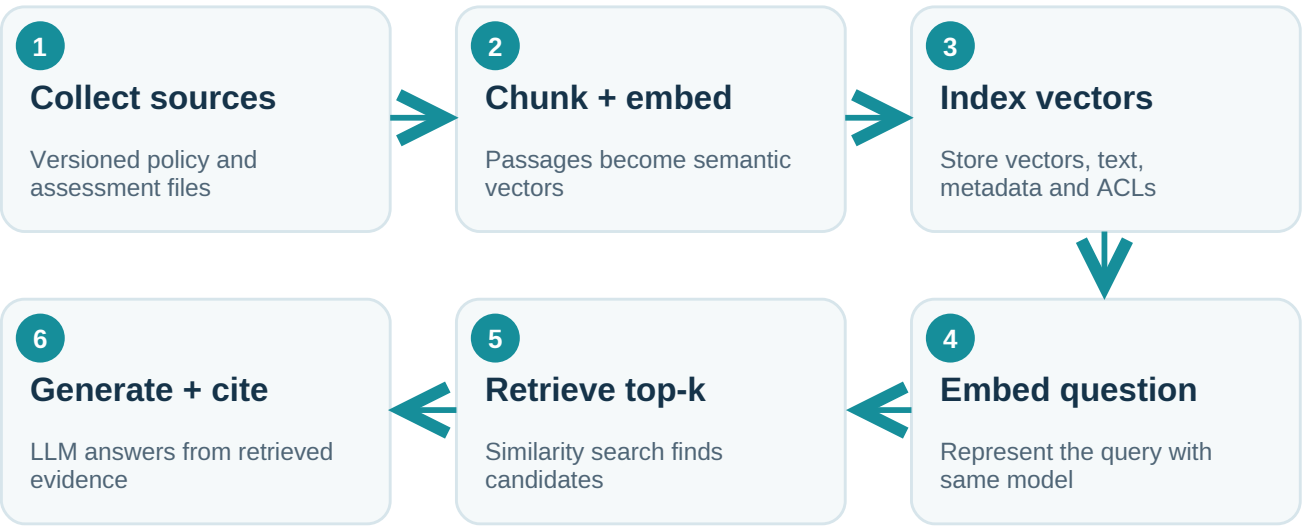


Naive RAG

From documents to evidence-grounded answers

HOW IT WORKS



THE QUESTION

What checks are required before a vendor accesses customer data?

WHY IT HELPS

Embeddings find meaning, not guaranteed truth. Source quality and access controls matter as much as the vector search.

WATCH OUT FOR

Small chunks lose context; large chunks mix topics. Similarity can find plausible but irrelevant passages.

WHEN TO USE IT

Use as a measurable baseline when the questions are narrow and first-pass retrieval is reliable.

