

AI Educational Chatbot Based on Large Language Models (LLMs) for Science Learning Question and Answer Program in High School

Yeni Raini^{1*}  0009-0005-7256-7457

Japhet Ndayisenga²  0000-0002-4899-4177

¹*Educational Technology, Ibn Khaldun University Bogor, Indonesia*

²*University of Burundi, Nigeria*

ABSTRACT

Artificial Intelligence (AI) technology based on Large Language Models (LLMs) in education offers significant potential to overcome the limitations of conventional learning, thereby enhancing personalized and adaptive learning experiences for students. The purpose of this research is to develop and evaluate the feasibility of a Chatbot. LLMs-based AI education specifically designed to support the Q&A activities of Science learning (Physics, Chemistry, and Biology) at the Senior High School (SMA) level. This AI chatbot was developed as a learning aid that can provide accurate answers, contextual explanations, and relevant examples in real-time to student questions at SMA Madania Bogor. The type of research used is Research and Development (R&D). The development model used specifically for software is Waterfall with the following stages: Requirements analysis, system design, implementation, testing, deployment, and maintenance. Data collection techniques include observation, interviews, questionnaires, and documentation. The method for developing media in this study is by designing a Chatbot architecture that has previously been analyzed data, selecting and fine-tuning LLMs that are suitable for the Science domain, and implementing an intuitive user interface, then the media is tested for accuracy and feasibility by expert review experts in the field of IT and Design and Chatbots that have been installed on a server or platform with continuous monitoring from the developer for program maintenance. The results of this study found that the AI Chatbot device based on Large Language Models (LLMs) is effective in improving students' question and answer skills in science learning activities at SMA Madania Bogor.

ARTICLE INFO

Keywords:

Chatbot Education, Large Language Models (LLMs), Q&A, Science Learning, High School

Article History:

Received: 2026-01-13

Revised: 2026-03-03

Accepted: 2026-03-14

Published: 2026-03-31


How to Cite in APA Style:

Raini, Y., & Ndayisenga, J. (2026). AI Educational Chatbot Based on Large Language Models (LLMs) for Science Learning Question and Answer Program in High School. *Educational Researcher Journal*, 3(1), 10–21. <https://doi.org/10.71288/educationalresearcherjournal.v3i1.133>

This is an open access article under [CC-BY-NC 4.0](https://creativecommons.org/licenses/by-nc/4.0/) license.



* Corresponding author: Yeni Raini yeni_raini@yahoo.co.id

 <https://doi.org/10.71288/educationalresearcherjournal.v3i1.133>

Introduction

Amidst the rapid adoption of Artificial Intelligence (AI) technology in various aspects, particularly in science learning (Physics, Chemistry, and Biology) at the high school level, students often face significant challenges, particularly in literacy and numeracy. This is evident from the 2022 PISA Study data, which shows that most 15-year-old students in Indonesia have abilities below PISA standards, including in their ability to apply abstract scientific concepts, indicating fundamental difficulties in conceptual understanding. On the other hand, the world of education is facing an inevitable wave of AI adoption following the developments of the digital era 4.0.

A survey shows that 95% of Indonesian students have utilized Generative AI in their learning process, making it the highest adoption rate among the 15 countries surveyed, while 65% of educators expressed interest in using AI to improve student learning outcomes. This phenomenon demonstrates the high demand for AI-based tools. Facing these challenges, technology adoption has become inevitable. The technological revolution driven by Artificial Intelligence (AI) offers transformative opportunities, particularly through the rapid development of Large Language Models (LLMs). The era of the industrial revolution 4.0 has positioned AI technology as a key agent of change in various sectors, including education. In recent years, the emergence of Large Language Models (LLMs), such as the GPT family, has revolutionized the ability of machines to understand, process, and generate natural language text that is highly coherent, contextual, and resembles human responses.

This research focuses on leveraging the revolutionary capabilities of Large Language Models (LLMs) to provide an Educational Chatbot solution that can offer on-demand Q&A support, instant feedback, and personalized explanations for science materials at SMA Madania Bogor. This is expected to overcome personalization and literacy barriers in students, utilize students' interest in technology, and contribute to improving conceptual understanding of science at the high school level. The problem formulation in this research includes several things, namely: How to develop an LLMs-based AI Educational Chatbot from an IT and design perspective? How does the LLMs-based AI Educational Chatbot play a role in student learning activities at SMA Madania Bogor? Is there a change in the effectiveness of student learning outcomes after the implementation of the LLMs-based AI Educational Chatbot media?

To address the problems identified above, this study uses three main approaches, namely: 1) Developing media with a special Waterfall software model through the stages of designing Chatbot architecture, selecting and fine-tuning LLMs that are suitable for the Science domain, and implementing an intuitive user interface; 2) Testing the feasibility of science materials and media developed through expert review tests by IT experts and Science teachers; 3) Implementing LLMs-based AI Educational Chatbot in the MIPA class of SMA Madania Bogor to determine the effectiveness of student learning outcomes.

In the past 2–3 years, AI chatbots based on LLMs have transformed the landscape of automated dialogic tutoring from simply answering questions to adaptive learning systems that provide layered explanations, formative feedback, and automated assessments. Early empirical evidence suggests that LLM chatbots can improve learning outcomes when designed with pedagogical principles (scaffolding, insightful questioning, formative interventions) and combined with technical controls to reduce *hallucinations* and bias. Several previous studies have shown that the use of LLMs as intelligent tutoring systems has a significant impact on the understanding of complex science concepts. A review of the literature on science education confirms that LLMs have great potential to expand access to learning through question-and-answer features and adaptive tutoring. However, there are critical notes regarding the risk of student dependency and the potential for factual errors if LLMs are used without explicit pedagogical strategies. These findings also highlight student psychological variables, where students with low self-efficacy tend to rely more frequently on LLMs for assistance.

The novelty of this research lies in the LLMs Chatbot, which has evolved from a simple conversational system into an intelligent tutoring system that combines pedagogical strategies, adaptive personalization, and scaffolding for high school science learning. This system also overcomes the biggest challenges in the scientific validity of answers, pedagogical alignment, curriculum integration, ethics/privacy, and authentic assessment. This differs from other research with conventional systems whose factual answers often have the risk of over-simplification or misrepresentation, especially when there is no structured pedagogical strategy or the accuracy of checked sources. Other modern research is increasingly aware of the risk of LLMs providing overly simplistic or even incorrect information if not properly guided. The novelty of this research lies in the integration of Large Language Models (LLMs) with specific and adaptive educational methodologies for the context of science learning at the high school level. The novelty is mapped into three items, as follows:

1. Pedagogy and Learning Methodology
2. Implementation of LLMs and AI Techniques
3. Local Content and Context



Figure 1. Research Novelty Chart

The novelty is evident in the integration of Large Language Models (LLMs) with specific and adaptive educational methodologies for the context of science learning at the high school level. These novel aspects are mapped into three main points: innovation in terms of pedagogy and learning methodology, development of the implementation of artificial intelligence (AI) techniques and the LLMs models used, and adaptation to the content and local context of students. This integration strategically addresses the limitations of previous research, such as weaknesses in complex formulas and the risk of AI hallucinations, by shifting the Chatbot's function from merely an answering machine to a learning assistant capable of detecting misconceptions and facilitating students' scientific thinking processes comprehensively.

Method

This study was conducted at SMA Madania Bogor, located in Bogor Regency, West Java, Indonesia. The research was carried out from January to November 2025, covering the stages of development, implementation, and evaluation. The study involved students from class XI MIPA 2 during the 2025–2026 academic year.

This study employed a Research and Development approach to design and evaluate an AI-based educational chatbot. The development process followed the Waterfall model, which consists of six sequential stages: requirements analysis, system design, implementation, integration and testing, deployment, and maintenance. This model was selected due to its structured and systematic approach, making it suitable for developing educational software with clearly defined requirements.

