

## Responsible NLP Checklist

Paper title: *CAMEC: Complexity-Aware Multi-Expert Collaboration for Reliable Chinese Medical Question Answering*

Authors: *Yukang Wu, Xiyuan Jia, Jiayi Wu, HongChen Yu, Yuhan Qiu, Wu Guohua*

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?

*See the Limitations and Ethics Statement sections.*

### 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?

*The datasets used in this work are publicly available benchmark datasets that do not contain personally identifying information. They consist of de-identified, exam-style medical questions and answers, and do not involve patient records or personal data.*

- 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?

*Reported in Section 4 (Experiments), where we describe the benchmark datasets used, including dataset sizes, evaluation splits, and evaluation protocols following prior work.*

### C. Did you run computational experiments?

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

*Discussed in Section 4 (Experiments), where we describe the experimental setup, model configurations, training procedures, and hyperparameter settings following prior work.*

- 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?

*Reported in Section 5 (Results and Analysis), where we present benchmark results using standard evaluation metrics and clearly specify that reported scores correspond to single-run evaluations following established benchmark protocols.*

---

*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.?

*(left blank)*

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)?

*(left blank)*

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?

*AI assistants were used in a limited and auxiliary manner for language polishing and improving clarity during manuscript preparation. All core scientific contributions, including method design, system architecture, experimental setup, evaluation, and analysis, were conceived, implemented, and validated by the authors. AI assistants did not generate experimental results, make scientific decisions, or influence the conclusions of this study.*