aimprosoft

HOW TO EVALUATE RAG EFFICIENCY

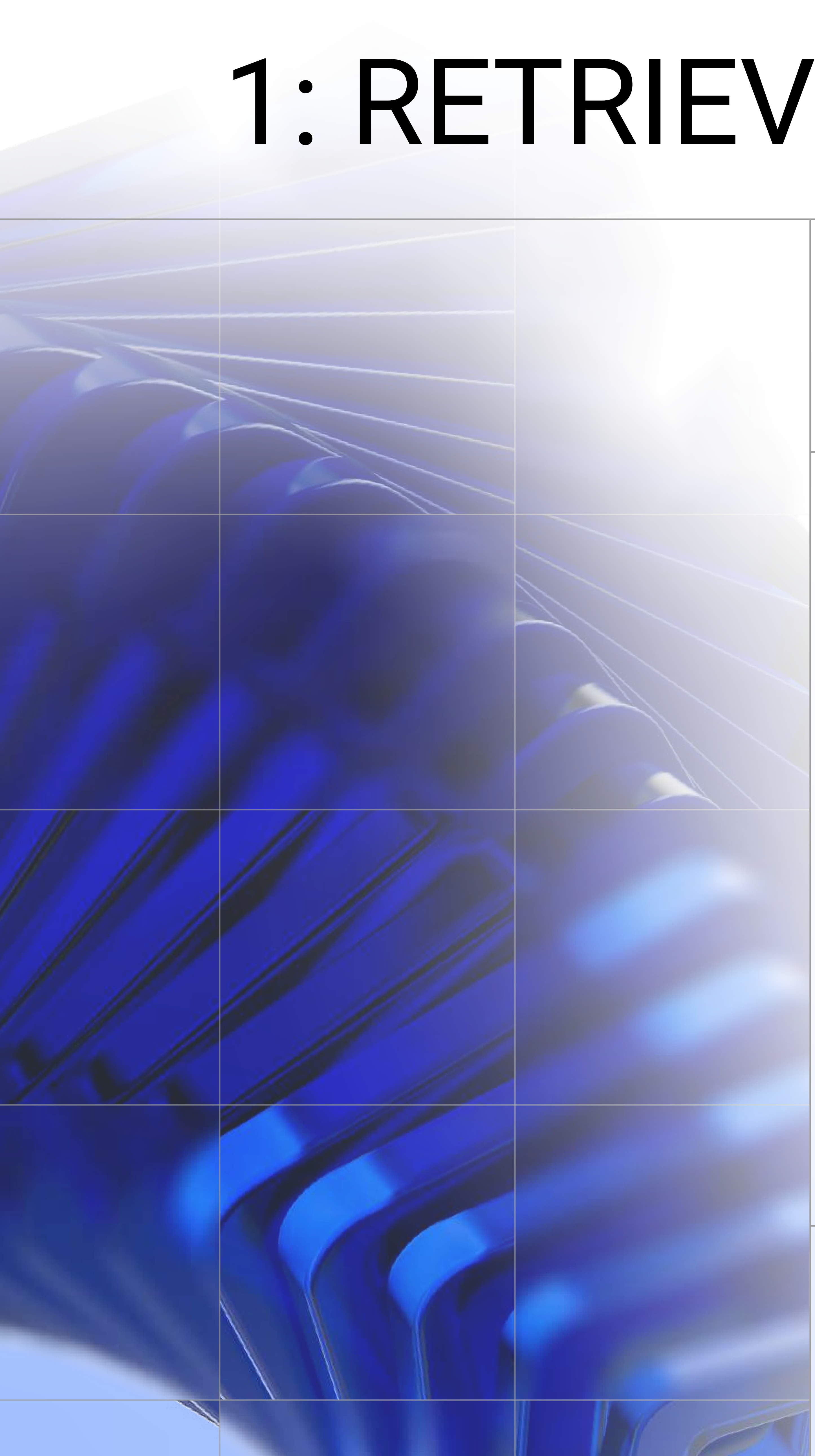
An executive assessment guide, so you can check RAG performance across critical dimensions and know exactly what to fix.

WHAT RAG EFFICIENCY MEANS

AN EFFICIENT, PRODUCTION-READY RAG SYSTEM:

- ✓ Retrieves the right enterprise knowledge consistently
- ✓ Answers only when evidence exists (admits "I don't know")
- ✓ Measurably reduces operational friction
- ✓ Remains secure and compliant under audit and scale
- ✓ Can explain every answer it generates

1: RETRIEVAL QUALITY



Core Question: Are the right documents being used?

Quick Check:

- Top results are current, approved versions (not drafts or outdated docs)
- Answers cite specific sources with clear attribution
- System uses metadata to filter results (date, owner, department)

If this is weak → What to do
→ **Tighten source control:** Restrict to approved documents only
→ **Add mandatory metadata:** Owner, version, date for every indexed file
→ **Fix ranking:** Prioritize authority + recency over similarity

2: ANSWER RELIABILITY

Core Question: Can you trust the system under pressure?

Quick Check:

- System says "I don't know" when the evidence is weak
- Every answer links back to source documents
- Refuses out-of-scope questions (legal advice, medical guidance, etc.)

If this is weak → What to do
→ **Set confidence thresholds:** Block answers when evidence is insufficient
→ **Enforce citations:** Require source + version for every response
→ **Define hard boundaries:** Block prohibited question types by design

3: OPERATIONAL IMPACT

Core Question: Is RAG changing how work gets done?

Quick Check:

- Teams spend less time searching for information
- Fewer escalations to experts for routine questions
- Embedded in daily workflows (not a standalone tool)

If this is weak → What to do
→ **Integrate into existing tools:** Slack, ticket systems, daily platforms
→ **Target high-friction use cases:** Support, compliance, operations first
→ **Measure process improvements:** Time saved, decisions accelerated

4: GOVERNANCE & CONTROL

Core Question: Will this pass an audit?

Quick Check:

- Access controls enforced at retrieval (matches source system permissions)
- Every answer traceable to a specific document + version
- Audit logs capture all retrieval events

If this is weak → What to do
→ **Mirror source permissions:** What users can't see in CRM, RAG can't retrieve
→ **Log everything:** Document IDs, versions, timestamps for every answer
→ **Get compliance sign-off:** Before production deployment

aimprosoft

NEED HELP GETTING THERE?

Schedule a data strategy call

contact@aimprosoft.com