



Junior Williams Consulting Inc.

SYSTEMS & ENTERPRISE ARCHITECTURE · CYBER & AI RISK

FIELD GUIDE · ENTERPRISE AI

Graph RAG is for connected evidence

A practical test for deciding when the graph is worth the extra machinery.

PREPARED BY

Junior Williams
Founder

PUBLISHED

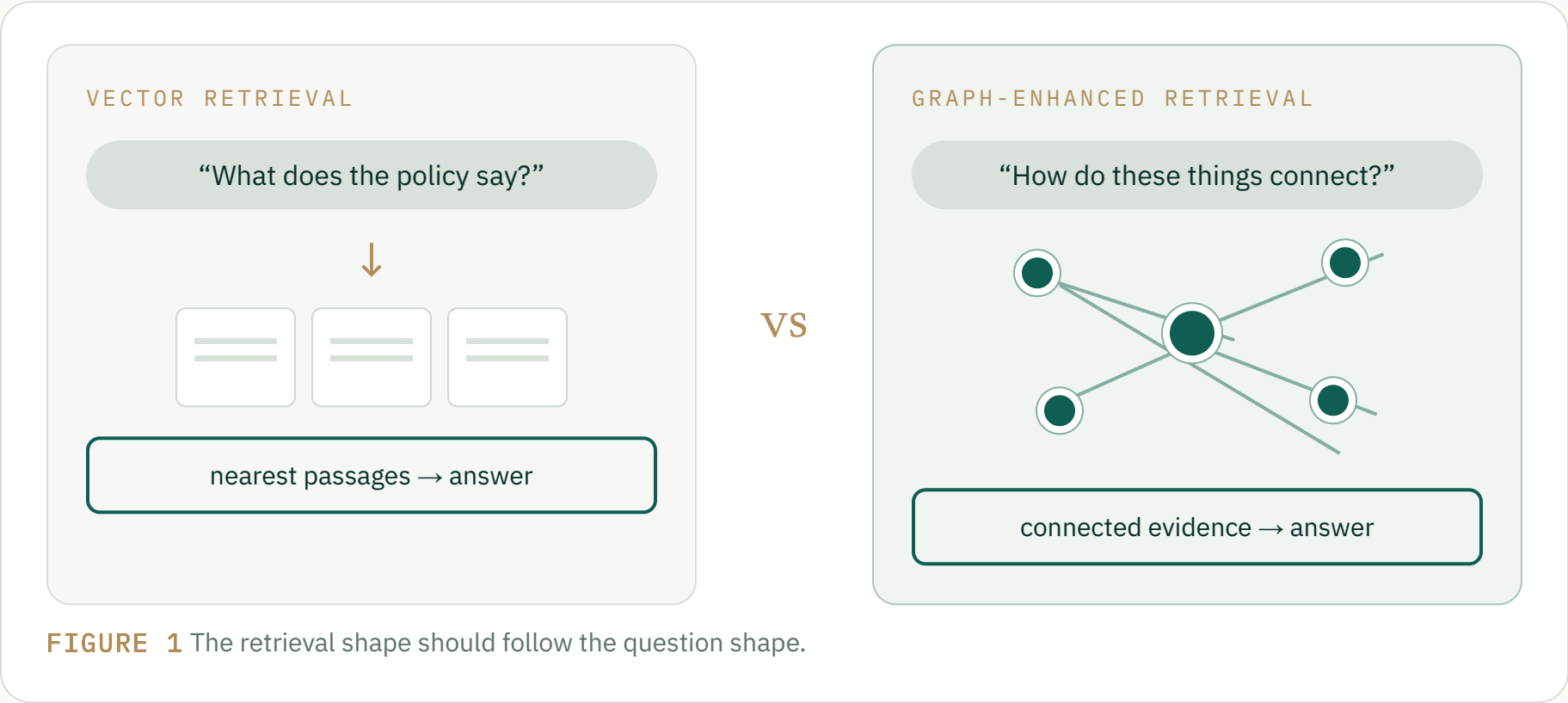
26 July 2026

CONTACT

Book an intro call
trustcyber.ca

Similarity cannot recover a relationship it never indexed

Vector retrieval is a strong baseline when the answer lives in a few relevant passages. Graph-enhanced retrieval is worth testing when the answer is distributed across entities, events and documents.