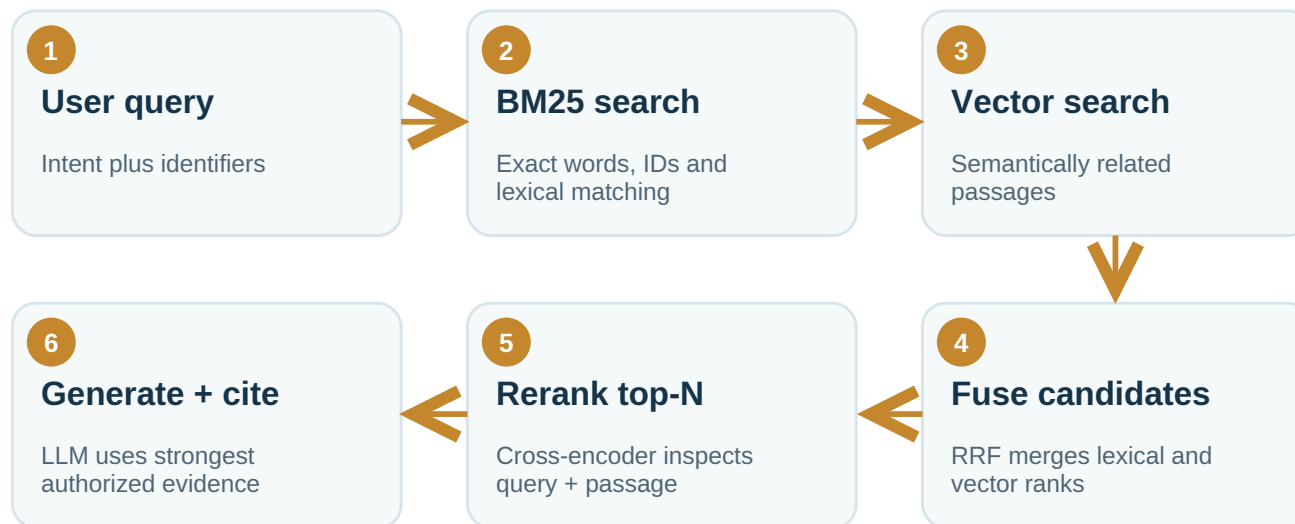
PROBLEM BEING SOLVED

How do we find relevant passages in a large policy and assessment corpus?

Advanced / Hybrid RAG

Exact terms + semantic meaning + stronger ranking

HOW IT WORKS



THE QUESTION

Locate requirements for TPRM-ENT-2026 and equivalent onboarding rules.

WHY IT HELPS

BM25 ranks term matches; vectors capture paraphrases. RRF joins rankings; a reranker improves candidate order.

WATCH OUT FOR

Reranking cannot recover a missing candidate. Enforce permissions before either retrieval branch returns content.

WHEN TO USE IT

Strong default for enterprise Q&A with policy IDs, product codes and natural language.

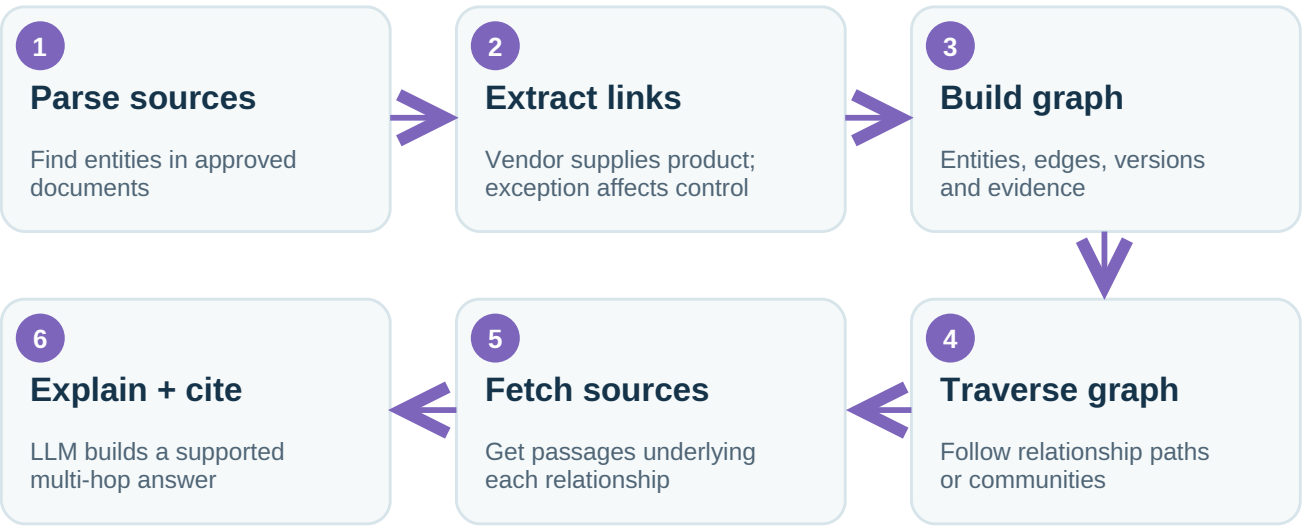
PROBLEM BEING SOLVED

How do we find a precise policy ID without missing semantically related language?