Data were collected using multiple techniques, including observation, interviews, questionnaires, and documentation. These methods were used to gather information on user needs, system performance, and user responses to the developed chatbot.

The development process began with requirements analysis to identify user needs and system specifications. This was followed by system design, which included the development of chatbot architecture and interaction flow. In the implementation stage, the chatbot system was developed by selecting and fine-tuning appropriate Large Language Models (LLMs) for science learning. The system was then integrated and tested to ensure functionality and performance.

After testing, the chatbot was deployed for use in a real classroom setting. The final stage involved maintenance, which included continuous monitoring, evaluation, and improvement of the system based on user feedback and system performance.

To evaluate the feasibility and effectiveness of the developed chatbot, expert validation and user testing were conducted. Expert validation involved IT experts and science teachers, while user testing involved students who interacted directly with the chatbot in learning activities.

Results

The results of this study indicate that the developed AI educational chatbot based on Large Language Models (LLMs) functions effectively as a learning support tool in science education at the senior high school level. The chatbot was designed to assist students in question-and-answer activities related to science subjects, including Physics, Chemistry, and Biology.

Functionally, the chatbot is able to respond to students' questions using communicative, contextual, and cognitively appropriate language. It provides interactive features such as concept explanations, example questions, and clarification of misconceptions. The results of functional testing show that all main features operated according to the design specifications, with a success rate of 100% in the defined testing scenarios.

The feasibility of the chatbot was evaluated through expert validation involving science subject experts and IT experts. The evaluation covered aspects of content accuracy, conceptual clarity, language use, and media design. The results indicate that the chatbot falls into the "very feasible" category, with high scores in all assessed aspects. Experts noted that the use of LLMs enables the chatbot to generate more adaptive, varied, and in-depth responses compared to conventional rule-based chatbots.

User testing was conducted with high school students in a classroom setting. The results show that students were able to use the chatbot independently without intensive guidance. Students reported that the chatbot helped them understand difficult scientific concepts, particularly those requiring reasoning and conceptual explanation. Questionnaire results also indicate high user satisfaction, especially in terms of ease of use, response speed, and clarity of explanations. Additionally, students expressed greater confidence in asking questions through the chatbot, as it reduces feelings of hesitation or embarrassment commonly experienced in classroom interactions.

Table 1. Pre-test and Post-test Results

Indicator	Pre-test	Post-test
Lowest Score	40	70
Highest Score	75	95
Mean Score	58.4	82.6
Standard Deviation	8.7	6.9

To measure learning effectiveness, a pre-test and post-test design was implemented involving 30 students from class XI MIPA 2. The test consisted of 20 multiple-choice questions assessing conceptual understanding and scientific reasoning, with a maximum score of 100. The results are presented in Table 1.

The results show a substantial improvement in students' learning outcomes after using the chatbot. The average N-Gain score was 0.62, which falls into the medium-high category according to Hake's criteria. This indicates that the chatbot is effective in improving students' understanding of science concepts.

Statistical analysis using a paired sample t-test shows a significant difference between pre-test and post-test scores ($p < 0.05$), confirming that the observed improvement is statistically significant.

From a system performance perspective, the chatbot demonstrated stable performance in handling various types of questions, including factual and conceptual inquiries. However, limitations were observed in responding to highly specific questions or those requiring visual or experimental representation.

Additional findings indicate that students with lower initial scores (≤ 55) experienced the most significant improvement, suggesting that the chatbot functions effectively as a digital scaffolding tool. Furthermore, the chatbot supports independent learning by allowing students to access learning assistance anytime, particularly during evening study hours.

