

Can AI Do Your Physics? What Large Language Models Can Do, How Students Use Them, and What It Means

Abstract

Generative AI increasingly appears to be part of how physics students work, particularly on the weekly problem sets they hand in. Recent surveys suggest this use is already widespread. This talk presents empirical results on what large language models can actually do in physics, and what those capabilities imply for teaching and assessment.

A first study evaluates current models, including GPT-4o and a reasoning model, on demanding Physics Olympiad problems. Performance is now strong enough that producing a correct solution can no longer be treated as reliable evidence of a student's understanding—an assumption that underpins much of how we grade homework, problem sheets, and unsupervised exams.

The same capabilities, however, can be turned toward supporting students as they learn, if used deliberately. A second study examines an LLM-based web application that generates feedback on physics problem-solving tasks. Here we assess not only how students perceive the feedback's usefulness and correctness, but also its actual correctness—two measures that need not agree.

Throughout, these findings—together with further current research—are used to draw out practical implications for the responsible use of AI in physics teaching, learning, and assessment, at both school and university level.