

HybridRAG Evaluation Report

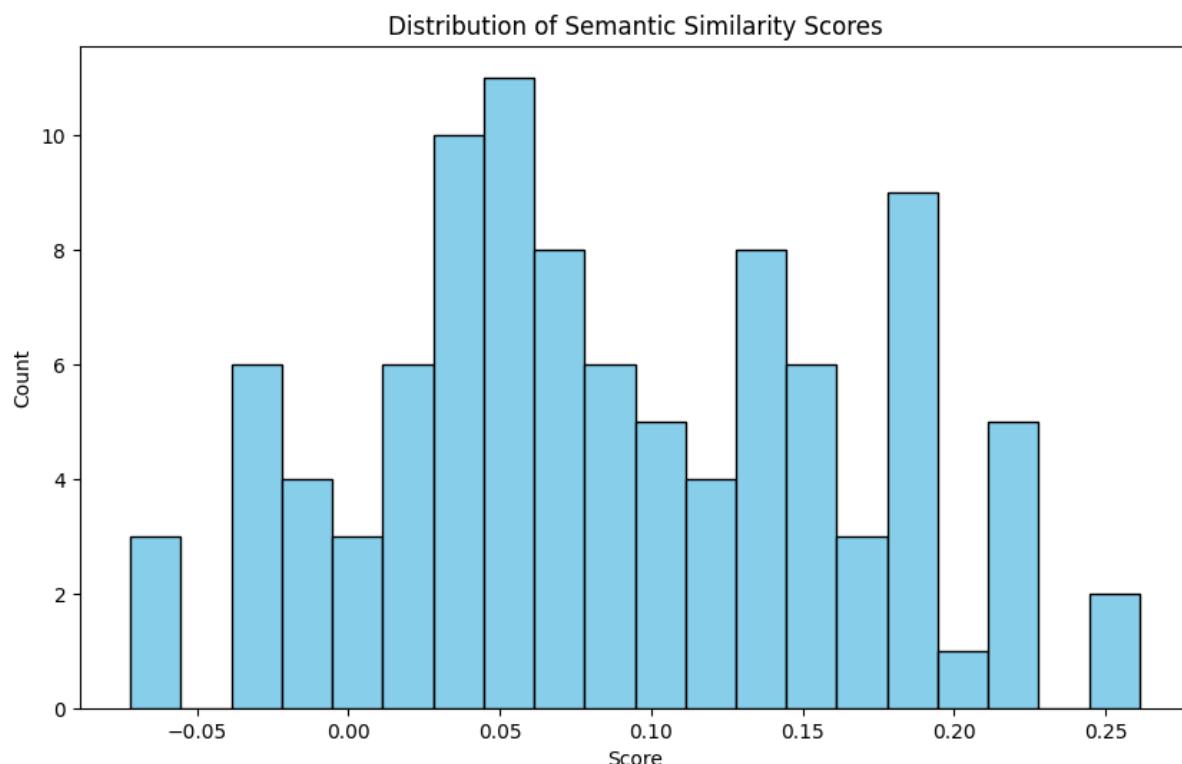
Generated automatically by the Evaluation Pipeline.

Executive Summary

Metric	Score
Average MRR (Hybrid)	0.2924
Average Semantic Similarity	0.0876
Average BLEU Score	0.0000
Average Faithfulness (NLI)	0.0000
Average LLM Judge Score (1-5)	2.74
Average Response Time	3.36 s

