

Responsible NLP Checklist

Paper title: *Vista-LLM: Decoupled Query-Guided Visual Token Pruning for Efficient Long-Video Large Language Models*

Authors: *Zhenyu Li, Zuchao Li, Ping Wang, Lefei Zhang, Haojun Ai*

How to read the checklist symbols:

- the authors responded 'yes'
- the authors responded 'no'
- 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 did not include a specific section on societal risks because our contribution is strictly limited to improving the computational efficiency of pre-existing models. Our method does not alter the model's output safety alignment or introduce new data sources that would warrant a separate risk analysis.

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?

We utilize standard public benchmarks (Video-MME, MLVU LongVideoBench, MVBench) and the Visual Genome dataset, which have already undergone anonymization and safety checks by their original creators. We did not perform additional PII filtering as we did not collect any new data.

- 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 4.1, Appendix A.1

C. Did you run computational experiments?

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

Section 4.1, Section 4.3, Appendix B.1, Appendix B.2

- 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?

Due to the substantial computational cost associated with Long-Video LLM inference, we report single-run results for our experiments, which is consistent with the baselines and standard practice in this research domain. However, we demonstrate robustness by evaluating on multiple diverse benchmarks and reporting average scores across different tasks within these benchmarks

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.?

This work does not involve human participants or annotators. Therefore, no instructions were required or reported.

- 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)?

No human participants were recruited or paid for this study. All evaluations rely on automated metrics and existing ground-truth labels.

- 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 utilize existing public datasets and did not collect new data from individuals, so no new consent procedures were necessary.

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

Since this research involves neither human subjects nor new data collection, Institutional Review Board approval was not applicable or sought.

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 solely for minor grammatical polishing, which we considered part of standard tool usage and thus did not explicitly disclose in the text.