

## Responsible NLP Checklist

Paper title: *TRM-Planner: Offline Target Planning and Distillation for Tiny Recursive Models*

Authors: *Euijin Baek, Housam Babiker, Mi-Young Kim, Randy Goebel*

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?

*Limitations; Ethical Considerations. The paper discusses potential risks related to increased offline computation and storage from caching teacher outputs, and also clarifies scope limitations such as the supervised-only setting and the fact that the cache/planner is not used at test time.*

### B. Did you use or create scientific artifacts? (e.g. code, datasets, models)

- N/A 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?

*(left blank)*

- 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 7.1 (Datasets) and the result tables. The paper reports dataset sizes and statistics, e.g., Sudoku-Extreme uses 1,000 training samples and 423,000 test samples; Maze uses 1,000 train and 1,000 test mazes; ARC-AGI-1 contains 800 tasks and ARC-AGI-2 contains 1120 tasks.*

### C. Did you run computational experiments?

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

*Section 7.4 (Cache and training setup) and Appendix B (Implementation details). The paper reports the offline cache-generation settings, student-side hyperparameters, optimizer/LR/batch-size settings, hardware, and seed information.*

- 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?

*Sections 7.4 and 7.5, plus Appendix C. The paper is explicit that main tables report representative runs (seed 0) unless otherwise noted, and Appendix C adds a controlled 3-seed study with mean std. The paper also clarifies when results are reported as Final versus post-hoc Peak diagnostics.*

*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.?  
(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?  
*Ethical Considerations. The paper states that an AI assistant was used for writing support (e.g., editing for clarity and suggesting phrasing/organization).*