


Opinion Article

ChatGPT and DeepSeek Evaluate One Another for Science Education

Konstantinos T. Kotsis ^{1,*} 

¹ Lab of Physics Education and Teaching, Department of Primary Education, University of Ioannina, Greece

* Correspondence: kkotsis@uoi.gr

<https://doi.org/10.59652/jetm.v3i1.439>

Abstract: This paper compares ChatGPT and DeepSeek in science education, highlighting their various uses. ChatGPT's advanced language processing creates a conversational learning environment that encourages interactive dialogue and immediate feedback, making it ideal for science discussions. However, it struggles with complex, research-based tasks, suggesting it may not be enough for advanced science topics. In contrast, DeepSeek is designed for technical and scientific tasks and provides a robust framework for understanding real-world applications through multimodal textual and visual interactions. DeepSeek excels at helping students understand and remember science's difficult visual concepts. Science simulations and in-depth research benefit from DeepSeek's Vision-Language Model and code generation and debugging optimization. The paper concludes that while each platform has its strengths, combining ChatGPT's interactive capabilities with DeepSeek's research-oriented features can transform science education by accommodating diverse learning styles and improving teaching.

Keywords: Physics Education, ChatGPT, DeepSeek, Interactive Learning, Multimodal Interaction

1. Introduction

As the education system changes quickly, technology has shaped teachers' teaching, especially in difficult subjects like science. AI tools like ChatGPT and DeepSeek can be combined to improve students' learning and comprehension. These platforms use machine learning algorithms to provide personalized help, but they work differently and have different uses. ChatGPT processes natural language well, allowing people to interact like tutors. DeepSeek, on the other hand, uses structured problem-solving and can pinpoint science concepts. This essay compares and contrasts these two important tools, examining their efficacy, educational implications, and potential to change science education. This study will examine how each tool makes learning more engaging and informative and how these findings affect teaching.

AI has improved teaching and learning by personalizing experiences and using resources more efficiently. AI technologies like chatbots like ChatGPT allow students to interact with content dynamically, making them important in education (Bekeš & Galzina, 2023; Kotsis, 2024a). ChatGPT provides immediate feedback and a conversational learning environment, helping students understand and remember more. DeepSeek's Vision-Language Model adds advanced features for real-world education. Lu et al. (2024) states that DeepSeek-VL's focus on real-world data ensures that the model effectively addresses science education issues, making difficult visual concepts easier for students to understand. These AI tools may be useful, but they demonstrate the importance of constantly assessing their flaws, especially in Named Entity Recognition, which is still underdeveloped in most chatbots (Shafee et al., 2024). Know how AI works now to maximize its use in education. This paper employs a comparative analysis method to evaluate the effectiveness of two AI models, ChatGPT and DeepSeek, in the context of physics education. The study focuses on understanding how each model contributes to learning by examining their unique features and applications.

2. Features of the platforms

Received: February 3, 2025

Accepted: February 10, 2025

Published: February 15, 2025



Copyright: © 2022 by the authors.
Submitted for open access publication
under the terms and conditions of the
Creative Commons Attribution (CC BY)
license
(<https://creativecommons.org/licenses/by/4.0/>).

2.1. ChatGPT

ChatGPT is a powerful science teaching tool because it uses natural language processing to engage users. It can answer simple and complex science questions because this computer is designed for complicated conversations. Teachers can tailor their explanations and help students understand better with this adaptability (Shaw et al, 2023; Kotsis, 2024b). ChatGPT's ability to connect to multiple data sources lets you add the latest information, which is crucial in a changing field. Researchers found that ChatGPT is inferior to specialized models in language-based tasks like binary classification and Named Entity Recognition (NER). Further improvements are needed to fully replace application-specific tools like DeepSeek in educational settings (Shafee et al., 2024). ChatGPT's language skills and constant improvement make it a flexible science learning tool (Lu et al., 2024).

