Team 5
Research Problem Survey

Revisiting RAG design principles
in agentic settings

17 November 2026

RESEARCH OBJECTIVE

We aim to derive design principles for Agentic RAG

Which RAG principles still hold when an agent controls retrieval?

Corpus structure

Pre-built relationships

Retrieval granularity

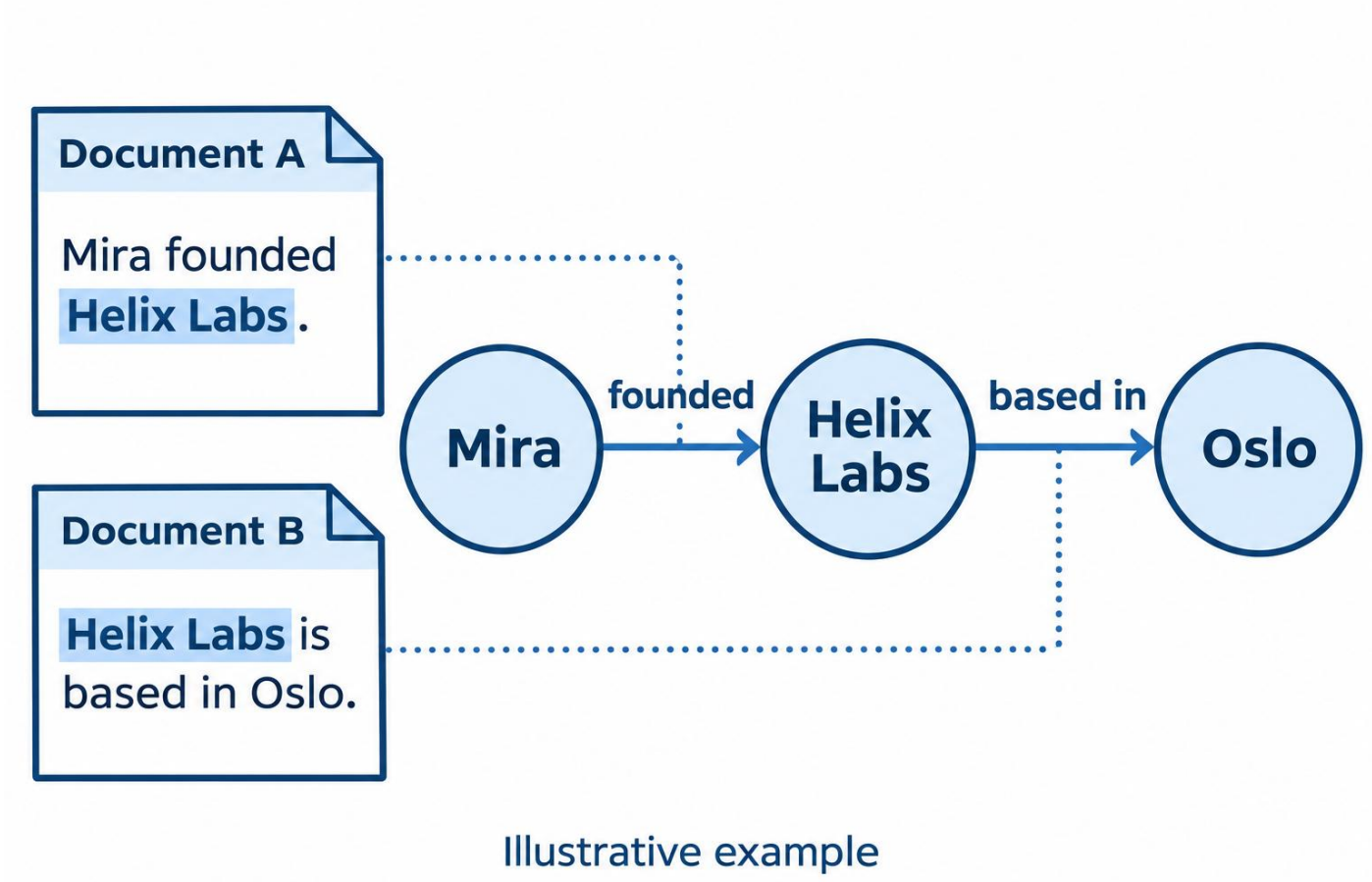
Units and context

Query rewriting

Evidence and constraints

Literature findings / Testable hypotheses / Conditional guidelines

Graph-based retrieval can benefit multi-hop reasoning



Han et al.

Some graph methods outperform flat retrieval on multi-hop tasks.

Xiang et al.

Benefits depend on task complexity and graph design.

Han et al. (2025; updated 2026), RAG vs. GraphRAG
Xiang et al. (2026), GraphRAG-Bench

Iterative search may change when graph structure pays off

Controlled comparison

	One pass	Agent loop
Flat text	Baseline	Dynamic hops
Graph	Graph retrieval	Graph + agent

Hold the model and corpus fixed. Report build cost separately.

Assumption

The agent can discover bridge entities during search.

Test

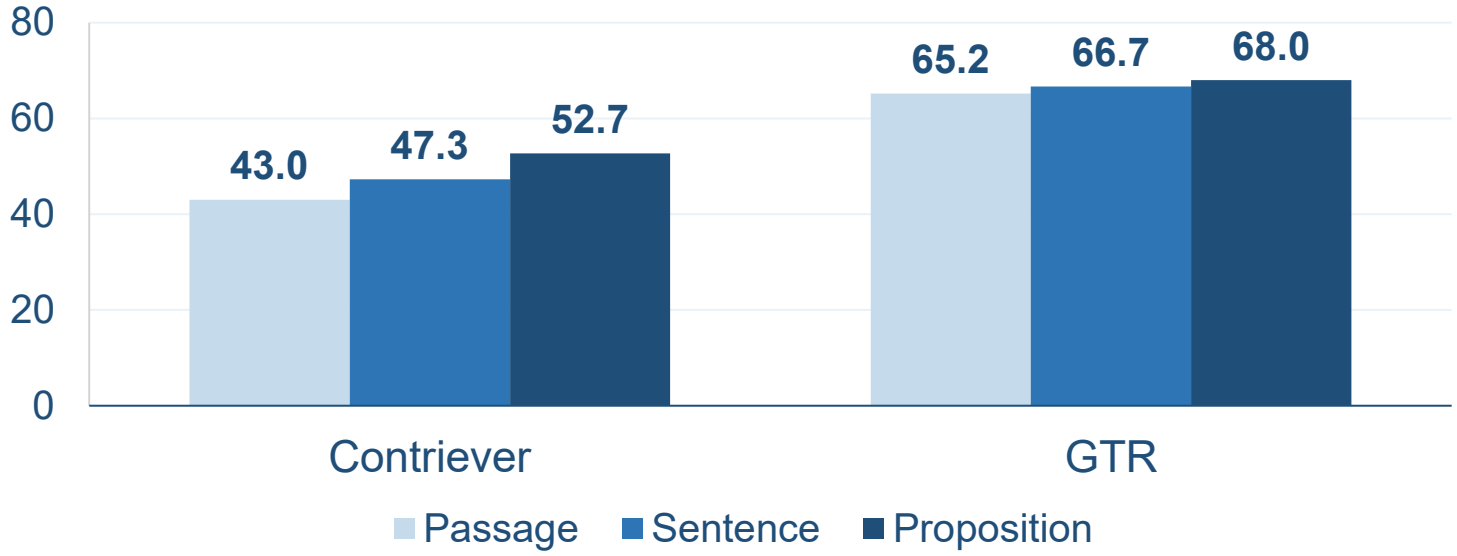
Does the graph benefit persist under matched budgets?

