

Statistical Benchmark Report

Generated: 2025-11-16 00:44

This report provides quantitative evaluation results for NDCG@k and Faithfulness, correlation diagnostics, outlier detection, and statistical visualizations.

1. Summary Statistics

files	10.0000
ndcg@k mean	0.9997
ndcg@k ci95 lo	0.9992
ndcg@k ci95 hi	1.0000
faith mean	0.4795
faith ci95 lo	0.3962
faith ci95 hi	0.5748
ndcg@k median	1.0000
faith median	0.4092
ndcg@k std	0.0006
faith std	0.1478
bootstrap iters	2000.0000
iqr k	1.5000
z thresh	3.0000
faith high thr	0.7000
faith mid thr	0.4000

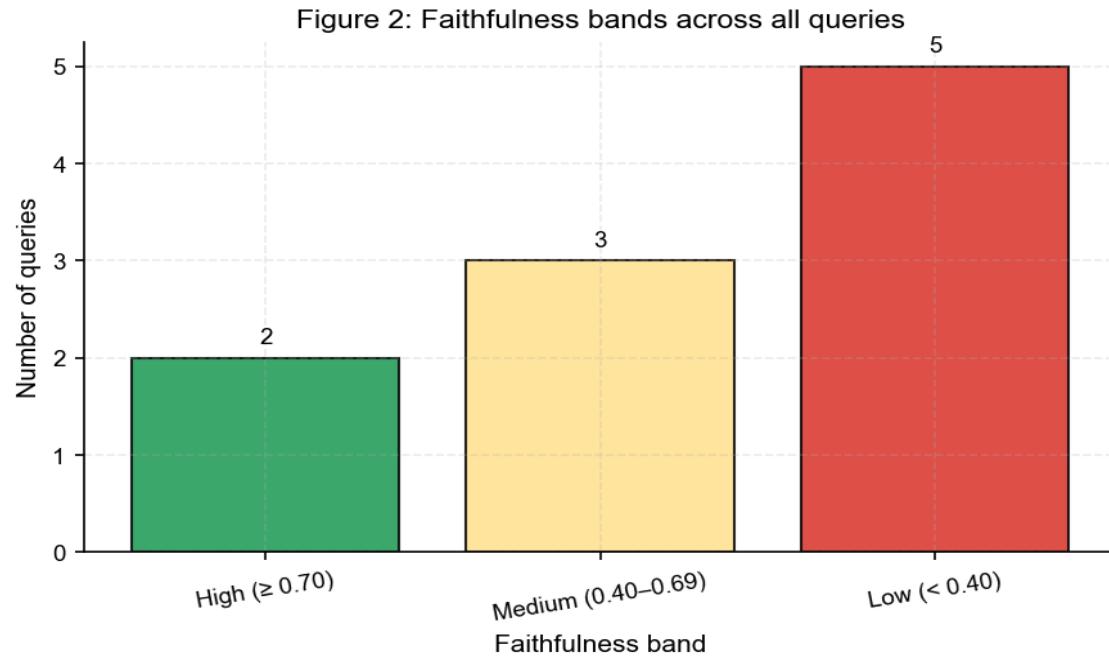
2. Correlation and Outlier Analysis

Metric Pair	Correlation	p-value
Pearson r	0.197	5.857e-01
Spearman ρ	0.138	7.030e-01

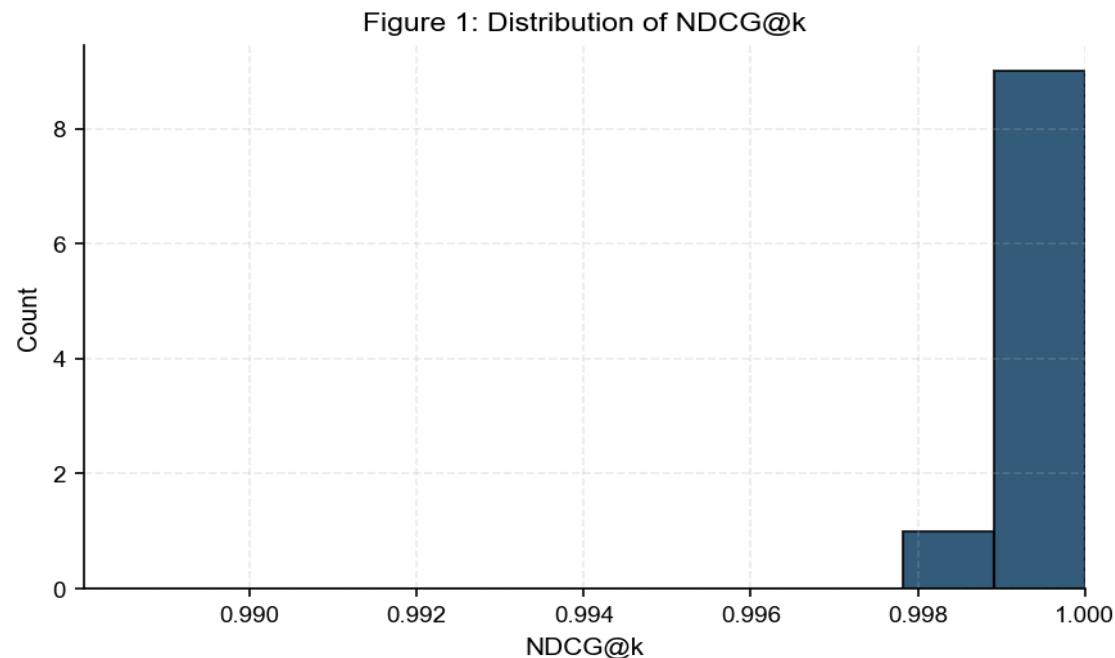
No outliers detected.

