



Contribution ID: 105

Type: Talk

ChatGPT vs. Physics Olympiad Participants: Insights for Rethinking Physics Problem Solving and Assessment

Friday 19 September 2025 11:30 (12 minutes)

Large language models (LLMs) are now widely accessible, reaching learners at all educational levels. Their rapid adoption has sparked concerns that students may use them in ways that bypass essential learning processes and undermine the integrity of established assessment formats. In physics education, where problem solving is at the heart of both teaching and assessment, these concerns are particularly pressing. To address them, it is important to understand how LLMs approach physics problems and what their capabilities and limitations mean for instruction and assessment.

In this talk, I will present findings from a study that compared the problem-solving performance of two advanced LLMs—GPT-4o and the reasoning-optimized o1-preview—with that of participants in the German Physics Olympiad. Using a set of well-defined Olympiad problems, we examined not only whether the models arrived at correct solutions but also how they reasoned through the problems, identifying characteristic strengths and weaknesses of LLM-generated solutions.

The results show that both models demonstrate advanced problem-solving capabilities, on average surpassing the performance of the human participants. Specifically, o1-preview outperformed both GPT-4o and the human benchmark. Prompting strategies seemed to have no to little effect on LLMs' performance. These findings highlight the rapidly evolving capabilities of LLMs and pose important challenges for physics education: How can assessments maintain their integrity when models can already outperform top students? And how can educators help learners engage critically and productively with these tools rather than simply relying on them?

I will conclude by discussing the implications of these findings for the design of summative and formative assessments in physics education and outline possible pathways for integrating LLMs into instruction in ways that support, rather than replace, meaningful learning.

Primary author: TSCHISGALE, Paul (IPN Kiel)

Presenter: TSCHISGALE, Paul (IPN Kiel)

Session Classification: Parallel

Track Classification: Other topic/subfield (please comment in the box below)