

## Responsible NLP Checklist

Paper title: *DiVE: Decoupling Intra-layer Visual Evidence for Mitigating Hallucinations in Large Vision-Language Models*

Authors: *Xinwei Li, Li Lin, Hui Jiao, Li Yao, Tien-Tsin Wong, Hanqian Wu*

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?

*No. Our paper focuses on a decoding-time mitigation method for reducing visual hallucinations in existing LVLMs, without introducing new data collection or new deployment-facing capabilities. The remaining concerns are largely inherited from the underlying LVLMs; we therefore discuss technical failure cases and limitations in Sec. 6 (Limitations) rather than adding a separate risk discussion.*

### 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?

*No. We do not collect new human data; we only evaluate on established public benchmarks (e.g., MS-COCO / A-OKVQA / GQA and MME-Hallucination).*

- 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?

*Yes. We report dataset/evaluation statistics and how the evaluation sets are formed in Sec. 4.1 Experiment Setup and Appendix A.2 (Benchmark), including the number of sampled questions/images for POPE, CHAIR, and MME-Hallucination.*

### C. Did you run computational experiments?

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

*Yes. The experimental setup and key hyperparameter values are provided in Sec. 4.1 (Experimental Setup). Additional sensitivity analyses are included in the Appendix B.1.*

- 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?

*Yes. In Sec. 4 we clarify the aggregation procedure for reported numbers, e.g., decoding each*

---

*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.*

*sample multiple times and averaging for CHAIR, and averaging runtime/overhead measurements over multiple runs (see Appendix B.2).*

**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.?

*N/A No human participants or annotators were involved.*

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)?

*N/A No recruitment or payment; no human annotators.*

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)?

*N/A No new data collection involving people; only public benchmarks.*

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

*N/A No human-subjects protocol.*

**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?

*No. We used AI assistants (e.g., ChatGPT) for limited support such as English polishing / LaTeX editing and minor code refactoring. All technical contributions, experiments, and conclusions were produced and verified by the authors.*