

The Impact of Language Models on Academic Evaluation: Rethinking Exam Design and Integrity

Friday, 20 March 2026 09:44 (22 minutes)

The rapid spread of large language models (LLM) in higher education has intensified discussions about their promise as instructional support tools and their risks as enablers of academic misconduct. Depending on how they are used, LLMs can assist instructors in developing more efficient learning and evaluation materials, as well as students to prepare for a test, or they can undermine assessment integrity when students or even educators rely on them uncritically.

This work investigates that tension through a quantitative study of two university courses - spanning software/computing for particle physics at a Master in Physics and applied machine learning at a Master in Bioinformatics, both at University of Bologna (Italy) - taught by the author over multiple years. A large archive of instructor-generated questions, routinely used to produce randomised multiple-choice exams, is compared against simulated exam sessions created by collecting answers produced by different LLMs to the same items. Contrasting human and synthetic performance across several academic years reveals distinctive statistical trends, offering insight into how closely LLMs mirror student behaviour, where they fail, and how their presence should influence future assessment design. The analysis further explores whether it is possible to construct LLM-resistant examinations, and how the models themselves may help shape more robust, learning-oriented evaluation strategies. The study ultimately underscores how generative AI is reshaping the landscape of academic responsibility.

As the name ChatGPT drifts phonetically between TeachGPT and CheatGPT, it neatly captures the tension explored in this work: the pedagogical promise and the integrity risks of language models are, quite literally, only a syllable apart.

Primary author: BONACORSI, Daniele (University of Bologna)

Presenter: BONACORSI, Daniele (University of Bologna)

Session Classification: Artificial Intelligence (AI) - IV

Track Classification: Track 10: Artificial Intelligence (AI)