

# HybridRAG Evaluation Report

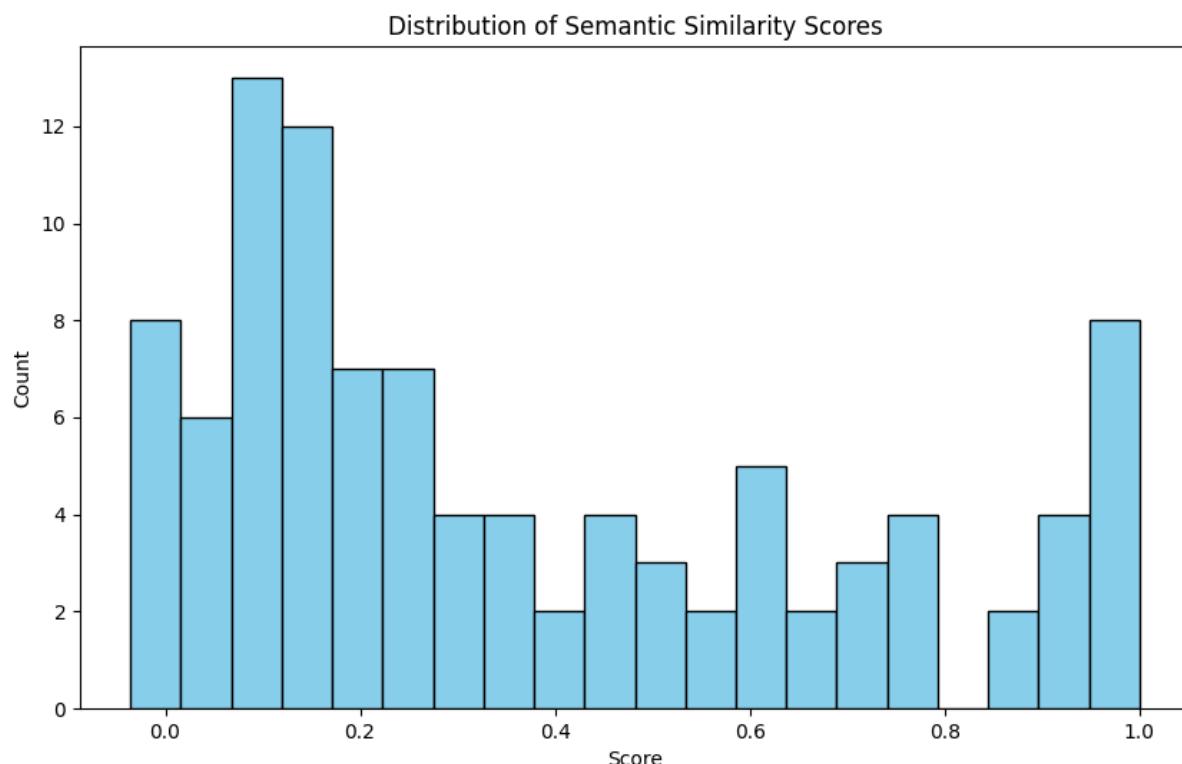
Generated automatically by the Evaluation Pipeline.

## Executive Summary

Metric	Score
Average MRR (Hybrid)	0.5689
Average Semantic Similarity	0.3703
Average BLEU Score	0.0817
Average Faithfulness (NLI)	0.5800
Average LLM Judge Score (1-5)	2.74
Average Response Time	0.87 s

