

Responsible NLP Checklist

Paper title: *Data-Centric Perspectives on Agentic Retrieval-Augmented Generation: A Survey*

Authors: *Jingwen Deng, Jihao Huang, Zhen Hao Wong, Hao Liang, Quanqing Xu, Bin CUI, Wentao Zhang*

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?

Section 4. Challenges and Open Problems

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?

As a survey, it reviews existing, published datasets rather than collecting new data. The responsibility for reviewing data content for Personally Identifying Info and offensive material lies with the original dataset creators, and auditing every cited dataset is beyond the scope of this survey.

- 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 3 and B

C. Did you run computational experiments?

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

N/A because this is a survey paper and did not involve computational experiments.

- N/A* 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?

N/A because this is a survey paper and did not involve computational experiments.

D. Did you use human annotators (e.g., crowdworkers) or research with human subjects?

- N/A* 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 because this is a survey paper and did not involve human annotators.

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.

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 because this is a survey paper and did not involve 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 because this is a survey paper and did not involve human annotators.

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

N/A because this is a survey paper and did not involve human annotators.

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?

LLMs such as ChatGPT were used in writing this paper to help polish the language. The writing of this survey were fully conducted under the authors' guidance. All authors are fully responsible for the accuracy and integrity of the writing.