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

Gokul Abisheak S
Department of Computer Science and Software Engineering
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
gokulabisheak12@gmail.com

Dharane S
Department of Computer Science and Software Engineering
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
dharanesegar08@gmail.com

Sujitha.S
Department of Computer Science and Software Engineering
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
sujithasrikanthan@gmail.com

Sajeevan S
Department of Computer Science and Software Engineering
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
sajeesiva12@gmail.com

K.T.S. Kasthuriarachchi
Department of Computer Science and Software Engineering
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
sanvitha.k@sliit.lk

Karthiga Rajendran
Department of Computer Science and Software Engineering
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
karthiga.r@sliit.lk

***Abstract*— The growing dependence on technology in education has facilitated the creation of customised e-learning solutions that address individual learning requirements. This study introduces BioMentor, a customised e-learning platform designed for Advanced Level (A/L) Biology students in Sri Lanka. The platform incorporates sophisticated artificial intelligence (AI) methodologies, such as Retrieval-Augmented Generation (RAG), transformer-based summarisation, adaptive quiz creation, automated answer assessment, and spaced repetition algorithms. The technology improves subject understanding, engagement, and retention via personalised learning pathways, hence enhancing overall academic achievement. Experimental findings demonstrate that BioMentor successfully addresses deficiencies in traditional education by adaptively modifying curriculum according to student performance and improving knowledge retention. The results underscore the capability of AI-driven educational platforms to transform domain-specific learning and provide scalable solutions for individualised education.**

Keywords—* *Personalized learning, Adaptive Learning, Retrieval-Augmented Generation (RAG), Transformer Models, Spaced Repetition, E-Learning, A/L Biology, Educational Technology

I. Introduction

In contemporary education, technology has emerged as an essential instrument for tailored and efficient learning. The Advanced Level (A/L) Biology curriculum presents distinct challenges for students owing to its comprehensive syllabus, sophisticated terminology, and complicated topics. In Sri Lanka, where English medium pupils adhere to a strictly regimented curriculum, conventional teaching and learning methods sometimes neglect specific student requirements. The dependence on rote memorisation, fixed content delivery, and insufficient adaptive educational technologies hinders learners from fully interacting with the material and realising their academic potential.

BioMentor: A tailored e-learning platform for advanced biology students in Sri Lanka, presented in English, serves as a transformative solution to overcome existing deficiencies. The platform offers a comprehensive, personalised learning environment through the integration of adaptive learning technologies. It utilises advanced approaches including spaced repetition algorithms, dynamic summarisation tools, adaptive quiz systems, and automated question-answering and evaluation processes to improve student performance. Each component targets distinct facets of the learning process—retention, understanding, engagement, and assessment—guaranteeing a comprehensive educational experience.

The platform incorporates optimised language models, Retrieval-Augmented Generation (RAG), and modular architectures to develop a scalable and adaptable solution. By customising education to the requirements of A/L Biology students, BioMentor not only reconciles traditional pedagogical approaches with contemporary educational needs but also establishes a benchmark for specialised e-learning platforms.

II. Literature Review

A. LLM-Based Abstractive Summarization with Voice Output implemented in various Software Architectures.

Abstractive text summarization has evolved with transformer-based architectures, overcoming the limitations of extractive methods that often lack coherence. Unlike extractive techniques, abstractive models generate paraphrased summaries, improving readability and contextual fluency. There are two primary types of summarization: long-document-based and topic-based. Long-document summarization processes large texts from formats like DOCX, PDF, and OCR-based scanned documents, extracting key ideas while maintaining coherence. Topic-based summarization, on the other hand, condenses information from various sources related to a specific subject, ensuring a broad yet concise overview. Transformer models like BART, T5, and Flan-T5 have significantly improved summarization quality. BART, a denoising autoencoder, reconstructs corrupted text for better fluency but is computationally intensive [1]. T5, trained on a vast dataset, is versatile across NLP tasks, while Flan-T5 enhances factual accuracy, making it ideal for educational summarization [2]. Retrieval-Augmented Generation (RAG) addresses long-document summarization by retrieving external context, improving factual consistency by up to 30% in specialized domains [3].

