Responsible NLP Checklist

Paper title: *VizoMem: A Visual-Textual Memory Framework for Efficient Long-Horizon Reasoning*
Authors: *Weijie Liang, Yuanfeng SONG, Xing Chen, Caleb Chen Cao, Sirui Han, Yike Guo*

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?
This work focuses on model efficiency and does not involve high-risk applications or sensitive data. All experiments are conducted on publicly available benchmark datasets. Potential risks are minimal and align with known limitations of large language models.

☑ 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?
This work uses publicly available benchmark datasets and does not involve new data collection. We rely on the data curation and filtering procedures provided by the original dataset creators. Therefore, we do not perform additional checks for personally identifying information or offensive content beyond those already established.

☑ 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 5.1 and Appendix C.2 report benchmark test data; Appendix B.1 reports the construction of training datasets.

☑ C. Did you run computational experiments?

☒ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
The experiments focus on validating the feasibility and efficiency of the visual representations; hyperparameter search was not the primary concern, and standard values were used.

☑ 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?
In Appendix C.1, we report the results of three independent runs of our method. In the main text, we report the mean across these 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?**

☒ N/A 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.?
All data are from publicly available datasets.

☒ N/A 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)?
All data are from publicly available datasets.

☒ N/A 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)?
All data are from publicly available datasets.

☒ N/A D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
All data are from publicly available datasets.

☑ **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 LLMs as evaluators to compute the LLM-as-a-Judge metric. Details of this usage are provided in Appendix B.3 and D.1.