ChatGPT and DeepSeek in the Physics Classroom: Affordances, Limitations, and Practical Applications

https://doi.org/10.69685/AQZM6859

Kotsis T. Konstantinos
Professor, Department of Primary Education, University of Ioannina, Greece
kkotsis@uoi.gr

Abstract

The rapid emergence of large language models (LLMs) is reshaping the landscape of physics education by offering new avenues for conceptual exploration and individualized support. This study examines the classroom affordances and limitations of two prominent LLMs—ChatGPT and DeepSeek—through a comparative analysis of their dialogic, representational, and reasoning capacities. Drawing on recent research in physics and science education, the article investigates how these tools can facilitate inquiry, scaffold problem solving, and enhance inclusive practices while also posing challenges related to bias, accuracy, and teacher authority. Particular attention is given to their complementary strengths: ChatGPT's conversational flexibility and accessibility, and DeepSeek's structured reasoning and technical precision. Pedagogical and ethical implications are discussed alongside recommendations for hybrid applications and teacher-led orchestration. The findings provide evidence-informed guidance for educators, curriculum designers, and policymakers seeking to integrate artificial intelligence into physics instruction without compromising critical thinking or equitable access.

Keywords: Artificial Intelligence, Physics Education, ChatGPT, DeepSeek, Educational Technology

Introduction

The incorporation of Artificial Intelligence (AI) in education has established a revolutionary environment where the distinctions between human cognition and machine assistance are perpetually redefined. In science and physics education, AI has evolved from a basic technological tool to a pedagogical entity that shapes the introduction, negotiation, and application of concepts in the classroom. Among the contemporary AI technologies, large language models (LLMs) like ChatGPT and DeepSeek illustrate divergent functionalities: ChatGPT engages in dialogic, open-ended exchanges, whereas DeepSeek provides structured, domain-specific assistance (Kotsis, 2025a). Comprehending these distinctions is crucial for assessing their use in the classroom, especially in physics education, where abstract theories frequently test students' ability to connect representations, experiments, and mathematical reasoning.

Historically, AI applications in education have focused on adaptive learning platforms, automated assessments, and intelligent tutoring systems (Luckin et al., 2016; Roll & Wylie, 2016). Nonetheless, the conversational capabilities of LLMs have created new opportunities for promoting student-centered interaction. In contrast to conventional tools, ChatGPT and DeepSeek emulate educational roles that surpass just material delivery, facilitating novel types of communication and support. ChatGPT enables students to articulate their thoughts through repeated questioning, thus facilitating conceptual inquiry. Conversely, DeepSeek prioritizes efficiency and systematic reasoning, generating solutions that are succinct and frequently customized for problem-solving endeavors. These complementary capabilities prompt an inquiry into the effective integration of such tools into classroom practices, ensuring that the function of educators is not undermined and that intricate discipline knowledge is not oversimplified.

Physics education, characterized by its dependence on models, experiments, and abstract conceptual frameworks, offers a particularly rich context for examining the role of LLMs. Studies have regularly shown that students find it challenging to align ordinary experiences with formal scientific ideas, particularly in mechanics, energy, and relativity (diSessa, 1993; Duit & Treagust, 2012). Educators have conventionally tackled these challenges by developing inquiry-based activities, employing analogies, or facilitating peer discussions. Recent research indicates that ChatGPT can be seamlessly incorporated into inquiry-based science curricula to enhance these pedagogical strategies (Kotsis, 2024a). AI tools now infiltrate this domain with the capacity to enhance these practices: they can facilitate inquiry, produce representations, and deliver instantaneous feedback. However, their presence raises inquiries regarding the overdependence on machine-generated explanations, the epistemic authority of artificial intelligence, and the extent to which these systems conform to the established objectives of promoting scientific literacy.

This research aims to analyze the classroom benefits and applications of ChatGPT and DeepSeek in physics teaching. This discussion places the two tools inside practical teaching contexts, rather than concentrating on theoretical frameworks and pedagogical ideologies, as explored in the first companion paper. It examines how educators and learners may utilize these AI systems for conceptual understanding, problem-solving, experimental reasoning, and reflective discourse. Furthermore, it assesses how the dialogic openness of ChatGPT and the structured precision of DeepSeek may either enhance or contradict conventional teaching methodologies. This approach is essential for educators, curriculum developers, and politicians who must address the ethical, epistemological, and pedagogical ramifications of integrating AI in educational institutions.

