Responsible NLP Checklist

Paper title: *Safety-Aware Dialogue System for Postoperative Oral Cancer Care with Structured Clarification and a Clinically Curated Dataset*
Authors: *Tzu-Chi Liu, Hui-Ying Yang, Shiow-Ching Shun, Yu-Chi Chen, Lu-Yen Anny Chen, Yong-Sheng Chen*

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. Potential risks and mitigation strategies are discussed in Section 7 (Limitations).

☑ **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 data used in this study does not include any personally identifying information or offensive content. As outlined in Section 3, all data were collected and processed in accordance with strict anonymization and ethical data handling procedures. The dataset contains no direct or indirect identifiers, such as names, contact information, or unique personal attributes. When human-related data was involved, it was either anonymized at the source or transformed into non-identifiable representations before analysis. Additionally, the data were carefully reviewed to ensure that there was no offensive, abusive, or sensitive content. Consequently, the dataset does not contain any personally identifying information, and no further de-identification steps were necessary beyond those mentioned in Section 3.

☑ 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?
Statistics of the dataset, including the number of QA pairs, data sources, and annotation details, are reported in Section 3.

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

☑ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
Experimental setups, model configurations, thresholds, and evaluation protocols are described in Sections 4 and 5.

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.

☑ 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?
Descriptive statistics, including mean expert ranking scores and false positive/negative rates aggregated over expert annotations, are reported in Sections 5.2 and 5.3.

☑ **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. While clinical experts were involved in annotation and evaluation, the paper does not include the full text of annotation or evaluation instructions; only high-level descriptions of expert review procedures are provided.

☒ 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. Clinical experts participated in data annotation and evaluation; however, the paper does not report detailed information about recruitment procedures or compensation, as expert involvement was part of institutional clinical collaboration rather than a formal paid recruitment process.

☑ 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)?
Yes. Consent and data use procedures are described in Section 3, where data collection under institutional ethical approval and de-identification are reported.

☑ D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
Yes. The data collection protocol was approved by the Institutional Review Board, as stated in Section 3.

☑ **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?
No. AI assistants were used only for language editing and grammar checking during manuscript preparation, without contributing to content generation, experimental design, analysis, or scientific decisions; therefore, no separate disclosure section was included.