

Statistical Benchmark Report

Generated: 2025-11-15 23:16

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	0.4923
faith ci95 lo	0.3763
faith ci95 hi	0.6309
ndcg@k median	1.0000
faith median	0.4285
ndcg@k std	0.0006
faith std	0.2045
bootstrap iters	2000.0000
iqr k	1.5000
z thresh	3.0000
faith high thr	0.8000
faith mid thr	0.5000

2. Correlation and Outlier Analysis

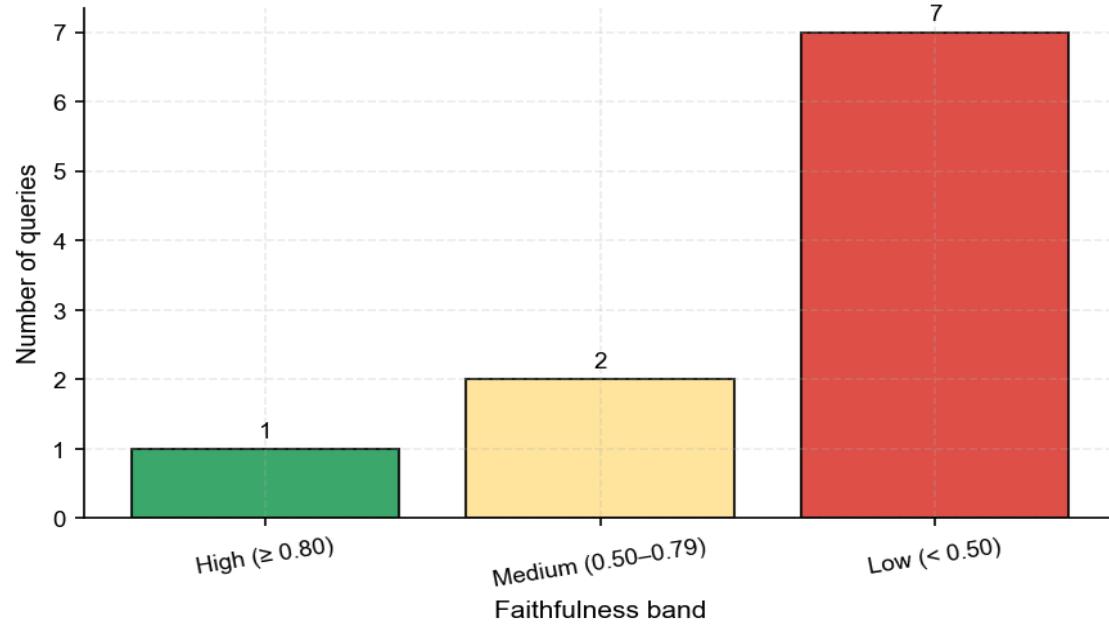
Metric Pair	Correlation	p-value
Pearson r	-0.143	6.942e-01
Spearman ρ	-0.415	2.328e-01

No outliers detected.