3. Visual Analytics

faithfulness bands global.png

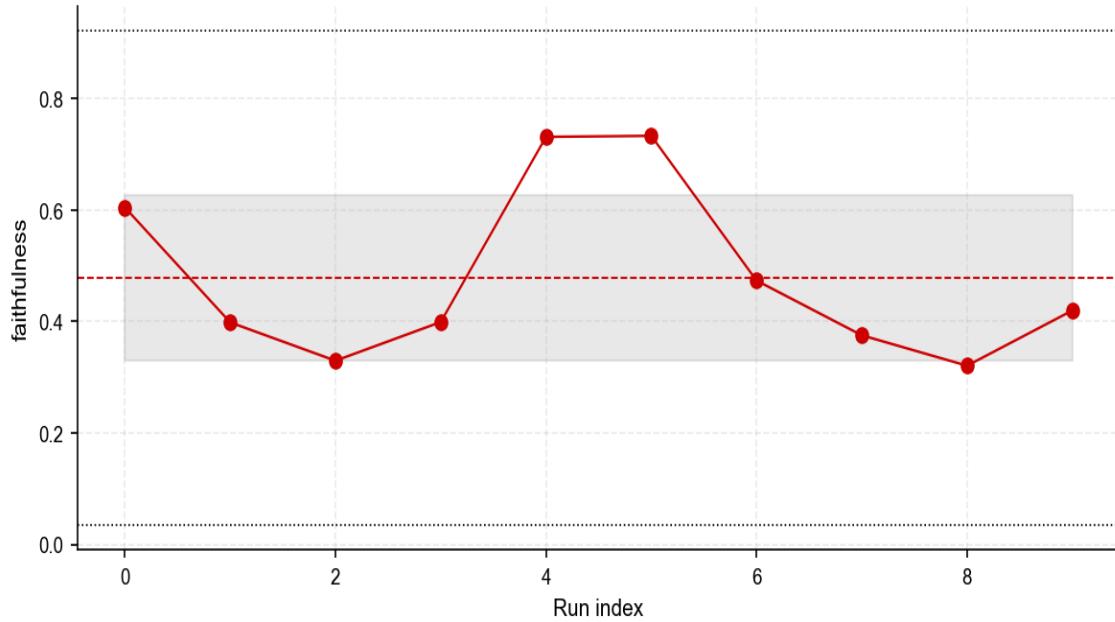


hist ndcg.png



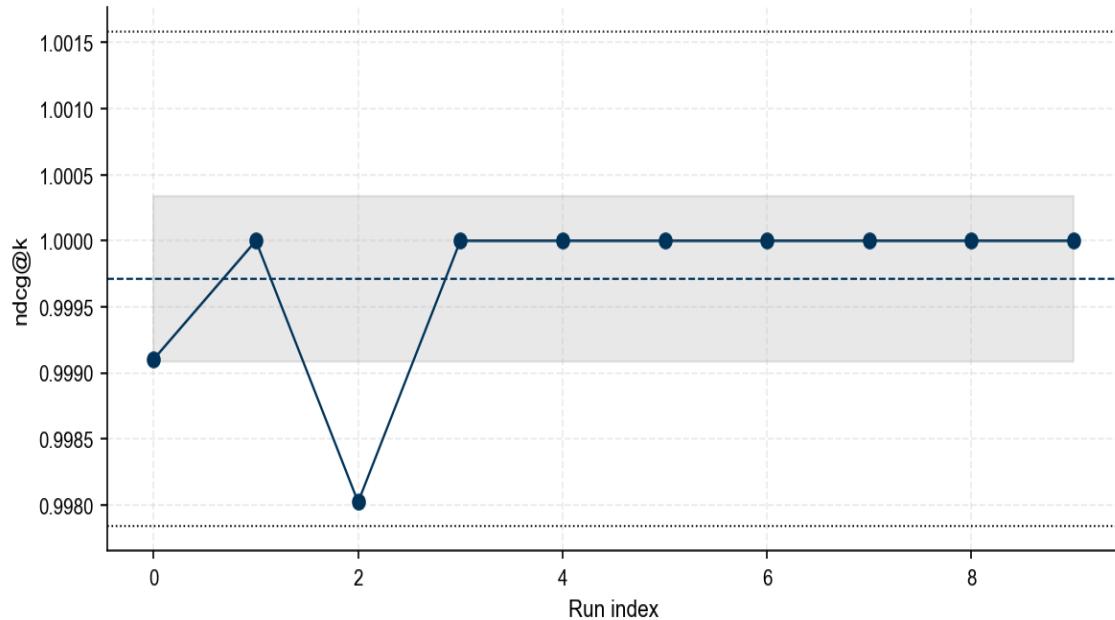
run order faithfulness.png

Figure 5: Run-order chart for faithfulness



run order ndcg@k.png

Figure 4: Run-order chart for ndcg@k



scatter ndcg vs faithfulness.png

Figure 3: Scatter NDCG@k vs Faithfulness ($r=0.197$, $p=0.138$)