This study emphasizes that AI tools are not neutral technology; rather, they function as instructional agents that influence students' experiences in science by prioritizing classroom practice. Their effective incorporation into pedagogy relies on harmonizing their capabilities with discipline goals and the overarching aims of education. In physics education, where fostering inquiry, critical thinking, and epistemic agency is paramount, the comparative utilization of ChatGPT and DeepSeek may yield insights into how AI can both enhance and limit learning. This work contributes to the discourse on how AI might enhance not only efficiency in learning but also depth, creativity, and genuine participation with scientific processes.

Consequently, this article presents a focused empirical and conceptual analysis of how large language models can be pedagogically embedded in physics education. Through a comparative examination of ChatGPT and DeepSeek, this study explores how their distinctive capabilities can support conceptual understanding, experimental reasoning, and inclusive classroom practice, while also addressing associated ethical and methodological considerations. The study aims to inform researchers, educators, and policy makers seeking to integrate AI tools into science curricula in ways that preserve inquiry, critical thinking, and equitable access to high-quality physics learning.

Classroom Affordances of ChatGPT and DeepSeek

The practical utility of big language models in physics education depends on the extent to which their capabilities are integrated into routine classroom activities. ChatGPT's significant addition is its ability to maintain conversational interaction that resembles real-time instruction. In whole-class discussions, small-group activities, or individual exercises, the model can elicit student ideas, interrogate reasoning with subsequent inquiries, and produce counterexamples or analogies that reveal and disrupt misconceptions before they solidify into erroneous schemas (Ding et al., 2023; Graesser, Chipman, Haynes, & Olney, 2005). Due to their immediacy and adaptability, these interactions are ideally suited for formative assessment cycles, wherein educators assess comprehension and learners obtain contingent feedback that directs subsequent actions, especially when tackling enduring misconceptions in physics (Kotsis, 2025b). Empirical research on dialogic education indicates that such a response facilitates metacognitive monitoring and self-explanation—two mechanisms closely linked to lasting conceptual advancements in physics (Chi & Wylie, 2014).

In addition to communication, ChatGPT provides an adaptable representation. It can rephrase an explanation at an alternative reading level, convert specialized terminology into common English, and suggest other representational formats—verbal, symbolic, or qualitative descriptions—to align with learners' representational fluency (Ainsworth, 2006). In laboratory-related tasks, students can utilize the model to devise measurement procedures, foresee potential sources of mistake, or create prediction–explanation–observation prompts that equip them for practical work. Such AI-supported scaffolding is particularly valuable in low-cost experimental settings, where simple homemade apparatuses can be used to teach fundamental thermal concepts through qualitative demonstrations (Stylos & Kotsis, 2021). Addressing students' persistent misconceptions about heat and temperature across educational levels further underscores the importance of carefully structured experimental activities (Kotsis et al., 2023). These practices conform to inclusive design principles: varying explanation depth, providing alternative phrasings, and utilizing scaffolded question prompts can mitigate obstacles for multilingual learners and students requiring additional practice with disciplinary discourse (Eden, Adeleye, & Adeniyi, 2024; Nam & Bai, 2023). Teachers can regulate classroom discussions by providing specific role instructions to ChatGPT (for instance, “ask me three questions that assess my understanding of Newton's third law, one at a time”), transforming an open-ended generator into a structured formative routine that enhances teacher facilitation rather than undermining it (Roll & Wylie, 2016).

