Name: Borun Li

Date: 11/4/2026

Document Tested: Sample SDF Document

Query Used: "What test methods were used for quality control?"

Part 1: Configuration Comparison Table

Config	Chunks Retrieved	Best Chunk (short excerpt)	Answer (shortened)	Confidence (1–5)	Notes
A	5	“"Quality control includes visual inspection, dimensional measurement, and functional testing..."”	“Quality control used visual inspection, dimensional measurement, and functional testing. Sampling was performed according to AQL.....”	4	
B	2 filtered from 3	“Quality control includes visual inspection, dimensional measurement, and functional testing…”	“Quality control used visual inspection, dimensional measurement, and functional testing. The QC procedure also required sampling according to AQL levels…”	3	

C	5	“Quality control includes visual inspection, dimensional measurement, and functional testing…”	“The quality control process used multiple test methods: visual inspection, dimensional measurement, and functional testing; sampling according to AQL...”	4	

Part 2: Analysis Questions

- Describe the three configurations you tested. What did you change in each?
  Config A (Baseline): Used `top_k=5`, with no similarity threshold and no reranker. It retrieved the top 5 chunks purely based on vector similarity.
  Config B (Filtered): Reduced retrieval size (`top_k=3`) and applied a strict similarity threshold (`threshold=0.75`) without a reranker. This was designed to filter out weak matches and only keep highly relevant chunks.
  Config C (Reranked): Kept a larger retrieval pool (`top_k=5`) and a lower threshold (`threshold=0.60`), but **enabled an LLM-based reranker**. This allowed the system to pull in a wider net of candidates but then intelligently re-sort them by true relevance
- Which configuration gave you the best result, and why?

- **Config C** gave the best result.
  By setting a moderate threshold (0.60) and a larger `top_k` (5), it ensured high recall (capturing important details like "reliability testing" and "ISO 9001" which Config B missed). However, unlike Config A which just dumped these chunks into the context, the **LLM reranker** pushed the highly relevant QC testing methods to the very top and pushed noisy chunks (like employee training) to the bottom. This resulted in an answer that was both comprehensive and highly focused, minimizing hallucinations or irrelevant information.