

Responsible NLP Checklist

Paper title: *Federated Retrieval-Augmented Generation: A Systematic Mapping Study*

Authors: *Abhijit Chakraborty, Chahana Dahal, 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?

Section 6 discusses risks related to privacy/utility trade-offs, latency overhead from encryption, and evaluation limitations.

☒ B. Did you use or create scientific artifacts? (e.g. code, datasets, models)

☒ B1. Did you cite the creators of artifacts you used?

Section 2 cites all datasets, benchmarks, and systems (e.g., MIRAGE, FeB4RAG, RAGAS).

☒ B2. Did you discuss the license or terms for use and/or distribution of any artifacts?

Section A.1 notes benchmarks/toolkits are open-sourced under permissive research licenses.

☒ B3. Did you discuss if your use of existing artifact(s) was consistent with their intended use, provided that it was specified? For the artifacts you create, do you specify intended use and whether that is compatible with the original access conditions (in particular, derivatives of data accessed for research purposes should not be used outside of research contexts)?

Section 2 clarifies artifacts were used in accordance with their intended purpose for benchmarking and evaluation.

☐ 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?

This study did not collect or use personal/offensive data; only published benchmarks and papers were analyzed.

☒ B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?

Section A.1 describes domains, coverage, and evaluation focus of the benchmarks.

☒ 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 2 reports the number of studies screened (50) and included (18).

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.

☒ **C. Did you run computational experiments?**

☐ C1. Did you report the number of parameters in the models used, the total computational budget (e.g., GPU hours), and computing infrastructure used?

No new model training was conducted.

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

No experiments required hyperparameter tuning.

☐ 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?

No computational results were produced; only literature synthesis.

☐ C4. If you used existing packages (e.g., for preprocessing, for normalization, or for evaluation, such as NLTK, SpaCy, ROUGE, etc.), did you report the implementation, model, and parameter settings used?

No ML packages were used beyond standard bibliographic tools.

☒ **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.?

No human annotation was conducted.

☐ 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)

☐ D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?

(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?

(left blank)