

**If your RAG system
doesn't work...**

**...it's probably NOT an
AI or technology issue.**

The real problems are often:

- ✘** Poor data quality
- ✘** Poor search implementation



Earlier, I shared tips on RAG search implementation 🙌

 Check out the link in the post

Now, let's focus on the other root cause:

👉 How to improve data quality with a 3-phase prioritization framework.



1



Phase 1: Quick Wins

Low effort, high impact improvements to build trust fast.



✓ Source Reliability & Whitelisting

Restrict ingestion to trusted, high-quality sources

💡 **Why important:** Ensures foundational data accuracy

📌 **Effort:** Low, requires initial curation only



✓ Metadata & Provenance Tracking

Store author, date, and version info with each document

💡 **Why important:** Improves traceability and accountability

📌 **Effort:** Low, can be automated during ingestion



✓ User Feedback Loops

Add "Flag/Report" option in the RAG interface

💡 **Why important:** Enables continuous user-driven improvement

📌 **Effort:** Low–Medium, requires UI & pipeline integration



2



Phase 2: Medium-Term Enhancements

Automate, monitor, and sustain data quality at scale.



✓ Data Quality Pipelines

Automate validation, deduplication, enrichment,
and schema checks

💡 **Why important:** Enforces systematic quality control

📌 **Effort:** Medium, requires ETL/ELT pipeline setup



✓ Continuous Data Auditing

Run scheduled scans for outdated, irrelevant, or broken docs

💡 **Why important:** Maintains data relevance and integrity

📌 **Effort:** Medium, requires monitoring dashboards



✓ Incremental Updates

Shift from bulk loads to small, validated batches

💡 **Why important:** Reduces error propagation and improves freshness

📌 **Effort:** Medium, requires ingestion pipeline redesign



3

Phase 3: Long-Term Maturity

Institutionalize governance and add safeguards for critical domains.



✓ Data Governance & Stewardship

Assign data owners and implement approval workflows

💡 **Why important:** Establishes clear ownership and accountability

📌 **Effort:** High, requires organizational alignment



✓ Human-in-the-Loop Validation

Manual review checkpoints for sensitive industries

(finance, healthcare, legal)

💡 **Why important:** Ensures accuracy for high-stakes decisions

📌 **Effort:** High, resource-intensive







My Recommendation:

- ✓ Start with Phase 1 Quick Wins –
low effort, high impact**
- ✓ Scale into Phase 2 Automation &
Auditing**
- ✓ Mature with Phase 3 Governance &
Human Validation**



I share practical tips on business, marketing, and AI

-  Follow me for more insights
-  Repost to help others
-  Save for future reference
-  DM for collaborations

#RAG #LLM #EnterpriseRAG #GenAI #EnterpriseAI