

We sincerely thank the reviewers for their thoughtful and constructive feedback on our paper, "**Evaluating RAG Pipelines for Arabic Lexical Information Retrieval: A Comparative Study of Embedding and Generation Models**". We carefully considered all suggestions and revised our manuscript accordingly. The reviewers' insightful comments helped improve the clarity, depth, and overall quality of our research. We also express our gratitude for the opportunity to present our work. Below, we provide detailed responses to each question and indicate where corresponding changes have been made in the paper.

Reviewer 1 Question	Response Summary	Revised Section
How might the RAG pipeline be adapted for other Abjad or Ajami languages, such as Hausa or Pashto?	We included a discussion on adapting RAG pipelines for languages like Pashto and Sindhi, highlighting shared morphological features.	Results and Discussion, under the subsection "Adapting the RAG Pipeline for Abjad and Ajami Languages"
Were there challenges with Arabic diacritics and morphology, and how might these inform adaptations?	We expanded the analysis of Arabic morphology and diacritics, noting common retrieval challenges and proposing techniques applicable	Results and Discussion
Could the sentence embedding methodology apply to similar languages with complex morphology?	We addressed this by discussing how embedding methods might generalize to morphologically rich languages such as Pashto, given that the sentence embedding model with the highest score was trained on Persian and Arabic. We suggested that training it further on Pashto-specific data might enhance performance.	Results and Discussion, under the subsection "Adapting the RAG Pipeline for Abjad and Ajami Languages"

Reviewer 2 Question	Response Summary	Revised Section
Can you provide more details about the dataset's size, diversity, and lexical coverage?	We have already added a detailed description of the evaluation dataset, including task types, number of examples, and lexical diversity.	Starting from line 267
What criteria were used to select models and fine-tune parameters?	1) We explained model selection criteria based on previous studies and newly developed Arabic LLM benchmarks. 2) There was no fine-tuning applied. The only parameter used was the temperature, set to zero. We clarified this in the generation methodology and added information about the prompts used.	1) Model selection clarification: Line 193 2) Parameter explanation: Line 184
Were there particular challenges the pipeline struggled with, and could examples be included?	We provided additional analysis for each task in the results and discussion section, with examples of model responses, highlighting tasks with low and high performance and explaining reasons for the differences.	Line 320
How might the methodology generalize to languages using Arabic script, such as Uyghur or Sindhi?	We included a discussion on adapting RAG pipelines for languages like Pashto and Sindhi, highlighting shared morphological features.	Results and Discussion, under the subsection "Adapting the RAG Pipeline for Abjad and Ajami Languages"
How does the study build on or differ from prior research in lexical retrieval?	We evaluated new models and previously known models in researches. Also assessed the performance of these models within RAG pipelines, focusing on their application in Arabic lexical tasks from our dataset.	