

# GymAdvisorAI - RAG vs GraphRAG Evaluation

Eval date (demo): 2026-01-22 | LLM enabled: True

Mode	Accuracy	Avg latency (ms)	Numeric MAE	P@5	R@5	Reasoning score	Match top-1
GraphRAG	0.25	1316.2	524.333	1.0	0.217	0.111	0.5
RAG	0.0	4093.7	598.333	0.0	0.0	0.0	0.0

## Notes

RAG baseline uses retrieval over textual docs (TF-IDF / optional embeddings). GraphRAG answers by querying the Neo4j knowledge graph and applying explicit constraints.

# Per-case (short)

Preview is truncated for readability.

Mode	Case	OK	Latency (ms)	Preview
graphrag	count_sessions_last_30	0	1575.3	You did 4 session(s) in the last 30 day(s).
graphrag	bench_tonnage_last_30	0	1494.8	Tonnage for bench press in the last 30 day(s): 0.0
graphrag	last_squat_date	0	1436.2	No recorded session for squat.
graphrag	safe_exercises_no_shoulder	1	1507.4	Exercises without risk 'shoulder': Back Squat, Barb
graphrag	plateau_reasoning_bench	0	1335.5	Not enough history for bench press to assess plate
graphrag	what_if_add_sets	0	0.9	What-if tonnage: 3x10 @ 50.0 = 1500.0
graphrag	matching_u2_best_plan	0	1573.9	Top 3 training plan matches for user 'u2': 1. Power
graphrag	matching_u3_best_plan	1	1605.9	Top 3 training plan matches for user 'u3': 1. Hyper
rag	count_sessions_last_30	0	6279.4	RAG baseline (retrieval=tfidf). Top snippets: PDF:
rag	bench_tonnage_last_30	0	3841.8	RAG baseline (retrieval=tfidf). Top snippets: PDF:
rag	last_squat_date	0	3785.6	RAG baseline (retrieval=tfidf). Top snippets: PDF:
rag	safe_exercises_no_shoulder	0	3532.6	RAG baseline (retrieval=tfidf). Top snippets: Exerc
rag	plateau_reasoning_bench	0	3529.9	RAG baseline (retrieval=tfidf). Top snippets: UserF
rag	what_if_add_sets	0	4258.9	RAG baseline (retrieval=tfidf). Top snippets: PDF:
rag	matching_u2_best_plan	0	3421.6	RAG baseline (retrieval=tfidf). Top snippets: PDF:
rag	matching_u3_best_plan	0	4099.5	RAG baseline (retrieval=tfidf). Top snippets: PDF: