Responsible NLP Checklist

Paper title: *Jigsaw-Puzzles: From Seeing to Understanding to Reasoning in Vision-Language Models*
Authors: *Zesen Lyu, Dandan Zhang, Wei Ye, Fangdi Li, Zhihang Jiang, Yao Yang*

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.

N/A A2. Did you discuss any potential risks of your work?
There are no potential risks associated with this work.

☑ **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 the relevant literature for the datasets and models used in Sections 3 and 4.

N/A B2. Did you discuss the license or terms for use and/or distribution of any artifacts?
The datasets and models are publicly available and licensed for research purposes, and we have cited the corresponding papers.

☑ 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)?
The use of all existing artifacts in this work is consistent with their intended purposes and license conditions. For the artifacts we create, we explicitly specify that they are intended for research use only, and their use remains compatible with the access restrictions of the original data, in particular that derivatives of research data will not be used outside of research contexts. Please check section 3 & 4.

N/A 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?
The data does not include personally identifying info or offensive content.

N/A B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?
No additional documentation was provided beyond the description in the paper, as the artifacts used are restricted to a single domain and research context, and do not involve sensitive demographic information.

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.

☑ 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?
Please cheak Sec 3.

☑ **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?
Please section 4.1.

☒ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
We do not need to conduct hyperparameter search and best-found hyperparameter values.

☑ 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?
Please check sec 4.2.

☒ 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 do not use some existing packages.

☑ **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.?
Please check section 3.2.

☐ 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)?
We do not report this event.

☑ 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)?
Please check sec 3.

☐ N/A D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
This work uses publicly available datasets and models that are licensed for research purposes; hence ethics review board approval is not applicable.

☐ N/A D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?
We do not report this event.

☒ **E. Did you use AI assistants (e.g., ChatGPT, Copilot) in your research, coding, or writing?**

☐ N/A E1. If you used AI assistants, did you include information about their use?
We did not use any AI assistants.