## Responsible NLP Checklist

Paper title: *MaGiX: A Multi-Granular Adaptive Graph Intelligence Framework for Enhancing Cross-Lingual RAG*
Authors: *Nguyen Manh Hieu, Vu Lam Anh, Hung Pham Van, Nam Le Hai, Linh Ngo Van, Nguyen Thi Ngoc Diep, Thien Huu Nguyen*

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?
*We discuss potential risks and limitations of our work in the Limitations section, directly following the Conclusion (Section 5).*

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

☑ B1. Did you cite the creators of artifacts you used?
*The datasets (ZaloWikipediaQA, ZaloLegal2021, NaturalQuestions, PopQA, MuSiQue) and baselines (BGE-M3, GraphRAG, LightRAG, HippoRAG 2) are cited in Section 4.1 (Experiment Setup), with full references provided in the References section*

☑ B2. Did you discuss the license or terms for use and/or distribution of any artifacts?
*The datasets and baselines used in our experiments (Section 4.1) are publicly available for research purposes under their respective licenses as cited in the References.*

☑ 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)?
*We use the datasets ZaloWikipediaQA, ZaloLegal2021, NaturalQuestions, PopQA, and MuSiQue, and baseline methods BGE-M3, GraphRAG, LightRAG, and HippoRAG2, all released by their authors for research purposes (see Section 4.1, Appendix C). The artifacts we contribute (cross-lingual KG, fine-tuned embeddings and Reranking Mechanism, Sec. 3.23.3-3.5) are likewise intended only for academic research.*

☒ 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 used in our work (ZaloWikipediaQA, ZaloLegal2021, NaturalQuestions, PopQA,*

*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.*

*MuSiQue) are publicly released benchmarks that do not contain personally identifying information or offensive content. Therefore, no additional anonymization was required.*

☑ B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?
*We document datasets in Section 4.1 and Appendix C (domains, languages, dataset statistics), and describe our artifactsthe cross-lingual KG and fine-tuned embeddingsin Sections 3.23.3, with their retrieval and usage mechanism detailed in Section 3.5.*

☑ 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?
*We report dataset statistics, including the number of queries and documents, in Section 4.1 and Appendix C (see Table 6).*

☑ **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?
*We did not explicitly report the number of parameters or total computational budget. Our work mainly relies on publicly available models (e.g., Gemini 2.0 Flash, BGE-M3) and standard GPUs for fine-tuning.*

☑ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
*Yes. We report our experimental setup and hyperparameter choices in Appendix C : Additional Experimental Settings*

☒ 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?
*We report single-run results in Section 4.2 and Appendix D. We did not include descriptive statistics (e.g., error bars, multiple runs) due to computational cost constraints, but we ensured consistent evaluation across all methods.*

☒ 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?
*We described our implementation details and evaluation metrics in Section 4.1 and Appendix C, but we did not explicitly report the package names used. Since our experiments relied on widely adopted frameworks, we considered the package choice standard and did not detail it further.*

☒ **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.?
*(left blank)*

N/A 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)*

N/A 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)*

☒ N/A D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
*(left blank)*

☒ N/A 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?
*Although not mentioned in the paper, we used AI assistants (e.g., ChatGPT) only for language refinement and grammar editing. They were not involved in experiment design, coding, or result generation.*