

Responsible NLP Checklist

Paper title: *SPASM: Stable Persona-driven Agent Simulation for Multi-turn Dialogue Generation*

Authors: *Han Luo, Guy Laban*

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?

Ethical Considerations (final unnumbered section after 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?

No. The paper states that the dialogues are fully synthetic and do not involve real users or personal data, so anonymization of human data was not required.

- 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 3.1 Dataset Construction; Section 3.2 Dataset Semantics; Abstract.

C. Did you run computational experiments?

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

Sections 3.13.5; Appendix J. We report the experimental setup, including backbone combinations, temperatures, decoding settings, prompt definitions, and ablation settings. We did not perform a hyperparameter search; experiments used fixed generation settings.

- 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?

Sections 3.23.5; Table 1; Table 2; Table 3; Table 4; Figure 2; Appendix L (Tables 56).

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.?

Appendix O

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)?

No. The manuscript reports the annotation protocol and agreement analysis for two trained annotators, but it does not include details about annotator recruitment or compensation. Because the human annotation was used as a limited validation step rather than as a large-scale crowdsourcing study, we focused on reporting the protocol and reliability of the annotations instead.

- 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)?

No. The manuscript reports the annotation protocol for two trained annotators and states that the underlying dialogues are fully synthetic and do not involve real users or personal data, but it does not describe a consent procedure for annotators or explain how their annotation data would be used. Because the human component is limited to validation of synthetic conversations rather than collection of participant-generated dialogue, the paper focuses on annotation protocol and reliability instead.

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

No. The study primarily uses fully synthetic dialogues generated by language models and does not involve collection of data from real users or personal data. The human component is limited to manual validation of synthetic conversations by two trained annotators, so the manuscript does not report formal ethics-review-board approval or exemption.

- 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?
(left blank)