In NLP deployment, monolithic architectures outperform microservices in real-time applications like e-learning due to lower latency [5]. Additionally, text-to-speech (TTS) integration, improves accessibility and retention in multimodal learning [4]. Combining transformer-based summarization, RAG, monolithic deployment, and TTS provides an effective framework for optimizing e-learning platforms. Advancements in LLMs and transformer architectures will further enhance educational summarization, improving contextual accuracy, factual consistency, and accessibility.

B. *LLM-Based Adaptive Quiz for Enhancing Biology MCQ Skills*

The automatic generation of multiple-choice questions (MCQs) has gained significant attention due to advancements in Natural Language Processing (NLP) and machine learning. Transformer-based models such as T5 and BERT have been widely used for Automatic Question Generation (AQG) because of their ability to generate semantically relevant questions from educational content with minimal human intervention [10][20]. Studies have shown that these models can produce high-quality MCQs that align with learning objectives, making them valuable tools for educational applications.

One of the primary challenges in AQG is ensuring that the generated questions vary in difficulty and are tailored to the learner's proficiency. Adaptive learning systems dynamically adjust quiz difficulty based on student performance, improving engagement and knowledge retention. Research has demonstrated that tracking student responses, including accuracy and response time, allows for better-targeted quizzes that address weak areas and enhance learning outcomes [11], [12].

Recent advancements have focused on domain-specific MCQ generation, particularly in subjects like Biology. Fine-tuning models like LLaMA on specialized datasets ensures that questions align with curriculum standards. Additionally, the incorporation of Retrieval-Augmented Generation (RAG) further enhances question relevance by retrieving contextually appropriate information [13], [14].

The growing research in AQG and adaptive learning highlights the potential of personalized quiz generation. By leveraging transformer-based models and dynamic difficulty adjustments, these systems improve student engagement and performance, forming the foundation for the proposed adaptive MCQ quiz platform for A/L Biology students.

C. *LLM-Based Answer Generation and Evaluation*

Fine-tuning transformer-based models has gained significant attention in Natural Language Processing (NLP) for educational applications. Large Language Models (LLMs), such as BERT and GPT-4, have been extensively utilized for automated question answering (QA) and content generation. Studies indicate that fine-tuning pre-trained models with domain-specific datasets significantly enhances their ability to provide accurate and context-aware responses [6]. Research has demonstrated that employing structured and essay-style training datasets can improve model performance in academic assessments [7].

Recent advancements in automated answer evaluation have explored deep learning and NLP-based methodologies. Transformer-based AI models, including BERT and GPT-3, exhibit strong capabilities in textual comprehension and evaluation [8]. However, these models often require extensive training data and may encounter challenges in handling domain-specific queries. To address these limitations, hybrid models integrating multiple evaluation techniques have been proposed to improve accuracy in educational assessments. These models incorporate structured and unstructured data analysis, yielding enhanced student response evaluation [9]. The integration of semantic similarity measures, lexical similarity techniques, and grammar verification mechanisms further strengthens the reliability of automated evaluation systems, contributing to more effective and insightful feedback in digital learning environments.

D. *Enhancing Vocabulary Memorization through Adaptive Spaced Repetition*

Spaced repetition is a learning method that optimizes review intervals to enhance long-term memory retention. The SM-2 algorithm, developed by Piotr Woźniak in 1987 for SuperMemo, is one of the most well-known implementations. It is based on Ebbinghaus's forgetting curve, which shows that memory retention declines exponentially without reinforcement. Spaced repetition mitigates this by strategically reviewing information at increasing intervals. Research by Cepeda et al. [15] confirms that distributed practice improves long-term retention more effectively than cramming, while Roediger and Butler [16] found that repeated testing with intervals enhances memory performance better than passive rereading.

SM-2 adjusts review intervals based on user performance, using an easiness factor (EF) to determine how often an item should be reviewed. This ensures difficult concepts are revisited more frequently, while easier ones have longer gaps. Pavlik and Anderson [17] demonstrated that SM-2 significantly improves efficiency in language learning and medical education. Digital platforms like SuperMemo and Anki implement SM-2, enabling adaptive study schedules. Kang [18] highlighted its role in modern educational technology, while Karpicke and Bauernschmidt [19] found that active retrieval combined with spaced repetition enhances memory better than passive review.

