HybridRAG Evaluation Report
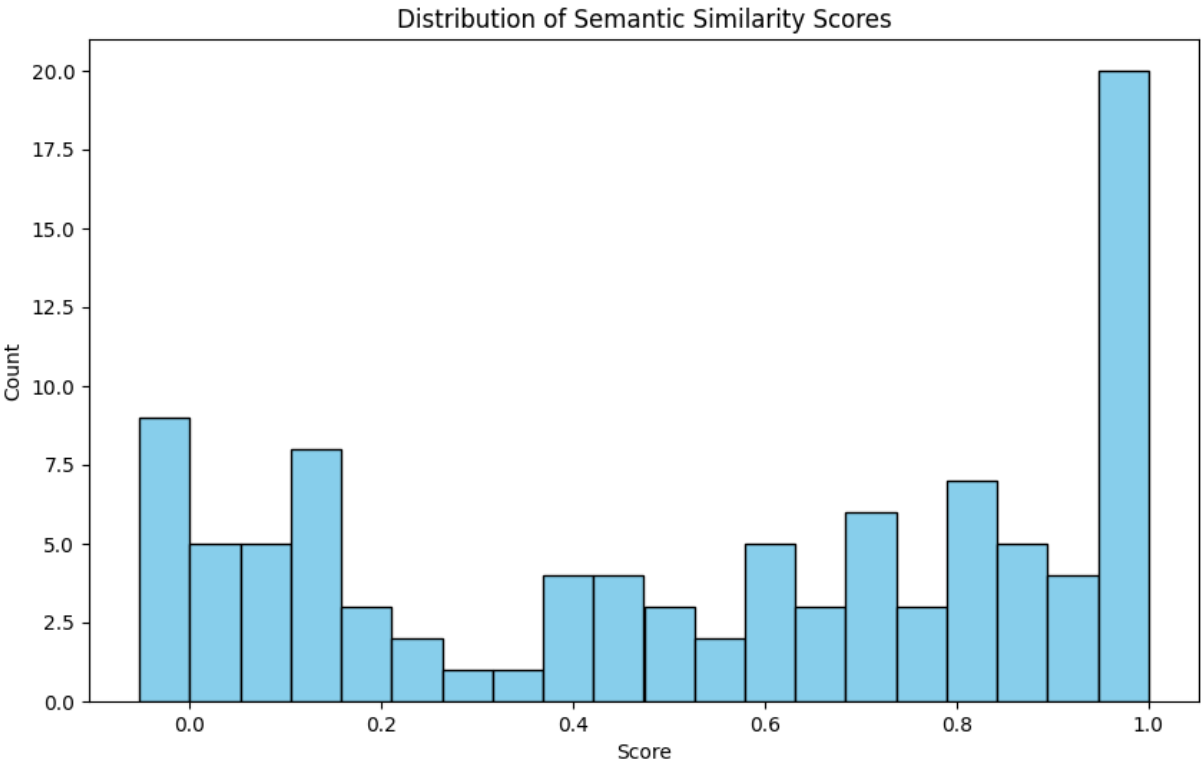
Generated automatically by the Evaluation Pipeline.