Trivedi et al. (ACL 2023), IRCoT
Han et al. (2025; updated 2026), RAG vs. GraphRAG

Finer retrieval units can improve retrieval accuracy

Passage Recall@5 (%)

Average across five open-domain QA datasets



Contriever

+9.7 pp

43.0% to 52.7%

GTR

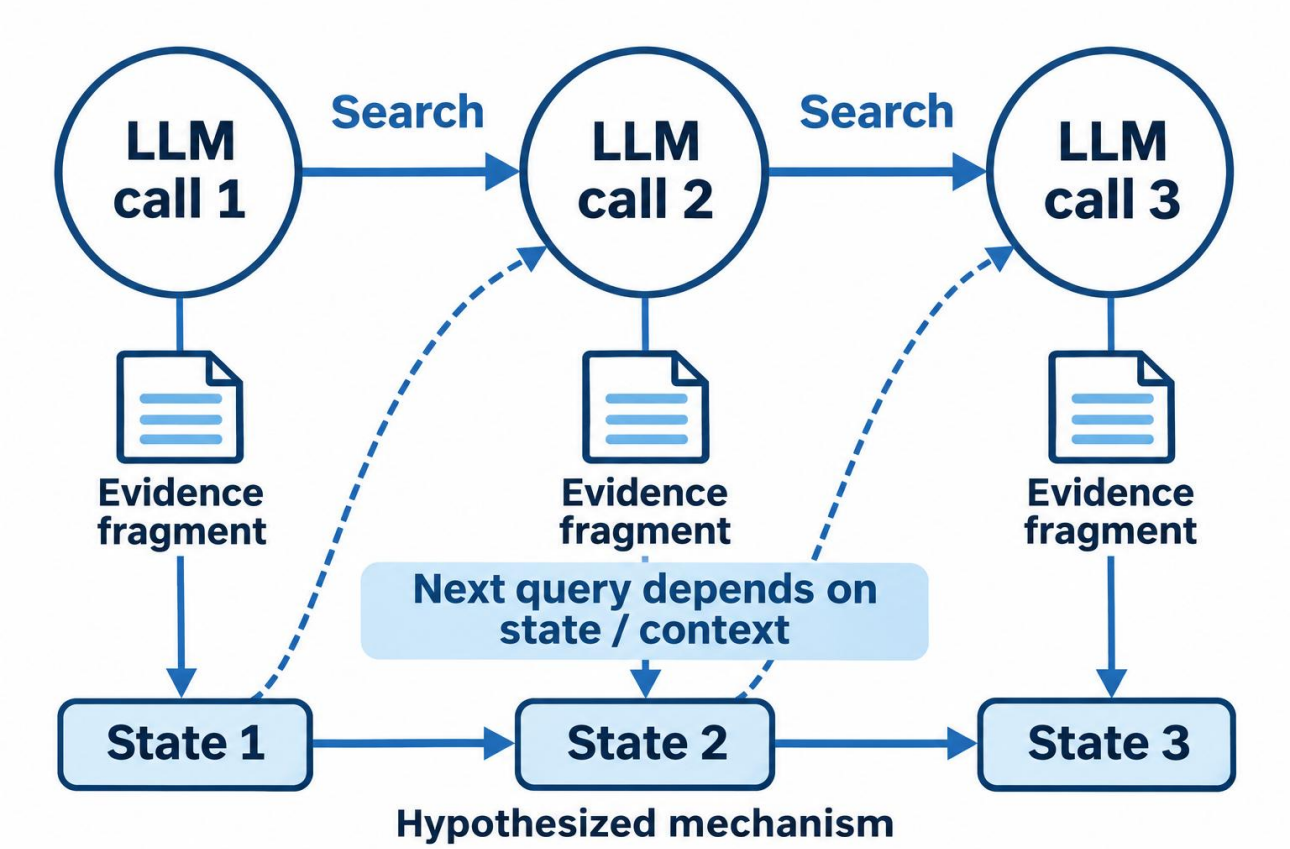
+2.8 pp

65.2% to 68.0%

Fine units score passages; evaluation returns five unique source passages.

Chen et al. (EMNLP 2024), Dense X Retrieval, Table 3

Fine-grained context may weaken agentic retrieval



More LLM calls

Smaller evidence units may require more search decisions.

State / context dependence

Lossy or overloaded state may degrade later queries and evidence recall.

Test: evidence recall and LLM calls under matched budgets.

Chen et al. (EMNLP 2024), Dense X Retrieval
Du et al. (2026), A-RAG

Query rewriting can improve retrieval alignment



Ma et al.