III. METHODOLOGY

A. *LLM-Based Abstractive Summarization with Voice Output implemented in various Software Architectures.*

The summarization component, built using the Flan-T5 Base model, was fine-tuned for structured, syllabus-aligned summaries of Sri Lankan A/L Biology content. Key steps included text extraction, preprocessing, model selection, fine-tuning, RAG framework integration, deployment, and evaluation using ROUGE scores, ensuring accuracy, coherence, and real-time performance optimization.

The summarization system supports topic-based and long-text summarization. Topic-based summarization extracts relevant information for specific queries, while long-text summarization handles large documents by segmenting them into chunks, summarizing each separately, and merging them coherently.

The training dataset, sourced from government-approved books and guides, ensured syllabus-aligned content for A/L students. A preprocessing pipeline, including text cleaning, segmentation, tokenization, normalization, and stop word removal, enhanced model efficiency and summary quality. The Keywords column was removed as it was irrelevant to abstractive summarization.

Several transformer-based models were evaluated before selecting Flan-T5 Base. BART offered fluency but was computationally expensive for real-time use [1], while PEGASUS required extensive training data and resources [2]. Flan-T5 Base was chosen for its instruction-following capabilities, efficiency, and domain adaptability [3]. Compared to Flan-T5 Large and XXL, the Base version balanced performance and efficiency, making it ideal for deployment.

A Retrieval-Augmented Generation (RAG) framework was integrated to enhance the contextual relevance of generated summaries. RAG overcomes limitations of standard summarization models by retrieving additional educational content from a pre-indexed knowledge base before generating summaries [4]. A FAISS-based retrieval system was implemented, allowing relevant information to be retrieved before being passed to the Flan-T5 model for summary generation.

To evaluate the effectiveness of the summarization model, ROUGE (Recall-Oriented Understudy for Gisting Evaluation) scores were used. ROUGE is a widely adopted metric in text summarization, measuring the similarity between generated summaries and reference summaries. It evaluates the quality of summaries by comparing word and phrase overlaps between generated and human-written summaries. Three key ROUGE metrics were used in the evaluation:

TABLE I. ROUGE METRICS

Metric	Description
ROUGE-1	Measures unigram overlap, assessing keyword retention.
ROUGE-2	Evaluates bigram overlap, analyzing phrase continuity and fluency.
ROUGE-L	Assesses longest common subsequence overlap, determining structural similarity.

The summarization component was deployed in both monolithic and microservices architectures. The monolithic architecture integrated the summarization model directly into the application, ensuring seamless communication between components without external API dependencies. This resulted in lower latency and simplified deployment. In contrast, the microservices architecture deployed the summarization module as a standalone service, communicating with the application via REST API requests. While this allowed for modular scalability, it introduced network latency and required additional infrastructure for API management and containerized deployment. Given the current system requirements, the monolithic approach was found to be more efficient for real-time summarization tasks [5].

Additionally, the summarization component included a text-to-speech (TTS) feature, using the gtts library to convert generated summaries into voice output, enhancing accessibility for auditory learners. The TTS engine ensured that this component provided both textual and auditory learning experiences, making educational content more inclusive [4].

B. *LLM-Based Adaptive Quiz for Enhancing Biology MCQ Skills*

This study presents an adaptive MCQ quiz generation platform for A/L Biology students, dynamically adjusting quiz difficulty based on student performance. The methodology includes dataset preparation, model fine-tuning, quiz generation, system architecture, and performance evaluation.

The dataset consists of MCQs from past A/L Biology exams and school-approved assessments, each containing five answer choices, a correct answer, and a difficulty label (easy, medium, or hard). Due to the absence of explicit topic labels, sentence embeddings were used to cluster questions based on contextual similarity. To ensure accuracy in difficulty classification, a subject matter expert manually validated the dataset. Previous studies on automatic question classification have employed similar strategies to enhance dataset reliability [10], [11].

For question generation, the LLaMA-2-7b-chat-hf model was selected for its efficiency in language processing tasks. The model was fine-tuned using Quantized Low-Rank Adaptation (QLoRA) to optimize computational efficiency while maintaining accuracy. Transformer-based architectures like T5 have demonstrated effectiveness in automated question generation, making them a suitable reference [10], [12]. Retrieval-Augmented Generation (RAG) was integrated to improve contextual accuracy by retrieving domain-specific information, ensuring that generated questions align with curriculum standards.