Discussion

The findings of this study provide strong evidence that AI-based educational chatbots powered by Large Language Models (LLMs) can significantly enhance students' conceptual understanding and engagement in science learning. The observed improvement in learning outcomes, as indicated by the medium-high N-Gain value and statistically significant test results, suggests that the chatbot effectively supports meaningful learning rather than mere memorization. This finding is consistent with previous studies indicating that LLM-based systems can function as intelligent tutoring tools that promote deeper conceptual understanding and adaptive learning (Hwang & Fu, 2025; Kasneci et al., 2023).

From a pedagogical perspective, the effectiveness of the chatbot can be attributed to its ability to implement scaffolding mechanisms within the learning process. The chatbot provides step-by-step explanations, contextual examples, and adaptive responses based on student queries, which aligns with constructivist learning theory and the concept of scaffolding (Vygotsky, 1978). In this sense, the chatbot functions as a digital learning assistant that supports knowledge construction through interaction rather than acting solely as an information provider. This supports the argument that AI-based systems can enhance active and student-centered learning environments (Hwang & Fu, 2025).

Moreover, the results indicate that the chatbot contributes to the development of learner autonomy and self-regulated learning. Students were able to access the chatbot independently and engage in learning activities without constant teacher supervision. The flexibility of accessing the chatbot at any time suggests that AI-based tools can extend learning beyond classroom boundaries and support continuous learning. This finding is in line with previous research highlighting the role of AI in fostering independent and personalized learning experiences (Kasneci et al., 2023; Labadze & Bokhua, 2024).

The findings also highlight important social and psychological dimensions of AI integration in education. Students reported increased confidence in asking questions through the chatbot, as it reduces the fear of making mistakes or being judged by peers. This suggests that AI-based learning environments can create more inclusive and supportive learning conditions. Similar findings have been reported in prior studies, which emphasize that AI chatbots can improve student engagement and reduce learning anxiety (Roviati et al., 2024).

From a technological perspective, the use of LLMs enables the chatbot to generate flexible, context-aware, and linguistically natural responses, distinguishing it from traditional rule-based systems. This adaptability allows the chatbot to handle diverse student queries and provide personalized learning support. However, the limitations identified in this study, particularly in responding to highly specific or visualization-dependent questions, are consistent with concerns raised in the literature regarding the limitations of LLMs, including the risk of inaccurate or misleading outputs (Kasneci et al., 2023; Pan et al., 2024).

An important finding of this study is the significant improvement observed among students with lower initial performance. This suggests that the chatbot serves as an effective digital scaffolding tool for students who require additional support. By providing accessible and adaptive explanations, the chatbot helps bridge learning gaps and supports differentiated instruction. This aligns with the principles of inclusive education and personalized learning supported by AI technologies (Hwang & Fu, 2025).

Despite these promising results, the implementation of AI-based chatbots in education also raises important considerations. The potential for over-reliance on AI systems may reduce students' critical thinking skills if not accompanied by appropriate pedagogical strategies. Additionally, the possibility of inaccurate or fabricated responses, often referred to as "hallucinations," highlights the need for continuous monitoring and validation of AI-generated content (Kasneci et al., 2023). Therefore, the role of teachers remains essential in guiding and contextualizing the use of AI in educational settings.

Overall, this study demonstrates that the integration of LLM-based AI chatbots in science education offers significant pedagogical and practical benefits, particularly in enhancing conceptual understanding, learner autonomy, and student engagement. However, the effectiveness of such systems depends on a balanced integration between technological capabilities, instructional design, and human oversight. Future implementations should adopt a hybrid approach that combines AI-driven learning support with teacher facilitation to maximize educational outcomes while minimizing potential risks.

Conclusion

This study concludes that the AI-based educational chatbot powered by Large Language Models (LLMs) is effective in improving students' understanding of science concepts at the senior high school level. The implementation of the chatbot enhances learning outcomes, supports independent learning, and facilitates more interactive and adaptive learning experiences. The findings also indicate that the chatbot functions not only as a question-and-answer tool but as a learning assistant that promotes conceptual understanding and student engagement.