2.2. DeepSeek

DeepSeek is the only one that combines vision and language, especially for science teaching. The DeepSeek-VL model works with web content and high-resolution images to improve usability in educational settings where visual data is important (Lu et al., 2024). This in-depth coverage lets students interact with complex physical concepts differently, helping them understand them. While quickly analyzing visual inputs, DeepSeek's strong pretraining methods maintain its language processing skills (Jiang et al., 2024). It helps explain complex science concepts (Shafee et al., 2024). DeepSeek is a unique tool for teachers and students seeking effective science education because it combines advanced language and visual skills

3. Comparing the two platforms

Understanding science concepts is crucial in education, especially on ChatGPT and DeepSeek. Specialized science knowledge includes basic theories, complex applications, and advanced problem-solving methods. It is important to clearly explain complex information to ensure students understand and participate (Hasnawati et al., 2022). Because it can start conversations, ChatGPT can explain according to different learning styles. However, technical issues may not be fully covered. DeepSeek uses carefully selected databases and expert knowledge to help with difficult science topics (Howard & Gugger, 2020). By comparing these two platforms, teachers can learn how specialized information is shared and how it affects science learning.

The natural language processing skills of ChatGPT and DeepSeek differ in science teaching areas. ChatGPT interacts well with learners and simplifies complex ideas by writing in a way that makes sense and is relevant (Kucukaydin, 2019). Its F1 score of 0.94 in binary classification tasks shows that it can understand and process educational user queries (Shafee et al., 2024) and perform well in language-based applications. However, Xin et al. (2024) say DeepSeek uses a hybrid vision-language model emphasizing language skills and high-resolution visual data. DeepSeek's unique mix allows for multimodal learning, which improves science comprehension by combining textual and visual elements.

Science teaching aids, especially AI models like ChatGPT and DeepSeek, are effective if they engage students and explain complex concepts. ChatGPT excels at language because it provides immediate feedback and encourages conversational learning. This makes it good at answering student questions with specificity. However, Lu et al. (2024) describe DeepSeek as a strong framework for understanding real-world applications by fully exhibiting a wide range of visual tasks. In science, picturing ideas helps you understand them. Shafee et al. (2024) found that AI isn't great at some tasks, but both models can make learning more interactive. These technologies can make science classes more engaging and active, encouraging students to explore scientific concepts.

4. The platforms evaluate one another

The evaluation of ChatGPT and DeepSeek for physics education was conducted by asking both platforms the same question: "What is better, DeepSeek or ChatGPT, for Physics Education?" This approach allowed for a direct comparison of their responses.

The reply from ChatGPT was:

It depends on what you are looking for. Here is a comparison of DeepSeek and ChatGPT based on key factors relevant to science education and general AI capabilities:

1. Language Understanding & Explanation Quality

ChatGPT (GPT-4) tends to provide more detailed, structured, and pedagogically sound

explanations, making it better for teaching science concepts and breaking down complex topics.

DeepSeek-V2 is strong in language tasks but generally less refined in educational explanations than GPT-4.

2. Problem-Solving & Mathematical Ability

ChatGPT (GPT-4 Turbo) handles symbolic math, calculus, and step-by-step science problem-solving quite well. It also integrates Python for computations.

DeepSeek-V2 has good mathematical reasoning but may not be as robust or consistent as GPT-4 in science-specific applications.

3. Coding & Simulations

DeepSeek-Coder is specifically optimized for code generation and debugging. It's very useful for science simulations (e.g., writing Python code for numerical methods, differential equations, or visualization in VPython/Matplotlib).

ChatGPT (GPT-4 Turbo) is also strong in coding but better at explaining code to students. If you're teaching science coding, ChatGPT might be a better tutor.

4. Content Generation (Problems, Exams, Lessons)

ChatGPT is superior at generating well-structured educational content (like quizzes, lesson plans, and structured problem sets).

DeepSeek can also generate problems, but it might need more refinement and checking for quality.

5. Availability & Cost

DeepSeek models are open-weight, meaning they can be used freely or integrated into systems without a paywall.

ChatGPT-4 (Turbo) requires a subscription for full access, but it has a more polished user experience.

Final verdict for science education is the following:

If you need in-depth explanations and structured learning tools, ChatGPT (GPT-4) is better.

If you want open-source AI for coding simulations or automated feedback, DeepSeek-Coder might be useful.