The quiz generation algorithm employs an adaptive strategy that initially presents a balanced mix of easy, medium, and hard questions. As students progress, difficulty levels adjust dynamically based on response accuracy, response time, and performance trends. This adaptive mechanism reinforces weaker concepts, promoting gradual skill enhancement. Prior research on intelligent tutoring systems highlights the effectiveness of such adaptive learning methodologies in optimizing student performance [12], [13].

The backend, implemented using Fast API, processes quiz generation, manages user data, and communicates with the fine-tuned model to generate contextually appropriate questions. A PostgreSQL database stores user profiles, quiz history, and performance analytics, enabling personalized quiz recommendations. The modular system design supports future enhancements while maintaining reliability. The backend dynamically adjusts question difficulty based on real-time student performance and ensures secure communication between the model and database. Performance metrics are logged to facilitate progress tracking and enable data-driven improvements.

To track student progress effectively, the platform provides real-time performance feedback at the end of each quiz session. Reports include an analysis of accuracy across difficulty levels, response time per question, and overall performance trends, helping students identify weak areas. The system uses this feedback to adjust future quiz difficulty, ensuring a personalized learning experience. Research in MCQ-based learning systems has shown that such feedback-driven strategies enhance engagement and knowledge retention [11], [13], [14].

System evaluation was conducted through structured user testing with A/L Biology students to assess the adaptive quiz mechanism. Metrics such as quiz score improvements, response time reduction, and user satisfaction were analyzed. User feedback guided refinements in the quiz algorithm and interface. Results indicate that the proposed platform enhances personalized learning, aligning with prior studies on automated MCQ generation and intelligent tutoring systems [12], [14].

C. LLM-Based Answer Generation and Evaluation

The proposed methodology enhances students' ability to answer structured and essay-type questions through two key systems: answer generation and evaluation. The answer generation system generates contextually relevant responses, while the evaluation system analyzes student answers for accuracy, relevance, and grammar, providing adaptive feedback for improvement.

The LLaMA 3 Instruct model was fine-tuned using Parameter-Efficient Fine-Tuning (PEFT) techniques, including Low-Rank Adaptation (LoRA) and adapters from Hugging Face's PEFT library, to reduce computational costs. This approach adapted a small subset of parameters while keeping most of the pre-trained model frozen, significantly reducing computational overhead. Additionally, quantized fine-tuning with the BitsAndBytes library enabled 4-bit Normal Float compression for efficient training on consumer GPUs. Supervised fine-tuning optimized structured and essay-type QA using cross-entropy loss.

Instruction-tuning was applied to align the fine-tuned LLaMA-3 8B Instruct model with educational Q&A requirements, ensuring effective generation of structured and essay-formatted responses in an academic setting.

Retrieval-Augmented Generation (RAG) was implemented to enhance contextual understanding. Semantic embeddings from the SentenceTransformer (multi-qa-mpnet-base-dot-v1) model enabled semantic mapping of queries. A FAISS index efficiently stored and retrieved the top-k relevant entries from structured Q&A and notes datasets, ensuring relevant context retrieval for user queries.

The fine-tuned LLaMA-3 model generates responses based on retrieved context, ensuring alignment with educational expectations. It is trained on structured and essay-based questions and deployed using Hugging Face's transformers pipeline for inference. Before response generation, FAISS retrieves concise Q&A pairs for structured questions, while essay-type questions retrieve detailed context from both Q&A and notes. Safe generation techniques, including logits clamping and softmax normalization, ensure numerical stability and prevent NaN/Infinity errors.

For structured answer generation, retrieved Q&A pairs construct concise 1-2 sentence responses with word limit constraints. Essay answer generation utilizes retrieved Q&A and notes for long-form responses with a minimum word count. Response diversity is controlled using temperature, top-p, and top-k sampling, while a repetition penalty prevents redundancy.

A structured multi-step approach is used to evaluate student answers. A model-generated answer serves as a benchmark, and student responses are assessed using SciBERT-based semantic similarity, TF-IDF cosine similarity, and Jaccard similarity. SciBERT computes semantic closeness, TF-IDF captures lexical overlaps, and Jaccard similarity quantifies keyword commonality. Language Tool ensures grammatical accuracy. A hybrid scoring model assigns predefined weights to similarity metrics and grammar assessment. Evaluated data is stored in MongoDB for performance analytics, trend identification, and personalized learning recommendations. Adaptive feedback highlights missing or extra keywords, grammar errors, and customized exercises to enhance learning. The final score (S) is computed as follows:

$$S_{final} = w_{scibert}S_{scibert} + w_{tfidf}S_{tfidf} + w_{jaccard}S_{jaccard} + w_{grammar}S_{grammar}$$

where w values represent the predefined weights for each metric.

