - Jest - JavaScript Unit Testing
 - Postman – API Testing
 - Unit test – Unit Testing tool for Python
- Deployment Tools
 - Docker - Containerization
 - Kubernetes – Orchestration
- Project Management – Jira

3.9 Project Requirements

3.9.1 Functional Requirements

The functional requirements for the proposed model are as follows:

- Collect dataset from Advanced Level Biology past papers.
- Update the database with government-provided biology resources for Advanced Level.
- Evaluate answers and provide suggestions and feedback for improvement.
- Generate answers for structured and essay-type questions.

3.9.2 Non-Functional Requirements

The non-functional requirements for the proposed model are as follows:

- Usability
- Availability
- Accuracy
- Performance

3.10 Project Scope

3.10.1 Project In Scope

1. Answer generation for structured and Essay questions.
2. Feedback and Evaluation.
3. Maintain a database of government-approved biology resources.
4. User Interface development.
5. If the LLAMA 3 model is unable to achieve, alternative models will be considered for question answering, feedback, and suggestions.

3.10.2 Project Out of Scope

1. Real-Time tutoring.
2. Offline access
3. The accepted inputs are only text.

3.11 Testing

This research proposes a set of features for the e-learning platform for English-medium Biology students and testing is the final stage to determine the efficiency of the proposed solution. This phase will be conducted in several phases, each of which will focus on different aspects of the platform and will be used to check whether the intended educational goals and objectives will be achieved.

In the next phase, the tool will be piloted to a group of A/L biology students in local and deployed setting. This will include acceptance testing, to check how well the system will perform in the real world and to get feedback from users. Most of the testing will be done manually and the tool's internal functions and assertions will be used to test certain areas of the tool. This will guarantee that the tool serves the functional purpose, is reliable and provides accurate answers, evaluation, and suggestions for the question

3.12 Timeline

The proposed timeline for the project is as follows

Tasks	2024								2025			
	May	June	July	August	September	October	November	December	January	February	March	April
Feasibility study												
Evaluate feasibility and background study												
Requirement gathering												
Background survey												
Literature review												
Requirement analysis												
Software requirement specification												
Functional and non-functional requirement												
Proposal presentation												
Project proposal Report												
Software design												
Designing wireframes												
ML component development												
Front end development												
Software integration												
Deployment & maintenance												
Progress presentation 1												
Research Paper												
Progress presentation 2												
Testing												
Final presentaion and viva												

Figure 5: Timeline

3.13 Risk Management Plan

Identified Risk	Risk level	Probability for occurrence of risk	Mitigation plan
Technical Failures or Bugs	High	High	Seek assistance from experts or consult with the supervisor and co-supervisor for their guidance.
Potential scope changes for the project	High	Medium	Evaluate the complete scope and identify the scope area which can be covered within the project timeline with the help of the supervisor and co-supervisor.
Lack of Field Knowledge	High	Low	Build foundational knowledge using online resources
Loss of data because of hardware damage or system failure	High	Medium	Store backup in cloud
Panel Requests changes	High	Medium	Evaluate the changes and identify the area which can be covered within the project timeline with the help of the supervisor and co-supervisor.

Table 2: Risk Management Plan

3.14 Communication Management Plan

The goal of the communications management plan is to make sure that throughout the project, the supervisor, co-supervisor, and other team members have access to the information they need to do their jobs well. Strategic planning and good communication are essential for a project's successful completion.

- **Email:** For formal updates, documentation, and detailed communications.
- **WhatsApp:** For informal, quick exchanges and real-time updates and for quick discussions and urgent matters

4 COMMERCIALIZATION

The commercialization of the Bio Mentor platform will be achieved through a combination of advertising and subscription methods.

1. Advertising: The platform will display targeted advertisements within the UI. These ads may include banner ads, sponsored content, and promotional offers relevant to students.
2. Subscription Methods: The site will provide a subscription to make revenue.

5 BUDGET

As the result of the suggested model is a software-based solution, the implementation does not include any hardware. The membership fees to the cloud provider for the virtual machines' processing power will be the main source of expense.

However, additional expenses will be expected as indicated in the graph below.

Type	Cost
Internet use and web hosting	LKR.10,000.00
Training Cost	LKR.30,000.00
Publication Cost	LKR.70,000.00
Stationery	LKR.1,000.00
Total	LKR.110,000.00

Table 3: Cost Management Plan

6 SUMMARY

The Bio Mentor is developed to overcome the problems of the Sri Lankan Advanced Level Biology students especially those studying in English medium schools. Most of the students fail to develop independent learning and self-assessment skills because they mostly depend on their teachers to give feedback on their work.

To overcome these problems, Bio Mentor intends to offer answers for question types such as structured and essay-based questions, and also to give feedback and recommendations on answers provided by the students, according to the needs of the A/L Biology students in Sri Lanka.

By focusing on these objectives, Bio Mentor's objectives are to enhance the educational results of Sri Lankan A/L Biology students many times and equip them with the necessary tools in the context of the rapidly developing educational sphere.

7 REFERENCES

[1] BERT-Based Model for Reading Comprehension Question Answering. [ONLINE]

<https://ieeexplore.ieee.org/document/9972639> [Accessed 12 May 2024]

[2] Question Answering System using NLP and BERT. [ONLINE]

<https://ieeexplore.ieee.org/document/10551101> [Accessed 22 May 2024]

[3] Question Answering Model Based Conversational Chatbot using BERT Model and

Google Dialogflow. [ONLINE] <https://ieeexplore.ieee.org/document/9972639> [Accessed

03 June 2024]

[4] BERT-Based Mixed Question Answering Matching Model. [ONLINE]

<https://ieeexplore.ieee.org/document/9972639> [Accessed 05 June 2024]