For the same question, the reply from DeepSeek was:

When comparing DeepSeek and ChatGPT for science education and general AI capabilities, several key factors should be considered. Here's a breakdown based on relevance to science education and general AI performance:

1. Science Education Capabilities

DeepSeek:

- o DeepSeek is designed with a focus on scientific and technical domains, including science. It may have specialized training in science concepts, making it better suited for explaining complex topics like quantum mechanics, relativity, or thermodynamics.
- o It might provide more accurate and detailed explanations of science problems, including step-by-step solutions and derivations.
- o DeepSeek could be optimized for educational use cases, offering tailored responses for students and educators.

ChatGPT:

- o ChatGPT (especially GPT-4) has a broad knowledge base, including science, but its strength lies in general-purpose conversational abilities rather than domain-specific expertise.
- o It can explain science concepts effectively but may sometimes lack the depth or precision required for advanced topics.
- o ChatGPT is better for introductory or intermediate science education but may struggle with highly specialized or cutting-edge science topics.

Winner: DeepSeek (for specialized science education).

2. General AI Capabilities

DeepSeek:

- o DeepSeek may excel in technical and scientific tasks but could be less versatile in general-purpose applications like creative writing, casual conversation, or non-scientific problem-solving.
- o Its focus on science and related fields might limit its ability to handle diverse topics outside its specialization.

ChatGPT:

- o ChatGPT is a general-purpose AI with strong capabilities across a wide range of

tasks, including creative writing, coding, and casual conversation.

- o It performs well in interdisciplinary tasks and can adapt to various contexts, making it more versatile for users who need an all-in-one AI tool.

Winner: ChatGPT (for general-purpose use).

3. Accuracy and Depth of Responses

DeepSeek:

- o Likely to provide more accurate and in-depth responses for science-related queries due to its specialized training.

- o It may include more technical details, mathematical rigor, and references to scientific literature.

ChatGPT:

- o While generally accurate, ChatGPT may occasionally provide oversimplified or incomplete answers for advanced science topics.

- o It relies on its broad training data, which may not always include the latest or most specialized science research.

Winner: DeepSeek (for accuracy and depth in science).

4. Ease of Use and Accessibility

DeepSeek:

- o May have a steeper learning curve for non-technical users due to its focus on scientific and technical content.

- o Its interface and responses might be more tailored to educators, researchers, or advanced students.

ChatGPT:

- o Designed for ease of use, ChatGPT is accessible to a wide range of users, including beginners.

- o Its conversational style and adaptability make it user-friendly for both educational and casual purposes.

Winner: ChatGPT (for accessibility and ease of use).

5. Up-to-Date Knowledge

DeepSeek:

- o If DeepSeek is regularly updated with the latest science research, it could provide more current information.

- o However, this depends on the frequency of updates and the scope of its training data.

ChatGPT:

- o ChatGPT's knowledge cutoff (e.g., September 2021 for GPT-4) limits its ability to provide the most recent developments in science.

- o It may not include breakthroughs or discoveries made after its training period.

Winner: DeepSeek (if regularly updated).

6. Interactivity and Engagement

DeepSeek:

- o May focus more on delivering factual and technical content, which could be less engaging for students who prefer interactive or conversational learning.

ChatGPT:

- o Excels in interactive and engaging conversations, making it more suitable for keeping students motivated and interested in learning science.

- o Its ability to simulate dialogues and adapt to user preferences enhances the learning experience.

Winner: ChatGPT (for interactivity and engagement).

Summary:

(1) DeepSeek is better for advanced science education and technical accuracy, making it ideal for educators, researchers, and students who need in-depth explanations and specialized knowledge.

(2) ChatGPT is better for general-purpose use, accessibility, and interactive learning, making it suitable for a broader audience, including beginners and those who need versatility.

If your primary focus is science education, DeepSeek is likely the better choice. ChatGPT is the winner for a more versatile AI with strong general capabilities.

ChatGPT's response highlights its strengths in providing detailed, structured, and pedagogically sound explanations, which are beneficial for teaching physics concepts and simplifying complex topics. On the other hand, DeepSeek-V2 is recognized for its proficiency in language tasks but is generally considered less refined in delivering educational explanations

compared to GPT-4. ChatGPT, specifically the GPT-4 version, was noted for its ability to provide detailed, structured, and pedagogically sound explanations, which are advantageous for teaching physics concepts and simplifying complex topics. In contrast, DeepSeek-V2 was acknowledged for its strength in language tasks but was generally considered less refined in educational explanations than GPT-4.