The advantages of DeepSeek are most apparent when training necessitates a rigorously structured and precisely articulated presentation. Physics subjects that entail significant intrinsic cognitive load—such as complex derivations in electromagnetism, boundary conditions in thermodynamics, or the relationship between graphical and analytical representations of motion—are enhanced by worked-example sequences that reduce extraneous cognitive demands while emphasizing problem schemas (Sweller, Ayres, & Kalyuga, 2011). With these scenarios, DeepSeek's methodical and clearly organized responses assist students with monitoring goal-subgoal relationships, maintaining symbol consistency, and linking vocal statements to formal expressions. Studies on intelligent tutoring systems demonstrate that organized coaching enhances procedural fluency and near-transfer performance, especially for novices developing problem schemas (VanLehn, 2011). DeepSeek prioritizes succinct, curriculum-aligned formulations, making it effective for producing practice sets with rubric-aligned solutions, generating exam-style feedback that identifies deviations from standard solution paths, and providing concise reference summaries that reinforce terminology during revision.

When educators combine both approaches, complementary advantages become actually applicable. A standard sequence commences with dialogic elicitation to externalize students' intuitive models and generate constructive cognitive conflict; it advances with focused, structured clarification that restructures concepts into disciplinary formats; and it concludes with succinct dialogic reflection to solidify new insights and foresee boundary conditions or exceptions. This alternation utilizes various learning mechanisms—engagement and sense-making on one side, and reduction of extraneous load and schema acquisition on the other—while maintaining the epistemic standards of physics as both inquiry and formal system. Teacher orchestration is crucial in transforming affordance into learning: lateral reading prompts and explicit AI literacy strategies assist students in assessing generated claims against reliable sources and empirical evidence, thereby reducing the risks of plausible yet erroneous outputs (Wineburg & McGrew, 2017; Zhao, 2023; Holmes et al., 2015). In inclusive classrooms, integrating ChatGPT's adaptable language with DeepSeek's consistent formulations can enhance involvement while maintaining conceptual alignment with curriculum objectives.

Recent evidence on the energy literacy of pre-service primary teachers highlights the ongoing need for such targeted preparation, particularly in science and environmental education (Stylos et al., 2023), underscoring the potential of AI tools to support both conceptual understanding and sustainable teaching practices.

Two advisory observations emerge from the affordance study. Initially, ChatGPT's dialogic flexibility, although productive, may meander without clear job limitations; educators ought to control prompts by delineating roles, procedures, or success metrics. Secondly, DeepSeek's proficiency in formal clarity may promote answer acquisition at the cost of comprehension if students are not consistently mandated to rationalize their procedures, examine boundary situations, or convert across representations. In both instances, the affordances are pedagogically effective only if classes are structured to utilize them for formative objectives, rather than to replace experimental practice, representational fluency, or instructor discretion.

Literature on Classroom Applications

The incorporation of AI systems like ChatGPT and DeepSeek into educational practices has produced an expanding corpus of studies examining their capacity to improve classroom instruction and learning outcomes. Recent studies emphasize that these technologies function as adaptable educational aides, offering prompt feedback, supporting inquiry, and replicating genuine scientific dialogue. AI dialogue systems have demonstrated the ability to enhance metacognitive involvement by prompting students to express their reasoning processes and confront misconceptions, which is fundamental to good science education (Holmes et al., 2022). Comparative analyses of AI-assisted and conventional instruction indicate that students gain advantages in both information acquisition and higher-order thinking skills (Ouyang et al., 2022; Kotsis, 2024a).

Teacher professional development has become a pivotal focus in this literature. The effective implementation of AI in educational settings relies on educators' trust and ability to integrate these tools into their teaching practices. Research emphasizes that teacher training should extend beyond mere technical proficiency with AI platforms, prioritizing the development of pedagogical content knowledge that integrates AI capabilities with curriculum objectives (Alam, 2023; Kotsis, 2024b). Evidence from Greek pre-service primary teachers underscores this need, revealing persistent gaps in self-efficacy for teaching physics that well-designed AI-supported training could help address (Stylos et al., 2022). Professional learning communities and specialized workshops have been recognized as useful methods to enhance teacher preparedness, especially when artificial intelligence is positioned as a co-instructor rather than a substitute for pedagogical authority (Zawacki-Richter et al., 2019).

