

## Responsible NLP Checklist

Paper title: *ASTRA: Adaptive Semantic Tree Reasoning Architecture for Complex Table Question Answering*

Authors: *Xiaoke Guo, Songze Li, Zhiqiang Liu, Zhaoyan Gong, Yuanxiang Liu, Huajun Chen, Wen Zhang*

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?

*Our work focuses on improving the accuracy of question answering over tabular data using standard public benchmarks. We do not foresee any specific ethical risks or negative societal impacts associated with the proposed methodology.*

### 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?

*Ethics Statement, Appendix A*

- 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?

*See Appendix C and Table 6.*

### C. Did you run computational experiments?

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

*See Section 4.1 and Appendix A.*

- 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?

*Due to the high computational cost of utilizing commercial LLM APIs (e.g., GPT-4o, DeepSeek-V3) on large-scale benchmarks, we report results from a single run for the main QA experiments (Table 2).*

---

*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 human subjects were involved. Manual verification was conducted by the authors.*

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 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 were involved.*

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

*No human subjects 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?

*N/A*