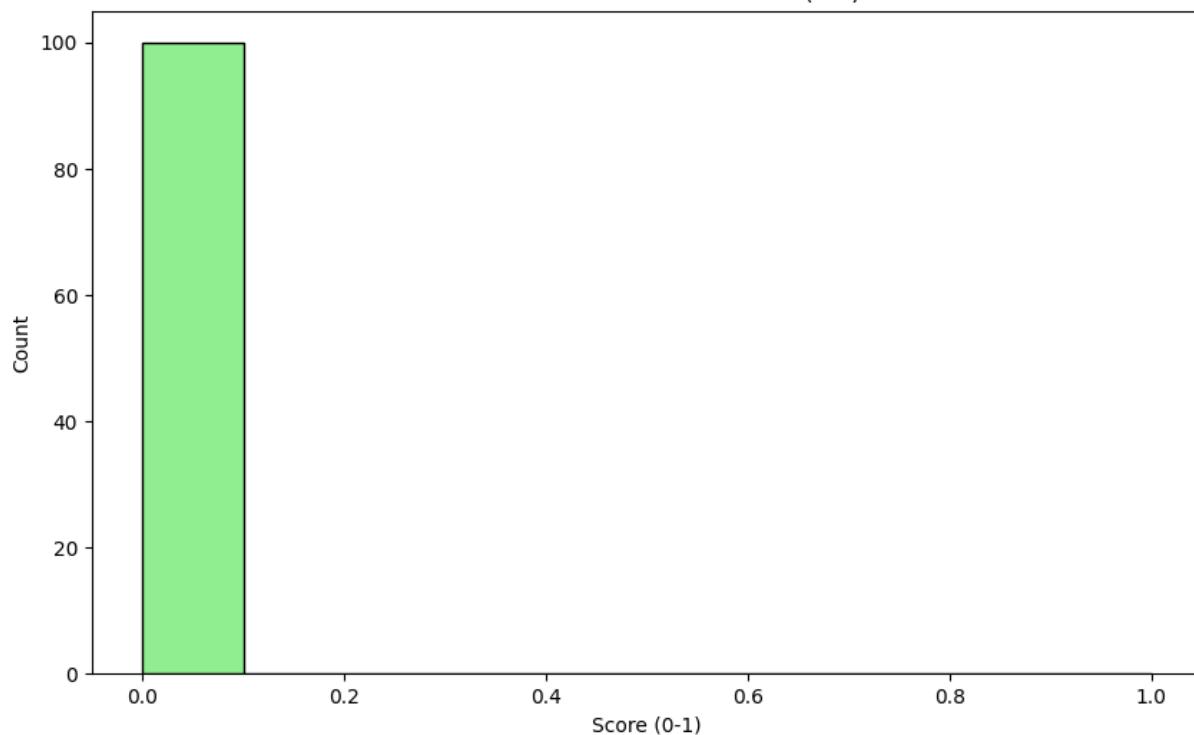
Visualizations

Metric Distributions



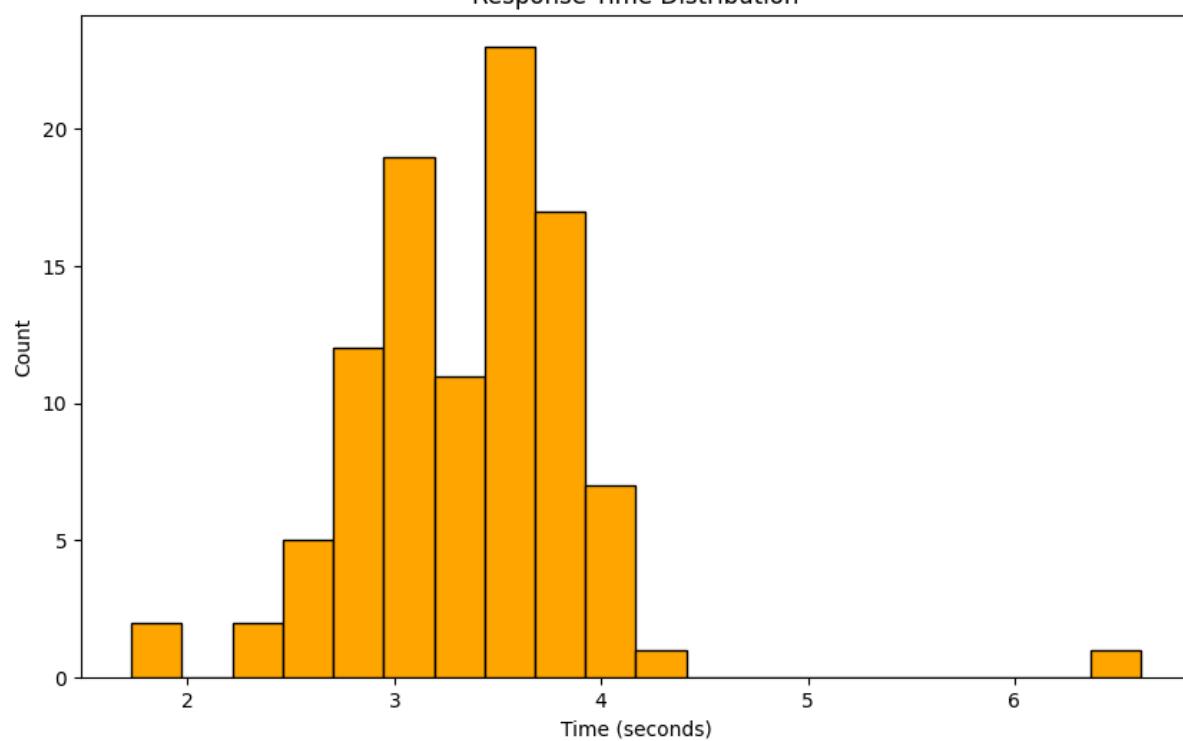
Faithfulness Distribution

Faithfulness Score Distribution (NLI)

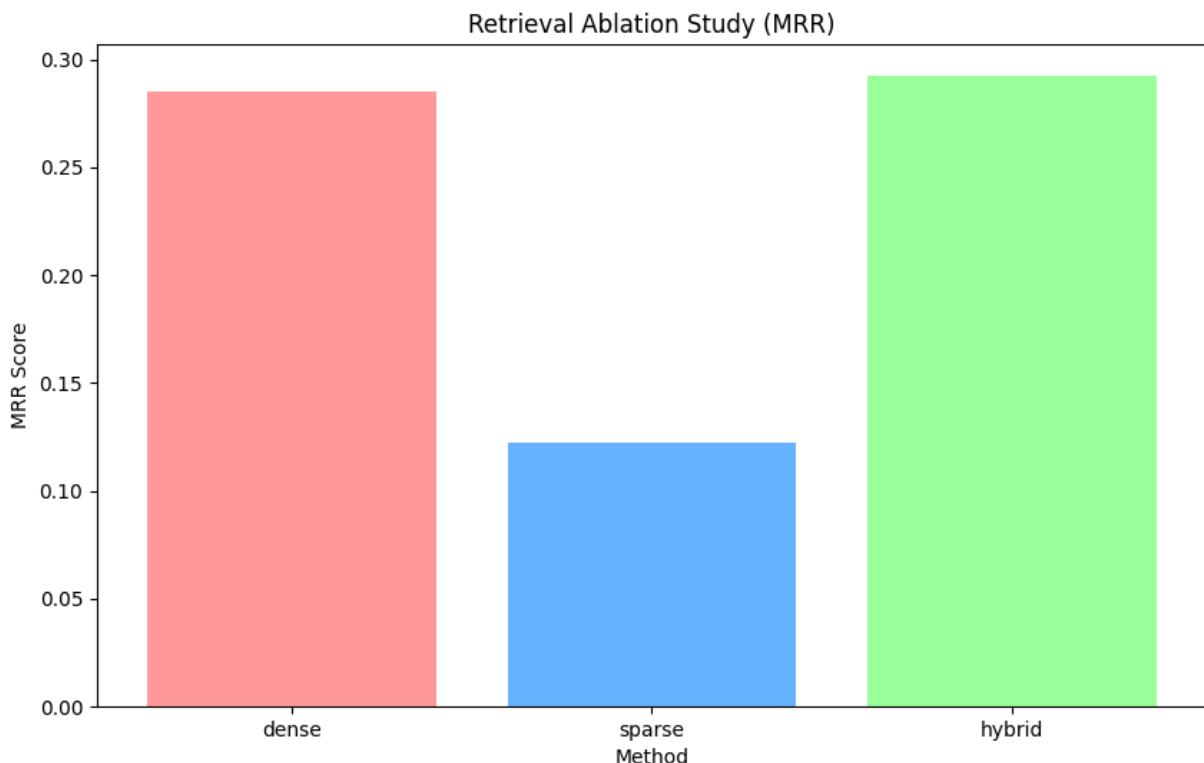


Response Time

Response Time Distribution



Retrieval Ablation



Ablation Study Results

Method	MRR	Hit Rate
Dense	0.2850	0.3500
Sparse	0.1220	0.2800
Hybrid	0.2924	0.4000

Error Analysis (Bottom 5 Semantic Scores)

ID	Question	Semantic Score	Issue
23	Was Romania less financially active by 1998/199 as Portugal was the fourth riches nation as listed as #1 in this category on Global Competitive Index. [[Cs3dod8A.1=(2b.3&hiD4ip9styIxyY	-0.0719	Low Similarity

72	How must people connect devices operating services on Oracle Compnant with Amazon servers to Oracle Opensystem deployment servers while building the desktop environments.? I was given two Linux Linux systems closfonasyanja2 for Oracle QFS. It required these environments that include which opensource container and Oracle Had	-0.0694	Low Similarity
92	Before being introduced along Route 81 the IPA and SMBs, would people choose other locations they did because both uses worked very badly or they would pay little with regard these things or have it worked perfectly every first reerg and on future missions to different geographic continenta on line 22. what about	-0.0563	Low Similarity
53	When took place Bhandranjagtakathandyunin 1994 release(first article title published only 5 hours ago (25 Nov 1995 in Times Books by Anelaih Safari) in "Nal tishaysh, Nachhethans yun, Ja	-0.0374	Low Similarity
100	The "Thiruvelkulatraanainva" refer	-0.0371	Low Similarity