

Responsible NLP Checklist

Paper title: *Chart-MRAG: Benchmarking Multimodal Retrieval Augmented Generation on Chart-based Documents*

Authors: *ymyang, Jiang Zhong, Li Jin, Xiao Sun, Jingwang Huang, gaojinpeng, Qing Liu, Yang Bai, Jingyuan Zhang, Rui Jiang, Qin Lei, Kaiwen Wei*

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?

All potential risks have been fully addressed in the Ethical Considerations section, including annotation details and payment terms.

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?

In the Ethical Considerations section, we declare that all data was collected online, retaining only relevant text and figures, manually verified and cleaned ; no personal identifying information or offensive content exists.

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?

Relevant data statistics are fully presented in Section 4 Chart-MRAG Bench. The proposed dataset serves as testing data only, so no train/test/dev split is required.

C. Did you run computational experiments?

^{N/A} C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?

Fully discussed in Section 5 Experiments, including hyperparameter search and optimal 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?

Descriptive statistics for results are detailed in Section 5 Experiments.

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?

- 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 used human annotators, and detailed information (including cooperating annotation company, number and region of annotators, annotation tasks, salary, time, verification and acceptance plans, and other potential risks) is provided in the Ethical Considerations section.

- 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 used human annotators, and detailed information (including cooperating annotation company, number and region of annotators, annotation tasks, salary, time, verification and acceptance plans, and other potential risks) is provided in the Ethical Considerations section.

- 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)?

Data is collected from public websites permitted for academic use. We explain how the data is used in Section 4 Chart-MRAG Bench.

- D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?

This work involves no ethical concerns, so IRB approval is not required.

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 employed GPT as LLM-as-a-Judge in the Section 3 CHARGE Framework and used it for manuscript polishing. All models are clearly stated and cited in the paper.