

Sample AI Agent Readiness Check

Technical risk/readiness review - controlled laboratory fixture

AR-2026-006 / Scenario 2

Public sample deliverable - V1.1

Date: 2026-09-01

Status	Meaning
Public disclosure posture	Controlled laboratory fixture. This is a public sample of deliverable structure and judgment presentation, not an external client engagement.
Readiness decision	CONDITIONALLY READY — FOR BOUNDED CONTROLLED EXECUTION ON THE ASSESSED PATH, SUBJECT TO PRESERVATION OF THE TESTED CONTROL CONDITIONS.
What this is not	Not a certification, audit opinion, compliance approval, safety guarantee, platform-wide assurance, or production approval.

This sample shows how an Independent AI Risk / Readiness Review can present evidence, findings, control conditions, readiness boundaries, and recommended actions to a technical or product decision owner.

1. Executive Summary

Decision question. Does the available evidence support the proposed next step for this agent within the reviewed scope and control conditions?

Short answer. The assessed path is conditionally ready for bounded controlled execution, provided the tested control conditions are preserved. This conclusion applies only to the assessed controlled path and does not establish system-wide or production-wide readiness.

Topic	Client-facing answer
What was reviewed	A controlled RAG + Agent execution path: retrieval boundary, decision basis, explicit approval, bound invocation, acknowledgement, fresh post-execution verification, and final evidence review.
What was established	Retrieval-boundary control operated successfully; the approved agent action executed on the intended target; the resulting state was independently verified; approval-to-execution lineage was preserved.
Controllability	Established for the assessed controlled path, subject to approval binding, exact target/material-parameter match, single controlled invocation, and fresh verification.
Readiness	Conditionally ready for bounded controlled execution on the assessed path, subject to preserving tested control conditions.
Main limitation	Bounded test coverage. One evaluator-run record did not retain command/stdout/stderr, creating a non-blocking reconstructability limitation.
What was not established	No general production readiness, autonomy-increase readiness, general RAG safety/quality, general agent reliability, universal control effectiveness, or human cognitive causality.

Authoritative decision

CONDITIONALLY READY — FOR BOUNDED CONTROLLED EXECUTION ON THE ASSESSED PATH, SUBJECT TO PRESERVATION OF THE TESTED CONTROL CONDITIONS.

Reader explanation

In plain language: the reviewed evidence supports a bounded next step on the specific controlled path that was assessed. The result should not be generalized to other actions, targets, failure modes, production-scale behavior, or increased autonomy without additional evidence.

2. System / Workflow Reviewed

The reviewed fixture tested a controlled AI workflow in which retrieval evidence informed a decision basis, a human approval was recorded, and one approved agent action was executed against a specific

target record in a controlled workflow system. A fresh post-execution read then checked whether the intended resulting state was present.

Workflow step	What the review looked for
Retrieval boundary	Did the retrieval path return only records inside the intended boundary for the assessed scenario?
Decision basis	Was the proposed action tied to retrieved evidence and the reviewed scenario?
Approval	Was the action explicitly approved before execution?
Bound invocation	Did the actual invocation match the approved operation, target and material parameters?
Acknowledgement	Was a target-system response received for the action?
Fresh verification	Was the resulting state independently checked after the action?

3. Scope and Assessment Boundary

In scope. The assessed controlled path for AR-2026-006 / Scenario 2: retrieval boundary, approval binding, exact target and parameter matching, single controlled invocation, acknowledgement, and fresh post-execution verification.

Out of scope. The review did not assess platform-wide readiness, production-scale behavior, broader autonomy, additional targets, different actions, failure conditions, general RAG quality, general RAG safety, general agent reliability, or legal/compliance certification.

Evidence class. Controlled laboratory fixture evidence. This is not an external client engagement and should not be presented as one.

4. Architecture / Workflow

The assessed path can be read as a bounded control sequence:

Control sequence	Client meaning
Reviewed retrieval	The assessed retrieval remained within the intended boundary for the reviewed scenario.
Decision basis	The proposed action was tied to a documented basis before approval.
Explicit approval	Execution depended on a recorded approval for a specific operation, target and material parameter.
Bound invocation	The action that was invoked matched the approved action.
Single controlled invocation	No silent retry or unapproved second action is part of the supported path.

