

## Responsible NLP Checklist

Paper title: *Candidate-Aware Retrieval and Reranking for Multiple-Choice Question Answering: Arabic as a Case Study*

Authors: *Yassine Bouziane, Youness Moukafih, Mounir Ghogho*

How to read the checklist symbols:

- the authors responded 'yes'
- the authors responded 'no'
- <sup>N/A</sup> 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 work remains in the technical domain and does not deploy the system for public or high-stake scenarios, no direct risks are present.*

### 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?

*The datasets employed in our work are established educational resources, curated and vetted by their original creators or distributing organizations to exclude personally identifiable information and offensive content.*

- 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, subsection 3.2*

### C. Did you run computational experiments?

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

*section 4*

- 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?

*Primarily reports general performance results to demonstrate the effectiveness of the candidate-aware retrieval and reranking method without detailed statistical analysis across multiple runs.*

---

*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.?  
*section 3, subsection 3.3*

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 did the evaluation by our self. The evaluation involved only a small amount of data, there was no need to recruit external participants.*

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)?  
*The datasets used in your research consist of publicly available educational resources, which are typically distributed under terms that already address consent and data use by the original providers*

D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?  
*There was no new data collection protocol implemented, seeking ethics review was not applicable or necessary.*

**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?  
*AI writing tools used solely for minor spelling and grammar corrections. Such minimal use does not require special disclosure or citation.*