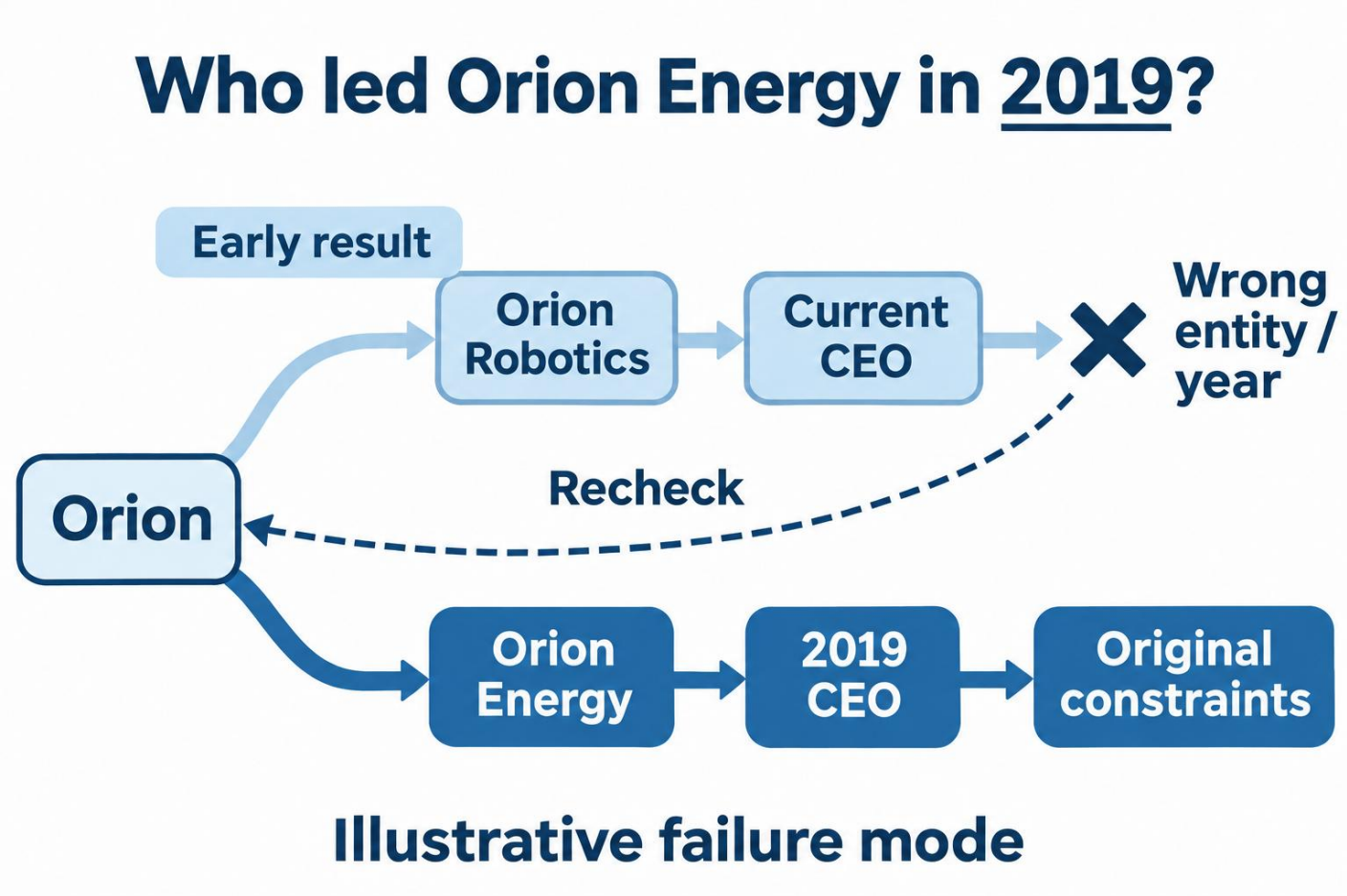
Rewriting queries improved downstream QA in their experiments.

IRCoT

Intermediate reasoning can guide subsequent retrieval.