Executive Summary
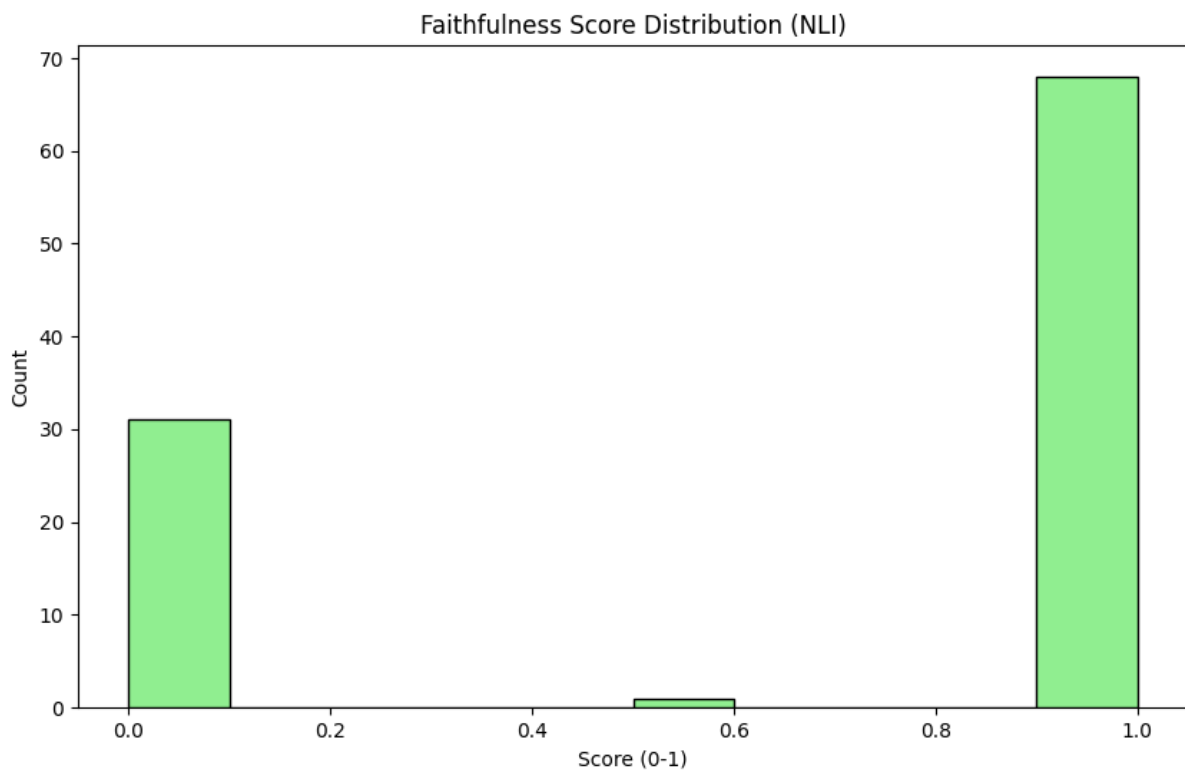
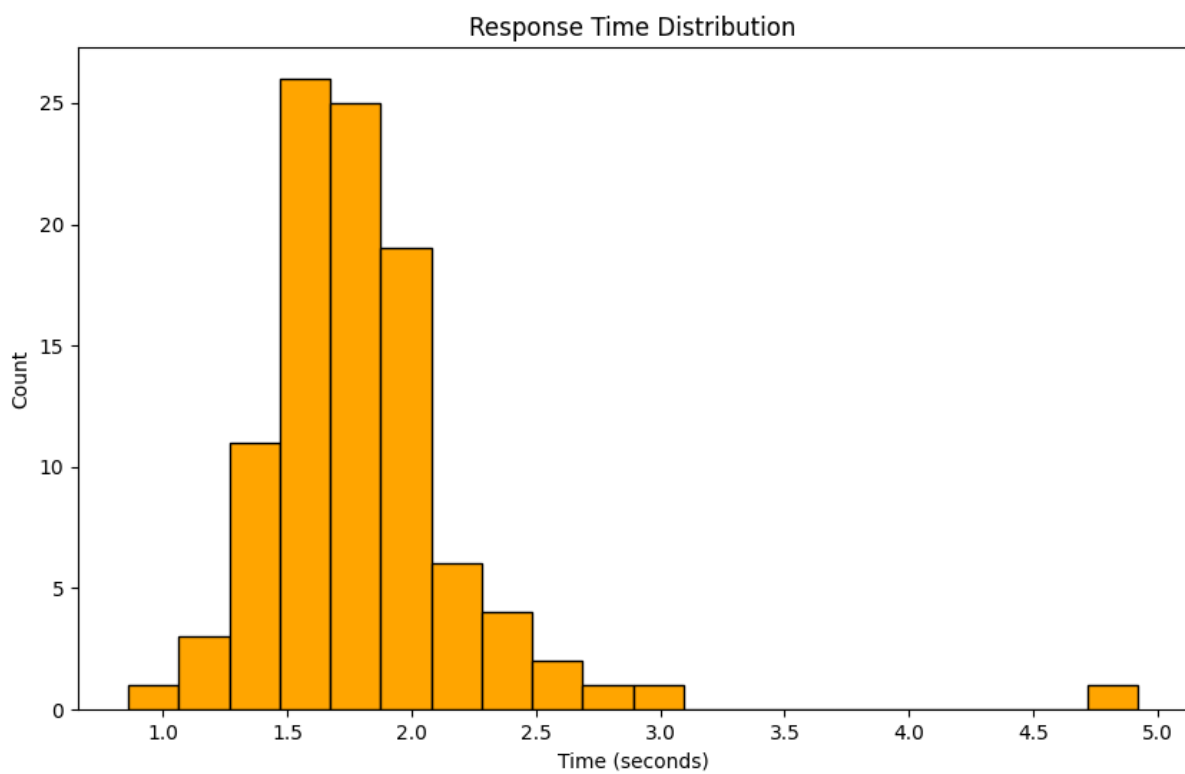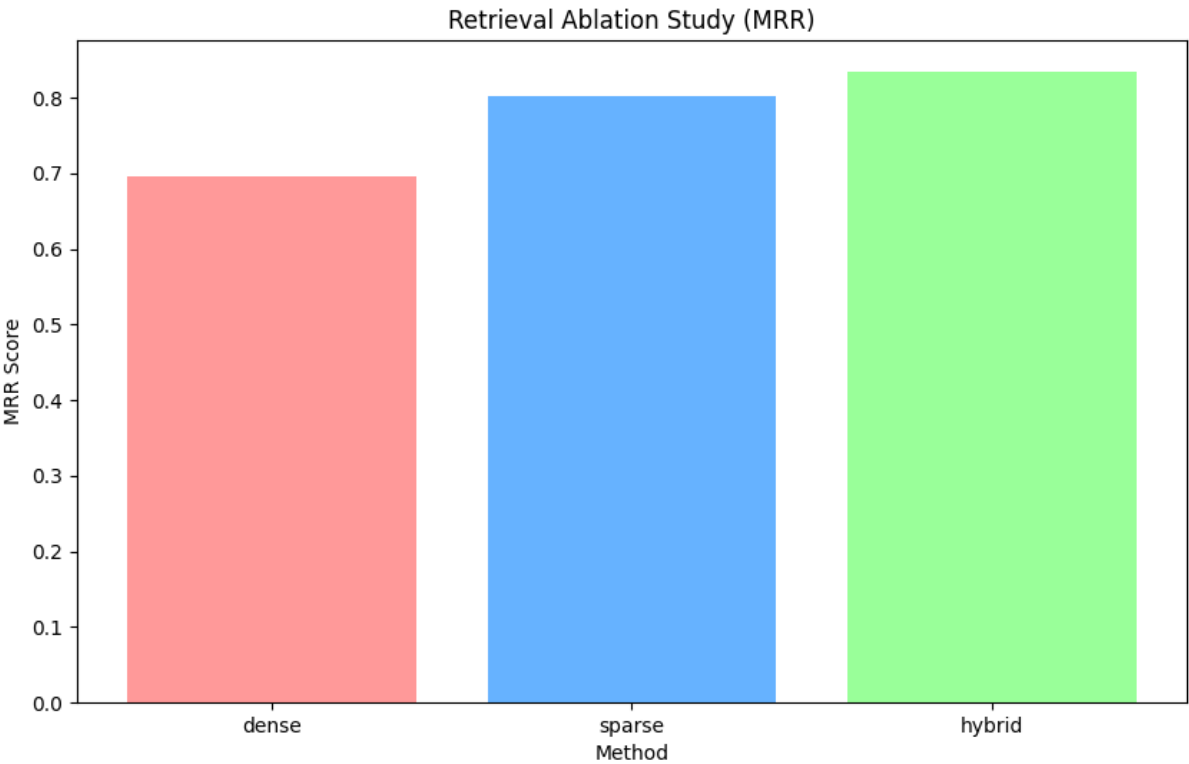| Metric	Score
Average MRR (Hybrid)	0.8342
Average Semantic Similarity	0.5398
Average BLEU Score	0.1770
Average Faithfulness (NLI)	0.6850
Average LLM Judge Score (1-5)	2.88
Average Response Time	1.81 s

Visualizations

Metric Distributions

Distribution of Semantic Similarity Scores

Faithfulness Distribution

Faithfulness Score Distribution (NLI)

Response Time

Response Time Distribution

Retrieval Ablation

Retrieval Ablation Study (MRR)

Ablation Study Results

Method	MRR	Hit Rate
Dense	0.6953	0.7800
Sparse	0.8014	0.8800
Hybrid	0.8342	0.8800

Error Analysis (Bottom 5 Semantic Scores)

ID	Question	Semantic Score	Issue
61	What is the main idea of the passage?	-0.0527	Low Similarity
87	What is the difference between the two encryptions?	-0.0454	Low Similarity
29	What is the purpose of the EU's data protection regulations?	-0.0395	Low Similarity
90	What is the difference between the two codes?	-0.0256	Low Similarity

| 24 | What is the difference between the two examples? | -0.0220 | Low Similarity |