

HybridRAG Evaluation Report

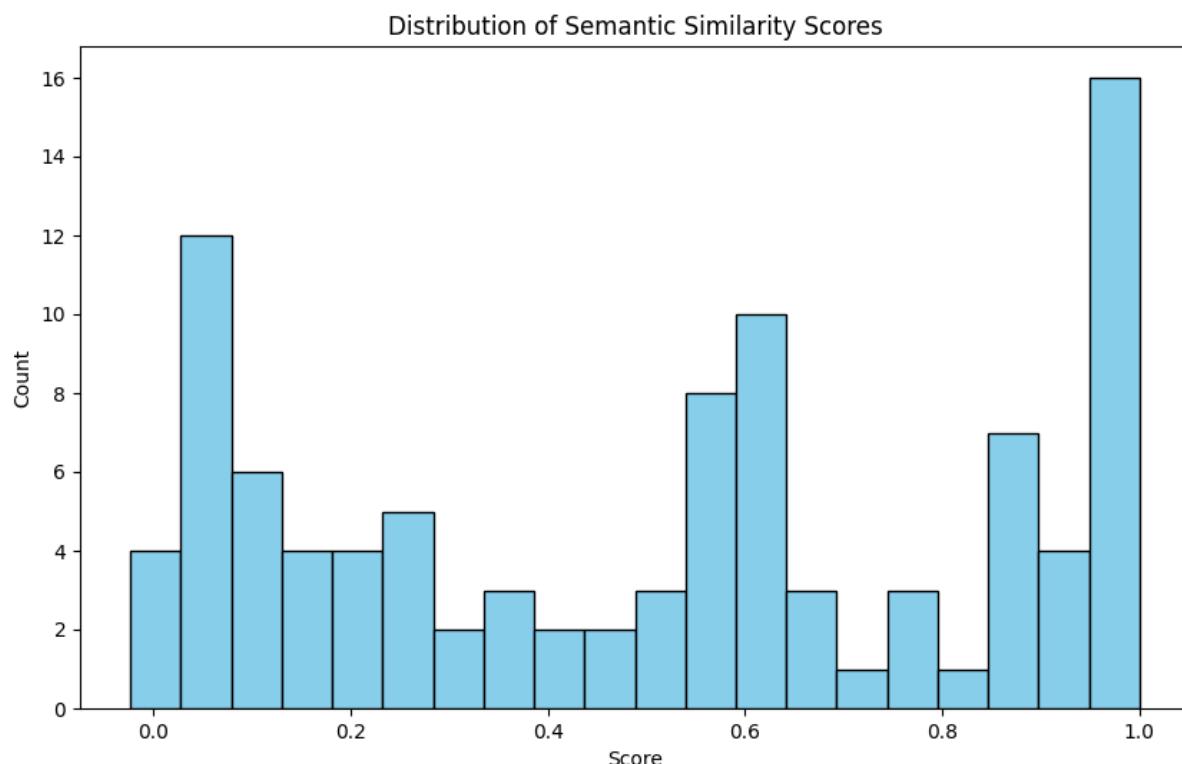
Generated automatically by the Evaluation Pipeline.

Executive Summary

Metric	Score
Average MRR (Hybrid)	0.8382
Average Semantic Similarity	0.5083
Average BLEU Score	0.1209
Average Faithfulness (NLI)	0.6800
Average LLM Judge Score (1-5)	2.88
Average Response Time	1.85 s

