Responsible NLP Checklist

Paper title: *SEA: Supervised Embedding Alignment for Token-Level Visual-Textual Integration in MLLMs*
Authors: *Yuanyang Yin, Yaqi Zhao, Yajie Zhang, Yuanxing Zhang, Ke Lin, Jiahao Wang, Xin Tao, Pengfei Wan, Wentao Zhang, Feng Zhao*

> 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?
7

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

☑ B1. Did you cite the creators of artifacts you used?
4 Experiments

☒ B2. Did you discuss the license or terms for use and/or distribution of any artifacts?
We did not include a license discussion because no new artifacts were released; all used models and datasets follow their original licenses.

☒ 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 did not explicitly discuss artifact use in the paper. However, all datasets and pretrained models employed in our work are open-source and distributed under licenses that permit academic research. Our use of these artifacts (e.g., fine-tuning open-source models on open datasets) is fully consistent with their intended purpose and license conditions.

☒ 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 did not explicitly discuss this aspect in the paper. However, all datasets used in our work are widely adopted open-source benchmarks that do not contain personally identifying information or offensive content, and they are released under licenses ensuring compliance with ethical and legal standards.

☒ B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?

We used publicly available, standard open-source datasets. No additional documentation was provided as these datasets are widely used and compliant with standard ethical guidelines.

☒ 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 used publicly available datasets with pre-defined train/dev/test splits; all relevant statistics are provided in the original dataset publications.

☑ **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?
4 Experiments

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

☑ 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?
4.4 Ablation Study

☒ 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 used standard open-source packages and evaluation metrics without modification; default settings were applied.

☒ **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.?
(left blank)

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)?
(left blank)

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)?
(left blank)

N/A D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
(left blank)

N/A D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?
(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?
We used AI assistants (e.g., ChatGPT) only for minor writing and grammar corrections. No AI was used for research, data analysis, or experimental results. We did not include a separate section reporting AI use in the manuscript, as the usage was limited to editorial improvements.