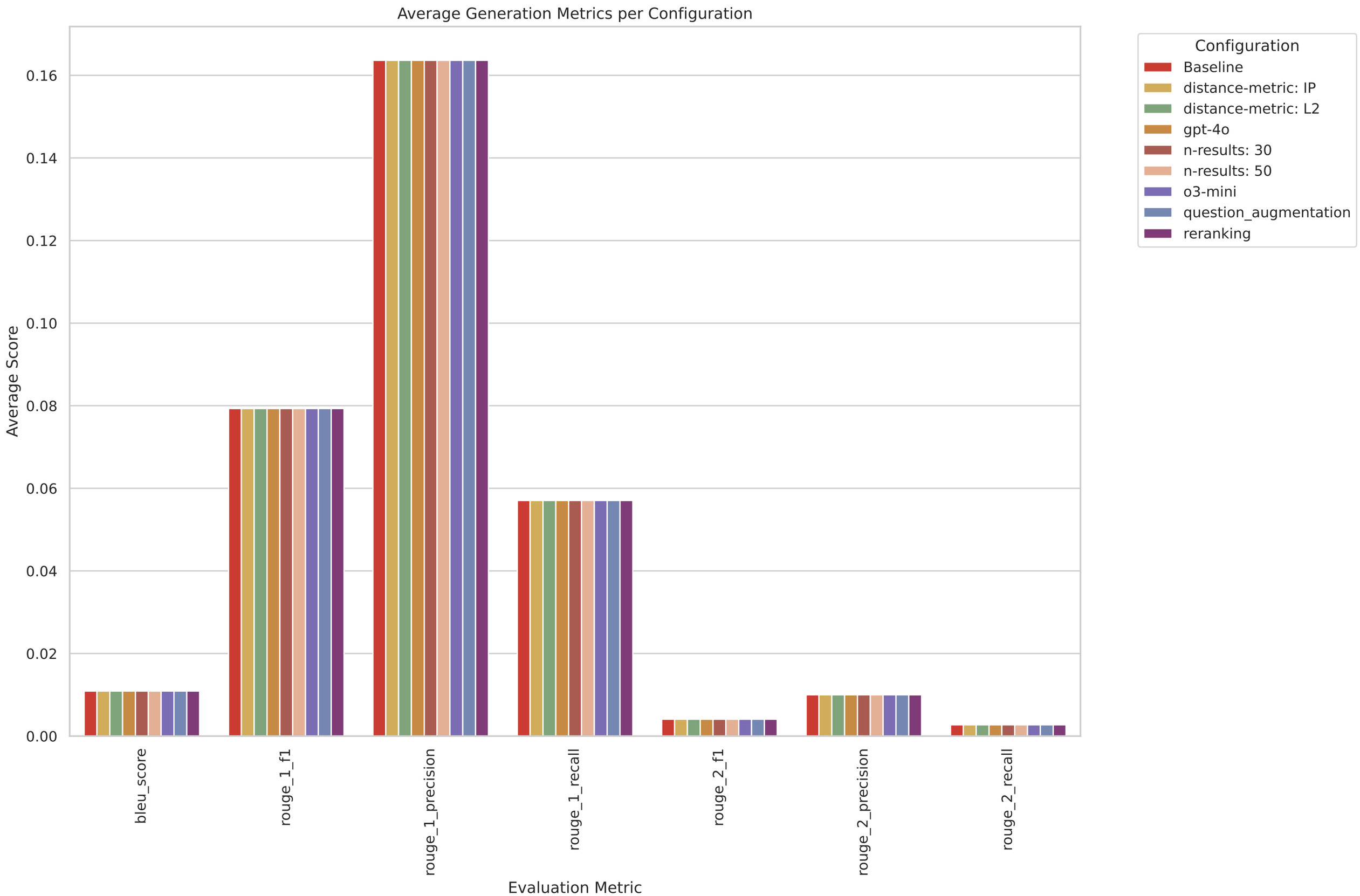
Average Generation Metrics per Configuration
Average Score
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16
bleu_score
rouge_1_f1
rouge_1_precision
rouge_1_recall
rouge_2_f1
rouge_2_precision
rouge_2_recall
Evaluation Metric
Configuration
Baseline
distance-metric: IP
distance-metric: L2
gpt-4o
n-results: 30
n-results: 50
o3-mini
question_augmentation
reranking