

Responsible NLP Checklist

Paper title: *ALGOGEN: Tool-Generated Verifiable Traces for Reliable Algorithm Visualization*

Authors: *Liaokunpeng, Yuexiao Ma, Yisheng Lin, Hualin Zeng, Xiawu Zheng, Rongrong Ji*

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?

Section "Ethical Considerations". We discuss potential risks related to misuse of automatically generated educational content, including the possibility of producing low-quality or misleading visualizations at scale, and emphasize responsible use and validation.

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?

Section "Ethical Considerations" and Appendix E. We state that the dataset is constructed from public LeetCode problems and does not include personal or sensitive data. No personally identifying information is used, and the dataset is intended for research purposes only.

- 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 (Dataset Construction) and Figure 3. We report dataset size (200 tasks), algorithm families, and difficulty distribution (Easy/Medium/Hard), along with detailed statistics and taxonomy.

C. Did you run computational experiments?

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

Section 4 (Experimental Setup). We describe the experimental setup, including model selection (Qwen, DeepSeek, GLM), generation strategy (three-stage generation with repair), and evaluation pipeline. Implementation details and prompts are further provided in the appendix.

- 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 (Results) and Tables 13. We report average performance, success rates, and aggregated evaluation metrics across 200 tasks, including mean results and comparisons across models and methods.

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.?

Not applicable. The study does not involve human participants or annotators.

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)?

Not applicable. No human participants were recruited or paid.

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)?

Not applicable. No human subject research was conducted.

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

Not applicable. No human subject research was conducted.

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 "Ethical Considerations". We state that AI assistants (e.g., ChatGPT) were used for language polishing, and that all content was reviewed and verified by the authors.