

Define the real work before the demonstration

Agree the inputs, roles and non-negotiables first. Do not substitute a polished demo dataset.

DECISION OR WORK TO SUPPORT

PROVIDER A

PROVIDER B

PROVIDER C

REPRESENTATIVE SOURCE SET

USERS AND ACCESS ROLES

REQUIRED DEPLOYMENT AND DATA BOUNDARY

WHAT THE ANSWER MUST ENABLE

PASS/FAIL GATES - REQUIRED, NOT DESIRABLE

Ask each provider the same five questions

Include natural-language questions, exact identifiers, ambiguous terms and a known failure case.

Record the test conditions

Use the same source snapshot, permission role and relevance judgement for each run.

#	Question / identifier	Expected source or result	A: evidence	B: evidence	C: evidence
1					
2					
3					
4					
5					

Judge usefulness and source trace, not just whether an answer sounds plausible.

Capture evidence beyond the search result

Enter M (met), P (partial), F (failed) or U (unavailable). Attach or cite the proof separately.

Area and test	Evidence to ask for	A	B	C
Content coverage	Representative formats, source versions, failed items			
Retrieval relevance	Exact, semantic and hybrid results; agreed relevance			
Permissions	Allowed/denied roles; source-to-result enforcement			
Source evidence	Citation or passage traced back to the source			
Model/data boundary	Data path, tenancy, retention and model choices			
Connections/workflow	Target connector; approval, version and failure path			
Operations	Monitoring, change, backup, support and ownership			

Keep dated screenshots, source links or test notes alongside this worksheet.

Make the gaps and assumptions visible

Do not turn a missing demonstration into a pass. Resolve mandatory failures before ranking extras.

PROVIDER A - REQUIRED GATES PASSED / TOTAL	PROVIDER B - REQUIRED GATES PASSED / TOTAL	PROVIDER C - REQUIRED GATES PASSED / TOTAL
UNRESOLVED GAP, ASSUMPTION OR FAILURE	OWNER / PROOF NEEDED / BY WHEN	

PROVISIONAL SHORTLIST AND RATIONALE