In the scientific domain, AI-supported virtual laboratories and simulation environments signify a notable advancement. AI-enabled platforms enable students to design experiments, manipulate variables, and view simulated outcomes, which is particularly beneficial in schools with restricted access to physical laboratory resources. Recent research has demonstrated that ChatGPT can directly aid educators by producing organized experiment worksheets for classroom use (Kotsis, 2024b). Such environments have demonstrated the capacity to enhance conceptual comprehension and experimental proficiency, while simultaneously alleviating cognitive stress via adaptive scaffolding (Yu et al., 2022). DeepSeek's generative reasoning abilities can assist students methodically in hypothesis testing and error analysis, while ChatGPT can offer explanations and analogies that render abstract concepts more concrete. Recent work provides concrete examples of classroom-ready lesson plans that operationalize these capabilities, demonstrating how AI tools can be embedded in everyday physics instruction to support inquiry and problem solving (Vakarou et al., 2024).

The literature indicates the utilization of AI in examination preparation and formative evaluation. Research indicates that AI tutors can create practice problems, provide feedback on written answers, and adjust difficulty levels according to student performance, thereby facilitating personalized learning and fair access to preparation resources (Kasneci et al., 2023). Educators can utilize these tools to enhance test preparation and to create alternative assessment methods that evaluate creativity, reasoning, and problem-solving skills, thereby expanding the definition of achievement in science education.

Ultimately, inclusion has been a prevalent subject in conversations regarding classroom AI. AI-driven tools can provide tailored training for pupils with varying linguistic, cognitive, and socio-emotional requirements. The multilingual capabilities of big language models facilitate communication for immigrant or refugee students, while adaptive prompts support engagement for pupils with learning challenges (Smutny & Schreiberova, 2020). Furthermore, AI tutors can offer accessible explanations through many modalities, including text-to-speech and simplified summaries, ensuring that learners with visual or hearing impairments are included in collaborative classroom discussions.

The literature indicates that the educational potential of ChatGPT and DeepSeek resides not just in their technological sophistication but also in their capacity to facilitate inclusive, inquiry-driven, and adaptive learning environments. Effective application relies on meticulously crafted implementation strategies that synchronize technical capabilities with the principles of equitable and meaningful education.

Pedagogical and Ethical Implications

The use of large language models like ChatGPT and DeepSeek in science and physics teaching presents substantial pedagogical advantages and urgent ethical challenges. From an educational standpoint, these tools can democratize access to sophisticated problem-solving techniques and individualized explanations. Students who typically find complex physics ideas challenging can now engage with AI for customized feedback, iterative questioning, and scaffolding that adjusts to their cognitive abilities (Zawacki-Richter et al., 2019). This affordance may facilitate a transition to constructivist learning environments, wherein learners collaboratively generate knowledge with AI serving as a facilitator rather than only an information supplier.

This educational potential is intertwined with ethical dilemmas, especially with equity, racism, and academic honesty. Despite ChatGPT's notable conversational fluency, apprehensions persist about factual mistakes and "hallucinations" in its responses, potentially misleading students in scenarios like virtual laboratories or test preparation (Kasneci et al., 2023). DeepSeek, designed with an emphasis on open-source openness, presents a divergent approach that prioritizes collaborative knowledge creation, yet remains susceptible to the dissemination of systemic biases inherent in its training data. These concerns underscore the ethical obligation for educators to present AI not as an infallible authority but as an imperfect partner whose results require rigorous examination.

A further ethical aspect pertains to inequalities in access. Although both ChatGPT and DeepSeek can facilitate personalized tutoring, disparities in access to digital infrastructure may exacerbate educational inequalities between affluent and underprivileged schools (Holmes et al., 2022). This necessitates a meticulous equilibrium between fostering AI-driven innovation and preventing the exacerbation of digital disparities.

Furthermore, the issue of authorship and originality in student work presents a significant ethical dilemma. The simplicity of incorporating AI-generated material into tasks obscures the distinction between genuine student learning and algorithmic assistance (Cotton et al., 2023). Consequently, educators must not only govern AI utilization but also foster novel paradigms of academic integrity, wherein transparency regarding AI participation is included in evaluation systems. This reconfiguration possesses instructional significance: by prompting students to reveal and contemplate their AI usage, educators can convert potential threats into possibilities for the advancement of critical digital literacy.