GraphRAG

Answer questions about connected entities

HOW IT WORKS



THE QUESTION

Which vendors support regulated systems that have open security exceptions?

WHY IT HELPS

Relationships—not just similar text—become first-class retrieval signals. Graph and source text work together.

WATCH OUT FOR

False links, stale graphs, cost of extraction, and incorrect permissions on edges can create misleading answers.

WHEN TO USE IT

Pilot when multi-hop relationships are a frequent, measured failure mode.

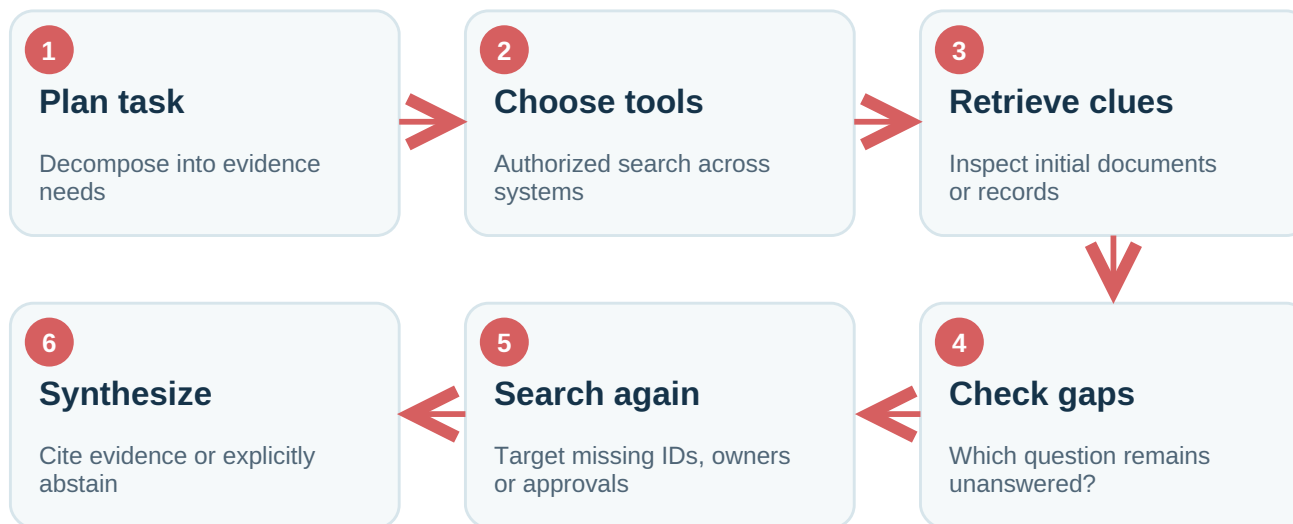
PROBLEM BEING SOLVED

How are vendors, products, contracts, findings and risk exceptions connected?

Agentic RAG

Plan, search, test, then search again

HOW IT WORKS



THE QUESTION

Find the policy, confirm the current assessment, then identify the exception approver.

WHY IT HELPS

An agent can decide what information it still needs, not simply accept the first search result.

WATCH OUT FOR

Tool loops can waste time and cost, follow wrong leads, or cross permission boundaries.

WHEN TO USE IT

Use bounded, read-only agents for complex multi-system investigations when baseline RAG fails.

PROBLEM BEING SOLVED

What if the answer requires several linked questions and multiple source systems?

Google's Sufficient-Context Agentic RAG

Check whether evidence is enough before answering

HOW IT WORKS



THE QUESTION

Which policy applies, which approval is needed, and who approved the exception?

WHY IT HELPS

Google's approach explicitly assesses context sufficiency and uses iterative, targeted retrieval for missing facts.