Ma et al. (EMNLP 2023), Query Rewriting
Trivedi et al. (ACL 2023), IRCoT

Query rewriting may cause search / goal drift



Hypothesis

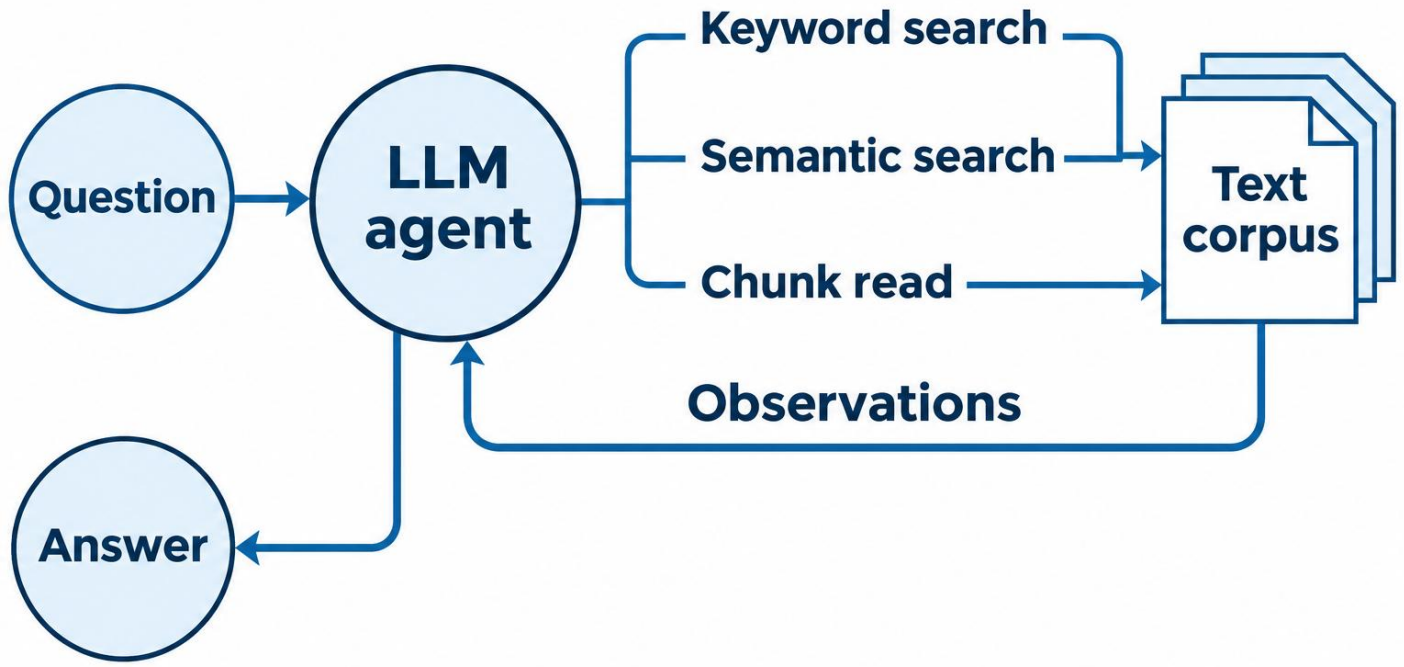
Early evidence shifts later queries away from the original retrieval goal.

Measure

Lost entity or time constraints, off-target searches, and evidence recall.

