

Responsible NLP Checklist

Paper title: *Reasoning Is Not All You Need: Examining LLMs for Multi-Turn Mental Health Conversations*
Authors: *Mohit Chandra, Siddharth Sriraman, Harneet Singh Khanuja, Yiqiao Jin, Munmun De Choudhury*

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?

We have discussed the potential risk with the ethical considerations in section 9

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?

We do not have any personally identifiable information or offensive content in our dataset.

- 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?

Section 3 provides the details for the dataset.

C. Did you run computational experiments?

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

We have described the details within Appendix section M and Table 30.

- 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 have provided descriptive statistics in the result section (section 5)

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.?

While we could not provide the full text of instructions (because we evaluated 6 different metrics), we have provided the comprehensive scoring guidelines for each metric within Tables 2429 (these

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.

guidelines were same for the annotators and LLM judge) and a figure of the annotation interface (Figure 7). Additionally, Appendix L mentions that the participant reviewed the consent form which outlined the task information, potential risks, and compensation.

- 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)?

Yes we have provided details about the recruitment platform, compensation and fairness of the compensation in Appendix section L.

- 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 we have discussed this aspect in Appendix section L.

- D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
While generating synthetic conversations from a publicly available dataset was exempt from Institutional Review Board (IRB) review, we obtained IRB approval for the human evaluation phase conducted via Prolific.

- 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 have included the use of AI for text-rephrasing in appendix section N.