The ethical framework of AI in educational settings intersects with professional identity and teacher autonomy. Educators may experience pressure to implement AI tools despite insufficient professional development or institutional directives. This contradiction necessitates astute pedagogical leadership that honors teachers' professional discretion while providing them with training to ethically and effectively handle AI (Lim et al., 2023).

The educational and ethical ramifications of ChatGPT and DeepSeek are interconnected. Their ability to improve individualized learning, diversity, and engagement must be evaluated against the risks of misinformation, injustice, and the diminishment of authorship. Consequently, the incorporation of AI in physics education must be approached from two perspectives: as a pedagogical opportunity for innovation and as an ethical obligation to ensure fairness, trust, and authenticity in educational settings.

Comparative Strengths and Limitations

The comparative comparison of ChatGPT and DeepSeek within the realm of physics education underscores both their potential benefits and limitations. Although both products provide significant advantages for facilitating learning, their design philosophies and fundamental functionalities diverge in ways that influence their success in the classroom.

The principal strength of ChatGPT resides in its conversational fluency and adaptability to various educational environments. Its capacity to maintain prolonged dialogue with students, address clarifying inquiries, and provide explanations at various levels of complexity renders it especially appropriate for physics classrooms, where conceptual comprehension frequently necessitates iterative discourse (Kasneci et al., 2023; Rudolph et al., 2023). Both educators and learners gain from its accessibility, as ChatGPT is extensively available across several platforms and incorporated into educational frameworks. Furthermore, its multilingual capabilities and user-friendliness establish it as an effective resource for international classrooms, particularly in environments where English is not the predominant medium of teaching (Holmes et al., 2022). Nonetheless, ChatGPT exhibits certain limitations. Its propensity to generate plausible yet scientifically erroneous statements—frequently referred to as “hallucinations”—constitutes a risk when used to teach physics content (Zhai, 2023). Moreover, the model's training architecture prioritizes general-purpose communication rather than specialized domain reasoning, thus constraining its ability to address complex or abstract physics problems (Dwivedi et al., 2023).

Conversely, DeepSeek seems to be refined for depth of thinking and technical precision. This distinction is particularly apparent in comparing analyses of how ChatGPT and DeepSeek address student mistakes in physics topics (Kotsis, 2025a). Preliminary findings indicate that it may thrive at structured problem-solving tasks, especially in mathematics and the computational facets of physics (Neha & Bhati, 2025). This establishes DeepSeek as a valuable resource for advanced learners necessitating intensive interaction with symbolic representations, mathematics, or sequential logical reasoning. Furthermore, its design philosophy prioritizes organized outputs, potentially diminishing uncertainty in intricate problem-solving scenarios. However, DeepSeek's relative novelty in the educational sector constrains the extent of empirical study about its application in classrooms, and its current implementation is relatively limited in comparison to ChatGPT (Wang & Kantarcioglu, 2025). These limits may hinder their implementation, especially in schools with inadequate technical infrastructure.

The strengths and weaknesses of both models indicate a complementary function in physics education, a viewpoint further highlighted by recent comparative studies of ChatGPT and DeepSeek in the realm of science education (Kotsis, 2025b). ChatGPT's linguistic proficiency and educational flexibility facilitate conceptual exploration and student involvement, but DeepSeek's analytical depth provides a more robust framework for complex problem-solving. Both exhibit inherent limitations: neither can entirely replace the pedagogical acumen of educators, and both are susceptible to reinforcing biases present in their training datasets (Bender et al., 2021). Their efficacy is contingent upon structured utilization, educator facilitation, and incorporation into meticulously crafted pedagogical approaches.

Discussion

The future of AI integration in physics education depends on the amalgamation of sophisticated language models with educationally sound design. ChatGPT and DeepSeek exemplify the nascent phase of AI-enhanced educational technologies, with their enduring efficacy reliant on methodical evaluation, collaborative design with educators, and continuous improvement. Although existing applications show potential in tutoring, assessment assistance, and conceptual clarification, empirical data from genuine classroom experiments are limited. In the absence of such validation, it is challenging to determine if the perceived affordances of these models can be reliably converted into quantifiable learning improvements (Zawacki-Richter et al., 2019; Holmes et al., 2019).

