Responsible NLP Checklist

Paper title: *GLIMPSE: Do Large Vision-Language Models Truly Think With Videos or Just Glimpse at Them?*

Authors: *Yiyang Zhou, Linjie Li, Shi Qiu, Zhengyuan Yang, Yuyang Zhao, Siwei Han, Yangfan He, Kangqi Li, Haonian Ji, Zihao Zhao, Haibo Tong, Lijuan Wang, Huaxiu Yao*

> 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?
We discussed our potential risks regarding data contamination and trustworthiness in LLM-generated response in Limitations and Experiments chapter.

☑ **B. Did you use or create scientific artifacts? (e.g. code, datasets, models)**

☑ B1. Did you cite the creators of artifacts you used?
We have cited more than 30 of models and benchmark papers in our Reference.

☑ B2. Did you discuss the license or terms for use and/or distribution of any artifacts?
We have confirmed that all copyrights are rightly used during our work in Appendix.

☑ B3. Did you discuss if your use of existing artifact(s) was consistent with their intended use, provided that it was specified? For the artifacts you create, do you specify intended use and whether that is compatible with the original access conditions (in particular, derivatives of data accessed for research purposes should not be used outside of research contexts)?
We have discussed the consistency of existing artifacts in Experiments section.

☑ 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 have discussed in detail our dataset curation in Methodology section.

☑ B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?
We have provided in Section 2 and 3.

☑ 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?
We have discussed in Experiment Chapter. (section 4)

☑ **C. Did you run computational experiments?**

☑ C1. Did you report the number of parameters in the models used, the total computational budget (e.g., GPU hours), and computing infrastructure used?
We have discussed in Experiment Chapter. (chap 4)

☑ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
We have discussed in Experiment Chapter. (chap 4)

☑ 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?
We have discussed in Experiment Chapter. (chap 4)

☑ C4. If you used existing packages (e.g., for preprocessing, for normalization, or for evaluation, such as NLTK, SpaCy, ROUGE, etc.), did you report the implementation, model, and parameter settings used?
We have discussed in Experiment Chapter. (chap 4)

☑ **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.?
We have discussed in Appendix 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)?
We have discussed in Appendix 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)?
We have discussed in Appendix A.

☑ D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
We have discussed in Appendix A and Section 4.

☑ D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?
We have discussed in Experiment Section(chap 4).

☒ **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 haven't used AI assistants during our project.