

Responsible NLP Checklist

Paper title: *Think Parallax: Solving Multi-Hop Problems via Multi-View Knowledge-Graph-Based Retrieval-Augmented Generation*

Authors: *Jinliang Liu, Jiale Bai, Shaoning Zeng*

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 6. The Limitations and Future Work section discusses risks including: dependence on KG quality, reliance on static knowledge graphs, potential bias introduced by shortest-path weak supervision, and over-retrieval for simple 1-hop queries.

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?

The datasets used (WebQSP, CWQ, BioASQ) are publicly available academic benchmarks for knowledge graph QA. They do not contain personally identifying information or offensive content.

- 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 details including benchmark names, task types, and evaluation splits are described. Hop-length breakdowns are reported in Table 1.

C. Did you run computational experiments?

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

Section 4.1 and Appendix B.1. Implementation details include encoder choice (BGE-M3), training epochs, optimizer (AdamW), learning rate schedule, batch size, PSR strength, and GPU hardware.

- 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?

Appendix D.1 and Table 7. Five independent training runs are reported with mean and standard deviation across all metrics, confirming stability (std < 2.0 points). Main tables report mean performance.

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?

AI assistants (e.g., ChatGPT) were used for grammar checking and proofreading of the manuscript. All research ideas, experiments, and results are original work by the authors.