



# Assessing Large Language Models for Automated Feedback Generation in Learning Programming Problem Solving

**Priscylla Silva**

Universidade de São Paulo & Instituto Federal de Alagoas

**Evandro Costa**

Universidade Federal de Alagoas





## Motivation

- Feedback is crucial for learning programming
- Individualized feedback



## Methodology

- Evaluated four LLMs: GPT-4o, GPT-4o-mini, GPT-4-Turbo, and Gemini-1.5-Pro
- Dataset: 45 real student solutions from introductory programming courses



## Key Findings

- 63% of feedback was correct, but 37% contained errors
- LLMs struggle with logical mistakes but perform well in syntax-related issues