

## Responsible NLP Checklist

Paper title: *RecMem: Recurrence-based Memory Consolidation for Efficient and Effective Long-Running LLM Agents*

Authors: *Zijie Dai, Shiyuan Deng, Sheng Guan, Yizhou Tian, Xin YAO, Xiao Yan, James Cheng*

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?

*We discuss dual-use risks, unintended harms from incorrect consolidation/retrieval, and the positive environmental impact from reduced LLM invocations (see Section 7 for details).*

### 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 collect any new human-subject conversational data. Our experiments use only publicly released benchmarks (LoCoMo and LongMemEval-S). LongMemEvals dataset creators report a rigorous human screening process to ensure the released evaluation set contains no personally identifiable information (PII) or offensive content, and the collection protocol was deemed IRB-exempt. LoCoMo is generated via a human-machine pipeline with synthetic personas (which may include fictional names/demographics) and human editing for consistency. We use the datasets as provided by their authors and do not attempt to deanonymize any content. We believe under this setting, there’s no need to include such details.*

- 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?

*Yes. Section 4.1 provide a high-level description of the two used datasets, while Appendix D provide a complete report for those two.*

### C. Did you run computational experiments?

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

*Yes, we provide a high-level description about experimental setup in Section 4.1. Appendix C provides a detailed hyperparameter sensitivity experiments while Appendix E thoroughly discuss our experimental settings.*

---

*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?

*Section 4 (Metrics) provides such details.*

**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.?

*(left blank)*

- 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)?

*(left blank)*

- 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)?

*(left blank)*

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

*(left blank)*

**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?

*We used AI only for language polishing (grammar correction and paraphrasing for clarity/readability). The AI assistants were not used to generate new scientific content, new ideas, experimental results, or literature search.*