Fresh verification	The result was checked after execution rather than inferred from acknowledgement alone.
--------------------	---

Important boundary: the technical path showed RAG PASS and Agent PASS as separate results. Those results do not by themselves create a readiness decision. The readiness decision comes from the approved human risk/readiness judgment applied to the evidence.

5. Key Findings

1. RAG retrieval-boundary control operated successfully

Element	Client-facing content
Finding	RAG retrieval-boundary control operated successfully for the assessed execution.
Evidence	Controlled retrieval-boundary evaluation = PASS.
Materiality	HIGH
Confidence	HIGH WITHIN ASSESSED BOUNDARY SCOPE
Risk / decision implication	The assessed retrieval boundary can be treated as operated successfully within the reviewed scope. This does not establish general RAG quality or general RAG safety.
Controllability	Supports the bounded evidence basis for the controlled path.
Recommended action	Preserve the relevant control condition and do not generalize beyond the assessed path.

2. Approved Agent action executed and independently verified

Element	Client-facing content
Finding	The approved Agent action was executed on the intended target and the intended resulting state was independently verified.
Evidence	Agent execution checks = PASS; 12/12 comparisons; fresh post-state verification = PASS.
Materiality	HIGH
Confidence	HIGH WITHIN ASSESSED EXECUTION SCOPE
Risk / decision implication	The reviewed action was not merely proposed or acknowledged; the intended post-action state was independently checked. This does not establish general agent reliability.
Controllability	Supports controllability only where approval binding, exact target/parameter match, single invocation and fresh verification are preserved.
Recommended action	Preserve the relevant control condition and do not

	generalize beyond the assessed path.
--	--------------------------------------

3. Approval-to-execution control linkage preserved

Element	Client-facing content
Finding	Approval-to-execution control linkage was preserved across decision basis, approval, invocation and verification. The review retained RAG and Agent as separate technical results and created no aggregate technical verdict. This finding does not establish human cognitive causality.
Evidence	Decision basis, approval, invocation, verification, control-lineage record, and technical assessment input.
Materiality	HIGH
Confidence	HIGH FOR TRACEABLE CONTROL LINEAGE
Risk / decision implication	The lineage from approval to execution remained traceable, but this does not prove human cognitive causality and does not create an aggregate technical verdict.
Controllability	Supports the conclusion that the assessed controlled path is controllable under the tested conditions.
Recommended action	Preserve the relevant control condition and do not generalize beyond the assessed path.

4. Execution-record completeness gap remains a non-blocking reconstructability limitation

Element	Client-facing content
Finding	One evaluator attempt did not retain the exact command, stdout, and stderr. The evaluator input, evaluator identity, result, exit code, timestamp, and hashes remained retained. The gap does not overturn the established PASS result, but it remains an explicit evidence limitation.
Evidence	Retained evaluator input, identity, result, exit code, timestamp, and hashes; missing command/stdout/stderr.
Materiality	MEDIUM FOR RECONSTRUCTABILITY
Confidence	HIGH
Risk / decision implication	This affects reconstructability discipline, not the established controlled-path result. Future runs should retain the complete execution record.
Controllability	The controllability conclusion remains bounded and must preserve this limitation.
Recommended action	Retain complete command/output/error records for future governed evaluator runs.

6. Risk Interpretation and Decision Implications

The positive result is decision-relevant because it shows that the assessed path can preserve retrieval boundary, approval binding, target/parameter integrity, single invocation, and fresh verification. These are the conditions that make the readiness decision bounded rather than general.

The main decision implication is not “the system is safe” or “the agent is production-ready.” The implication is narrower: the available evidence supports bounded controlled execution on the assessed path if the tested control conditions remain in place.

The reconstructability gap matters because a complete governed execution record should retain command, output and error streams for later review. The gap is non-blocking in this case because other retained evidence supported the result and final validation closed the review with an explicit non-blocking gap.

