

## Responsible NLP Checklist

Paper title: *PUPPET: Neural-Symbolic Standardized Patients for Mental Health*

Authors: *Chen Xu, Yu Ji, Zhenyu Lv, Yang Yi, Yizhe Yang, Luyao Ji, Chaoyi Chen, Xianyang Wang, Tian Lan, Zhihua Wang, Juan Wang, Xunde Dong, Fuze Tian, Qunxi Dong, Bin Hu*

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?

*Ethical Statement section and Limitations section. We explicitly discussed potential risks and confirmed that appropriate mitigation strategies are in place in the Ethical Statement section.*

### 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?

*Ethical Statement.*

- 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?

*4.1, C.2*

### C. Did you run computational experiments?

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

*Section 4 (Experimental Setup) and Appendix.*

- 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?

*Section 4 (Results) and Appendix. We reported statistical significance (e.g., p-values) and summary statistics for the evaluation results.*

### 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.?

*Section 4.2.1 and Appendix B.3.1.*

---

*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.*

- 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)?  
*Appendix B.3.1 and Appendix B.3.2 (Evaluation Setup).*
- 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)?  
*Ethical Statement.*
- D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?  
*Appendix B.3.1 and Appendix B.3.2.*
- 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 C (Use of AI Assistants). We explicitly declared the scope of our AI usage in the appendix, confirming that it was strictly limited to language polishing and formatting, and not used for coding or research ideation.*