

Responsible NLP Checklist

Paper title: *CoT-Edit: Reinforcement Learning of Chain-of-Thought Reasoning for Code Edit Suggestion*

Authors: *Wuya Chen, Yihao yang, Yue Lin*

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?

No. Our paper presents a technical methodology without discussing deployment implications or societal risks. The work is presented as a research contribution to improve code editing techniques, and we do not claim deployment readiness or provide deployment guidelines. Any deployment decisions would require separate safety and risk assessments beyond the scope of this technical paper.

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 training and evaluation data consists of internal code repositories that have undergone standard anonymization procedures as part of our organization's data handling practices. All personally identifying information (e.g., author names, email addresses, internal paths) was removed prior to use in this research. The data does not contain offensive content as it consists solely of source code from professional software

- 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 4.1.1 provides detailed statistics for our dataset construction, including the number of examples, data sources, and split configurations used for training and evaluation.

C. Did you run computational experiments?

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

Yes. Appendix B provides detailed experimental setup.

- 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?

Yes. Sections 4.1.1 and 4.1.2 report summary statistics for our experimental results.

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.?

No. Participants in our study were internal team members who received instructions through direct verbal and written communication rather than standardized instruction documents. The informal nature of the internal study did not require formalized instruction protocols or risk disclaimers. However, we ensured consistent communication of task requirements across all participants.

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)?

No. Our dataset consists of anonymized internal company data. Per company policy, de-identified production data can be used for research and model improvement purposes without individual consent. All data was properly anonymized to remove personally identifiable information.

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

No. Our study used anonymized internal company data and did not involve new data collection or human subjects research. Ethics review board approval was not required for this retrospective analysis of de-identified data.

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?

Yes. We used GitHub Copilot as a coding assistant to help implement parts of our experimental code.