

Responsible NLP Checklist

Paper title: *BMAM: Brain-inspired Multi-Agent Memory Framework*

Authors: *Yang Li, Jiaxiang Liu, Yusong Wang, Yujie Wu, Mingkun Xu*

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?

Section 7 (Ethics Statement) discusses user consent, data ownership, and long-term data retention concerns for persistent-memory agent systems.

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?

All four benchmarks used (LoCoMo, LongMemEval, PrefEval, PersonaMem) are publicly released datasets of synthetic or anonymised dialogues. No personally identifying information was collected or released by this work. Dataset citations are in Section 4.1.

- 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 4.1 (Table 1) reports the scale of all four benchmarks: LoCoMo (10 groups, 1986 QA pairs), LongMemEval (500 questions), PersonaMem (20 users, 589 QA), and PrefEval (1000 questions). Appendix A.1 provides module and agent capacities.

C. Did you run computational experiments?

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

Section 4.1 describes the evaluation protocol. Appendix D (Hyperparameters) lists all brain-region capacities, embedding model, LLM backbone (gpt-4o-mini), temperature settings, and algorithmic parameters (I_{max} , τ_{temp} , learning rate, RRF constant k).

- 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?

Appendix A reports Wilson 95% confidence intervals and two-proportion z-tests for the brain-region ablation (Table 16 / significance-stats). All LoCoMo numbers are averaged over 8 independent runs; ablation experiments are averaged over 3 runs.

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.?
(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?
AI assistants (ChatGPT-class LLMs) were used for language polish on portions of the manuscript (grammar, phrasing) and for exploratory code scaffolding during development. All scientific content, experimental design, results, and conclusions are the authors' own. No AI assistant authored any research content or ran any experiments.