From a practical perspective, the chatbot can be utilized as a complementary learning medium to support classroom instruction, particularly in assisting students who experience difficulties in understanding scientific concepts. It also provides opportunities for teachers to optimize instructional time by reducing repetitive explanations and focusing on higher-order learning activities.

However, this study is limited to a single group of students within one school context, which may affect the generalizability of the findings. Therefore, future research is recommended to implement this AI-based chatbot in different schools and subject areas to examine its broader applicability and effectiveness across diverse learning environments. Such studies are important to ensure that the benefits of AI integration in education can be generalized and adapted to various educational contexts.

References

- Adam, SH, & Ali, SU (2025). Utilizing artificial intelligence as a learning medium in the era of education 4.0. *Socius: Journal of Social Sciences Research*, 2(11), 751-755. <https://doi.org/10.5281/zenodo.15756165>
- Adiya, AZDN, Anggraeni, DL, & Albana, I. (2024). Comparative analysis of the use of software development methodologies (Waterfall, Prototype, Iterative, Spiral, Rapid Application Development (RAD)). *Mercury: Journal of Information Systems and Informatics Engineering Research*. 2 (4),122–134. <https://doi.org/10.61132/mercurius.v2i4.148>
- Ainun. (2021). *Waterfall Method: Definition, Advantages & Stages of the Waterfall Model* . Accessed from <https://salamadian.com/metode-waterfall/>
- Azizah, Aulia et al. (2024). Comparison of ChatGPT and Gemini Response Capabilities to Conceptual Questions About Optics . *Tadulako Journal of Physics Education* , 12 (1), 15-26. <https://doi.org/10.22487/jpft.v12i1.3510>
- Bassil, Y. (2012). A Simulation Model for the Waterfall Software Development Life Cycle. *International Journal of Engineering & Technology* . 2(5),1-7. <https://arxiv.org/pdf/1205.6904>
- Fadila, Farid & Raini Y. (2023). Development of Wordwall Educational Game-Based Evaluation Instruments in Mathematics Subjects in Class Iv Sdn Laladon 02 Bogor Regency, *Educational Researcher Journal* . 1(3), 1-13 <https://doi.org/10.71288/educationalresearcherjournal.v1i3.18>
- GoodStats. (2025). *95% of Indonesian Students Use AI in the Learning Process*. Retrieved from <https://data.goodstats.id/statistic/95-mahasiswa-ri-gunakan-ai-dalam-proses-pembelajaran-Flm7A>
- Hadi, MS, & Muttaqin, M. (2024). Utilization of Artificial Intelligence in Science Learning in the Digital Era. Jakarta: PT Raja Grafindo Persada.
- Hasebrook, J. P. (2025). Detecting Student Intent for Chat-Based Intelligent Tutoring Systems. In *Explainable AI for Intelligent Tutoring Systems* (pp. 71-87). Springer Nature.
- Hwang, G. J., & Fu, Q. K. (2025). A Comprehensive Review of AI-based Intelligent Tutoring Systems: From Traditional Methods to LLM-Powered Applications. *International Journal of Cardiology Cardiovascular Risk and Prevention*, 1-10. <https://doi.org/10.1016/j.compbimed.2025.110594>
- Kangiwa, H., et al. (2024). *Integration of AI Technology in STEM Learning in Secondary Schools*. *Scientific Journal of Mathematics, Earth and Space*. 4;2 (6) <https://doi.org/10.62383/bilangan.v2i6.320>
- Kasneci, E., et al. (2023). ChatGPT for good? On opportunities and challenges of Large Language Models for education. *Learning and Individual Differences* , 103.102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Labadze, K., & Bokhua, N. (2024). From intent-based Chatbots to LLM-powered teaching assistants. *Results in Engineering* . 22.101968. <https://doi.org/10.1016/j.nlp.2024.100101>
- Lewis, P., Perez, E., Piktus, et al. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks . *Advances in Neural Information Processing Systems* (NeurIPS).
- Noy, N. F., & McGuinness, D. L. (2001). *Ontology Development 101: A Guide to Creating Your First Knowledge Graph*. Stanford University.
- OECD. (2022) *Results Factsheets Indonesia*. OECD (Organization for Economic Co-Operation and Development) Publication. Accessed from https://www.oecd.org/en/publications/pisa-2022-results-volume-i-and-ii-country-notes_ed6fbcc5-en/indonesia_c2e1ae0e-en.html
- Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., & Wu, X. (2024). Unifying Large Language Models and Knowledge Graphs: A Roadmap. *IEEE Transactions on Knowledge and*