WATCH OUT FOR

A sufficiency evaluator can be overconfident or costly. Google's "up to 34%" is a benchmark-specific result.

WHEN TO USE IT

Evaluate for high-stakes multi-hop queries with persistent partial-answer failures.

PROBLEM BEING SOLVED

How do we prevent a relevant-but-incomplete answer from looking definitive?

Karpathy's LLM Wiki

Persistent, source-linked organizational knowledge

HOW IT WORKS



THE QUESTION

What changed in the vendor assessment policy since last quarter?

WHY IT HELPS

A maintained, interconnected wiki acts as a reusable knowledge layer rather than only raw-chunk search.

WATCH OUT FOR

Summary drift, contamination, missed updates and ACL leakage. AI summaries must not replace source authority.

WHEN TO USE IT

Experiment where topics recur and knowledge evolves; pair with retrieval and review controls.

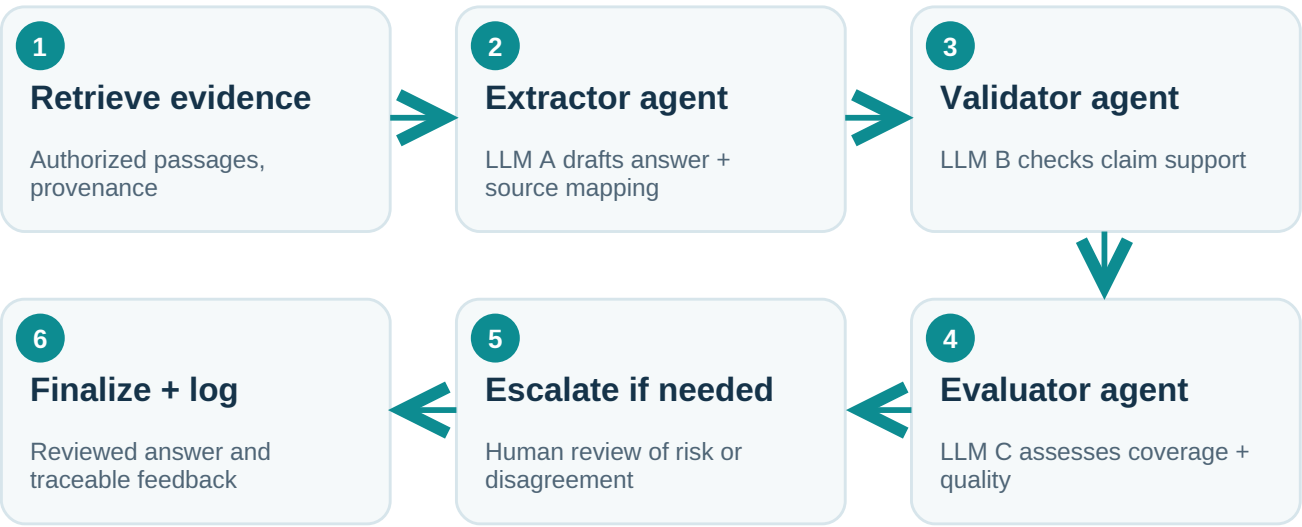
PROBLEM BEING SOLVED

Why reconstruct the same vendor-policy knowledge from scratch for every question?

Question Assistant: Three-Agent Quality Control

Extractor · Validator · Evaluator · Human-in-the-loop

HOW IT WORKS



THE QUESTION

Does the proposed answer fully follow the vendor assessment policy and cite evidence?

WHY IT HELPS

Three differently modeled checks protect answer quality; this is not the same as an agent that autonomously retrieves.

WATCH OUT FOR

Automated validators can share errors. Human-reviewed gold tests and explicit escalation rules remain essential.

WHEN TO USE IT

Pair with any retrieval method when answer trust, auditability and human accountability matter.

PROBLEM BEING SOLVED

How do we ensure retrieved evidence becomes a defensible assessment answer?