ENTERPRISE KNOWLEDGE AI

RAG Assessment

A structured diagnostic for retrieval quality, evidence grounding, citation alignment and operating performance - before you redesign or scale your RAG stack.

ASSESSMENT GOAL

**When the answers look plausible,
but the failure mode is unclear.**

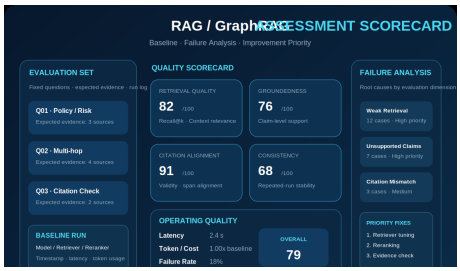
01 Why RAG Assessment?

Diagnose the failure mode before changing the architecture.

Start with measurement, not a rebuild.

- **Missing evidence** The right evidence is not retrieved, or the context is incomplete.
- **Evidence mismatch** Claims and citations do not align with the supporting source.
- **Operational uncertainty** Latency, cost or repeatability prevents a PoC from scaling.

PRINCIPLE Benchmark first. Change only what the evidence supports.



Demo screen | Baseline · Failure Analysis · Improvement Priority

02 What do we assess?

Use the same question set and criteria to isolate failure modes.

01 Retrieval Quality

Recall@k · Precision@k · Context Relevance

Does the system retrieve the evidence required to answer?

02 Groundedness & Citation

Claim Grounding · Citation Validity

Do claims stay within the evidence and cite the right span?

03 Failure Analysis

Retrieval · Evidence · Answer Failure

Which layer fails, and where should remediation start?

04 Efficiency & Consistency

Latency · Token/Cost · Repeatability

Can the system operate consistently within cost and latency constraints?

03 How does the assessment work?

1 Baseline
Establish the current RAG baseline

2 Compare
Compare like-for-like under the same conditions

3 Verify
Verify claims, citations and evidence

4 Prioritize
Define remediation and a focused PoC

04 Deliverables and validation scope

Assessment Output

- **Baseline Scorecard** A measurable current-state benchmark
- **Failure Taxonomy** Failure modes by retrieval, evidence and answer
- **Prioritized Remediation Plan** Actions ranked by expected impact
- **Focused PoC Recommendation** Scope, KPIs and measurable acceptance criteria

START HERE Benchmark the current system before redesigning the architecture.

LIMITED VALIDATION SNAPSHOT

RAG Assessment v1.2.1

6/6 Golden6 E2E	22 Fixed Claims	44 Evidence Eval.
38/44 Citation Scope	24/24 Provenance	23/23 Unit Tests

SCOPE NOTE

38/44 is not an overall RAG accuracy score. Human Gold approval and an independent held-out benchmark are still pending; the figures represent bounded technical validation.