

Responsible NLP Checklist

Paper title: *Does RLVR Extend Reasoning Boundaries? Investigating Capability Expansion in Vision-Language Models*

Authors: *Minghe Shen, Zhuo Zhi, Chonghan Liu, Shuo Xing, Zhengzhong Tu, Che Liu*

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?

This work focuses on benchmark-based evaluation of visuallanguage reasoning over static map representations and synthetic environments. It is not intended for real-world deployment or safety-critical applications. As such, we do not identify specific potential risks associated with this work.

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 datasets used in this work consist of static map images and synthetic or publicly released benchmark data, which do not contain information that names or uniquely identifies individual people. No user-generated personal data is collected or used. As such, we did not perform additional anonymization steps.

- 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?

subsection 4.1

C. Did you run computational experiments?

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

subsection 4.1

- 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?

subsection 4.3 and 4.4

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 work 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)?

This work does not involve recruitment or payment of 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)?

This work does not involve data collected from individuals or user-generated content. The datasets consist of synthetic environments and publicly released benchmark data not associated with identifiable persons. Therefore, issues of individual consent do not apply.

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

No human subjects were involved in this study, so ethics review board approval was not required.

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?

AI assistants were used for code debugging and language polishing. All experimental design, implementation, and analysis were conducted by the authors.