Ma et al. (EMNLP 2023), Query Rewriting
Trivedi et al. (ACL 2023), IRCoT

A-RAG provides a base for agent-controlled retrieval



Existing capabilities

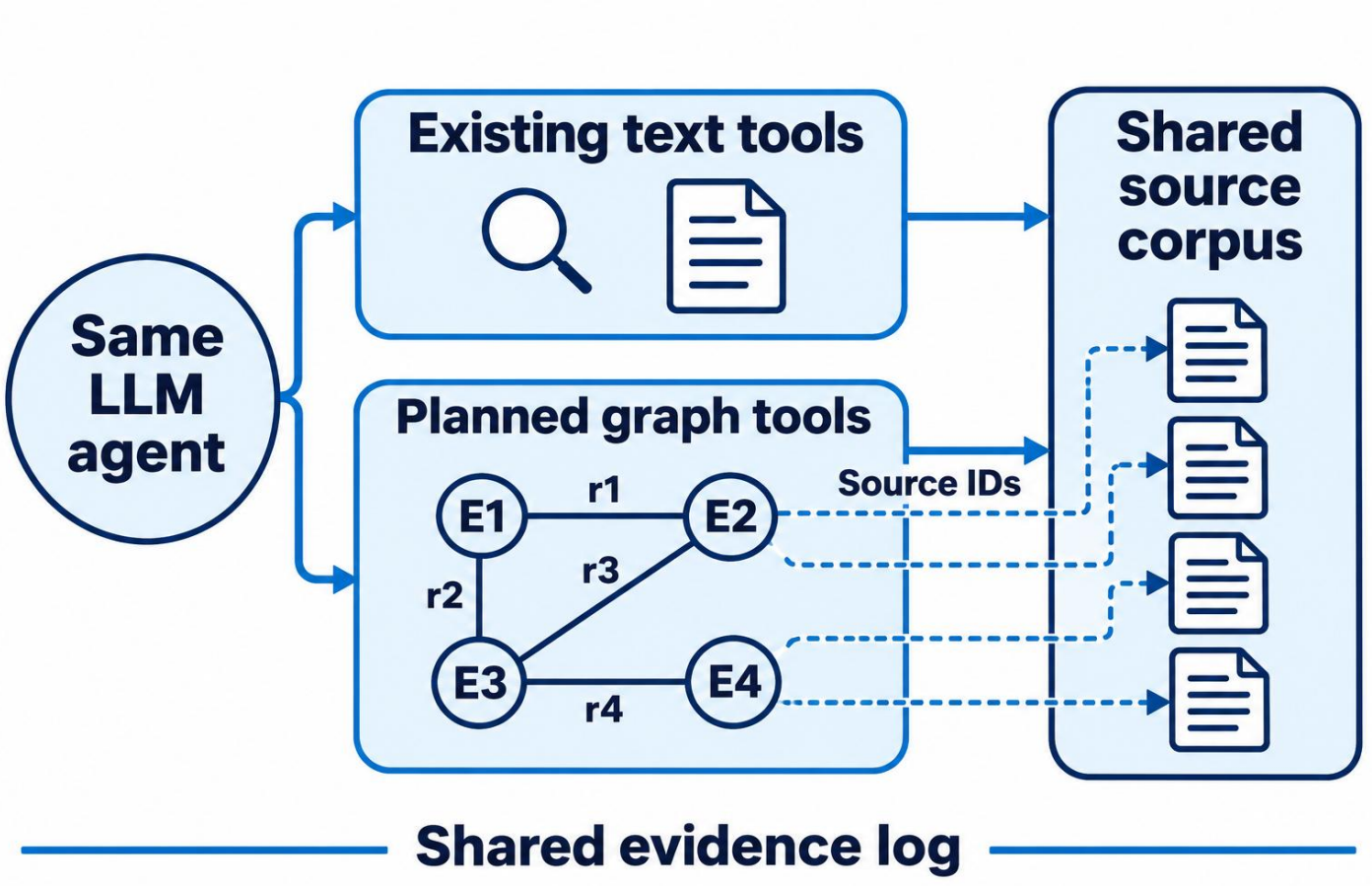
Autonomous tool selection and iterative execution.

Why this base

A shared controller for our controlled comparisons.

