Statistical Benchmark Report

Generated: 2025-11-16 21:36

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

1. Summary Statistics

files	100.0000
ndcg@k mean	1.0000
ndcg@k ci95 lo	1.0000
ndcg@k ci95 hi	1.0000
faith mean	0.5845
faith ci95 lo	0.5457
faith ci95 hi	0.6194
ndcg@k median	1.0000
faith median	0.6286
ndcg@k std	0.0002
faith std	0.1874
bootstrap iters	2000.0000
iqr k	1.5000
z thresh	3.0000
faith high thr	0.5000
faith mid thr	0.2500

2. Correlation and Outlier Analysis

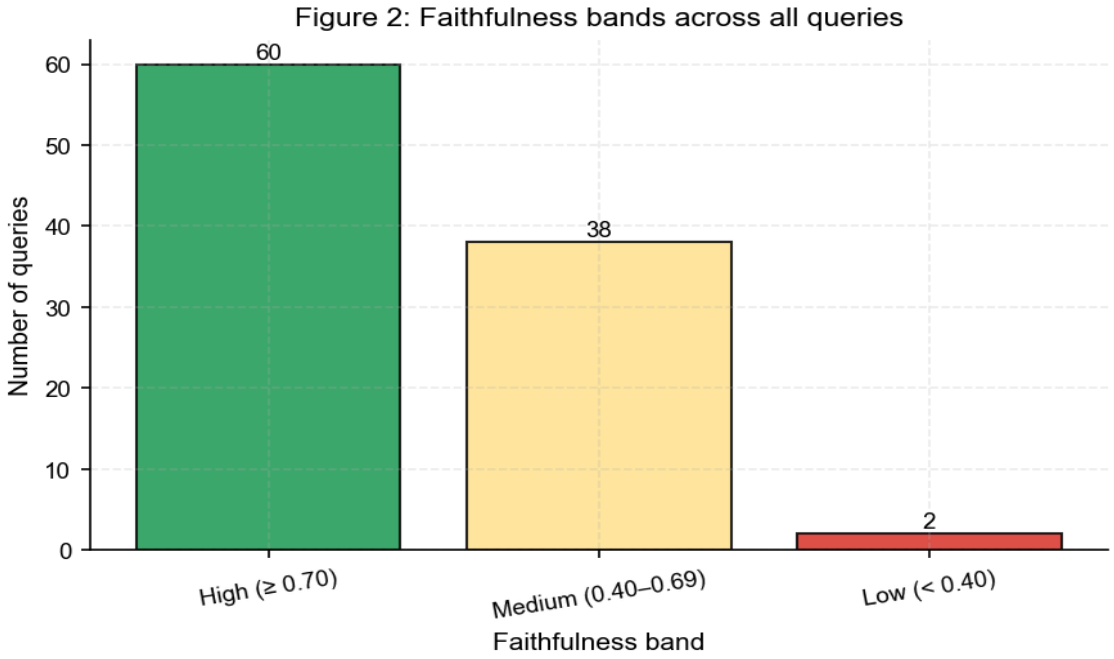
Metric Pair	Correlation	p-value
Pearson r	0.122	2.280e-01
Spearman ρ	0.131	1.954e-01

Detected Outliers

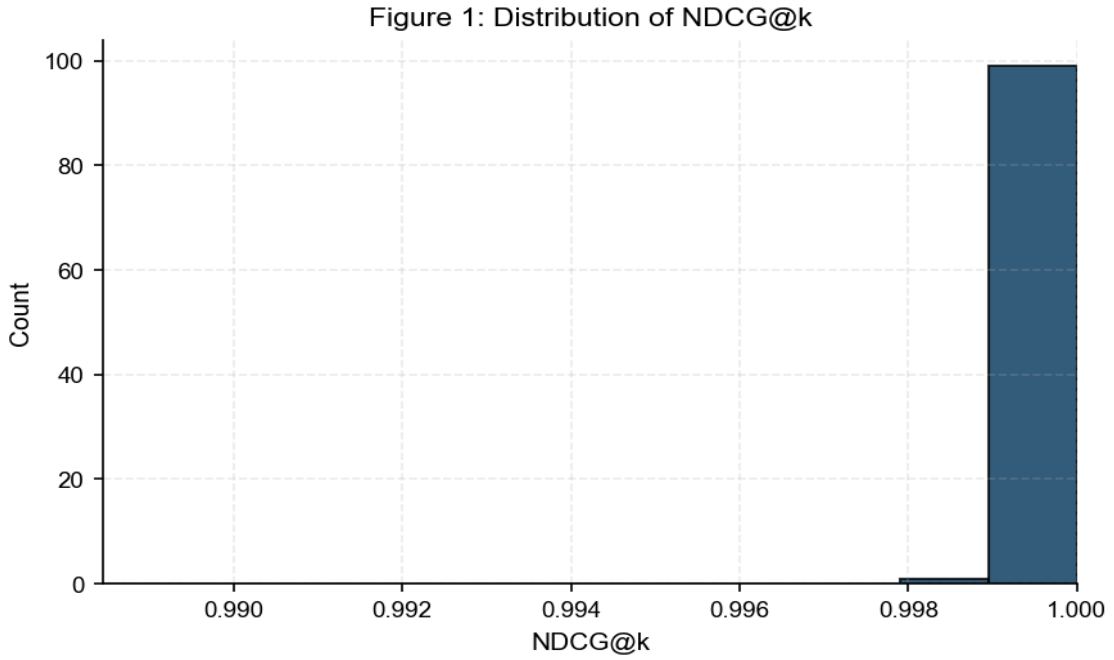
Query	NDCG@k	Faithfulness	z_ndcg	z_faith
Compare_and_contrast_the_main_theoretical_perspectives_on_summarize_the_emergenc	0.998	0.358	-9.84	-1.20