Grammar checking ensures language accuracy, while adaptive feedback provides insights into missing/extra keywords, grammar errors, and personalized exercises. This approach integrates retrieval-augmented generation and answer evaluation techniques, enhancing student engagement and academic performance.

D. Enhancing Vocabulary Memorization through Adaptive Spaced Repetition

Adaptive Spaced Repetition (ASR) aimed at improving memory retention through the integration of cognitive concepts, and gamification. The solution utilises a FastAPI backend for effective API management and a React frontend for an engaging user experience. The amalgamation of these technologies guarantees seamless data transfer, rapid reaction times, and a user-friendly interface for learners.

The system's foundation is the SM-2 algorithm, an established spaced repetition method that adjusts review intervals based on user performance. The system modifies review schedules based on answer accuracy, retrieval duration, and review attempt frequency. The FastAPI backend analyses user performance data and adjusts scheduling in real-time to guarantee optimal review timeliness. Technology personalises learning strategies and optimises knowledge retention by continuously monitoring user performance and suitably spacing reviews.

$$EF` = EF + (0.1 - (5 - q) \times (0.08 + (5 - q) \times 0.02))$$

This equation updates the Ease Factor (EF), which determines how quickly the review intervals grow. The term (5−q) represents how difficult the recall was (where q is a rating from 0 to 5). If the recall was poor (q is low), the equation decreases EF more significantly, making future reviews happen sooner. If recall was easy (q is high), the change is minimal, allowing longer intervals. The constants 0.1, 0.08, and 0.02 fine-tune the adjustment, ensuring that difficult items are reviewed more frequently while easy ones are spaced out more.

The backend, constructed with FastAPI, oversees user interactions, retains learning progress, and organises review sessions. Essential elements comprise User Authentication, employing JWT-based mechanisms for secure access;

Progress Tracking, utilising a NoSQL database (MongoDB) to archive user progress, review intervals, and performance metrics; and an Adaptive Scheduling API, which ascertains optimal review timelines based on personalised performance data. The backend facilitates seamless transitions between study sessions, enabling users to resume from their previous point with tailored information.

React-powered frontend with interactive flashcards offers a dynamic learning experience. Flashcards feature annotated illustrations and pronunciation guides, catering to visual and auditory learners. Gamification elements like leaderboards and badges enhance engagement and motivation. A simple interface allows for easy progress tracking and challenge-based learning.

The system perpetually alters review schedules in real time through User Performance Analysis, modifying difficulty and frequency according to retrieval success. Progress Visualisation is facilitated via dashboards that offer insights into learning trends, strengths, and shortcomings, enabling students to track their progress over time.

Gamification features are effortlessly integrated to promote continuous learning. Streaks and Achievements incentivise users with badges for persistent practice, so encouraging beneficial study habits. Leaderboards establish competitive rankings, motivating students to remain engaged by contrasting their achievement with that of their classmates. Adaptive Challenges progressively escalate in complexity, maintaining an equilibrium between challenge and learning efficacy. These gamification components are meticulously crafted to offer both intrinsic and extrinsic motivation, ensuring sustained user engagement over time.

User engagement metrics and performance analysis will be employed to evaluate the effectiveness of the system in terms of learning outcomes and retention enhancements. Data will be gathered on session frequency, accuracy rates, and time allocated to various modules. Through the analysis of this data, the system can enhance its adaptive scheduling methods to deliver a more customised learning experience. Furthermore, user feedback will be collected to improve the system's usability and efficacy.

