



RAG Evaluation Report (Text Mining-2025)

Members / Evaluators

1. Sahaya Gnanadurai
2. Javed Ahmed Memon
3. Kashyap Kalita

Aggregated Mean Scores

context_precision: 0.800
context_recall: 0.700
faithfulness: 0.826
answer_relevancy: 0.799
answer_correctness: 0.681

RAG NOTES

NOTES: AGENTIC RAG: TEXT MINING LEGAL DOCUMENTS OVERVIEW This Agentic RAG is a legal document question-answering system comparing three AI agent architectures for European legal documents (inheritance, divorce law across Estonia, Italy, Slovenia). Uses RAGAS metrics to evaluate which approach works best. THE THREE APPROACHES SINGLE AGENT: Simple pipeline: Question → Route to DB → Retrieve → Answer MULTI-AGENT: Parallel agents answer separately, then supervisor combines them HYBRID LEGAL: Legal-specialized with metadata filtering KEY FINDING: All agents retrieve equally well precision, but problems occur AFTER retrieval: Single Agent: No issues after retrieval → Clean answers Multi-Agent: Synthesis adds hallucinations Hybrid: Hard filters remove good documents Reasons: - Best performance - Highest faithfulness - Zero hallucinations - Production-ready now - Best for legal domain TECHNICAL STACK LLM: GPT-4o-mini (temperature 0.2 - deterministic) Embeddings: sentence-transformers/all-MiniLM-L6-v2 Vector Store: FAISS Framework: LangChain Evaluation: RAGAS (5 metrics) CONCLUSION Architectural simplicity wins over complexity. Single Agent emerges as clear winner with best overall performance that makes it unsuitable for legal applications where accuracy is critical.

