

Responsible NLP Checklist

Paper title: *Mixture of Length and Pruning Experts for Knowledge Graphs Reasoning*

Authors: *Enjun Du, Siyi Liu, Yongqi Zhang*

How to read the checklist symbols:

- the authors responded 'yes'
- the authors responded 'no'
- 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?

A2 Elaboration: See Section 6 (Limitations), where we discuss risks such as biased path selection under incomplete/noisy KGs and scalability trade-offs on very large graphs.

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

- B1. Did you cite the creators of artifacts you used?

See Section 4.1 (Experimental Setup, Datasets) for dataset attributions and the References for all cited creators of datasets and baseline methods.

- B2. Did you discuss the license or terms for use and/or distribution of any artifacts?

See Section 4.1 (Datasets) and Appendix A (Training/Implementation Details). We use publicly released research benchmarks under their original licenses/policies and do not redistribute data beyond references to the official sources.

- B3. Did you discuss if your use of existing artifact(s) was consistent with their intended use, provided that it was specified? For the artifacts you create, do you specify intended use and whether that is compatible with the original access conditions (in particular, derivatives of data accessed for research purposes should not be used outside of research contexts)?

See Section 4.1 (Datasets) and Appendix A. Our use is for academic research and evaluation, consistent with the datasets intended use; no out-of-scope deployment is performed.

- 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 benchmarks are standard knowledge graph datasets (e.g., WN18RR, FB15k-237, NELL-995, YAGO3-10, Family, UMLS) that do not include PII or offensive content; no personal data collection is involved (see Section 4.1).

- B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?

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.

Section 4.1 (Datasets) describes dataset domains and properties; Appendix A provides additional details about settings and usage.

- 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 (Datasets) reports dataset statistics/splits and the evaluation metrics; additional statistics and settings are provided in Appendix A.

C. Did you run computational experiments?

- C1. Did you report the number of parameters in the models used, the total computational budget (e.g., GPU hours), and computing infrastructure used?

Appendix A (Training Details) documents the computing infrastructure (e.g., NVIDIA GPUs and memory footprint) and training/inference time analysis (also see Section 4.2/Figure 3 for per-epoch timing comparison). Parameter counts and implementation specifics are summarized in the implementation details.

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

Section 4.1 and Appendix A describe the experimental setup, hyperparameters, and training/inference procedures (including the MoE gating, pruning settings, and path length ranges).

- 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 4.2 (Results) and Appendix A report comparative results across datasets with standard ranking metrics (MRR, Hits@K) and include training/inference efficiency analyses; we make clear the evaluation protocol and how metrics are aggregated.

- C4. If you used existing packages (e.g., for preprocessing, for normalization, or for evaluation, such as NLTK, SpaCy, ROUGE, etc.), did you report the implementation, model, and parameter settings used?

Appendix A (Implementation/Training Details) includes package/framework versions and settings used for preprocessing, model training, and evaluation.

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.?

No human subjects or annotators were involved.

- 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)?

No human subjects or annotators were involved.

- 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)?

No human subjects or annotators were involved.

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

No human subjects or annotators were involved.

- D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?

No human subjects or annotators were involved.

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 AI assistants (e.g., ChatGPT-style tools) only for proofreading and wording improvements in the manuscript. They were not used to generate experiment content, code logic, or results.