Responsible NLP Checklist

Paper title: *Thought-Action Graph Reasoning: Faithful and Efficient Reasoning of Large Language Models via Reusing Past Experience*
Authors: *Zhixiao Qi, Feng Huang, Yunqi Zhang, Shijie Zhang, Qingqing Sun, Yongfeng Huang, Minghu Jiang, Shuai Chen, Tianyi Zhang*

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?
Yes. We discussed potential risks associated with large language models, such as incorrect reasoning paths, in the "Limitations" section.

☑ **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?
No. We used exclusively established open-source academic datasets (e.g., WebQSP, CWQ, GrailQA). These datasets are already widely vetted by the research community and do not contain any personally identifying information or offensive content, thus no additional anonymization steps were required.

☑ 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 relevant statistics in "B Experiments" section for the data that we used .

☑ **C. Did you run computational experiments?**

☑ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
We discuss the experimental setup in "5.4 Further Analysis" section.

☑ 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 descriptive statistics about our results in "5.1 Experimental Setup", "5.2 Main Results" and "A TAG Construction" sections.

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.?
No. This study does not involve any human participants or new manual annotations. We exclusively used publicly available datasets and pre-trained models, so no instructions were given to participants.

☒ 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. This study does not involve any human participants or new manual annotations. We exclusively used publicly available datasets and pre-trained models, so no instructions were given to participants.

☒ 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. This study exclusively uses publicly available, third-party datasets. These datasets were previously released under specific open-source licenses, and their original creators were responsible for obtaining the necessary consent from participants. Our study did not involve any direct interaction with human subjects.

☒ D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
No. This study is exempt from ethics review board approval because it exclusively utilizes publicly available datasets that do not involve new data collection from human subjects.

☑ **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?
Yes. We disclose the use of AI assistants in our paper. Specifically, we used GPT-4o-mini to assist in the construction of the Thought-Action Graph (TAG) as described in "4.1 TAG Construction" section. Additionally, GPT-4o-mini was used as a baseline model for comparative analysis in the "5 Experiments" section.