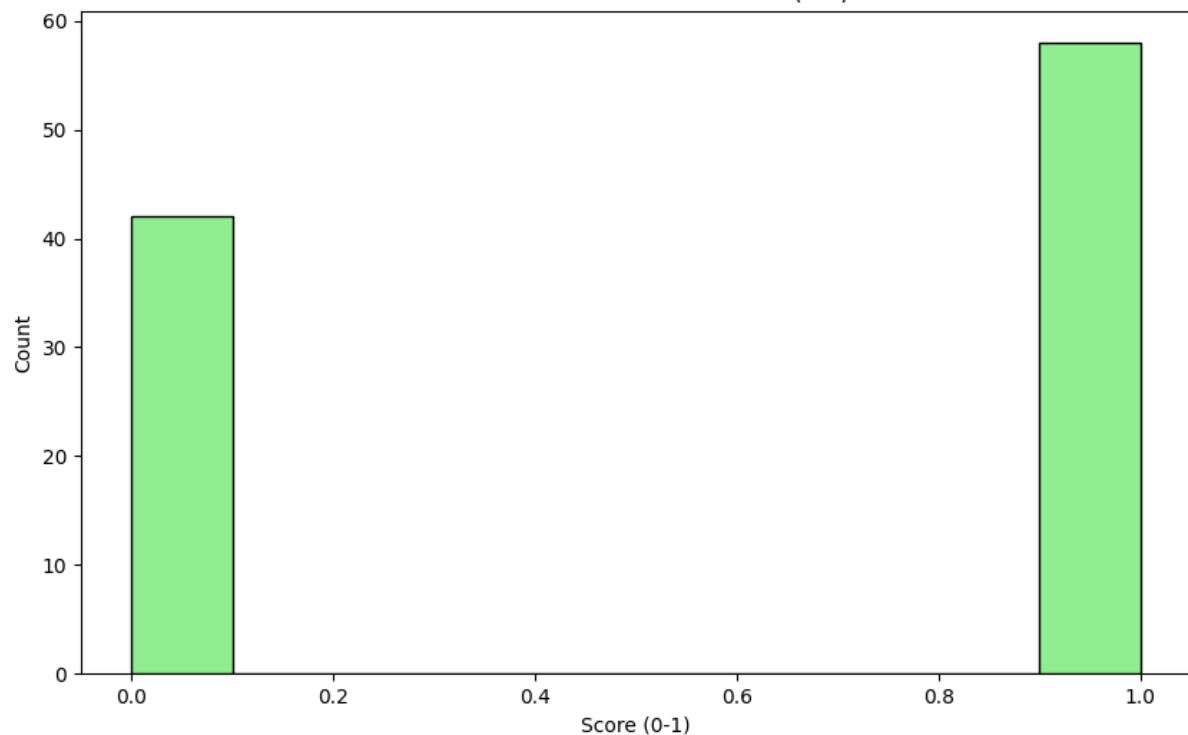
## Visualizations

### Metric Distributions



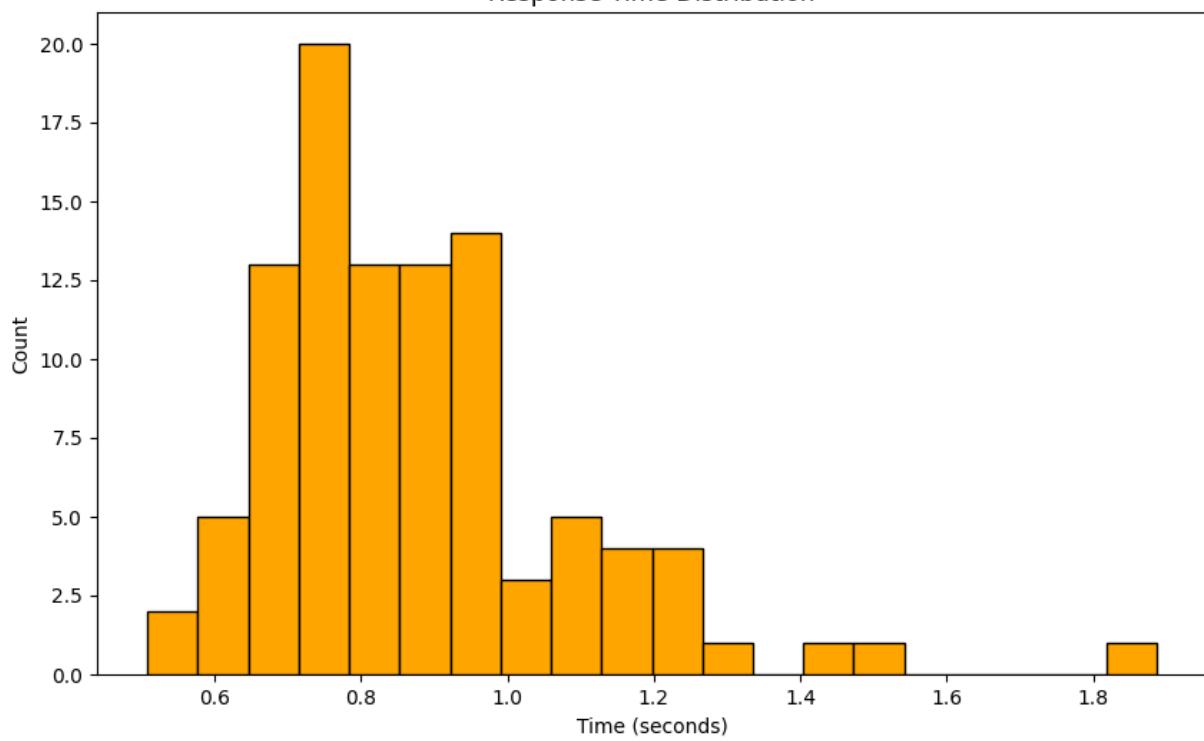
### Faithfulness Distribution

Faithfulness Score Distribution (NLI)

