

CHECKLIST

Enterprise RAG Implementation Checklist

Sixteen checks across the four decisions that actually de-risk a retrieval rollout. Reuse it at kickoff, mid-pilot and pre-production go-live so gaps stay explicit rather than being discovered by end users.

How to use it. RAG is a system, not a feature flag. These items are ordered so that data and governance decisions happen before effort goes into prompt polish. Walk each section with one accountable owner from IT or data, a product sponsor, and someone who can speak to compliance. **Tick an item only when evidence exists** — sample queries, logs or a signed policy — not when a ticket is open.

PROJECT	OWNER	REVIEW STAGE	DATE
---------	-------	--------------	------

01 Data readiness

- ☐ Identify source systems and data owners
- ☐ Validate document quality and freshness
- ☐ Define update cadence and ingestion strategy
- ☐ Classify sensitive data and access rules

02 Retrieval quality

- ☐ Choose chunking strategy by content type
- ☐ Set baseline retrieval benchmarks
- ☐ Implement reranking where needed
- ☐ Track citation coverage and relevance

03 Security & governance

- ☐ Apply role-based retrieval access
- ☐ Enable audit logs for user queries
- ☐ Define retention and data handling policy
- ☐ Review compliance requirements early

04 Production rollout

- ☐ Run pilot with measurable success metrics
- ☐ Monitor latency, cost, and quality KPIs
- ☐ Set fallback behavior for low-confidence answers
- ☐ Create iteration loop from user feedback

SEQUENCING TRAPS

Four ways teams lose months

Tuning before plumbing. Flashy model tuning while canonical URLs, deduplication and ACL tests are skipped — then the embedding model gets the blame.

Offline accuracy on a frozen snapshot. Production feeds drift. Tie releases to data version hashes and shadow-evaluate new index builds before cutover.

Measuring only whether the prose felt helpful. Pair qualitative review with task success: did the user finish the workflow, and was every material claim tied to a retrievable source?

Anonymous 'platform' ownership. Name a product manager for outcomes and an engineering lead for reliability, or this becomes an unmaintained search box with a chat skin.

BEFORE GO-LIVE

Write the definition of done once

Which metrics must sit in green for one full week; which failure screens are acceptable; and who can authorise a rollback. Ambiguous launch criteria produce thrash between engineering and stakeholders — write them down once and reuse them for the next domain you onboard.

Working through this and want a second pair of eyes? Srishti GenAI builds governed retrieval and agent systems for enterprise teams — grounded retrieval, human approval at defined points, traceable decisions. Reach us at srishtigenai.com/contact.