ADScAI
Applied Data Science & Artificial Intelligence Symposium

Developing a Conversational AI-Driven Physics Tutor with Multimodal Capabilities to Foster Critical Thinking Among G.C.E. Advanced Level Students in Sri Lanka

Helani U. Seekkubadu
Software Engineering Teaching Unit, Faculty of Science
University of Kelaniya
Kelaniya, Sri Lanka
udarahseekku@gmail.com

Isuru U. Hewapathirana
Software Engineering Teaching Unit, Faculty of Science
University of Kelaniya
Kelaniya, Sri Lanka
ihewapathirana@kln.ac.lk

Keywords—AI-driven Physics Tutoring, MLLMs, Critical Thinking in Physics Education, Personalized Problem-Solving Guidance, Educational AI for A/L Physics

I. Introduction

In Sri Lanka, G.C.E. A/L Physics students face various challenges, including limited access to quality educational resources and higher private tuition costs. Although many AI-driven educational tools have emerged as potential solutions, most do not explicitly focus on the local A/L Physics curriculum and often provide direct answers instead of step-by-step guidance, hindering conceptual understanding and critical thinking. To address this, Phy6Bot was designed and developed as a multimodal conversational AI assistant using existing MLLMs to deliver curriculum-aligned, interactive learning experiences while promoting students' critical thinking when solving Physics problems. Further, its accuracy and effectiveness were evaluated by demonstrating it to the relevant students and teachers and obtaining their feedback. Due to time and resource constraints, this study focuses solely on the Magnetic Field unit, serving as a proof-of-concept. Future iterations aim to expand across the full A/L Physics curriculum.

II. Literature Review

AI-driven educational tools enhance learning through personalized experiences, real-time feedback, and smart content delivery [1]. The introduction of Large Language Models (LLMs) has further transformed education by enabling virtual tutors like ChatGPT to assist students with assignments, answer questions, and provide instant support in natural language [2]. However, applying LLMs to complex subjects like Physics presents challenges, including reasoning inaccuracies, errors in extracting problem parameters, hallucinated facts [3], inconsistent responses [4], and difficulties processing Physics-related multimodal inputs like diagrams and graphs. While multimodal LLMs exist, such as GPT-4V, which can interpret diverse data types [5, 6], they are not specifically trained for Physics-related visuals or tailored to curricula like the Sri Lankan A/L syllabus. Additionally, many AI tutoring systems provide direct answers instead of guiding students through step-by-step reasoning, potentially weakening their conceptual understanding and critical thinking. To bridge these gaps, this study developed Phy6Bot, a conversational AI tutor aligned with the Sri Lankan A/L Physics curriculum. By integrating multimodal capabilities, Phy6Bot supports text and image inputs, offers structured guidance, and fosters critical thinking, ensuring a more effective and equitable Physics learning experience.

III. Materials and Methods

Phy6Bot's architecture consists of three key modules: Input Processing, Problem-Solving, and Guiding (Fig. 1). The Input Processing Module interprets user queries in text or image formats, using an MLLM to extract core questions and summarize visuals. Questions are then classified into units and subunits of the A/L curriculum for targeted guidance. After that, the question is forwarded to the Problem-Solving Module to determine the question type. If the question is theoretical, it is directed to the Theory Provider, which retrieves relevant Physics content from the A/L curriculum using the Retrieval-Augmented Generation (RAG) technique before passing it to the Guiding Module. Non-theoretical questions are processed through a structured sequence of components: extracting and evaluating given conditions, retrieving relevant theories and formulas through RAG, formulating stepwise solutions, and using the solver to integrate Wolfram Alpha for precise mathematical computations. The Summarizer then consolidates the output and forwards it to the Guiding Module. The Guiding Module enhances interactive learning through personalized feedback. It includes the Guider, which supports incremental problem-solving by providing hints, follow-up questions, and adaptive feedback; the Judge, which evaluates responses; and the Mistake Identifier, which detects errors and informs the Guider to ensure tailored guidance. Unlike ChatGPT, which directly provides answers, Phy6Bot encourages independent problem-solving through progressive assistance. Phy6Bot's design is influenced by existing AI-driven tutoring frameworks such as MathyAI's Module Classifier [7] and Lei et al.'s Multi-Agent System for Condition Mining (MACM) [8]. A prototype was demonstrated to A/L students and teachers, who provided feedback on usability and effectiveness. Evaluation used questionnaires with Likert-scale and open-ended questions, with responses analyzed descriptively and thematically.

IV. Results and Discussion

The student questionnaire included 16 Likert scale questions across five key evaluation categories. TABLE I

https://doi.org/10.31705/ADScAI.2025.15

presents overall mean values obtained for each category, all are above 3.5, reinforcing Phy6Bot's effectiveness in the Sri Lankan A/L curriculum. Teacher feedback was based on 15 Likert scale questions, mainly on effectiveness and accuracy. Results confirmed Phy6Bot's capability in problem identification, condition extraction, theory retrieval, and generating correct solution steps. Most aspects scored 4.0 or higher, indicating strong satisfaction with its problem-solving, guidance, mistake detection, and correction. Both students and teachers rated Phy6Bot user-friendly and engaging, with mean scores of 3.78 and 4.0, respectively. Its accuracy in identifying formulas and theoretical concepts was rated 4.3 (students) and 4.0 (teachers), while computational correctness received 3.9 and 4.0. Additionally, step-by-step guidance was rated 4.2 by students and 3.5 by teachers, reflecting its structured approach to problem-solving.

