1. **Limited Real-Time Adaptability in RAG Pipelines**

**Gap:** Most RAG studies (e.g., LangChain documentation, Springer Metadata QA) focus on predefined document sets and do not effectively support dynamic, real-time retrieval across formats (like PDFs, HTML, and voice).

1. **Lack of Unified Metadata and Chunking Frameworks**

**Gap:** Metadata usage and semantic chunking are often treated separately. The Springer Metadata QA paper highlights metadata’s value, while chunking (2310.14021) is typically used without semantic alignment.

1. **Inadequate Support for Voice and Multimodal Input in RAG**

**Gap:** Whisper (2306.07944) is robust for STT, but integration into QA pipelines remains underdeveloped. Papers like 2502.15264 point to STT errors affecting downstream QA but don’t offer solutions.

1. **Underutilization of Hybrid and Postgres-Based Vector Retrieval**

**Gap:** Hybrid retrieval (2408.05141) shows promise but lacks implementation examples using relational databases like PostgreSQL.

**5. Conversational Memory Not Fully Integrated with RAG**

**Gap:** Papers like 2410.15944 discuss short/long-term memory in theory, but there’s limited application of memory chaining in document-based RAG systems.