3. Visual Analytics

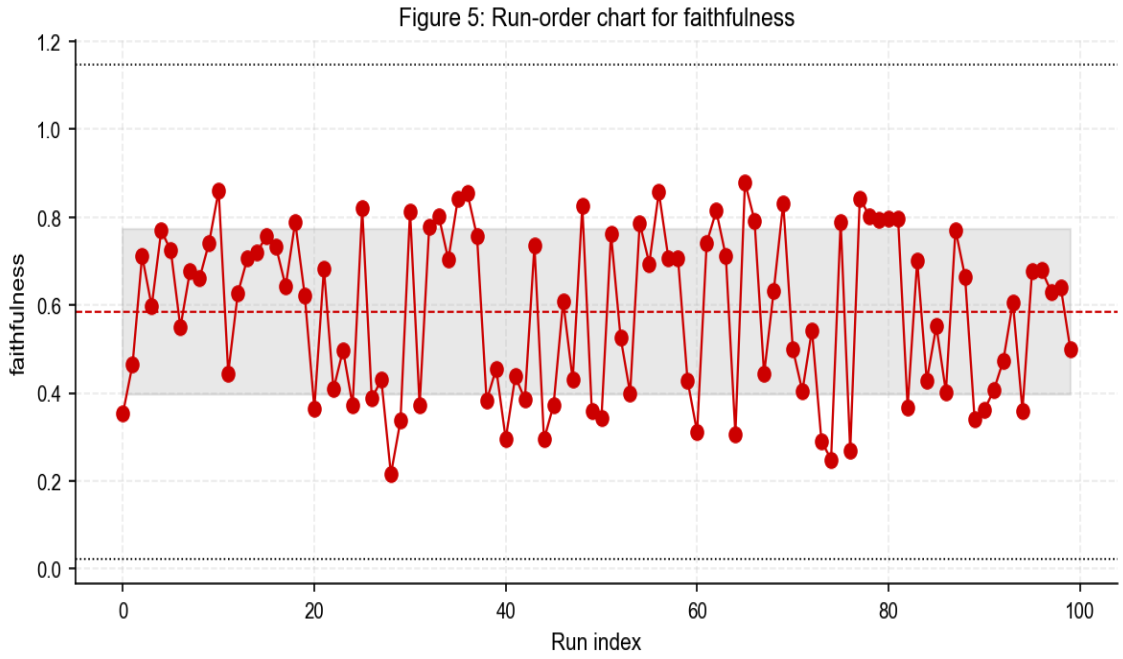
faithfulness bands global.png



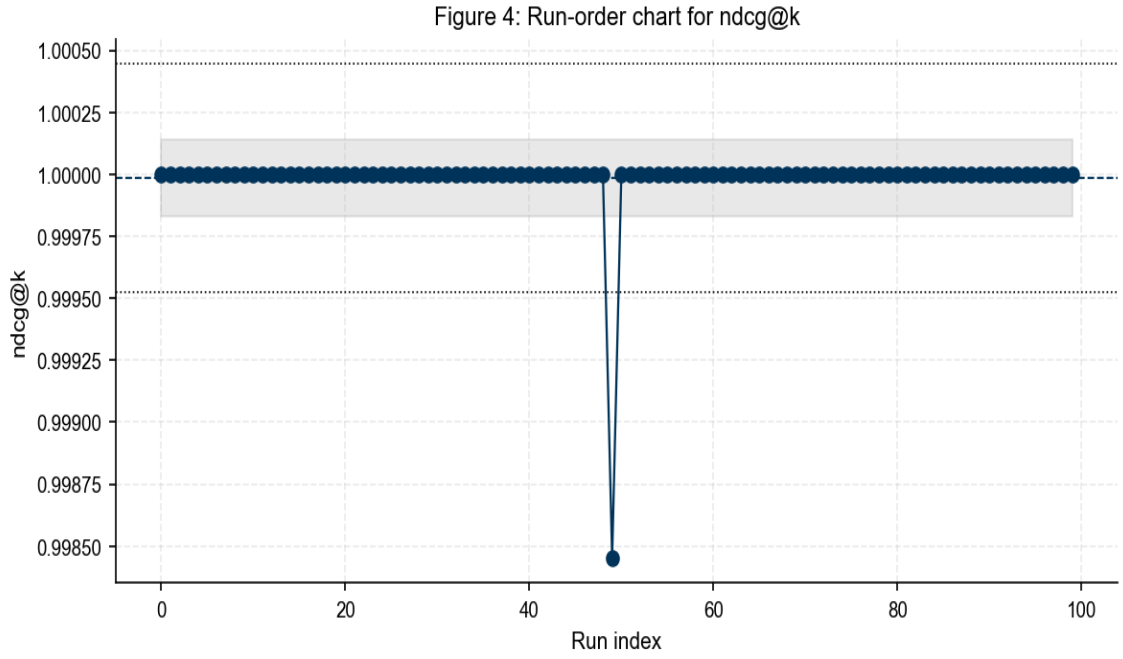
hist ndcg.png



run order faithfulness.png



run order ndcg@k.png



scatter ndcg vs faithfulness.png