TABLE I. Summary Statistics of Student Responses

Evaluation Category	***Mean Likert Scale Value***
Guiding Effectiveness	3.84
Accuracy & Computational Performance	4.10
User-Friendliness & Interaction Quality	3.78
Educational Impact & Conceptual Understanding	3.90
Future Usage & Adoption	3.70

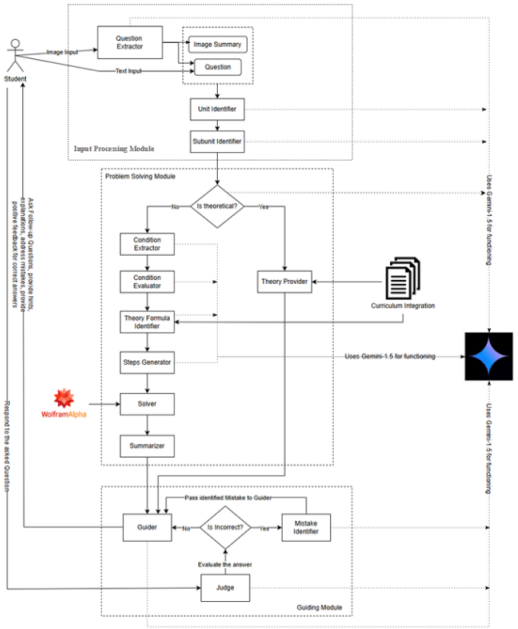

Fig 1- Phy6Bot's Architecture Diagram.

Thematic analysis highlighted key benefits, including increased confidence, clear step-by-step explanations, user-friendliness, and improved understanding of magnetic field concepts. One student noted, "***It gave explanations I never knew, even though I used those concepts for problem-solving. This is good for learning physics theories as it gives clear and detailed explanations***" emphasizing Phy6Bot's role in enhancing conceptual understanding. However, students identified limitations such as lack of explicit feedback on mistakes, repetitive questioning, and difficulties in inputting Physics symbols. Suggested improvements focused on refining feedback mechanisms, enhancing user experience, and adding more learning features. For instance, a student proposed, "***It would be good if Phy6Bot could guide MCQs and essay-type questions and generate similar questions in areas where I have made many mistakes. Also, a feature that allows us to upload a video related to physics practical, where it asks questions related to that practical***." These insights underscore the need to expand Phy6Bot's capabilities to support diverse question formats, improve interactive learning resources, and refine feedback mechanisms to create a more seamless user experience. Teachers praised Phy6Bot for promoting independent learning and critical thinking, with one stating, "***It doesn't spoon-feed answers, which is helpful in developing critical thinking skills***." This aligns with the goal of AI-driven educational tools in enhancing problem-solving abilities rather than merely providing solutions. However, both students and teachers noted that step-by-step guidance, while valuable, could be time-consuming during A/L exam preparation. Teachers also recognized its potential beyond A/L education, particularly in undergraduate studies and deeper conceptual exploration. Compared to tools like ChatGPT-4 and Gemini, Phy6Bot offers a distinct advantage by fostering critical thinking through guided reasoning rather than providing direct answers. Its integration of Wolfram Alpha further strengthens mathematical and algebraic accuracy in its problem-solving.

V. Conclusion

Phy6Bot demonstrates an AI-driven approach to enhancing A/L Physics education in Sri Lanka by fostering critical thinking and interactive problem-solving. Its modular design effectively processes and guides students through Physics problems, improving conceptual understanding and computational accuracy. Future work will expand the evaluation scope to cover the full A/L Physics syllabus, ensuring Phy6Bot's effectiveness across diverse topics. Enhancing multimodal reasoning with advanced models like GPT-4o, GPT-o1, Claude Sonnet 3.7, and 4, along with large-scale evaluation incorporating a broader sample and diverse question set, will further strengthen its reliability as an AI-powered educational tool. Additionally, integrating adaptive learning techniques will personalize tutoring based on individual learning patterns, improving engagement and retention. Beyond A/L Physics, Phy6Bot could also be extended to other educational levels, broadening its impact on AI-assisted learning.

References

[1] F. Degni, "AI in Education- Transforming Learning and Teaching Methods," ResearchGate, Jun. 28, 2024.

[2] OpenAI, "GPT-4 Technical Report," Mar. 2024, arXiv:2303.08774.

[3] T. Song, Y. Fan, C. Feng, K. Song, C. Liu, and D. Jiang, "Domain-specific ReAct for Physics-integrated iterative modeling: A case study of LLM agents for gas path analysis of gas turbines," 2024, arXiv:2406.07572

[4] M. N. Dahlkemper, S. Z. Lahme, and P. Klein, "How do Physics students evaluate artificial intelligence responses on comprehension questions? A study on the perceived scientific accuracy and linguistic quality of ChatGPT," Physical Review, vol. 19, no. 1, Jun. 2023.

[5] S. Wu, H. Fei, L. Qu, W. Ji, and T.-S. Chua, "NExT-GPT: Any-to-Any Multimodal LLM," Sep. 13, 2023, arXiv:2309.05519.

[6] S. Yin et al., "A Survey on Multimodal Large Language Models," arXiv.org, Jun. 23, 2023, arXiv:2306.13549.

[7] P. Zhao and A. Gonzalez, "MathyAI: An LLM-based Educational Virtual Assistant Application for Elementary Math Education CS 224V: Conversational Virtual Assistants with Deep Learning," 2022.

[8] B. Lei, Y. Zhang, S. Zuo, A. Payani, and C. Ding, "MACM: Utilizing a Multi-Agent System for Condition Mining in Solving Complex Mathematical Problems," 2024, arXiv:2404.04735.