A primary research focus is the collaborative construction of AI systems alongside educators. Educators are particularly equipped to integrate AI capabilities with curriculum goals, support critical thinking, and ensure compatibility with student growth stages. Research in educational technology indicates that tools developed in partnership with practitioners are more likely to promote uptake and pedagogical significance (Luckin, 2017; Roschelle et al., 2020). In the realm of physics, co-design may encompass the incorporation of AI into laboratory modules, the creation of context-sensitive problem-solving activities, and the integration of AI into inquiry-based learning frameworks.

A critical focus for future research is the creation of hybrid models that integrate the advantages of big language models with specialized simulations. The conversational fluency of ChatGPT and the technical reasoning capabilities of DeepSeek could be synergistically combined with physics engines, interactive simulations, or adaptive learning platforms. Hybrid models may offer students both explanations and the chance to test theories in dynamic virtual environments (Kozma, 2003; Yu et al., 2022).

The ethical and governance aspects will influence future trajectories. As these tools are integrated into educational systems, the issues of transparency, accountability, and algorithmic bias necessitate meticulous supervision. Recent research indicates that forthcoming AI in education should incorporate inherent methods for explainability and accountability to preserve confidence within the classroom environment (Williamson & Eynon, 2020; Holmes & Porayska-Pomsta, 2023).

The long-term goal for AI in physics education necessitates reimagining its incorporation into the curriculum, not as an ancillary component, but as an essential resource for fostering scientific reasoning and digital literacy. This necessitates multi-national comparative research, ongoing longitudinal investigations, and policy frameworks that promote fair access. The prospects of the discipline depend on the technical sophistication of ChatGPT and DeepSeek, as well as their educational organization and institutional adoption (Holmes et al., 2023).

Conclusion

The use of artificial intelligence tools like ChatGPT and DeepSeek in physics teaching presents a significant opportunity as well as a formidable obstacle. This paper examines their classroom affordances, applicability in various learning environments, ethical and pedagogical implications, as well as their strengths and limits. A depiction of transformative potential arises, moderated by the necessity for critical reflection and continuous empirical assessment.

The accessibility, conversational fluency, and swift adoption of ChatGPT yield immediate advantages for physics educators and learners, especially in areas such as fostering inquiry, assisting multilingual students, and providing instant feedback on problem-solving methodologies. The democratization of access to advanced language models renders it a broadly accessible gateway to AI-enhanced learning. In contrast, DeepSeek's technical and reasoning-oriented features indicate potential for content-rich and sophisticated learning environments, especially in tackling intricate physics ideas that require methodical reasoning. Collectively, these tools underscore the potential for synergistic applications, with ChatGPT facilitating dialogic learning and DeepSeek augmenting analytical rigor (Kasneci et al., 2023).

However, the constraints are equally significant. Both models encounter difficulties with precision, fabrication, and contextual awareness. Their utilization must be facilitated by adequately qualified educators who are cognizant of AI's advantages and trained to identify its limitations, so as to guide students appropriately. Furthermore, ethical considerations like equitable access, cultural prejudice, and data privacy must be fundamental to any adoption plan. Failing to address these issues, AI may intensify existing disparities instead of mitigating them.

The future of AI in physics education will hinge on the collaboration between researchers and educators in co-designing classroom interventions, ensuring new tools enhance genuine scientific learning rather than supplanting human pedagogical discernment. Hybrid models integrating natural language interaction, multimodal functionalities, and physics-oriented reasoning engines may characterize the forthcoming generation of instructional AI. This vision necessitates comprehensive empirical research across varied educational systems and learner demographics, along with legislative frameworks that ensure ethical execution.

In conclusion, ChatGPT and DeepSeek should not be regarded as substitutes for the human aspects of teaching and learning, but rather as advancing tools that, when judiciously used, can enhance the physics classroom. Their potential extends beyond merely automating responses or producing explanations; it enhances the educator's capacity to foster inquiry, creativity, and critical thinking. A judicious adoption, guided by pedagogical theory, ethical considerations, and ongoing research, will be crucial for the responsible utilization of AI in physics education.

