ENTERPRISE KNOWLEDGE AI

RAG Assessment

RAG·GraphRAG의 검색 품질, 근거 정합성, Citation, 운영 품질을 분리해 진단하고 개선 우선순위를 제시합니다.

ASSESSMENT GOAL

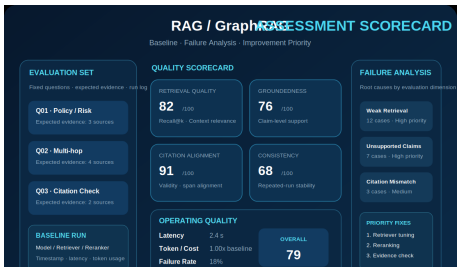
**"답은 그럴듯한데,
어디서 실패하는지 모를 때"**

01 왜 RAG Assessment가 필요한가?

재구축 전에 실패 원인을 먼저 분해합니다.

이런 경우에 먼저 진단합니다

- **검색 실패** 필요한 근거가 검색되지 않거나 핵심 문맥이 누락됩니다.
- **근거 불일치** 답변 Claim과 Citation이 실제 Evidence와 맞지 않습니다.
- **운영 불확실성** Latency·Cost·Repeatability 문제로 PoC가 실제 운영까지 확장되지 않습니다.



실제 데모 화면 예시 | Baseline · Failure Analysis · Improvement Priority

02 무엇을 진단하나요?

같은 질문셋과 같은 평가 기준으로 RAG 실패 원인을 분해합니다.

01 Retrieval Quality

Recall@k · Precision@k · Context Relevance

필요한 근거가 검색되는지, 핵심 문맥이 누락되지 않는지 확인합니다.

02 Groundedness & Citation

Claim Grounding · Citation Validity

답변 Claim이 Evidence에 의해 지지되는지, Citation 범위가 맞는지 봅니다.

03 Failure Analysis

Retrieval · Evidence · Answer Failure

검색·근거·답변 실패를 분리해 어디부터 고칠지 우선순위를 정합니다.

04 Efficiency & Consistency

Latency · Token/Cost · Repeatability

반복 실행 일관성, 응답 지연, 토큰·비용을 함께 비교합니다.

03 Assessment는 어떻게 진행되나요?

1 Baseline
고객 질문·데이터로
현재 기준선 정의

2 Compare
Basic / Hybrid 등
동일 조건 비교

3 Verify
Claim·Citation·Evidence
근거 검증

4 Improve
개선 우선순위·Focused PoC
범위 도출

04 고객이 받는 결과와 검증 범위

Assessment Output

- **Baseline Scorecard** 현재 RAG의 검색·근거·Citation·운영 기준선
- **Failure Taxonomy** Retrieval / Evidence / Answer 실패 유형 분류
- **Priority Fixes** 개선 우선순위와 재검증 항목
- **Focused PoC Recommendation** 고객 데이터로 비교할 PoC 범위와 KPI

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

검증 범위 주의

38/44는 전체 RAG 정확도가 아닙니다. Human Gold 승인과 독립 Held-out Benchmark는 완료 전이며, 현재 수치는 제한된 평가 데이터 기반 기술 검증입니다.