5. Conclusions

In conclusion, teaching science with ChatGPT and DeepSeek shows that each platform has pros and cons. ChatGPT's conversational interface and large knowledge base make it easy for students to ask questions, making learning interesting and customizable. However, DeepSeek's structured approach and problem-solving focus teach students critical thinking and analysis to master difficult science concepts. These tools can help students succeed in school. ChatGPT engages students with interactive dialogue, while DeepSeek challenges them to learn through practice. Ultimately, combining the two platforms can help teachers adapt their science lessons to students' learning styles, improving the teaching process.

Science instruction differs greatly between ChatGPT and DeepSeek. Their differences show their strengths and uses. ChatGPT, an advanced language model, can write like a person and explain in a conversational manner. It is ideal for getting students to discuss science concepts. However, it may not be deep or specific enough to solve complex science class problems. DeepSeek, however, is for research. Its powerful search engine organizes academic papers, technical resources, and detailed user answers. DeepSeek can help students find scholarly content and conduct in-depth research. Finally, ChatGPT promotes interactive learning, and DeepSeek aids in-depth research. Both help science education in different ways

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

References

- Bekeš, E. R., & Galzina, V. (2023). Exploring the Pedagogical Use of AI-Powered Chatbots Educational Perceptions and Practices. In *2023 46th MIPRO ICT and Electronics Convention (MIPRO)* (pp. 636-641). IEEE. <https://doi.org/10.23919/mipro57284.2023.10159734>
- Hasnawati, H., Syazali, M., & Widodo, A. (2022). Analysis of Understanding Science Concepts for Prospective Elementary School Teacher Candidates. *Jurnal Penelitian Pendidikan IPA*, 8(6). <https://jppipa.unram.ac.id/index.php/jppipa/article/view/2438>
- Howard, J., & Gugger, S. (2020). *Deep Learning for Coders with fastai and PyTorch*. O'Reilly Media.
- Jiang, Z., Yang, Z., Chen, J., Du, Z., Wang, W., Xu, B., & Tang, J. (2024). VisScience: An Extensive Benchmark for Evaluating K12 Educational Multi-modal Scientific Reasoning. *arXiv*. <https://doi.org/10.48550/arXiv.2409.13730>
- Kotsis, K. T. (2024a). Integrating ChatGPT into the Inquiry-Based Science Curriculum for Primary Education. *European Journal of Education and Pedagogy*, 5(6), 28–34. <https://doi.org/10.24018/ejedu.2024.5.6.891>
- Kotsis, K. T. (2024b). ChatGPT in teaching physics hands-on experiments in primary school. *European Journal of Education Studies*, 11(10), 126-143. <http://dx.doi.org/10.46827/ejes.v11i10.5549>
- Kucukaydin, M. A. (2019). Concept teaching in science classrooms: A critical discourse analysis of teachers' talk. *Journal of Education in Science Environment and Health*, 5(2), 209-226. <https://doi.org/10.21891/JESEH.568813>
- Lu, H., Liu, W., Zhang, B., Wang, B., Dong, K., Liu, B., ... & Ruan, C. (2024). Deepseek-vl: Towards real-world vision-language understanding. *arXiv*. <https://doi.org/10.48550/arXiv.2403.05525>
- Shafee, S., Bessani, A., & Ferreira, P. M. (2024). Evaluation of LLM Chatbots for OSINT-based Cyberthreat Awareness. *arXiv*. <https://doi.org/10.48550/arXiv.2401.15127>
- Shaw, D., Morfeld, P., & Erren, T. C. (2023). The (mis)use of ChatGPT in science and education. *EMBO Reports*, 24, e57501. <https://doi.org/10.15252/embr.202357501>
- Xin, H., Guo, D., Shao, Z., Ren, Z., Zhu, Q., Liu, B., Ruan, C., Li, W., & Liang, X. (2024). DeepSeek-Prover: Advancing Theorem Proving in LLMs through Large-Scale Synthetic Data. *arXiv*. <https://doi.org/10.48550/arXiv.2405.14333>