

Responsible NLP Checklist

Paper title: *RepoShapley: Shapley-Enhanced Context Filtering for Repository-Level Code Completion*

Authors: *Yu Huo, Kun Zeng, Siyu Zhang, YUQUAN LU, Cheng Yang, Yifu Guo, Xiaoying Tang*

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?

No. We do not include a dedicated discussion of broader societal or deployment risks. The paper studies repository-level code completion on public code repositories and focuses on retrieval control and completion quality rather than deployment to end users. We discuss technical limitations in the Limitations section.

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?

No. Our paper does not include a dedicated discussion of PII or offensive-content screening. The training data is derived from the permissively licensed source-code subset of The Stack, and we discuss licensing, opt-out provisions, and non-redistribution of raw source files in Appendix E.

- 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?

Yes. Section 4.1 reports the dataset scale and split statistics, including 290k Python repositories used for curation, 7.5k repositories used to construct 50k labeled instances, top-10 retrieved chunks, and a repository-level 95%/5% train/validation split.

C. Did you run computational experiments?

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

Yes. Section 4.1 reports the experimental setup and default training settings, including model families, 2 epochs, learning rate 210, 5% warm-up, max sequence length 4096, global batch size 512, and 8 NVIDIA H100 GPUs. Appendix B further describes hyperparameter optimization and the search space.

- 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,

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.

etc. or just a single run?

No. We report benchmark scores, ablations, and sensitivity analyses, but we do not report error bars, confidence intervals, or multi-run summary statistics such as mean and standard deviation, nor do we explicitly state results as averages over multiple random seeds.

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.?

(left blank)

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)?

(left blank)

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)?

(left blank)

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?

Used AI assistants for English proofreading and minor clarity edits.