

Responsible NLP Checklist

Paper title: *PathwiseRAG: Multi-Dimensional Exploration and Integration Framework*

Authors: *Hengrui Zhang, Pin-Siang Huang, Zhen Zhang, Peican Lin, Yao-Ching Yu, Bo Hu, Yulu Du*

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

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

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

We cite all creators of datasets in Section 4.1 and Foundation models and baseline systems are cited in Section 2.

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

While we use publicly available datasets and models with standard academic licenses, we did not explicitly discuss license terms in the paper. All datasets are available under academic use licenses, and our implementation will be released under MIT license.

- 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)?

Sections 4.1

- 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?

These datasets are designed for research purpose and don't contain any information that names or uniquely identifies individual people or offensive content

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

Section 4.1

- 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

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.

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.4

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

Section 4.1 and Appendix A.2

- 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 and 4.3

- 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?

Section 4.1 and Appendix A.2

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.?

Not applicable as our study did not involve human participants. We evaluated our system using existing benchmark datasets (HotpotQA, StrategyQA, ComplexWebQuestions, Natural Questions, TriviaQA) that are publicly available for research purposes and did not collect new human data or annotations.

- 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)?

Only use benchmark and metrics

- 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)?

Not applicable as we used existing publicly available benchmark datasets (HotpotQA, StrategyQA, ComplexWebQuestions, Natural Questions, TriviaQA) that were created and released for research purposes. All datasets were used in accordance with their intended research use and license terms as mentioned in Section 4.1.

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

Our research utilized existing benchmark datasets and did not involve human subjects research requiring ethics review board approval. The datasets (HotpotQA, StrategyQA, ComplexWebQuestions, Natural Questions, TriviaQA) were accessed and used according to their established academic licenses and terms of use as referenced in Section 4.1.

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

Not applicable as we did not collect new annotations for this study. We used existing benchmark datasets whose characteristics are documented in their original publications: HotpotQA (Yang et al., 2018), StrategyQA (Geva et al., 2021), and ComplexWebQuestions (Talmor and Berant, 2018), Natural Questions (Kwiatkowski et al., 2019), TriviaQA (Joshi et al., 2017). These datasets and their annotation processes are well-documented in the NLP community.

E. Did you use AI assistants (e.g., ChatGPT, Copilot) in your research, coding, or writing?

N/A E1. If you used AI assistants, did you include information about their use?

We don't need it.