

Responsible NLP Checklist

Paper title: *SAIR-Comb : A Structure-Aware Iterative Refinement Framework for Combinatorics Autoformalization*

Authors: *Weijie Jiang, Gaolei He, Beibei Xiong, Jianlin Wang, Zhengfeng Yang*

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?

This work is purely theoretical in nature, focusing on mathematical reasoning and formal language analysis. It does not involve human subjects, sensitive data, or applications with foreseeable societal risks. Therefore, a discussion of potential risks is not applicable to this study.

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?

The data used in this study consist of two parts: open-source datasets and internally annotated data. The open-source data are obtained from publicly accessible sources and consist solely of mathematical problems, containing no personally identifiable information or inappropriate content; thus, no additional anonymization is required. The internal annotation data were labeled exclusively by in-house experts, with no external participants involved, and underwent rigorous manual review to ensure the absence of any personally identifiable information or offensive material.

- 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?

Appendix B; Appendix C

C. Did you run computational experiments?

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

We reported the training hyperparameter in Appendix C, but did not detail the hyperparameter search process.

- 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

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.?

The data used in this study comprise two parts: open-source datasets and internally annotated data. The open-source datasets are obtained from publicly accessible sources and are used in strict compliance with their respective open-source licenses, with no violations or copyright infringements involved. The annotation work for the internal data was conducted exclusively by project team members, with no external participants recruited. All annotating members were fully informed and provided voluntary consent prior to their involvement. Therefore, no instruction texts or risk disclosures for external participants are provided in the paper.

- 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 recruit external participants; all annotation work was performed by internal project team members. Therefore, no external recruitment or related compensation fairness issues were reported.

- 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)?

All third-party data used in the study, including datasets and benchmarks, are open-source. Their licenses inherently grant permission for academic use, which fulfills the authorization requirement and makes separate consent discussion unnecessary for such data.

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

All data were obtained from publicly available opensource datasets or generated through annotation by internal project team members. The data content consists solely of mathematical formalreasoning tasks and does not constitute humansubject research requiring ethical oversight. Therefore, the study falls outside the scope of ethics review board approval.

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?

Section 4; Section 5; We also employ AI assistants to polish the paper.