• BUILD VERIFICATION / WORKSHEET 1 OF 5

Check AI advice before you build on it

Define one result and preserve the working reference

Use this pack before you rely on AI-generated instructions for a business workflow. Complete the five worksheets, test the assumption most likely to stop the work, and keep the evidence beside the advice. Pages 6–8 provide fictional tests, video context and dated Microsoft checks.

ONE OBSERVABLE OUTCOME AND WHO WILL ACCEPT IT

Project brief	Your answer
Project / candidate version / date	
Business owner / acceptance reviewer	
User, trigger and approved inputs	
Output, destination and completion evidence	
In scope: exact product and component	
Out of scope: writes, integrations, decisions	
Allowed data / authorised test accounts	
Constraints: access, tools, time, spend, retention	

WORKING REFERENCE: VERSION, LOCATION AND HOW TO REPEAT ITS BASELINE CASE

PROTECTED BASELINE: WHAT MUST STILL WORK, BACKUP LOCATION AND RESTORE OWNER

Before changing anything: retain an unchanged copy, reproduce one baseline case and share only the relevant sanitised reference. Record what it proves and what it leaves untested.

Version 1.0 · Sources checked 5 October 2026 · Method and templates, not a tenant-tested implementation. The expense examples are synthetic designs with no observed results.

• BUILD VERIFICATION / WORKSHEET 2 OF 5

Test the largest risk first

Choose a test that could disprove the assumption

Pick the failure that would stop the business outcome. Make the smallest authorised test that can expose it before adding easier features. Repeat this sheet when the largest unresolved risk changes.

ASSUMPTION AND WHY ITS FAILURE WOULD BLOCK THE OUTCOME

SUPPORTING SOURCE OR REFERENCE AND WHAT REMAINS UNCERTAIN

SMALLEST TEST: STEPS, ISOLATED ENVIRONMENT, SYNTHETIC INPUTS AND ROLES

EXPECTED OBSERVABLE RESULT AND MAXIMUM WAIT

EVIDENCE: RESPONSE, DURABLE RECORD, SIDE EFFECTS, RUN ID AND TIMESTAMPS

ACTUAL RESULT AND EVIDENCE LOCATION

Test status: ☐ Not run ☐ Pass ☐ Fail ☐ Inconclusive
Decision: ☐ Continue ☐ Change design ☐ Reduce scope ☐ Stop

OWNER, NEXT ACTION AND STOP CONDITION

A synthetic test to try

Submit EX-001 revision 1 twice while the first response is delayed, reusing its logical request ID. Also submit a legitimate report with a different ID. Expected: one accepted package for EX-001, no duplicate approval or delivery, and the separate report can proceed. Actual: Not run.

Inspect the durable effects. A disabled button or a status flag alone does not prove duplicate prevention. Reconcile an ambiguous timeout before retrying. See [3] on page 8.

• BUILD VERIFICATION / WORKSHEET 3 OF 5

Keep the advice and evidence

One record per consequential claim

Preserve the original wording. Put corrections beside it. Separate the advice, the implementation and the result you observed. Copy this page for each claim; attach longer excerpts rather than shrinking the writing space.

Advice ID: _____ Date: _____
Tool / model / version, or unknown: _____

EXACT QUESTION AND THE MINIMUM RELEVANT CONTEXT

EXACT ADVICE OR LOCATION OF THE UNCHANGED SOURCE EXCERPT

REQUIREMENT THIS ADVICE IS MEANT TO SATISFY

OFFICIAL SOURCE, SECTION, CHECKED DATE AND APPLICABILITY TO THIS ROUTE

WHAT THAT SOURCE SUPPORTS AND WHAT IT DOES NOT ESTABLISH

PROPOSED COMPONENT, CONFIGURATION VERSION AND CHANGE OWNER

TEST ID, EXPECTED RESULT AND OBSERVED RESULT OR EVIDENCE LINK

Evidence status: ☐ Plausible only ☐ Documentation supported
☐ Reproduced here ☐ Failed ☐ Unresolved

DIFFERENCE OR REMAINING GAP, RISK OWNER AND NEXT DECISION

Documentation supported does not mean tested. Reproduced here applies to the recorded inputs, identity and environment. Mark paraphrases as paraphrases. An inconclusive test cannot be counted as a pass.

• BUILD VERIFICATION / WORKSHEET 4 OF 5

Record the acceptance test

Write the expected result before running the case

Use the synthetic catalogue on page 6 as a starting point. Replace its expectations with approved business rules where needed. Copy this page for each required test. Test only in an authorised isolated environment.

