

Testing Approach

Key Components

- Evaluation Tool: LangSmith
- Test Dataset Generation: CrossLLM
- Evaluation Metrics: Various metrics provided by LangSmith

Implementation Details

1. Test Dataset Generation

- Utilizing CrossLLM to create a diverse and representative test dataset for RAG evaluation.

2. Evaluation Process

- Employing LangSmith to run comprehensive evaluations on the RAG system.
- Validating the system's performance using multiple metrics provided by LangSmith.

3. Key Metrics Analyzed

LangSmith provides a range of metrics for evaluation. Some of the key metrics observed include:

a) LCEL context-f3306126

- Score: 0.40
- Runtime: 0.94s
- Execution time: 0.96s

b) doc-relevance-31106ccd

- Score: 0.56
- Runtime: 0.92s
- Execution time: 2.66s

c) hallucination-9deeeb79

- Score: 0.82
- Runtime: 0.95s
- Execution time: 0.96s

These metrics provide insights into various aspects of the RAG system's performance, including context relevance, document relevance, and the presence of hallucinations in generated responses.

Conclusion

By combining CrossLLM for test data generation and LangSmith for comprehensive evaluation, this implementation provides a robust framework for assessing and improving RAG system performance across multiple dimensions.