3. Visual Analytics

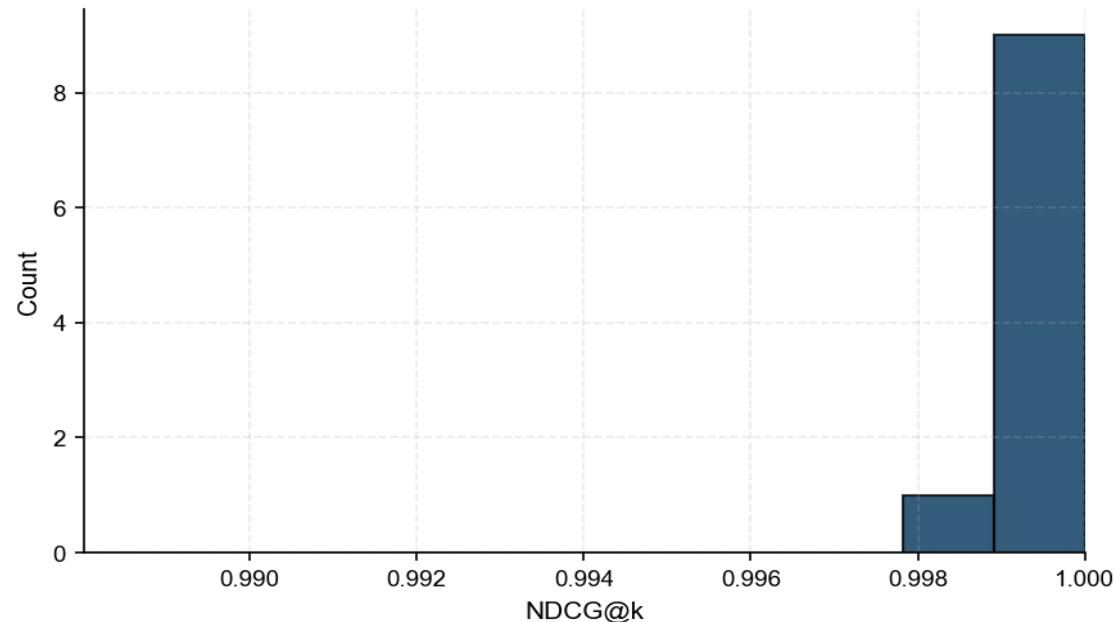
faithfulness bands global.png

Figure 2: Faithfulness bands across all queries



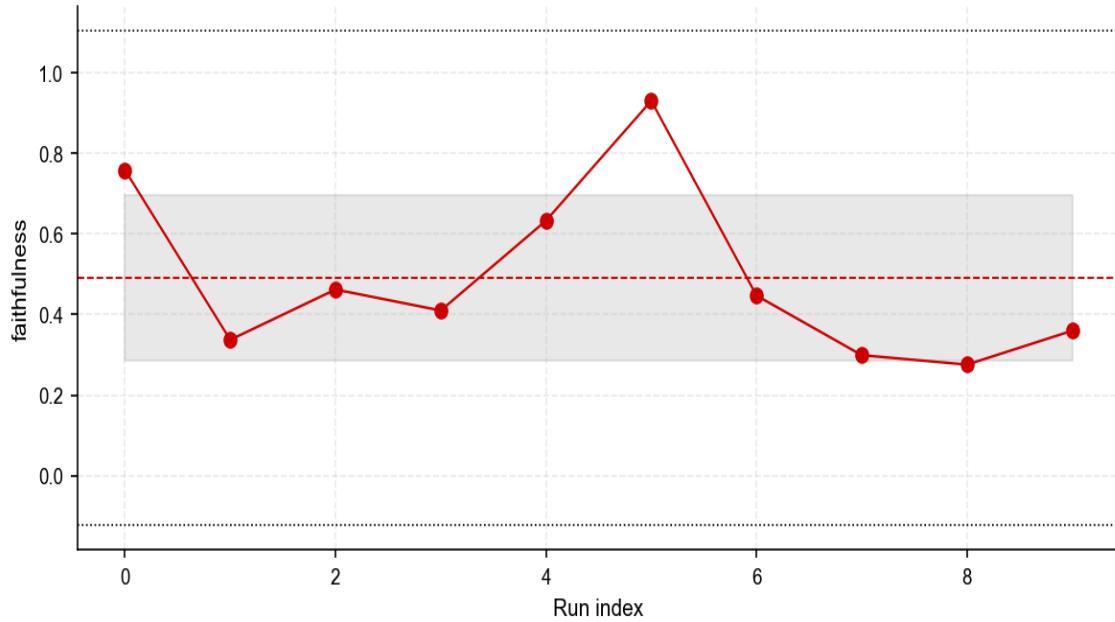
hist ndcg.png

Figure 1: Distribution of NDCG@k



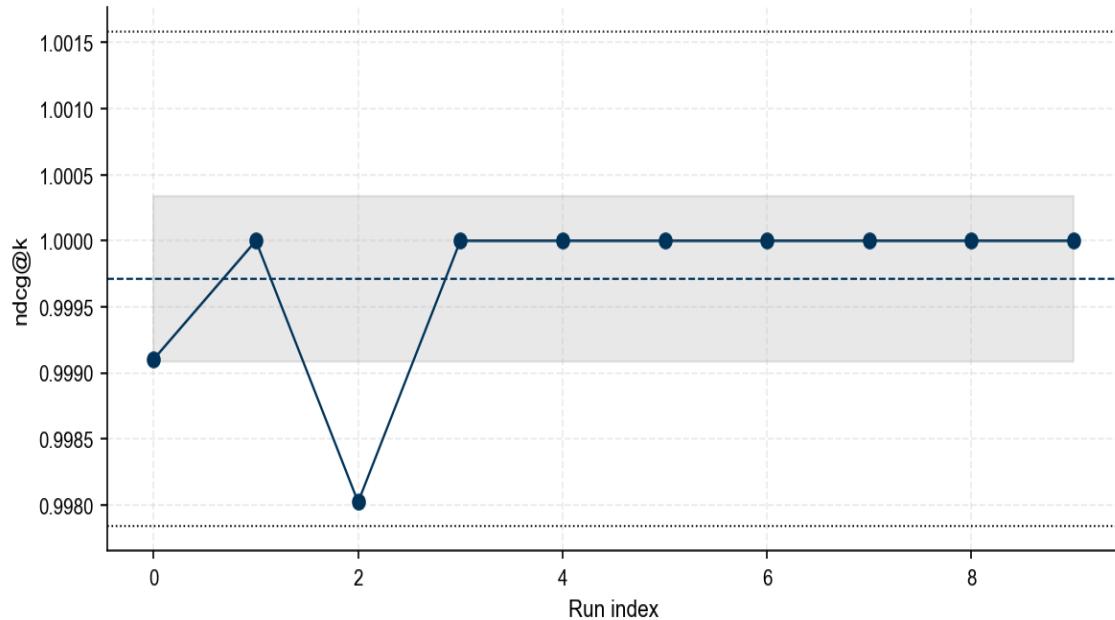
run order faithfulness.png

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 ($r=-0.143$, $p=-0.415$)