Du et al. (2026), A-RAG
A-RAG: official open-source implementation

We will add access to structured corpus representations



Entity and relation tools

Links to source passages

Shared trajectory logging

A-RAG: official open-source implementation

Three settings test multi-hop and domain-specific reasoning

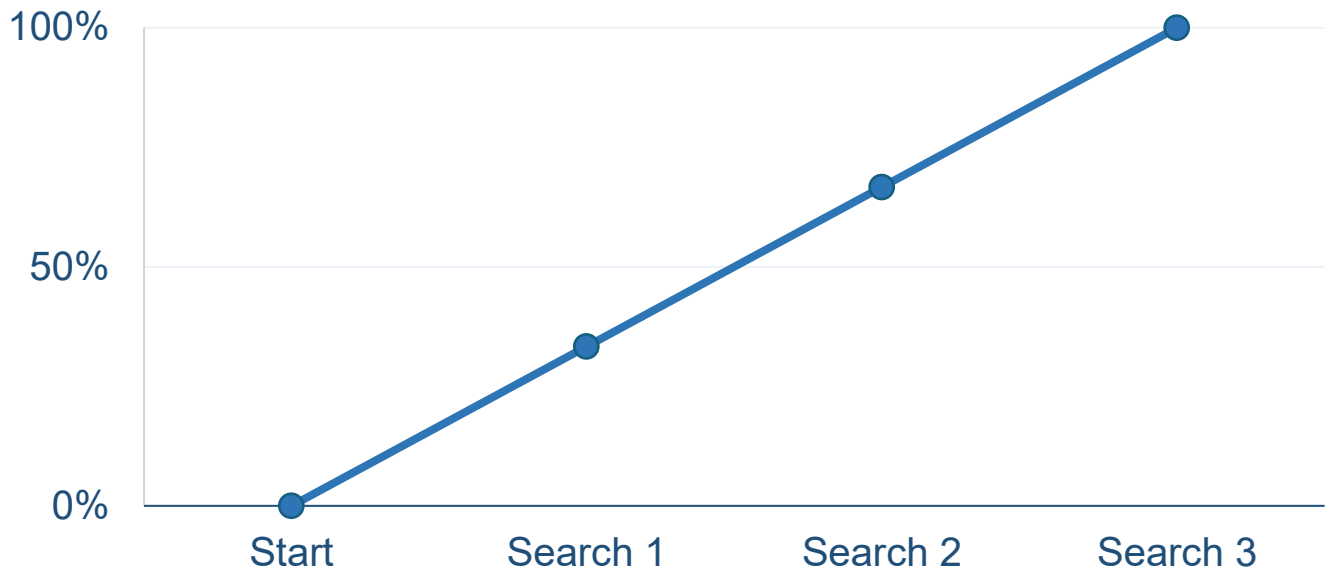
Setting	Reasoning focus	Evidence
HotpotQA	Multi-hop QA	Supporting facts
GraphRAG-Bench Novel	Cross-passage synthesis	Reference evidence
GraphRAG-Bench Medical	Domain-specific reasoning	Reference evidence

Use identical corpora and question sets across methods.

Yang et al. (EMNLP 2018), HotpotQA
GraphRAG-Bench: datasets and evaluation

Evaluation covers correctness, evidence and cost

Cumulative evidence recall



Illustrative trajectory: three required facts

Answer

LLM-judge accuracy
HotpotQA EM / F1

Evidence

Gold-evidence coverage across all search steps

Cost

Calls, tokens and latency

Yang et al. (EMNLP 2018), HotpotQA
Xiang et al. (2026), GraphRAG-Bench

Our contribution will be conditional design guidelines

Structure	When does a graph justify its cost?
Granularity	When should the agent read more context?
Rewriting	When should the agent revise or backtrack?

Work in progress. These are hypotheses to test.

References

Han et al. (2025; revised 2026). RAG vs. GraphRAG: A Systematic Evaluation and Key Insights. arXiv:2502.11371.

Xiang et al. (2026). When to use Graphs in RAG: A Comprehensive Analysis for Graph Retrieval-Augmented Generation. ICLR. arXiv:2506.05690.

Chen et al. (2024). Dense X Retrieval: What Retrieval Granularity Should We Use? EMNLP, pp. 15159–15177.

Ma et al. (2023). Query Rewriting in Retrieval-Augmented Large Language Models. EMNLP, pp. 5303–5315.

Trivedi et al. (2023). Interleaving Retrieval with Chain-of-Thought Reasoning for Knowledge-Intensive Multi-Step Questions. ACL, pp. 10014–10037.

Du et al. (2026). A-RAG: Scaling Agentic Retrieval-Augmented Generation via Hierarchical Retrieval Interfaces. arXiv:2602.03442.

Yang et al. (2018). HotpotQA: A Dataset for Diverse, Explainable Multi-hop Question Answering. EMNLP.