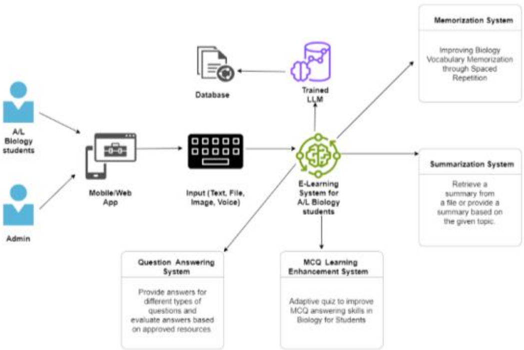

Fig. 1. Overall System Diagram

IV. Results and Analysis

A. *LLM-Based Abstractive Summarization with Voice Output implemented in various Software Architectures.*

The summarization component was implemented using monolithic and microservices architectures, evaluated on response time, resource usage, deployment efficiency, and debugging and troubleshooting. The monolithic approach integrated all functionalities within a single application, ensuring faster execution without inter-service communication. In contrast, the microservices architecture separated components into independent services, enhancing scalability and modularity. The table below compares both architectures based on key evaluation metrics.:

TABLE II. Comparison of Monolithic and Microservices Architectures

Feature	Monolithic Architecture	Microservices Architecture
Response Time	85% faster, no API overhead.	Slower, inter-service delays.
CPU and Memory Usage	Lower (~34% CPU, 28-36% RAM).	Higher (`43-62% CPU, 37-40% RAM).
Deployment Speed	47% faster (37.8 min).	Slower (71.5 min)
Debugging	Easier, centralized logs.	Harder, distributed logs.
Infrastructure	Simple, single container.	Complex, multiple containers.
Fault Tolerance	Lower, single failure risk.	Higher, independent services.

The performance of the summarization component was evaluated using ROUGE (Recall-Oriented Understudy for Gisting Evaluation) scores, a widely used metric for assessing summarization quality [3]. Unlike traditional accuracy-based evaluation metrics that require exact text matching, ROUGE allows for more flexible evaluation by measuring word and phrase overlap between generated summaries and reference summaries. The Flan-T5 Base model was tested on multiple biology-related educational documents, and the results are summarized in the following table:

TABLE III. ROUGE Score Evaluation Metrics

Metric	Score
ROUGE-1 (Unigram Overlap)	0.74
ROUGE-2 (Bigram Overlap)	0.40
ROUGE-L (Longest Common Subsequence)	0.54

The ROUGE scores indicate that the Flan-T5 Base model successfully retains critical information while ensuring summary coherence and readability. The high ROUGE-1 score (0.74) reflects the strong keyword retention capability of the model, while the ROUGE-L score (0.54) confirms that the model-generated summaries maintain good sentence structure alignment. The ROUGE-2 score (0.40), which evaluates bigram phrase continuity, suggests that the summarization model effectively preserves fluency while paraphrasing complex information [4].

B. *LLM-Based Adaptive Quiz for Enhancing Biology MCQ Skills*

The model was evaluated using a dataset of MCQ Questions and answers from the Sri Lankan Advanced Level Biology curriculum past papers. The training and evaluation pipeline incorporated parameter-efficient fine-tuning, retrieval-augmented generation (RAG) for contextual retrieval. However, assessment relied solely on domain experts, primarily Sri Lankan A/L Biology teachers, as no computational parameters exist for objectively evaluating response quality. Given the complexity and subjectivity of Biology answers, only human evaluation is feasible.

C. LLM-Based Answer Generation and Evaluation

The model was evaluated using a dataset of structured and essay-type questions from the Sri Lankan Advanced Level Biology curriculum. The training and evaluation pipeline incorporated parameter-efficient fine-tuning, retrieval-augmented generation (RAG) for contextual retrieval, and a hybrid scoring mechanism. However, assessment relied solely on domain experts, primarily Sri Lankan A/L Biology teachers, as no computational parameters exist for objectively evaluating response quality. Given the complexity and subjectivity of Biology answers, only human evaluation is feasible.

D. Enhancing Vocabulary Memorization through Adaptive Spaced Repetition

Adaptive Spaced Repetition (ASR) improves learning efficiency by dynamically modifying review intervals according to individual performance, hence optimising memory retention. The findings demonstrate that ASR markedly enhances recall rates relative to conventional fixed-interval repetition, as it emphasises challenging topics for more frequent review while extending the intervals for easy ones. Research indicates that learners utilising ASR have enhanced long-term retention, lower cognitive load, and superior learning outcomes across multiple fields, such as language acquisition and medical education. Nonetheless, its efficacy is contingent upon precise difficulty assessment and user involvement. Subsequent research should enhance adaptation algorithms to better individualise learning.

