

## Responsible NLP Checklist

Paper title: *Synapse: Empowering LLM Agents with Episodic-Semantic Memory via Spreading Activation*  
Authors: *Hanqi Jiang, Junhao Chen, Yi Pan, Ling Chen, Weihang You, Yifan Zhou, Ruidong Zhang, Yohannes Abate, Tianming Liu*

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 6 (Ethical Considerations)*

### 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?  
*Section 2 (Related Work) and Section 4.1. All used artifacts, including the LoCoMo benchmark (Maharana et al., 2024), pre-trained LLM backbones (GPT-4o-mini, GPT-4o, Qwen-1.5B, Qwen-3B), and the sentence encoder all-MiniLM-L6-v2, are properly cited.*

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?  
*Appendix A.3. We report per-category instance counts on LoCoMo: Multi-Hop (N=841), Single-Hop (N=282), Temporal (N=321), Open-Domain (N=96), totaling  $N_{total}=1540$  non-adversarial samples. Statistical significance tests are conducted on  $N=500$  instances (Table 1 caption).*

### C. Did you run computational experiments?

C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?  
*Section 4.1 (Implementation Details) and Appendix C (Hyperparameter Sensitivity Analysis, Table 6). We report all key hyperparameters, including embedding dimension (384), propagation steps  $T=3$ , retention =0.5, temporal decay =0.01, hybrid retrieval weights ={0.5, 0.3, 0.2}, inhibition  $M=7$ , spreading factor  $S=0.8$ , and LLM temperature=0.1. A systematic sweep over each hyperparameter is reported in Appendix Table 6.*

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?

Appendix D.1 (Table 8). We report mean F1 scores with standard deviations across 3 independent runs (all std 0.5). Statistical significance is verified via paired t-tests on instance-level scores ( $p < 0.05$ ,  $N = 500$ ), as noted in the Table 1 caption.

**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 (e.g., ChatGPT, Claude) were used for grammar proofreading, writing style refinement, and coding assistance (e.g., boilerplate code generation, debugging suggestions). All core scientific contributions including the SYNAPSE architecture design, spreading activation formulation, experimental methodology, results analysis, and manuscript content are the original work of the authors. No AI assistant was used for generating experimental results, analysis, or scientific claims.*