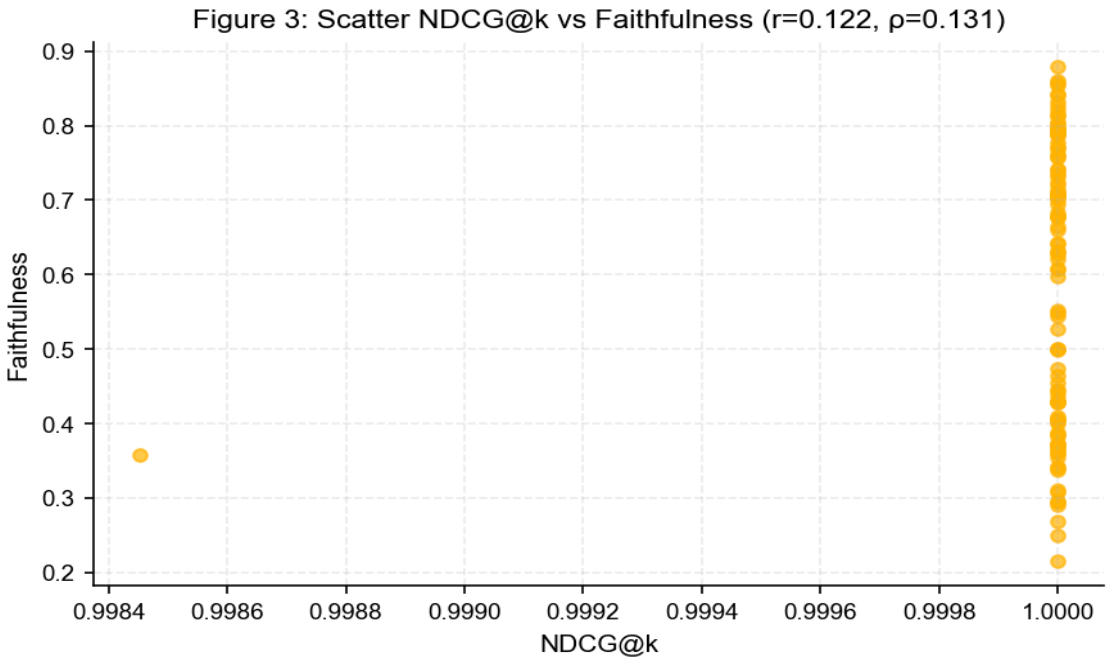


Figure 3: Scatter NDCG@k vs Faithfulness (r=0.122, ρ=0.131)
Faithfulness
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.9984
0.9986
0.9988
0.9990
0.9992
0.9994
0.9996
0.9998
1.0000
NDCG@k