

Responsible NLP Checklist

Paper title: *Schema-Guided Response Generation using Multi-Frame Dialogue State for Motivational Interviewing Systems*

Authors: *Jie Zeng, Yukiko Nakano*

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?

Mentioned in the Risks section of the Ethics Statement.

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?

Mentioned in the Information and Consent section of the Ethics Statement and Appendix B.4.

- 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?

The number of samples created for the dialogue strategy pool of the proposed method is provided in Appendix A.2. The number of strategy rules for the DIIR baseline and the number of samples used for the few-shot baseline are reported in Appendix B.3.

C. Did you run computational experiments?

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

The LLM model used is mentioned in Section 4.3 and the parameters used in the generation are mentioned in the details of each module (Appendix A.1, A.2, A.3, A.4). Information regarding the baselines is provided in Appendix B.2 and B.3.

- 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 present the descriptive statistics of the dialogues obtained in the user study in Table 4. Additionally, we report the results of the annotation analysis and subjective evaluation in Sections 5.1 and 5.2.

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.?

Detailed instructions for the user study participants, including the experimental overview and precautions, are provided in Appendix B.4. The full text of the instructions is displayed in Figure 10.

- 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)?

Mentioned in the details of the user study section in Appendix B.4.

- 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)?

Mentioned in the Informed Consent section of the Ethics statement and Appendix B.4

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

Mentioned in the Ethical Review section of the Ethical Considerations.

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?

Mentioned in Appendix E.