Test ID: _____ Date: _____ Tester: _____
Candidate version: _____ Account role: _____

REQUIREMENT AND RISK COVERED BY THIS TEST

ENVIRONMENT, CONNECTION IDENTITY, INPUT ID AND PREREQUISITES

STEPS TO REPRODUCE AND EXPECTED BUSINESS RESULT

ACTUAL RESULT AND TIME OBSERVED

EVIDENCE LINKS: RESPONSE, SAVED RECORD, RUN ID AND SIDE EFFECT COUNTS

Result: ☐ Pass ☐ Fail ☐ Blocked ☐ Not run ☐ Inconclusive

ISSUE, OWNER, NEXT ACTION AND RETEST DATE

Set the release gates before testing

☐ Every required test has evidence that matches the accepted business result

☐ No unauthorised data access or changes; no duplicate effect in required replay and concurrency tests

☐ Completion wording matches saved state; support can trace the run; fallback and in-flight handling are rehearsed

A weighted average cannot cancel a permission or data-integrity failure. Add load tests if throughput matters. Record real approval deadlines; a short simulation cannot prove a multi-day production run.

• BUILD VERIFICATION / WORKSHEET 5 OF 5

Decide whether to release

Keep the boundary and recovery plan visible

Candidate / date: _____ Reviewer: _____
Required tests passed: ____ / ____ Failed: ____ Unresolved: ____

PILOT OR RELEASE BOUNDARY: AUDIENCE, VOLUME, DURATION AND ALLOWED ACTIONS

EVIDENCE REVIEWED, ACCEPTED RESULTS, MATERIAL ERRORS AND HUMAN REWORK

REMAINING RISK AND OWNER; SPEND EXPOSURE, MONITORING AND STOP THRESHOLD

Decision: ☐ Hold ☐ Limited pilot ☐ Release within this boundary ☐ Revert

REASON, CONDITIONS FOR CHANGING THE DECISION AND APPROVALS STILL NEEDED

Prepare the stop and restore procedure

☐ Name who stops intake and how users reach the manual fallback

☐ Inventory running, pending and scheduled work; decide how each completes, is cancelled or is reconciled

☐ Preserve authorised logs and business state; restore the approved baseline

☐ Reconcile partial writes, approvals and notifications; restoring code does not undo these effects

☐ Run the baseline smoke test, recheck access and record who may reopen intake

RECOVERY PROCEDURE LOCATION, STOP OWNER AND SUPPORT CONTACT

NEXT REVIEW DATE OR TRIGGER: SOURCE, MODEL, CONNECTOR, PERMISSION OR FAILURE

Microsoft note: turning a cloud flow off prevents new runs; pending and in-progress runs continue [5]. Plan their handling explicitly. This pack's fictional workflow stays on Hold because no tests have been executed.

• SYNTHETIC EXPENSE TESTS / ALL NOT RUN

Test the failure paths

Twelve original examples to adapt to your business rules

Fixture: EX-001 revision 1; fictional employee Avery, reviewer Morgan and peer Riley; CAD 42.00 + CAD 18.00 = CAD 60.00; two placeholder receipts; request ID REQ-001. These are new test designs, not the video's 11 questions. No real reimbursement or external email is authorised by this pack.

ID / case	Expected business outcome	Capture as evidence
T01 · Valid report	One correct saved package; CAD 60.00, two receipts, stable ID and matching revision	Response, package, IDs and state history
T02 · Missing required data	Clear validation; no package treated as valid and no silent defaults	Validation, stored state, approval count
T03 · Same request repeated	Same logical request recognised; no second business effect	Request ID, record and delivery counts
T04 · Concurrent submission	One valid claim or a clear conflict for the same report revision	Both run IDs, times, state, effect counts
T05 · Delayed write	Honest pending response until completion; final result visible through an authorised status route	Response/write times, final state and ID
T06 · Timeout then retry	Original state reconciled first; replay does not duplicate a completed effect	Client error, original run and saved result
T07 · Intended identity	User and connection identity perform only their approved operations	Actor, connection and access outcomes
T08 · Unauthorised peer	App and direct resource access denied as policy requires; no leakage in errors or logs	Separate-account checks and resource result
T09 · Connector failure	Visible technical failure; no false success or onward approval; operator can diagnose	Error path, business state, correlation ID
T10 · Partial write then recovery	Existing effects reconciled; recovery completes or explicitly compensates without duplicates	Before/after records and recovery outcome
T11 · Edit or replay old revision	Approver sees the intended versioned package; stale actions cannot overwrite a newer accepted revision	Package version and stale-action result
T12 · Stop with work in flight	Intake stops; pending work is accounted for; baseline or manual fallback works; audit evidence retained	Run inventory, restore steps, smoke test

For T03–T06, also submit a separate legitimate report with a different request ID. Duplicate controls must not block valid new work. For T08, use approved test identities and routes, never another person's credentials.

