

Responsible NLP Checklist

Paper title: *GAM: Hierarchical Graph-based Agentic Memory for LLM Agents*

Authors: *Zhaofen Wu, Hanrong Zhang, Fulin Lin, Wujiang Xu, Xinran Xu, Yankai Chen, Henry Peng Zou, Shaowen Chen, Weizhi Zhang, Xue Liu, Philip S. Yu, Hongwei Wang*

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 potential risks in the Ethical Considerations section, including privacy concerns arising from persistent memory storage, risks of incorrect or biased memory formation, and challenges in ensuring user control over stored information.

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 utilized established public benchmarks (LOCOMO and LongDialQA) that are widely used in the research community. These datasets are derived from public TV scripts and synthetic dialogues, which are generally considered free of non-public Personally Identifiable Information (PII). Therefore, no additional steps for anonymization were 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?

Yes. We report dataset details in the Experimental Setup section, including dataset characteristics and evaluation settings.

C. Did you run computational experiments?

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

Section 4.1 (Experimental Settings) details the models and settings used, and Appendix B (Hyperparameter Sensitivity Analysis) discusses the selection of key hyperparameters (factors and retrieval size).

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,

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.

etc. or just a single run?

Section 4.2 (Main Results) reports average scores for key metrics, and Appendix B illustrates performance trends across parameter variations.

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.?

This study is based entirely on automatic evaluation using existing public datasets and did not recruit any human subjects or annotators.

- 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)?

This study did not involve human participants, so no recruitment or payment took place.

- 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)?

We utilized only public benchmark datasets that do not require additional consent procedures. No new human data was collected for this study.

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

Institutional Review Board (IRB) approval was not required as the study involved only public data and computational experiments without human subject intervention.

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 ChatGPT exclusively for linguistic polishing, such as checking for grammatical errors and improving text fluency. All scientific claims and intellectual content are original to the authors.