

Responsible NLP Checklist

Paper title: *TRACE: Evaluating Execution Efficiency of LLM-Based Code Translation*

Authors: *Zhihao Gong, Zeyu Sun, Dong HUANG, Qingyuan Liang, Jie M. Zhang, Dan Hao*

How to read the checklist symbols:

- the authors responded 'yes'
- the authors responded 'no'
- the authors indicated that the question does not apply to their work
- the authors did not respond to the checkbox question

For background on the checklist and guidance provided to the authors, see the [Responsible NLP Checklist](#) page at ACL Rolling Review.

A. Questions mandatory for all submissions.

- A1. Did you describe the limitations of your work?

This paper has a Limitations section.

- A2. Did you discuss any potential risks of your work?

See Section 6.1 and Limitations

B. Did you use or create scientific artifacts? (e.g. code, datasets, models)

- B4. Did you discuss the steps taken to check whether the data that was collected/used contains any information that names or uniquely identifies individual people or offensive content, and the steps taken to protect/anonymize it?

Our work focuses on code translation tasks using code data generated by large language models, with problems sourced from the GeeksforGeeks programming platform. The dataset contains only program code and problems, without any personally identifiable or offensive content.

- B6. Did you report relevant statistics like the number of examples, details of train/test/dev splits, etc. for the data that you used/created?

See Section 3.5

C. Did you run computational experiments?

- C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?

See Section 3.4

- C3. Did you report descriptive statistics about your results (e.g., error bars around results, summary statistics from sets of experiments), and is it transparent whether you are reporting the max, mean, etc. or just a single run?

See Section 5

D. Did you use human annotators (e.g., crowdworkers) or research with human subjects?

- D1. Did you report the full text of instructions given to participants, including e.g., screenshots, disclaimers of any risks to participants or annotators, etc.?

The taxonomy and human sanity check were conducted by two of the authors, who are very familiar with the work.

The [Responsible NLP Checklist](#) used at ACL Rolling Review is adopted from [NAACL 2022](#), with the addition of [ACL 2023](#) question on AI writing assistance and further refinements based on ARR practice. [ACL 2026](#) used a subset of ARR checklist form.

- D2. Did you report information about how you recruited (e.g., crowdsourcing platform, students) and paid participants, and discuss if such payment is adequate given the participants' demographic (e.g., country of residence)?

All experiments were conducted by the authors of the work. We did not recruit additional participants.

- D3. Did you discuss whether and how consent was obtained from people whose data you're using/curating (e.g., did your instructions explain how the data would be used)?

The original Transcoder-Test benchmark (created by META) has been widely adopted and refactored by a lot of previous work, including the work from Microsoft Research Asia (<https://arxiv.org/pdf/2501.18460>). As such, we believe the use of the benchmark in our work is reasonable for research purpose .

- D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
(left blank)

E. Did you use AI assistants (e.g., ChatGPT, Copilot) in your research, coding, or writing?

- E1. If you used AI assistants, did you include information about their use?

We only use AI assistants (e.g., GPT and Gemini) to help with some simple writing and coding tasks in this paper, such as grammar checking and figure plotting. We carefully reviewed the AI-assisted content and verified its correctness. As such, we believe that the use of these AI techniques does not pose a threat to the innovation or analysis of this work.