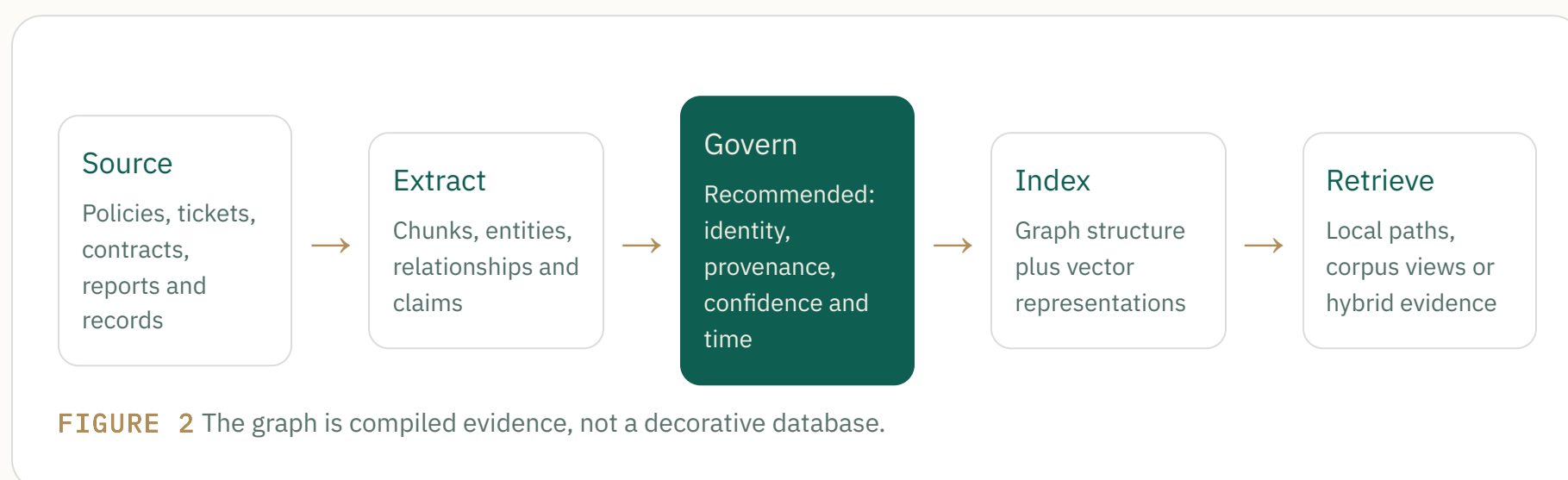


A supplier-risk question may require a path from vendor to service, control exception, business owner and renewal date. A corpus-wide question may require themes that no single passage contains. Similarity still helps, but it cannot create the missing structure.

02 • KNOW WHAT YOU ARE BUILDING

A graph index turns prose into traversable evidence

The method family is broader than any one implementation. Microsoft GraphRAG, LightRAG and custom knowledge-graph systems make different trade-offs, but the production work has the same hard parts.



INDEX TIME

Structure is created

A production pipeline can extract entities and relationships. It should also attach source receipts, reconcile identities and prepare graph and vector indexes.

QUERY TIME

Evidence is assembled

The application selects a retrieval mode and gathers connected evidence. Production controls should preserve source references and bound the model context.

Entity resolution is where a clever demo becomes an enterprise data problem.

Use it when the answer depends on connection or coverage

Graph-enhanced retrieval is justified by the question set. These patterns are stronger signals than corpus size or vendor enthusiasm.

01

Multi-hop dependency

Which services, controls and owners are exposed if this supplier fails?

RISK PATHS

IMPACT ANALYSIS

02

Corpus-wide sensemaking

What recurring themes, clusters or systemic issues appear across the collection?

GLOBAL QUESTIONS

03

Entity-centred investigation

What evidence, events and decisions relate to this customer, asset, person or obligation?

CASE WORK

TRACEABILITY

04

Cross-document contradiction

Where do sources disagree, and which statement is current, authorized and supported?

CONFLICTS

TEMPORAL STATE

Decision rule: if the answer can be supported by two or three nearby passages, vector retrieval may be enough. If the answer requires a path, a population, a cluster or a timeline, test graph-enhanced retrieval.

The graph can make a wrong answer look organized

Structure increases reach. It also increases the damage caused by bad identity, stale relationships and weak evidence controls.

IDENTITY

False merges

Two entities with similar names collapse into one node. The system produces a coherent path through unrelated evidence.

PROVENANCE

Untraceable edges

A relationship survives extraction without a usable source receipt. Reviewers cannot tell whether it was observed or inferred.

AUTHORIZATION

Cross-boundary retrieval

The graph exposes connections across security domains because access rules were applied to documents but not to nodes, edges and summaries.