V. CONCLUSION

This study presents BioMentor, an AI-powered e-learning tool aimed at assisting Sri Lankan A/L Biology students by tackling issues inherent in conventional teaching. BioMentor personalises the learning experience by incorporating advanced AI techniques, like abstractive summarisation, adaptive quizzes, automated answer assessment, and spaced repetition, to improve engagement and retention. The system's use of transformer models and Retrieval-Augmented Generation (RAG) exhibits substantial enhancements in content relevance and contextual comprehension.

The trial findings confirm the platform's efficacy in customising educational resources to meet individual student requirements. The adaptive quiz system modifies difficulty in response to performance, while automatic answer evaluation guarantees precise assessments with tailored feedback. Moreover, the incorporation of spaced repetition enhances long-term retention, solidifying learning results.

Future study may investigate the enhancement of BioMentor's functionalities to encompass new subjects and the incorporation of real-time data to optimise adaptive learning methodologies. The research highlights the revolutionary effect of AI in education, facilitating more accessible and individualised digital learning settings.

REFERENCES

[1] A. Venkataramana et al., "Abstractive Text Summarization Using BART," IEEE MysuruCon, 2022.

[2] S. D. Sudharson et al., "An Abstractive Summarization and Conversation Bot Using T5," ICAICCIT, 2023.

[3] K. Maurya et al., "NLP-Enhanced Long Document Summarization," InCACCT, 2024.

[4] TalkifyPy Team, "TalkifyPy: The Pythonic Voice Assistant," ACET, 2024.

[5] P. Sun et al., "Monolithic vs. Microservices Architecture in NLP Applications," IEEE Transactions on Software Engineering, 2023.

[6] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.” *arXiv preprint arXiv:1810.04805.*

[7] Brown, T., Mann, B., Ryder, N., et al. (2020). “Language Models are Few-Shot Learners.” *NeurIPS 2020.*

[8] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." *arXiv preprint arXiv:1810.04805.*

[9] Dutta, H. S., & Das, B. (2020). "Hybrid AI Models for Educational Applications." *Journal of Artificial Intelligence Research, vol. 67, pp. 231-245*.

[10] A. S. K. Shukla, D. Arora, and A. K. Sharma, "Automatic Question Answer Generation Using T5 and NLP," IEEE ICCCA, 2019.

[11] P. Kumar, N. Agarwal, and R. Nath, "Generation of Multiple-Choice Questions From Textbook Contents of School-Level Subjects," IEEE ICCCA, 2019.

[12] S. Kumar and M. Gupta, "Automatic Question Generation for Intelligent Tutoring Systems," IEEE ICCCA, 2019.

[13] A. R. Patel, P. K. Jha, and S. Roy, "MCQGen: A Large Language Model-Driven MCQ Generator for Personalized Learning," IEEE ICCCA, 2019.

[14] R. Sharma and K. Singh, "Generation of Multiple-Choice Questions From Indian Educational Text," IEEE ICET, 2023.

[15] N. J. Cepeda, H. Pashler, E. Vul, J. T. Wixted, and D. Rohrer, "Distributed practice in verbal recall tasks: A review and quantitative synthesis," *Psychological Bulletin*, vol. 132, no. 3, pp. 354–380, 2006.

[16] H. L. Roediger and A. C. Butler, "The critical role of retrieval practice in long-term retention," *Trends in Cognitive Sciences*, vol. 15, no. 1, pp. 20–27, 2011.

[17] P. Pavlik and J. R. Anderson, "Using a model to compute the optimal schedule of practice," *Journal of Experimental Psychology: Applied*, vol. 14, no. 2, pp. 101–117, 2008.

[18] S. H. Kang, "Spaced repetition promotes efficient and effective learning: Policy implications for instruction," *Policy Insights from the Behavioral and Brain Sciences*, vol. 3, no. 1, pp. 12–19, 2016.

[19] J. D. Karpicke and A. Bauernschmidt, "Spaced retrieval: Absolute spacing enhances learning regardless of relative spacing," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 37, no. 5, pp. 1250–1257, 2

[20] Y. Tomikawa, A. Suzuki, and M. Uto, "Adaptive Question–Answer Generation With Difficulty Control Using Item Response Theory and Pretrained Transformer Models," *IEEE Transactions on Learning Technologies*