Expected results are design requirements, not observed behaviour. All twelve tests start Not run. Use page 4 to record each actual result, evidence, issue owner and retest. Add an overdue-approval case when approvals are in scope.

• RELATED VIDEOS / SOURCE LIMITED CASES

Apply the lessons carefully

Use the cases to choose checks for your own workflow

Preserve the working reference

In the August retrospective, the creator describes a four-week, unshipped Copilot Studio customer-service build and an earlier working version. The reported lessons include broad project scope, an untested largest risk, withholding the working reference and too many sign-offs. These are the creator's retrospective, not independently reproduced measurements.

Apply it: name one product boundary on page 1, preserve the working reference and reproduce a baseline. On page 2, test the assumption that could block the outcome. Use meaningful checkpoints for scope, evidence and release while keeping required safeguards.

V1 · I Let Claude Build My Copilot Agent. It Went Off the Rails!! · 13 Aug 2026

Keep a record of what happened

The September description reports 11 hard questions about an expense workflow spanning Copilot Studio, Power Automate, Power Apps and SharePoint. Its chapters point to useful checks. The exact questions, original advice and final test outputs were not located in accessible sources, so this pack does not reproduce them or assign a score.

Apply it: take one exact recommendation and fill in page 3. Then compare the response, the saved record and the run history. A chapter title tells you where to investigate; it does not establish the result.

V2 · The AI Learning Advisor Isn't What You Think (Tested on a Real Build) · 29 Sep 2026

Go straight to the relevant chapter

6:17 · Evidence scoring

8:48 · The build sheet

9:38 · Queue versus write

10:48 · Error handling

11:58 · Submission lock

12:38 · Duration

14:21 · Employee profile

Evidence boundary: creator descriptions and the supplied chapter list were reviewed, not a full audiovisual record. The original build repository, run logs and later success claims were not independently checked. A related draft uses a different architecture and is not treated as final-video evidence. No screenshots or reproduced product results are claimed.

• MICROSOFT REFERENCE CHECK / 05 OCT 2026

Verify the exact route

Documentation is a starting point for a tenant test

Accepted is different from committed

HTTP-triggered pattern: Microsoft's asynchronous Response returns 202 with a Location for status. This acknowledges processing; it does not establish that the business record is saved. Check the matching durable result before saying the work is complete. [1]

Copilot Studio standard-harness tool flow: the documented route uses When an agent calls the flow and Respond to the agent, with asynchronous response Off and a 100-second action limit. Later actions may continue after the reply. The reply can report an accepted or pending state; it must not imply that later writes have finished. Do not copy the HTTP toggle setting into this route. [2]

Verify failure and access boundaries

Microsoft recommends idempotent handling of duplicate inputs [3] and documents failure, timeout and retry handling [4]. For this workbook, inspect the durable effect and recovery result. A green run, a locked form or a caught error alone does not prove the business outcome.

Record the caller and the actual connection identity [6]. Check app, flow and underlying data access separately [7]. Use representative non-owner accounts. App filtering or a hidden control alone is not evidence that the resource is protected.

Know what keeps running

The general flow-duration limit is 30 days, including pending approvals. General synchronous request limits are 120 seconds, subject to connector caveats; this does not replace the 100-second tool-flow requirement above. Turning a cloud flow off leaves pending and in-progress runs running. Include those runs and existing side effects in the stop plan. [5]

Microsoft's long-running approval pattern uses Dataverse and separate flows. If Dataverse is outside your approved scope, resolve that design choice before release. A shorter test does not prove indefinite waiting. [8]

Official references

[1] Microsoft Learn · HTTP asynchronous responses

[2] Microsoft Learn · Create an agent flow as a tool

[3] Microsoft Learn · Troubleshoot triggers and duplicate effects

[4] Microsoft Learn · Error handling in Power Automate

[5] Microsoft Learn · Flow limits and turning flows off

[6] Microsoft Learn · Flow sharing and run-only connections

[7] Microsoft Learn · Sharing canvas apps and underlying resources

[8] Microsoft Learn · Long-running approvals

Sources checked 5 October 2026. No flow, connector configuration, tenant permissions or example test was executed for this pack. Recheck the source and your environment after changes. Corrections: hello@automatealready.ca.