TIME

Stale state

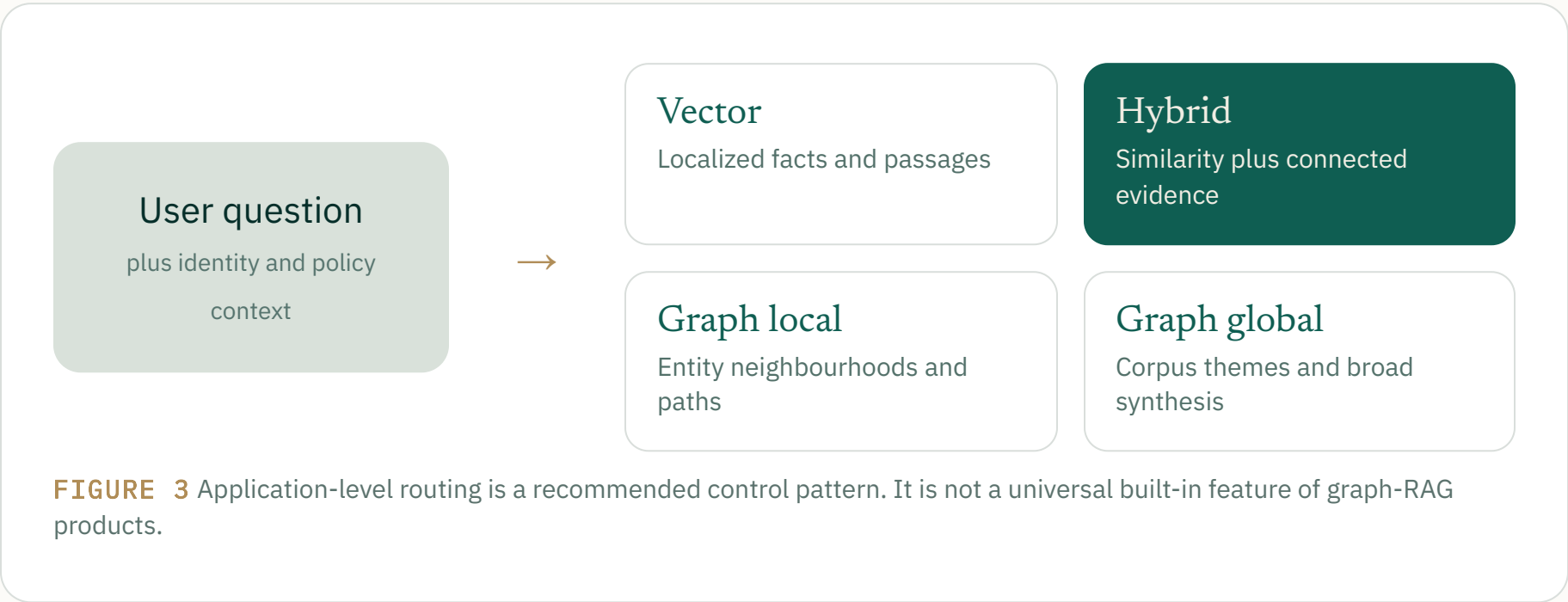
An old owner, status or dependency remains active because the index has no effective-date model or deletion path.

A relationship should carry a source, confidence class, time boundary and authorization context.

LLM extraction can supplement deterministic structure. It should not silently replace identifiers, approved relationships or record-level controls that already exist.

Route by question. Preserve more than one retrieval mode.

A production design can keep vector search for local factual questions and use graph retrieval only when the question requires connected or corpus-wide evidence.



POLICY

Filter first

Apply identity, tenant and record rules before evidence enters model context.

EVIDENCE

Keep receipts

Return source passages with every graph-derived relationship or summary.

EVALUATION

Test by class

Measure local, multi-hop and global questions separately against a simpler baseline.

Treat the graph as derived data. Rebuild it from authoritative sources, and version the rules that created it.

Make the graph prove it deserves production

Use a bounded corpus and a real question set. Compare graph-enhanced retrieval with the strongest credible vector baseline.

PILOT HYPOTHESES Starting points only. Validate every row on the target corpus and question set.

DECISION SIGNAL	VECTOR	HYBRID	GRAPH
Localized factual queries	Baseline	Test	Test
Multi-hop relationships	Baseline	Test	Test
Corpus-wide synthesis	Baseline	Test	Test
Freshness and deletion	Measure	Measure	Measure
Evidence explainability	Verify	Verify	Verify

PILOT SCOPE

Start narrow

- 20–50 governed documents
- 25–40 real questions
- Known entities and expected paths
- One accountable data owner

PASS CONDITION

Demand measured lift

- Better answers on target query classes
- Usable evidence traceability
- Acceptable update cost and latency
- No authorization regression

Do not approve the production architecture from a polished demo. Approve it from a question-class evaluation, an operating-cost model and evidence that the data can support stable identity.

Use the smallest architecture that answers the real questions

Graph-enhanced RAG can improve retrieval for some connected-evidence and corpus-wide questions. Results depend on the corpus, query set, baseline and implementation. The resulting graph is derived data that must be resolved, secured, refreshed and evaluated.

- 01 [Edge et al. — From Local to Global: A GraphRAG Approach to Query-Focused Summarization](#)
Microsoft Research paper describing entity graphs, communities and global sensemaking evaluation.
- 02 [Microsoft GraphRAG — Query Engine overview](#)
Official documentation for Basic, Local, Global and DRIFT search modes.
- 03 [Microsoft GraphRAG — Indexing overview](#)
Official indexing workflow for entities, relationships, claims, communities, summaries and embeddings.
- 04 [Guo et al. — LightRAG: Simple and Fast Retrieval-Augmented Generation](#)
Graph-based indexing, dual-level retrieval and incremental update approach.
- 05 [Microsoft Research — LazyGraphRAG](#)
A deferred-LLM approach and vendor-reported cost-quality results on a bounded evaluation.

The decision is architectural.

Begin with the query classes, evidence boundaries and operating model. Choose the graph technology after those constraints are visible.