AI SECURITY WORKING RESOURCE

RAG Boundary Review Worksheet

Check which documents reach the model, how access is enforced and what citations reveal.

Purpose

Use this workbook for RAG systems over customer, tenant, regulated, internal, or mixed-trust documents. Review retrieval and prompt assembly as one authorization path.

System / program	
Review owner	
Environment	[ ] Prototype [ ] Pilot [ ] Production
Review date / next review	

How to use this worksheet

[ ] Use one row per system, tool, credential, source, action, or test. Add rows in Word when needed.

[ ] Use factual system state. Do not place passwords, private keys, access tokens, or unnecessary personal data in the file.

[ ] Assign an owner and verification step to every blocker. Re-review after architecture, permission, or model changes.

[ ] This worksheet supports engineering review. It does not replace product-specific testing or a formal security assessment.

Handling note: keep credentials, customer identifiers, and restricted system details out of shareable copies.

WORKSHEET 01

Source and retrieval architecture inventory

Create one row per source collection. Record the source authorization model and how those rules are carried into ingestion, indexing, retrieval, and generation.

Source / owner	Users and tenants	Source authorization	Index / namespace	Trust and sensitivity

Review prompts

[ ] Separate public, internal, customer, regulated, and administrator-only collections.

[ ] Document deletion, re-indexing, permission-change, and stale-content behavior.

[ ] Flag collections that combine tenants, trust levels, or incompatible retention rules.

WORKSHEET 02

Tenant and document authorization tests

A user should never retrieve a chunk they could not open directly in the source system. Test allow and deny cases with known restricted material.

Test identity and query	Expected	Observed chunks / filters	Result and artifact	Fix owner

Review prompts

[ ] Include cross-tenant, cross-role, deleted-document, stale-permission, and missing-metadata cases.

[ ] Record the actual retrieved document and chunk identifiers, not only the generated answer.

[ ] Test omitted, empty, malformed, stale, and over-broad metadata filters.

WORKSHEET 03

Chunk lineage and metadata integrity

Trace each chunk to its source and stored permission metadata. Recheck current source access before the content is used.

Chunk / document ID	Source and section	Tenant / ACL reference	Ingest and permission state	Citation behavior

Review prompts

[ ] Preserve source document ID, section or page, tenant, trust label, ingest time, and permission reference.

[ ] Detect orphaned chunks, duplicate identifiers, stale ACLs, and documents deleted from the source.

[ ] Confirm citations do not reveal restricted titles, paths, labels, or neighboring chunks.

WORKSHEET 04

Trust labels and prompt assembly

Retrieved content can contain malicious instructions. Label it as source data and enforce server-side authorization for data access and resulting actions.

Control check	Status and notes
Prompt assembly distinguishes system policy, user request, retrieved data, and tool output.	[ ] Yes [ ] No [ ] N/A
Documents cannot redefine tool permissions, approval requirements, or the system role.	[ ] Yes [ ] No [ ] N/A
Source labels help interpretation; every resulting tool request receives independent authorization.	[ ] Yes [ ] No [ ] N/A
The agent cannot use retrieved secrets, hidden metadata, or internal instructions outside the user's authority.	[ ] Yes [ ] No [ ] N/A
Source citations remain attached after ranking, summarization, reranking, and answer generation.	[ ] Yes [ ] No [ ] N/A
Regression tests include indirect prompt injection and cross-document instruction conflicts.	[ ] Yes [ ] No [ ] N/A

Prompt assembly gaps, trust-label exceptions, and owner

WORKSHEET 05

Boundary findings and release decision

Record conditions that must hold before release. A cross-tenant retrieval or a source-authorization bypass is a release blocker.

Finding or condition	Severity	Required change	Owner	Due / verified

Decision

Release posture	[ ] Ready [ ] Limited pilot [ ] Blocked [ ] Formal assessment needed
Conditions to proceed	
Decision owner / date	
Next verification	