References

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16*(3), 183–198. https://doi.org/10.1016/j.learninstruc.2006.03.001

Alam, A. (2023). Harnessing the Power of AI to Create Intelligent Tutoring Systems for Enhanced Classroom Experience and Improved Learning Outcomes. In: Rajakumar, G., Du, K.L., Rocha, Á. (eds) Intelligent Communication Technologies and Virtual Mobile Networks. ICICV 2023. Lecture Notes on Data Engineering and Communications Technologies, vol 171. Springer, Singapore. https://doi.org/10.1007/978-981-99-1767-9_42

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 60*(5), 1–12. https://doi.org/10.1080/14703297.2023.2190148

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, *20*(1), 1-22. https://doi.org/10.1186/s41239-023-00392-8

Ding, L., Li, T., Jiang, S., & Gapud, A. (2023). Students' perceptions of using ChatGPT in a physics class as a virtual tutor. *International Journal of Educational Technology in Higher Education, 20*(1), 1–18. https://doi.org/10.1186/s41239-023-00434-1

diSessa, A. A. (1993). Toward an Epistemology of Physics. *Cognition and Instruction*, *10*(2–3), 105–225. https://doi.org/10.1080/07370008.1985.9649008

Duit, R.H., & Treagust, D.F. (2012). Conceptual Change: Still a Powerful Framework for Improving the Practice of Science Instruction. In: Tan, K., Kim, M. (eds*) Issues and Challenges in Science Education Research, Moving forward* (pp. 43-54). Springer, Dordrecht. https://doi.org/10.1007/978-94-007-3980-2_4

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., . . . Wright, R. (2023). Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Eden, C. A., Adeleye, O. O., & Adeniyi, I. S. (2024). A review of AI-driven pedagogical strategies for equitable access to science education. *Magna Scientia Advanced Research and Reviews, 10*(2), 044–054. https://doi.org/10.30574/msarr.2024.10.2.0043

Graesser, A. C., Chipman, B., Haynes, B. C., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education, 48*(4), 612–618. https://doi.org/10.1109/TE.2005.856149

Holmes, N. G., Wieman, C. E., & Bonn, D. A. (2015). Teaching critical thinking. *Proceedings of the National Academy of Sciences, 112*(36), 11199–11204. https://doi.org/10.1073/pnas.1505329112

Holmes, W., & Porayska-Pomsta, K. (2023). The ethics of artificial intelligence in education. London: Routledge.

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education promises and implications for teaching and learning*. Center for Curriculum Redesign.

Holmes, W., Persson, J., Chounta, I. A., Wasson, B., & Dimitrova, V. (2023). *Artificial intelligence and education: A critical view through the lens of learning science*. Council of Europe.

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., ... & Koedinger, K. R. (2022). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education*, *32*(3), 504–526. https://doi.org/10.1007/s40593-021-00239-1

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., . . . Kasneci, G. (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

Kotsis, K. T., Stylos, G., Houssou, P., & Kamaratos, M. (2023). *Students' Perceptions of the Heat and Temperature Concepts: A Comparative Study between Primary, Secondary, and University Levels. European Journal of Education and Pedagogy, 4(1),* 136–144. https://doi.org/10.24018/ejedu.2023.4.1.577

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 Develops Physics Experiment Worksheets for Primary Education Teachers. European Journal of Education Studies, 11(5). 1-20. http://dx.doi.org/10.46827/ejes.v11i5.5274

Kotsis, K. T. (2025a). Comparing ChatGPT and DeepSeek in Addressing Misconceptions about Physics Concepts. European Journal of Contemporary Education and E-Learning, 3(2), 191-206. https://doi.org/10.59324/ejceel.2025.3(2).17

Kotsis, K. T. (2025b). ChatGPT and DeepSeek Evaluate One Another for Science Education. EIKI Journal of Effective Teaching Methods, 3(1), 98-102. https://doi.org/10.59652/jetm.v3i1.439

Kozma, R. B. (2003). The material features of multiple representations and their cognitive and social affordances for science understanding. *Learning and Instruction, 13*(2), 205–226. https://doi.org/10.1016/S0959-4752(02)00021-X

Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *The International Journal of Management Education*, *21*(2), 100790. https://doi.org/10.1016/j.ijme.2023.100790

