

Responsible NLP Checklist

Paper title: *ODUTQA-MDC: A Task for Open-Domain Underspecified Tabular QA with Multi-turn Dialogue-based Clarification*

Authors: *Zhensheng Wang, ZhanTeng Lin, Wenmian Yang, KunZhou, Yiquan Zhang, Weijia Jia*

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.

- ^{N/A} A2. Did you discuss any potential risks of your work?

All data used in this study are publicly available and do not involve any ethical concerns, personal information, or protected data. Therefore, the research poses no risk to individuals or groups and does not require ethical approval.

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?

Our data do not contain any personally identifiable 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?

"3.1 Datasets Construction" and Appendix "A.7 Dataset Statistics"

C. Did you run computational experiments?

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

Appendix "B. Supplementary experiments"

- 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 did not report error bars or multirun summary statistics because LLM inference especially with chain of thought reasoning consumes very high computational resources and time. Due to these resource and time constraints, our results are based on single runs; comprehensive statistical analysis will be added in future work.

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.?

Appendix "B.3 Ablation Study" and Appendix "A.3 LLM-based Query Rewriting"

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)?

All data used in this study were obtained from publicly accessible sources under their standard terms of use. Because no private or personally identifiable information was collected, no additional consent was required.

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?

Appendix "A.3 LLM-based Query Rewriting"