# RMIT-ADM+S at the SIGIR 2025 LiveRAG Challenge
# G-RAG: Generation-Retrieval-Augmented Generation

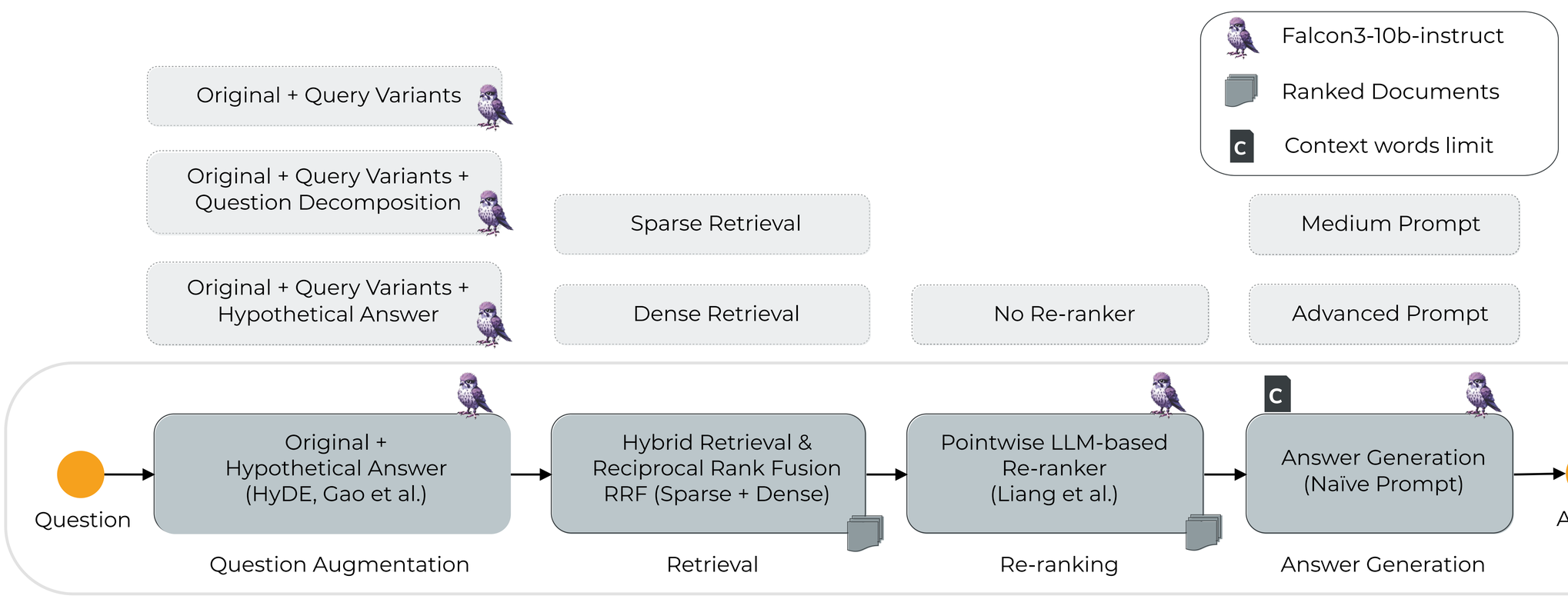
Kun Ran, Shuoqi Sun, Khoi Nguyen Dinh Anh, Damiano Spina, Oleg Zendel

RMIT UNIVERSITY

ADM+S ARC Centre of Excellence for Automated Decision-Making and Society

SIGIR 2025 LiveRAG Challenge

github.com/rmit-ir/GRAG-LiveRAG


## In-House Evaluation

### Question-Answer (QA) Pairs Generation

- QA pairs are generated by **DataMorgana**.
- Questions are based on **up to two** documents.
- Generated varying size Q&A sets.

### Manual Examination

- Reviewing Q&A pairs manually revealed some general challenges, such as "ryse" being mis-corrected to "rise." It was also used to adjust the evaluation prompts to agree with our (human) judgments.

