

Responsible NLP Checklist

Paper title: *AIRCoder: Adaptive Integration of Multi-dimensional Retrieval for Repository-level Code Completion*

Authors: *Chuanqi Shi, Miao Gao, Zhiqiang Gao*

How to read the checklist symbols:

- the authors responded 'yes'
- the authors responded 'no'
- ^{N/A} 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?

We discuss the potential risks and technical limitations in the following sections: - Section 5.4: We evaluate the risk of "hallucinations" (introducing non-existent entities) during LLM-based dependency extraction. - Limitations Section: We discuss potential risks regarding computational expense, scalability challenges in high-concurrency environments, and failures in capturing semantic dependencies when naming conventions are inconsistent.

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

- ^{N/A} 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 utilizes two widely used and publicly available repository-level code completion benchmarks: CrossCodeEval and RepoEval.

- 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?

Section 4.1.

C. Did you run computational experiments?

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

Section 4.1 and Appendix C.

- 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?

Section 5.2

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.

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.?

This study is purely computational and did not involve any human subjects or manual annotation.

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)?

This study did not involve any human 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)?

Our research utilizes established, publicly available benchmarks, specifically CrossCodeEval and RepoEval.

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

Our research utilizes established, publicly available benchmarks, specifically CrossCodeEval and RepoEval.

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?

In Research & Coding: We used Qwen3-8B as a core component of our AIRCoder framework for LLM-based dependency extraction. The prompts and reliability analysis are documented in Section 3.2 and Appendix A. In Writing: We used Gemini solely for grammatical polishing and improving the flow of the manuscript.