Luckin, R. (2017). Towards artificial intelligence-based assessment systems. *Nature Human Behaviour*, *1*(3), 1-3. https://doi.org/10.1038/s41562-016-0028

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence Unleashed: An Argument for AI in Education*. Pearson Education. https://discovery.ucl.ac.uk/id/eprint/1475756

Nam, B. H., & Bai, Q. (2023). ChatGPT and its ethical implications for STEM research and higher education: A media discourse analysis. *International Journal of STEM Education*, *10*(1), 1–24. https://doi.org/10.1186/s40594-023-00452-5

Neha, F., & Bhati, D. (2025). DeepSeek Models in STEM Education: Capabilities, Applications, and Challenges. *TechRxiv. Authorea Preprints*. https://doi.org/10.36227/techrxiv.174198447.71647707/v1

Ouyang, F., Zheng, L., & Jiao, P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. *Education and Information Technologies*, *27*(6), 7893–7925. https://doi.org/10.1007/s10639-022-10925-9

Roll, I., & Wylie, R. (2016). Evolution and Revolution in Artificial Intelligence in Education. *International Journal of Artificial Intelligence in Education, 26*(2), 582–599. https://doi.org/10.1007/s40593-016-0110-3

Roschelle, J., Martin, W., Ahn, J., & Schank, P. (2017). *Cyberlearning community report: The state of cyberlearning and the future of learning with technology*. SRI International. http://hdl.handle.net/20.500.12265/3

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT in education: Applications and challenges. *Education and Information Technologies, 28*(5), 7669–7687. Menlo Park, CA: SRI International. https://doi.org/10.1007/s10639-023-11610-8

Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the Facebook Messenger platform. *Computers and Education, 151*, 103862. https://doi.org/10.1016/j.compedu.2020.103862

Stylos, G., & Kotsis, K. T. (2021). *Use of a Simple Homemade Apparatus to Teach Basic Thermal Concepts – Six Qualitative Demonstrations/Experiments. The Physics Teacher, 59,* 477–479. https://doi.org/10.1119/10.0006134

Stylos, G., Kamtsios, S., & Kotsis, K. T. (2022). Assessment of Greek pre-service primary teachers' efficacy beliefs in physics teaching. Journal of Science Teacher Education, 34(1), 44–62. https://doi.org/10.1080/1046560X.2021.2023959

Stylos, G., Gavrilakis, C., Goulgouti, A., & Kotsis, K. T. (2023). *Investigating energy literacy of pre-service primary school teachers in Greece. Interdisciplinary Journal of Environmental and Science Education, 19(4),* e2318. https://doi.org/10.29333/ijese/13725

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer. https://doi.org/10.1007/978-1-4419-8126-4

Vakarou, G., Stylos, G., & Kotsis, K. T. (2024). AI for enhancing physics education: Practical tools and lesson plans. *The International Journal of Science, Mathematics and Technology Learning, 31(2),* 159–176. https://doi.org/10.18848/2327-7971/CGP/v31i02/159-176

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197–221. https://doi.org/10.1080/00461520.2011.611369

Wang, C., & Kantarcioglu, M. (2025). A Review of DeepSeek Models' Key Innovative Techniques. *ArXiv*. https://doi.org/10.48550/arXiv.2503.11486

Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology, 45*(3), 223–235. https://doi.org/10.1080/17439884.2020.1798995

Wineburg, S., & McGrew, S. (2017). Lateral reading and the nature of expertise: Reading less and learning more when evaluating digital information. *Teachers College Record, 119*(13), 1–40. https://doi.org/10.2139/ssrn.3048994

Yu, J., Denham, A. R., & Searight, E. (2022). A systematic review of augmented reality game-based Learning in STEM education. *Educational technology research and development*, *70*(4), 1169–1194. https://doi.org/10.1007/s11423-022-10122-y

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – Where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), 39. https://doi.org/10.1186/s41239-019-0171-0

Zhai, X. (2023). *ChatGPT user experience: Implications for education*. SSRN. 1–18. https://doi.org/10.2139/ssrn.4312418

Zhao, T. (2023). AI in Educational Technology. *Preprints*. https://doi.org/10.20944/preprints202311.0106.v1