

## Responsible NLP Checklist

Paper title: *TempTool-RL: Tool-Augmented Reinforcement Learning for Temporal Knowledge Graph Question Answering*

Authors: *Zicheng Huang, Yajuan Tong, Xinhui Tu, Tingting He*

How to read the checklist symbols:

- the authors responded 'yes'
- the authors responded 'no'
- <sup>N/A</sup> 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.*

- <sup>N/A</sup> A2. Did you discuss any potential risks of your work?

*(left blank)*

### 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 are standard, publicly available benchmarks constructed from temporal knowledge graphs and do not contain personal or user-generated data. We therefore did not explicitly discuss checking for personally identifying or offensive content in the paper.*

- 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?

*We report dataset statistics and data splits in Section 4 (Experiment), with additional details provided in Appendix C.*

### C. Did you run computational experiments?

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

*We describe the experimental setup and report the hyperparameter settings in Appendix C.*

- 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?

*We report results averaged over multiple runs, as described in Appendix C.*

*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.?

*We conducted a small-scale human evaluation only to assess the quality of the answers, which is not part of the proposed method and is independent of the main experimental results. We did not include the full text of the instructions given to the participants 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)?

*The small-scale human evaluation was conducted only as a supplementary analysis and did not constitute a core component of the study. We did not report details regarding participant recruitment or compensation in the paper.*

- 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)?

*The small-scale human evaluation was only used as a supplementary check of answer relevance and did not involve collecting or releasing any human-generated dataset. We did not discuss consent procedures for the participants in the paper.*

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

*The small-scale human evaluation was conducted only as a supplementary analysis, and we did not seek or report ethics review board approval or exemption in the paper.*

**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?

*We used an AI assistant to support writing and editing, including improving clarity and grammar, and assisting with LaTeX formatting. All scientific content, experimental design, results, and conclusions were developed and verified by the authors.*