

Responsible NLP Checklist

Paper title: *Decomposed Prompting Does Not Fix Knowledge Gaps, But Helps Models Say I Dont Know*

Authors: *Dhruv Madhwal, Lyuxin David Zhang, Dan Roth, Tomer Wolfson, Vivek Gupta*

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?

Our work consists of an analysis of prompting interfaces for closed-book QA and a simple disagreement-based abstention rule that encourages models to answer I dont know when different prompts disagree. All experiments use existing public open- and closed-source LLMs and established QA benchmarks (Bamboogle, FRAMES, MuSiQue, CRAG, HotpotQA, Mintaka). We do not collect new data, train or deploy new systems, or target high-stakes domains (e.g., medical or legal). Because the method is intended to reduce hallucinations rather than enable new capabilities, we do not identify any specific additional risks beyond standard, well-known concerns about LLMs, and therefore did not include a separate risk section.

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 analysis uses only standard, publicly released and renowned QA datasets (Bamboogle, FRAMES, MuSiQue, CRAG, HotpotQA, Mintaka), and we do not collect new data or annotate user-generated content. The proposed abstention method operates solely on model-generated answers and their agreement across prompts and does not require any personal or sensitive information. Because neither our datasets nor our analysis/method introduce new PII or offensive content beyond what is already handled by the original dataset creators, we did not add an additional dedicated discussion.

- 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?

3, Appendix B

C. Did you run computational experiments?

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

3

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?

3, 4, Appendix A, Appendix D

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.?

(left blank)

- 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)?

(left blank)

- 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)?

(left blank)

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

(left blank)

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?

Used for writing and grammar corrections