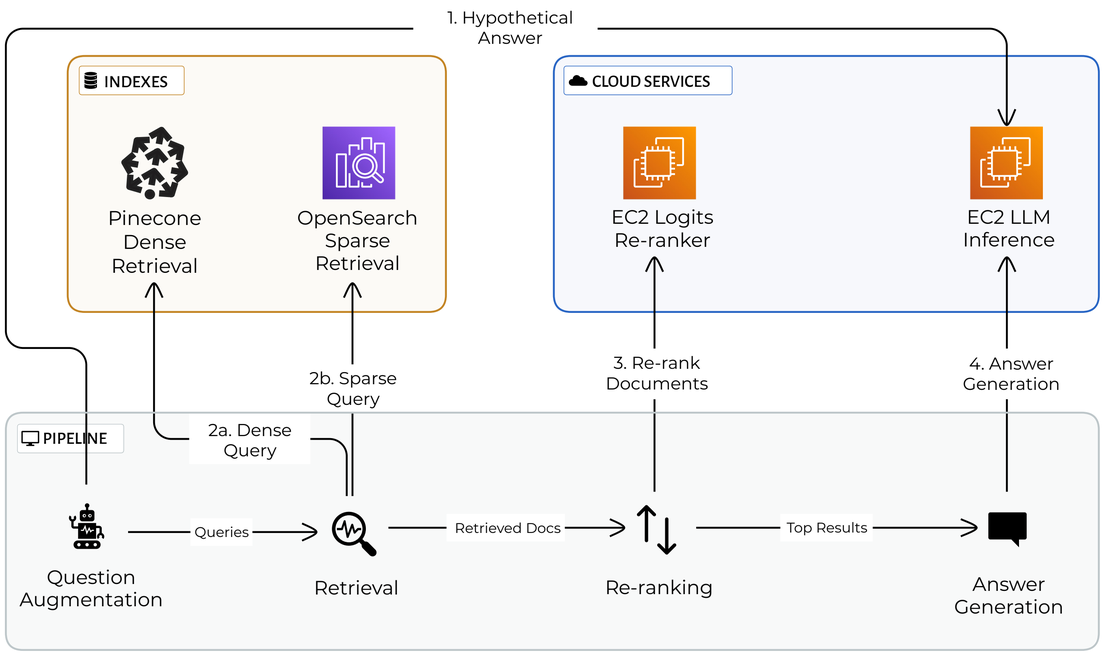
### LLM-as-a-Judge

- Inline with the challenge evaluation, we used **Claude Sonnet-3.5**.
- To assess **Correctness** (Relevance), reference answers from DataMorgana were used.

### Automated Evaluation Pipeline

- **Grid of Points (GoP)** for obtaining *champion* system configuration.
- **ANOVA** for directing the further *system optimization*.

## Pipeline and Services




- Parallelization: Parallel point-wise re-ranker for faster retrieval.
- Modular architecture: Each component is configurable and replaceable.
- Portable Components: An LLM re-ranker based on FastAPI and Transformers, and a vLLM-based LLM inference service deployed on an EC2 instance.
- All scripts and code are available.

## Results Analysis

### N-Way ANOVA Result

| Factor | SumSq | DF | MeanSq | F | PR(>F) | Partial $\omega^2$ |
|---|---|---|---|---|---|---|
| Question Augmentation | 0.319 | 2 | 0.160 | 125.692 | **<0.001*** | 0.757 |
| Query Variants Generation Prompt : Retrieval | 0.006 | 1 | 0.006 | 4.910 | **0.030*** | 0.047 |
| Query Variants Generation Prompt | 0.003 | 1 | 0.003 | 2.389 | 0.126 | 0.017 |
| Context Words Length Limits | 0.003 | 1 | 0.003 | 2.005 | 0.161 | 0.013 |
| Retrieval | 0.002 | 1 | 0.002 | 1.863 | 0.176 | 0.011 |
| Number of Documents Retrieved : Retrieval | 0.002 | 1 | 0.002 | 1.610 | 0.208 | 0.008 |
| Answer Generation Prompt | 0.001 | 1 | 0.001 | 1.106 | 0.296 | 0.001 |
| Context Words Len. Limits : Question Augmentation | 0.001 | 2 | 0.000 | 0.326 | 0.723 | -0.017 |
| Answer Gen. Prompt : Question Augmentation | 0.000 | 2 | 0.000 | 0.157 | 0.855 | -0.022 |
| Context Words Len. Limits : Answer Gen. Prompt | 0.000 | 1 | 0.000 | 0.049 | 0.825 | -0.012 |
| Query Var. Gen. Prompt : Num. of Docs. Retrieved | 0.000 | 1 | 0.000 | 0.023 | 0.881 | -0.013 |
| Number of Documents Retrieved | 0.000 | 1 | 0.000 | 0.012 | 0.914 | -0.013 |
| Residual | 0.098 | 77 | 0.001 | – | – | – |

* indicates statistically significant differences

- The 'Question Augmentation' component **significantly** impacts system performance.
- 'Query Variants Gen. Prompt' and 'Retrieval' jointly affect performance, but individually, neither has a significant effect.
- *No augmentation* is included in the *Champion* configuration based on GoP.
- Optimizing the system focused on improving mechanisms related to 'Question Augmentation.'

### HyDE Implementation

| RAG System | Correctness (Relevance) | Faithfulness |
|---|---|---|
| G-RAG (Champion + HyDE) | **8** | 12 |
| GoP Champion | 7 | **14** |
| Ties | 85 | 74 |

## Key Takeaways

- Test driven design
  - Fit for purpose: alignment with the goal
  - Statistical analysis through all stages
  - Enabled to prioritize parameter optimization
- Diverse team with complementary skills

Future work:
- Dynamic question augmentation
- Distillation and fine-tuning of smaller LLMs

This research was conducted by the ARC Centre of Excellence for Automated Decision-Making and Society (ADM+S, CE200100005), and funded fully by the Australian Government through the Australian Research Council and was undertaken with the assistance of computing resources from RACE (RMIT AWS Cloud Supercomputing). This work was conducted on the unceded lands of the Woi wurrung and Boon wurrung language groups of the eastern Kulin Nation. We pay our respect to Ancestors and Elders, past and present, and extend that respect to all Aboriginal and Torres Strait Islander peoples today and their connections to land, sea, sky, and community.

RMIT UNIVERSITY RACE Hub