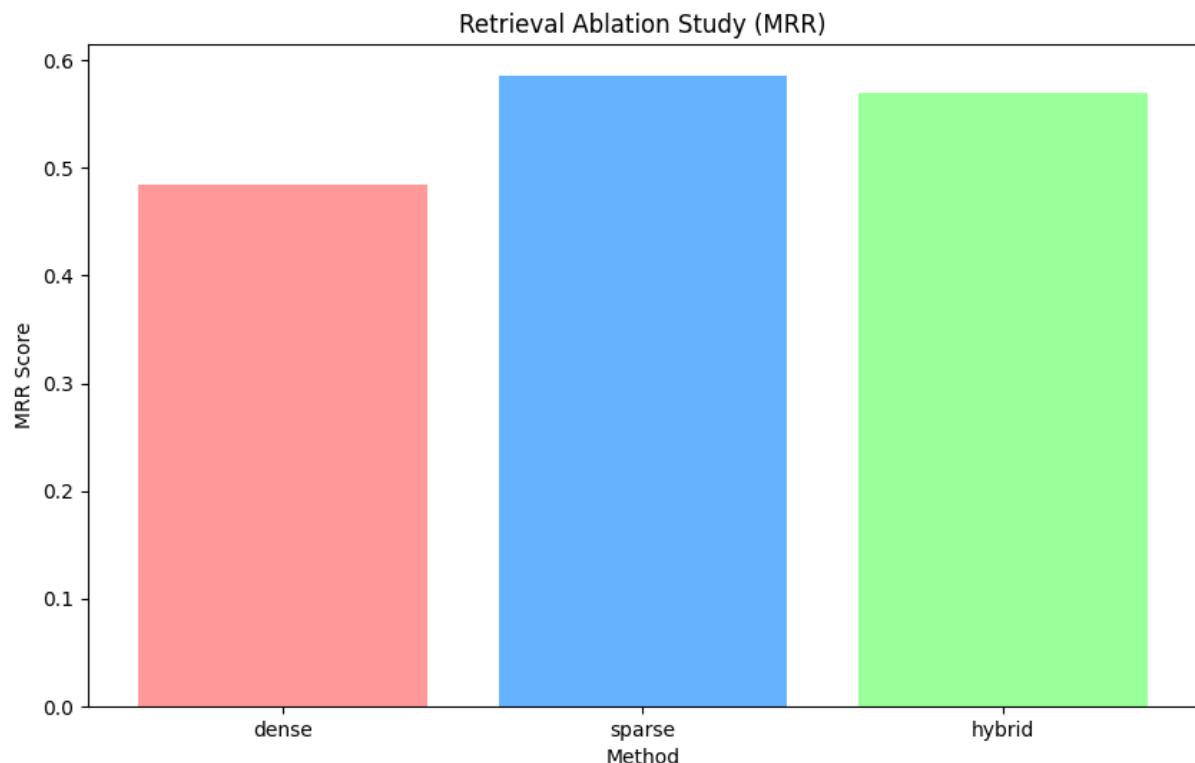


Response Time

Response Time Distribution



Retrieval Ablation



#### Ablation Study Results

Method	MRR	Hit Rate
Dense	0.4847	0.6700
Sparse	0.5855	0.7000
Hybrid	0.5689	0.7500

#### Error Analysis (Bottom 5 Semantic Scores)

ID	Question	Semantic Score	Issue
94	What is the difference between Kam and Davida?	-0.0357	Low Similarity
34	What is the difference between SOA and SOA Collaboration?	-0.0304	Low Similarity
79	What is the name of the book that was published in Switzerland?	-0.0150	Low Similarity
45	What is the best way to delete data?	-0.0044	Low Similarity

21	What is the function of $m$ and $n$ ?	0.0056	Low Similarity
----	---------------------------------------	--------	----------------