🔥 Hot Take:

RAG is just a glorified search

Are you using chunking, embeddings, and vector search…

and wondering why your RAG results are inaccurate?

The issue isn’t RAG.
It’s your search implementation.

RAG is 80% search, 20% generation.

Get search right, and everything changes.

🧠 RAG = Retrieval + Prompt

But **retrieval ≠ just vector search**

Here’s a practical breakdown of 5 types of search implementations you can combine in your RAG stack.

You’re not limited to one — hybrid or Agentic RAG is the future.

① Vector Search (Semantic Search)

⭐ **What:** Finds semantically similar content using embeddings

✅ **Best for:** natural language, fuzzy match

📌 **Example:** **“How do I restart my modem?” → “Router reboot instructions”**

💡 **DB:** Pinecone, Weaviate, Qdrant, Postgres (pgvector)

🧠 **Why:** Works when users don’t use exact terms

② Key-Value Search

⭐ **What:** Retrieves values using exact keys (like a dictionary)

✅ **Best for:** structured data, catalog lookups

📌 **Example: “SKU123” → Product info**

💡 **DB:** Redis, DynamoDB, MongoDB

🧠 **Why:** Fast and reliable for ERP & product systems

③ Full-Text Search

⭐ **What:** Matches exact or partial terms with token scoring

✅ **Best for:** technical terms, known phrases

📌 **Example:** “Error 504” → IT knowledge base article

💡 **DB:** Elasticsearch, MySQL, Postgres (pg_search)

🧠 **Why:** Powerful for structured, predictable language

④ GraphRAG

⭐ **What:** Uses a knowledge graph for relationship-aware answers

✅ **Best for:** org charts, entity links, workflows

📌 **Example: “Who reports to whom in Project X?”**

💡 **DB:** Neo4j, Amazon Neptune, ArangoDB

🧠 **Why:** Goes beyond words — understands connections

→

⑤ Metadata Filtering

⭐ **What:** Filters results based on structured fields

✅ **Best for:** narrowing down by tags, dates, source

📌 **Example:** **“Internal docs from Q1 2024”**

💡 **DB:** PostgreSQL, MySQL, MongoDB, Cassandra

🧠 **Why:** Adds relevance and control

If your RAG feels weak, it's not the LLM.

It's your search stack.

Search is **80% of RAG's performance.**

Prompting is just the last 20%.

💬 Want help with your RAG setup?

DM me. Too many teams are stuck on vector search when they need more.

🔔 Follow for more AI tips

♻️ Repost to help others

🔖 Save this for reference

#GenAI #RAG #GraphRAG #VectorSearch

#KeyValueSearch #SemanticSearch #EnterpriseAI

#RetrievalAugmentedGeneration #AIArchitecture