Evaluate a defined workflow, not a market promise.

A practical guide for reviewing one AI-supported building task before assigning more authority.

Name the intended use, accountable owner and decisions affected by an incorrect output. Define the information the workflow can use and the actions it may propose or perform. A product demonstration, a benchmark result and a vendor description provide different kinds of evidence; none alone establishes the suitability of a local deployment.

Review record	Minimum entry	Decision it supports
Use and consequence	Task, users, asset scope, affected people and plausible failure effects	Whether the proposed scope is acceptable to investigate
Evidence	Source/version, claim, uncertainty and reviewer	Whether a claim is supported for this use
Authority	Recommend, approve, execute, stop and restart roles	Who can authorize a change
Evaluation	Representative inputs, reference answers, criteria and retained results	Whether a bounded release is justified
Continuing review	Owner, change triggers, monitoring and exit procedure	Whether use remains appropriate

NIST's voluntary AI Risk Management Framework and Generative AI Profile provide useful cross-sector foundations. The process in this guide is an application to building decisions, not a certification or a universal checklist required by NIST. [N1, N2]

This guide replaces broad market-superiority and unverified platform-performance assertions with a reviewable method. It does not rank vendors or imply that an eight-product sample describes the whole market.

Make every material claim traceable.

An evidence register should let another reviewer reproduce what was checked.

Claim category	Acceptable record	What to avoid
Published capability	Product/version, exact public source, access date and relevant passage	Treating a roadmap or marketing slogan as an observed feature
Test observation	Input/version, procedure, result, environment and reviewer	Generalizing a demonstration into live operational protection
Outcome or ROI	Baseline, period, comparable population, total cost and attribution limits	Using estimates or fictional examples as client results
Relationship or recognition	Owning record and permission for the exact public wording	Invented academic affiliations, analyst recognition or patent status
Legal or standards claim	Applicable jurisdiction/edition and qualified review	Saying a tool or generic checklist ensures compliance

REPRODUCIBLE VENDOR REVIEW

Record the exact vendor/product, version or service date, evaluated task, access level, source URLs, observed evidence and unresolved questions. If identity must remain confidential, keep the underlying register available to authorized reviewers and limit the public conclusion accordingly. An anonymized scorecard without its evidence cannot support a broad public ranking.

Distinguish an unavailable answer from a failed test and a feature that is outside scope. Report the reviewed population and date. Do not convert absence of accessible evidence into a claim that a competing product lacks the capability.

> Citation presence is the start of a check, not its conclusion. Confirm that the source actually supports the claim, and retain the limits of its population, environment and period. [N2]

Separate advice from operational authority.

A recommendation should not silently become a command, a record change or an approval.

Role	Responsibility	Fictional boundary
AI-supported workflow	Prepare a source-linked recommendation and uncertainty	Read approved exports; propose maintenance priorities
Evidence reviewer	Check sources, identity and reference cases	Accept, correct or hold the recommendation
Decision owner	Accept scope, tests, residual uncertainty and release conditions	Approve a supervised read-only trial
Designated operator	Control permitted operational actions and stop work	Retain all live BAS approvals and override rights
IT/OT access owner	Authorize access and verify agreed safeguards	No control-system connection or write permission in this example

Define prohibited actions, escalation contacts, backup roles and restart conditions. Test access in an agreed environment without compromising operational reliability. A simulated denied write confirms that particular test result; it does not establish protection of a live BAS or authorize connecting to it. [N3]

CHANGE MEANS A NEW DECISION

Adding a data source, user group, model version, integration or execution right may invalidate the original assessment. Record the change, assess its effect and rerun the relevant checks. Retain the prior accepted configuration and a way to disengage the workflow.

NIST SP 800-82 Rev. 3 is the final 2023 OT guide cited here. Revision 4 was an initial public draft in September 2026. Draft material should be identified as such rather than described as the final standard.

Use tests that expose missing evidence.

Fictional example: 12 recommendation cases for a supervised read-only maintenance trial.

Cases	Observation	Result
01-08	Required asset identity and sources support the recommendation	8 pass
09-10	Evidence is missing; the workflow correctly holds and escalates	2 pass as correct holds
11-12	Recommendations omit required supporting references	2 fail
Release conditions outside those cases	Asset reconciliation and operator handover remain incomplete	Trial release remains on hold

One complete trace: A-01 locates AHU-07 in the East Wing; W-01 locates AHU7 in the West Wing. Expected behavior: flag the conflict and hold the affected recommendation. Observed behavior: hold with both references retained. Result: pass for this case. The underlying identity issue still requires owner resolution.

A USEFUL RELEASE RECORD

Decision: hold the trial. **Owner:** facility manager. **Conditions to reopen:** resolve or formally exclude affected identities, correct unsupported outputs, rerun affected tests and accept the operator checklist and escalation route. Record actual names and calendar dates in a real project.

Ten passes out of 12 do not override the two failures or the other release conditions. These fictional counts demonstrate decision logic; they establish no operating savings, incident reduction or live-control assurance.

Monitor the work the owner accepted.

Acceptance is tied to a version, scope and operating context. Keep the evidence useful after release.

Monitor	Record	Trigger for review
Source and asset changes	Identity conflicts, stale records and unresolved issues	Inputs no longer match the accepted scope
Output quality	Unsupported outputs, missed issues and correct holds	Failure pattern changes or agreed tolerance exceeded
Operational burden	Checking, correction, support and exception effort	Total effort outweighs useful value
Authority and access	Changed permissions, attempted actions and approvals	New execution rights or boundary failures
Outcome and accountability	Completed work, closure evidence and observed result	Benefit unmeasured, owner absent or handover fails

Agree who can pause the workflow and how the team returns to an accepted process. Review unfavorable evidence as carefully as successes. Continuing use, expansion, correction and retirement are all legitimate outcomes of an evidence review.

SOURCES AND SCOPE

[N1] NIST AI Risk Management Framework 1.0, AI 100-1. January 2023. Voluntary cross-sector guidance.

[N2] NIST Generative AI Profile, AI 600-1. July 2024. Context-dependent suggested actions.

[N3] NIST SP 800-82 Rev. 3, Guide to OT Security. Final, September 2023; building automation is within scope.

[C2] Cognitive Corp, Building Decision Toolkit v2.0. 31 October 2026. Five-page working form with fictional examples.

Version note: older market-positioning, quantified case and societal benefit-model material is not evidence for this guide. No institutional partnership, analyst endorsement, patent portfolio, validated ROI or guaranteed compliance is asserted.