Qeios

Peer Review

Review of: "An Empirical Study on Automatically Detecting AI-Generated Source Code: How Far Are We?"

Lamia Berriche[1]

1. Prince Sultan University, Saudi Arabia

The paper is technically sound and makes a valuable contribution. The authors tested multiple detectors across diverse datasets, programming languages (Python, Java, C++), and LLMs (ChatGPT, Gemini Pro, GPT-4, Starcoder2). However:

- The semantic similarity metric (cosine similarity) is reported, but the analysis lacks depth. It misses exploring whether structural vs. stylistic features are the main discriminators.
- Partially AI-assisted code is not evaluated.
- I recommend using techniques like SHAP or LIME to interpret which features (from code or AST) contribute most to classification.
- I recommend including these two references: 1- Aljumah, S. and Berriche, L., 2022. Bi-LSTM-based neural source code summarization. *Applied Sciences*, *12*(24), p.12587. 2- Berriche, L. and Larabi-Marie-Sainte, S., 2024. Unveiling ChatGPT text using writing style. Heliyon, 10(12).

Declarations

Potential competing interests: No potential competing interests to declare.

doi.org/10.32388/ZEA7J0