- Data Engineering.
- Pressman, R. S. (2015). *Software Engineering: A Practitioner's Approach* (8th ed.). McGraw-Hill Education
- Raini, Yeni, Suyitno, M, Paristiowati, M. (2024). The Development of Augmented Reality-Based Marker-Based Encyclopedia with Sets Approach (Science, Environment, Technology and Society) to Improve Digital Skills of Educational Technology Students, Proceedings Beyond Intelligence IFERP. <https://dx.doi.org/10.2139/ssrn.5147870>
- Raini, Yeni & Abidin Zainal. (2024). Marker-Based Module using the SETS Approach (Science, Environment, Technology and Society) Learning Media Subject Educational Technology Study. Proceedings of the 2nd Ibn Khaldun International Conference on Applied and Social Sciences (IICASS) ; https://doi.org/10.2991/978-2-38476-299-6_12
- Raini, Yeni et al. (2025). Educational Technology: Theory and Application . Padang: Azzia Karya Bersama.
- Raini, Yeni. (2024). Development of the EX2I Learning Model in Biomedical Courses Using the Steam Approach Based on Augmented Reality (Doctoral Dissertation, Jakarta State University).
- Raini, Yeni., Muslim, S., & Paristiowati, M. (2024). Development of Basic Biomedical Learning Model with STEAM (Science, Technology, Engineering, Art and Mathematics) Approach Based on Augmented Reality. *Journal Library of Progress-Library Science, Information Technology & Computer*. 44(3). <https://doi.org/10.48165/bapas.2024.44.2.1>
- Ramdoni, Lalu et al. (2025). Optimizing Student Information System Services with Telegram Integration: Chatbot Retrieval-Augmented-Generation Based on Large Language Model. *Journal of Information Technology, Computers and Applications (JTika)*. 7(1), 121-131. <https://doi.org/10.29303/jtika.v7i1.459>
- Richey, R. C., & Klein, J. D. (2007). *Design and Development Research: Methods, Strategies, and Visualization*. Routledge.
- Roviati, S., Priyanto, A., & Sumardi. (2024). Utilization of ChatGPT as a Learning Media for Biographical Texts in Indonesian Language Learning for Class X. *Al-Qalam Journal*. 25 (1), 456-463. <https://doi.org/10.64365/muradik.v1i4.121>
- Satria, Fitrah, Hersanto Fajri et al. (2023). UI/UX Design for Mobile-Based DKI Jakarta Regional Tax Census Application Using User-Centered Design Method . *Multidisciplinary Scientific Journal*, ;2(9),1-19. <https://jurnal.arkainstitute.co.id/index.php/nautical/article/view/618/1097>
- Sommerville, I. (2011). *Software Engineering* (9th ed.). Boston, MA: Pearson Education
- Sugiyono. (2019). *Educational Research Methods: Quantitative, Qualitative, and R&D Approaches*. Bandung: Alfabeta.
- Telkom University. (2025) *Waterfall Method in System Development: Definition, Stages, and Examples*. Accessed from <https://mif.telkomuniversity.ac.id/metode-waterfall-dalam-pengembangan-sistem-definisi-tahapan-dan-contohnya/>
- Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press.
- Widyasari, E., Murdiyasa, B., & Supriyanto, E. (2024). Educational Revolution with Artificial Intelligence: Opportunities and Challenges. *Alwatzikhoebillah Journal*. 10(2). <https://doi.org/10.37567/jie.v10i2.3405>