# Comparative Analysis of Generative AI Performance in University Programming Courses

## About

This study evaluates how effectively five different Generative AI (GenAI) tools can complete exercises and quizzes in five university programming courses. The research objective was to measure and compare the performance of these AIs across different courses to understand their capabilities in an academic setting. The methodology involved using the GenAI tools to solve all weekly exercises, mimicking a student who relies solely on AI to pass.

## Problem

The rise of powerful Generative AI presents a significant challenge to university education, particularly in programming, as students can potentially use these tools to complete assignments without genuine learning. This creates a gap in verifying student knowledge and skills, forcing educators to reconsider how courses are designed and assessed. The study addresses the urgent need to understand the exact capabilities of these AIs to adapt educational practices and maintain academic integrity.

## Study Outcome

- Generative AI tools were able to solve 43-90% of programming assignments and 80-100% of quizzes across the tested courses.
- The results indicate that a student could pass the exercise portions of most of these courses by solely relying on readily available GenAI tools.
- GitHub Copilot consistently performed the best, while the free version of ChatGPT (GPT-3.5) and Codeium were often the worst performers.
- AIs struggled most with complex tasks, exercises that built upon previous solutions, and tasks requiring interaction with external files or APIs.
- The study concludes that programming courses must incorporate assessments in controlled environments (e.g., monitored exams) to accurately verify student learning and skills.

## Keywords

Generative AI • Programming Education • Artificial Intelligence • Academic Integrity • Cheating • University Courses • ChatGPT