7. Controllability Assessment

Authoritative controllability statement

ESTABLISHED FOR THE ASSESSED CONTROLLED PATH — SUBJECT TO PRESERVED APPROVAL BINDING, EXACT TARGET / MATERIAL-PARAMETER MATCH, SINGLE CONTROLLED INVOCATION, AND FRESH POST-EXECUTION VERIFICATION.

Client-readable control conditions

Condition	Meaning for the assessed path	Must be preserved
Approval binding	Execution is tied to a specific prior approval.	Yes.
Exact target match	The action is applied to the intended target, not an adjacent object.	Yes.
Material-parameter match	The material action parameters match what was approved.	Yes.
Single controlled invocation	No unapproved retry or duplicate action is included in the supported path.	Yes.
Fresh verification	The final state is checked after execution rather than inferred from acknowledgement alone.	Yes.

8. Readiness Decision

Authoritative readiness decision

CONDITIONALLY READY — FOR BOUNDED CONTROLLED EXECUTION ON THE ASSESSED PATH, SUBJECT TO PRESERVATION OF THE TESTED CONTROL CONDITIONS.

What this means

The assessed path can move forward only as bounded controlled execution under the tested control conditions. The decision does not authorize unrestricted production use, broader autonomy, additional action types, new targets, or use outside the assessed path.

What this does not mean

- It does not certify the system.
- It does not approve production-wide readiness.
- It does not establish general agent safety or reliability.
- It does not establish general RAG quality or RAG safety.
- It does not justify an autonomy increase by itself.
- It does not assess risk outside the reviewed path.

9. Recommendations

Theme	Client-readable recommendation	Current status
Preserve tested controls	Keep explicit approval binding, exact target/parameter binding, single controlled invocation, no silent retry, and fresh post-execution verification.	Active condition
Do not generalize readiness	Do not treat this result as production-wide or system-wide readiness.	Active claim boundary
Do not use alone for autonomy increase	Broader autonomy, additional targets, different actions, failure conditions and production-scale behavior require separate evidence.	Active decision boundary
Improve execution-record completeness	For future runs, retain exact evaluator command, stdout and stderr as part of the governed execution record.	Active evidence/reconstructability improvement
Final validation / reconstructability closure	Final validation was subsequently completed. Keep this as provenance; do not present it as an outstanding client action.	Satisfied closure condition

10. Evidence Basis and Limitations

Evidence basis

The review relied on a controlled laboratory evidence package containing retrieval-boundary evaluation, decision-basis material, recorded approval, bound invocation, target-system acknowledgement, fresh post-execution state verification, comparison checks, final technical verdict material, and final validation records.

Evidence area	Client-facing role
Retrieval-boundary evaluation	Supports the finding that retrieval-boundary control operated successfully within assessed scope.

Decision basis and approval record	Supports the finding that execution was tied to a specific approved action.
Invocation and acknowledgement record	Shows the action was sent and a response was received; acknowledgement alone is not treated as final-state proof.
Fresh post-execution verification	Supports the finding that the intended resulting state was independently verified.
Comparison and verdict records	Support the controlled-path integrity checks.
Final validation record	Confirms final validation completed with explicit non-blocking reconstructability gap.

Limitations and caveats

- Residual uncertainty remains from bounded test coverage and the evaluator attempt-2 execution-trace completeness gap.
- Risk outside the assessed path is not assessed.
- General RAG quality PASS is not supported.
- General RAG safety is not supported.
- General Agent reliability is not supported.
- Unrestricted production readiness is not supported.
- Readiness for autonomy increase is not supported.
- Readiness outside the assessed path is not supported.
- Human cognitive causality is not established.
- Universal effectiveness of the tested controls is not supported.

The execution-record completeness gap is bounded: one evaluator attempt did not retain command/stdout/stderr. These missing items must not be reconstructed or implied to have been retained.

Publication caveat. This document is a public sample based on a self-built controlled laboratory fixture. It illustrates the form and discipline of a client-facing deliverable. It is not an external independent client engagement and does not authorize public claims beyond the stated sample scope.