

Responsible NLP Checklist

Paper title: *DORA: A Dual-Objective Reinforcement Learning Framework for Effective and Efficient Multimodal Agentic Search*

Authors: *Guangming Qin, Yuhao Deng, Yukun Zhao, Zhenyang Li, Junfeng Wang, Dawei Yin, Ye Yuan, Guoren Wang, Yizhou Yan, Chengliang Chai, Lei Cao*

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?

Our research focuses on developing a reinforcement learning framework (DORA) for improving search efficiency and answer accuracy in multimodal question answering. The method is purely algorithmic and foundational, targeting the optimization of model reasoning behavior rather than deployment in real-world applications. It does not involve sensitive data, human subjects, or user-generated content, and does not produce outputs intended for public dissemination (e.g., news, social media, or user profiles). Therefore, it poses minimal risk of malicious or unintended harmful effects such as disinformation, surveillance, or privacy leakage.

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?

Our study does not collect new data. We use only publicly available, curated benchmark datasets (Enc-VQA, InfoSeek, SK-VQA), which have been widely used in prior research and do not contain any personally identifiable information (PII) or offensive content. These datasets consist of synthetic or anonymized questions and images without naming or uniquely identifying individuals. Therefore, no additional anonymization or filtering was required.

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

C. Did you run computational experiments?

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

Section 4

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.

- 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?

Section 4

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 employ publicly available open-source datasets that are already annotated, requiring no additional human labeling or data collection.

- 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 employ publicly available open-source datasets that are already annotated, requiring no additional human labeling or data collection.

- 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 employ publicly available open-source datasets that are already annotated, requiring no additional human labeling or data collection.

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

Our work relies solely on publicly available benchmark datasets (Enc-VQA, InfoSeek, SK-VQA), which are open-source and widely used in prior research. Since we did not collect new human data, no ethics review board approval was 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?

Yes. We used AI assistants (e.g., ChatGPT) only for writing assistance, such as language polishing and improving clarity of the manuscript. All technical content, model design, experiments, and analysis were developed and validated by the authors.