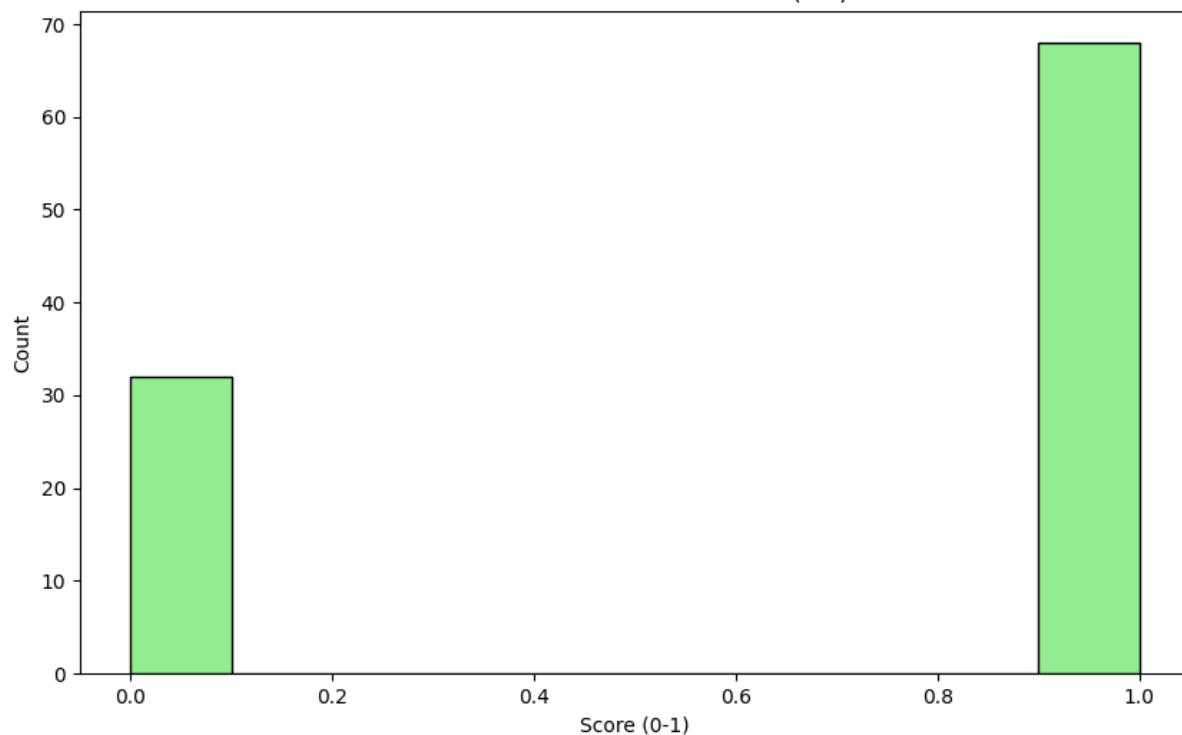
Visualizations

Metric Distributions

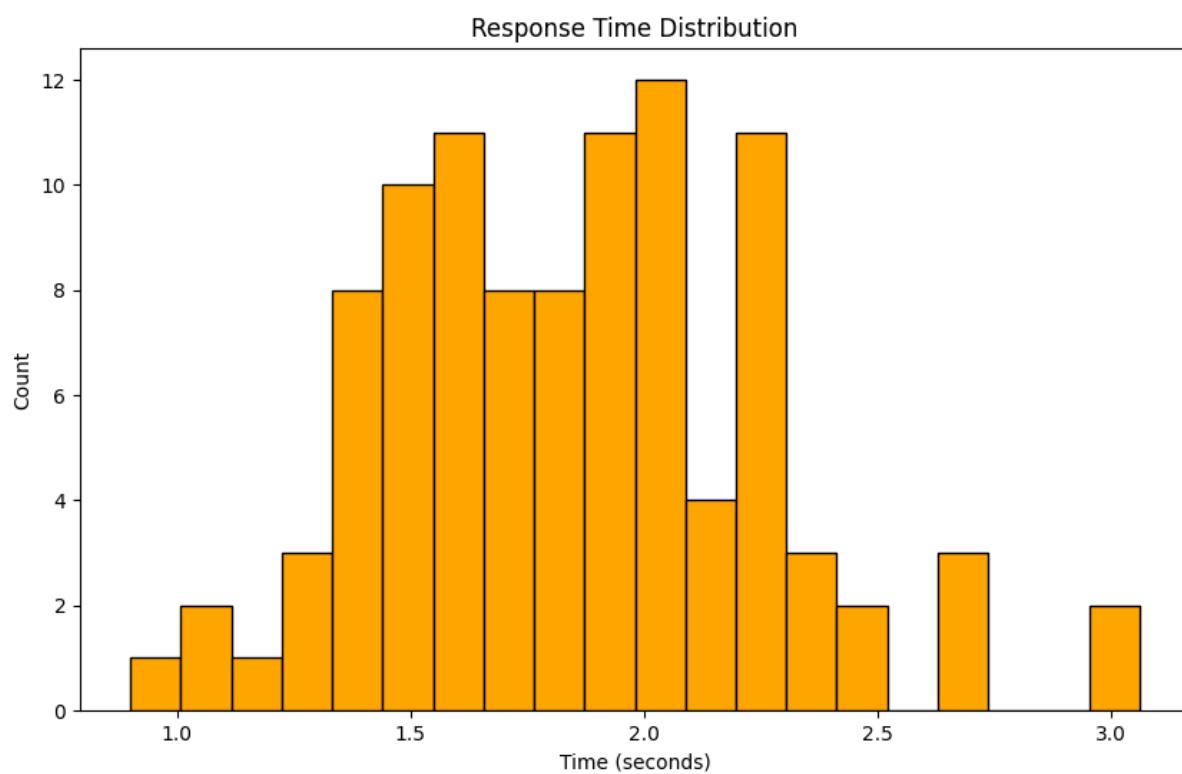


Faithfulness Distribution

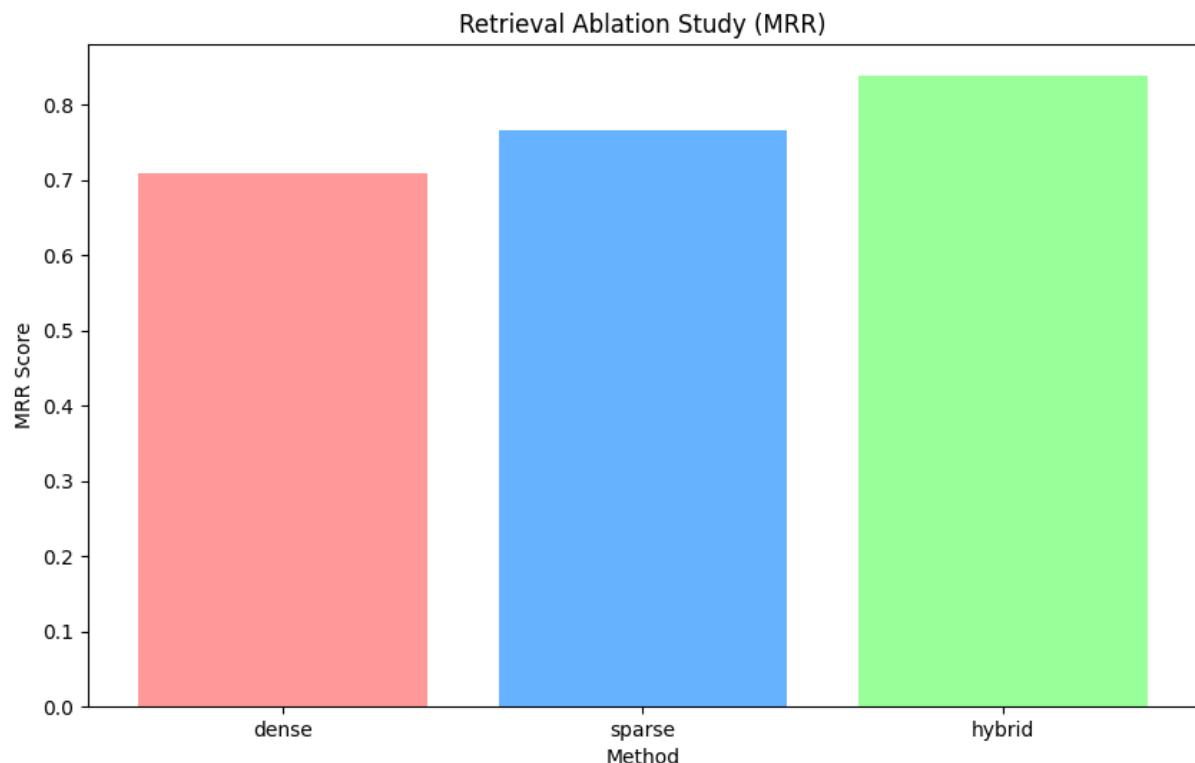
Faithfulness Score Distribution (NLI)



Response Time



Retrieval Ablation



Ablation Study Results

Method	MRR	Hit Rate
Dense	0.7094	0.8300
Sparse	0.7666	0.8500
Hybrid	0.8382	0.9000

Error Analysis (Bottom 5 Semantic Scores)

ID	Question	Semantic Score	Issue
52	What is the name of the Byzantine agreement protocol?	-0.0234	Low Similarity
57	What is the most likely reason that an eavesdropper will not learn the exchanged key?	-0.0030	Low Similarity
43	What is the difference between the two words?	-0.0022	Low Similarity

2	What is the name of the conference that Bill Gates and Craig Mundie attended?	0.0114	Low Similarity
37	What was the name of the agreement that was signed by the United States and Britain?	0.0279	Low Similarity