

Local release rehearsal

1

EXPERIMENTAL OPEN LAB, NOT INCLUDED IN THE PAID HANDOFF REVIEW

Rehearse a workflow change without touching an external system

This experimental open-lab CLI runs supported baseline and candidate workflows with synthetic fixtures in separate Docker environments. HTTP actions are redirected only to a local capture service. This screen verifies the three generated files in your browser and presents the differences in plain language.

2

What does it prove?

Observed means that the transformed synthetic request reached the local capture service. It does not prove node execution order, approvals, retry or loop behaviour, production safety, or a penetration-test result. A person makes the release decision.

How the n8n AI Security Regression Gate is built

When an n8n workflow changes, a security behaviour that must stay the same can break. This gate tries the change in a local, closed environment without touching a live system, and writes down what was observed together with its limits.

1 What this work is

A static review shows what is written in a file. It does not show whether an unsigned request is still rejected after a change, or whether an action goes out without waiting for approval. I wanted to see that before release, without touching any real system.

The result is an open-source command-line project called n8n-ai-agent-security-lab. It is MIT licensed, runs on Node.js 20 and needs no extra package installation. Workflow files stay on your computer. The tool sends no request to an AI service.

Static audit	Reports risky paths in an exported workflow. It runs nothing.
Regression gate	Sends the synthetic requests in a contract file to an isolated staging webhook and checks the behaviour that must not change.
Local release rehearsal	Runs the baseline and candidate in separate Docker environments and records the action difference.
GitHub PR gate	Reviews the workflow files in a repository on every pull request without running them, and leaves evidence files.

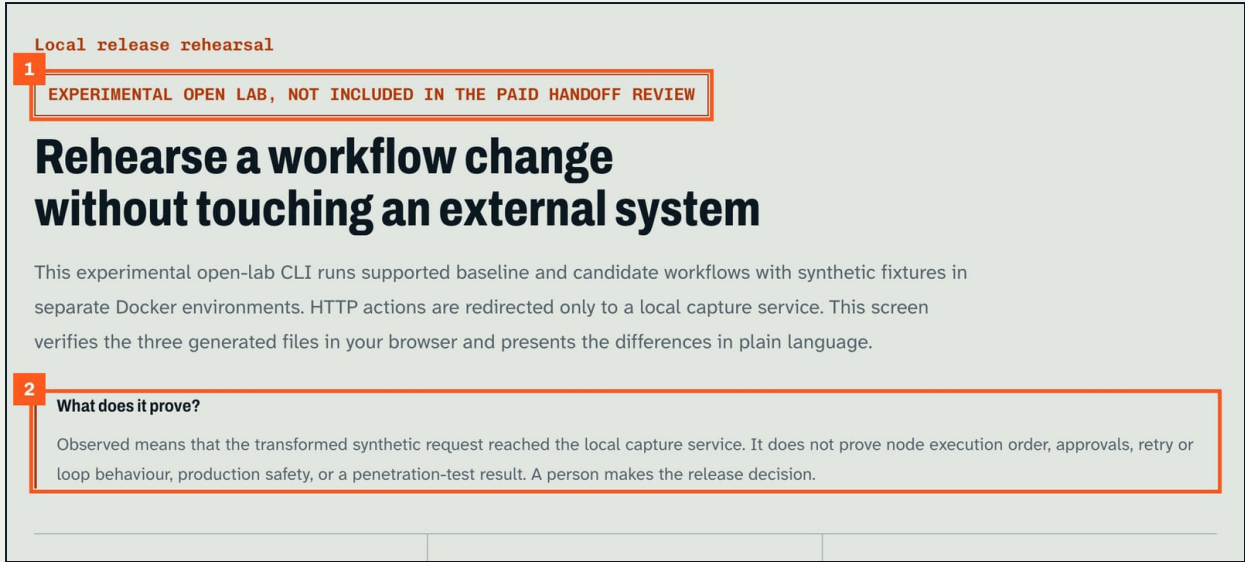
A few words

Regression	A behaviour that used to work correctly breaks because of a new change.
Baseline	The workflow version in use today, approved before.
Candidate	The new version you want to publish.
Staging	A test environment that is not connected to real systems.
Fixture	Sample input made for a test, with no real data in it.
Contract	A small JSON file that says how the staging workflow must answer each request.
Capture service	A local listener that accepts HTTP requests during the rehearsal instead of the real destination.

2 What runs where

All execution happens on your computer. The CSINT site only verifies the three files made by the rehearsal, in browser memory. The site runs no Docker and no n8n.

1. Input	2. Local CLI	3. Docker	4. Three files	5. Browser	6. Person
One baseline, one candidate and one to three sanitised fixtures.	Reads the files within limits and checks the supported scope.	A separate, externally closed n8n environment per version.	summary.json, report.md, evidence-manifest.json.	SHA-256 bindings are verified on the site, with no upload.	A person makes the release decision.



The top of the experimental rehearsal tab on the ai-agent-audit page.

- 1

Lab label
The page marks this part as an experimental open lab. It is not included in the paid handoff review.
- 2

What it proves
The meaning of observed sits above the result, so the limit is read before the result. The screen never gives automatic approval.

3 The regression gate and its eight checks

The gate runs the static audit first. Then it sends the tests in the contract to a staging webhook in order. Each test checks the expected status code and JSON fields. A separate state endpoint uses a counter to confirm that no action went out before approval.

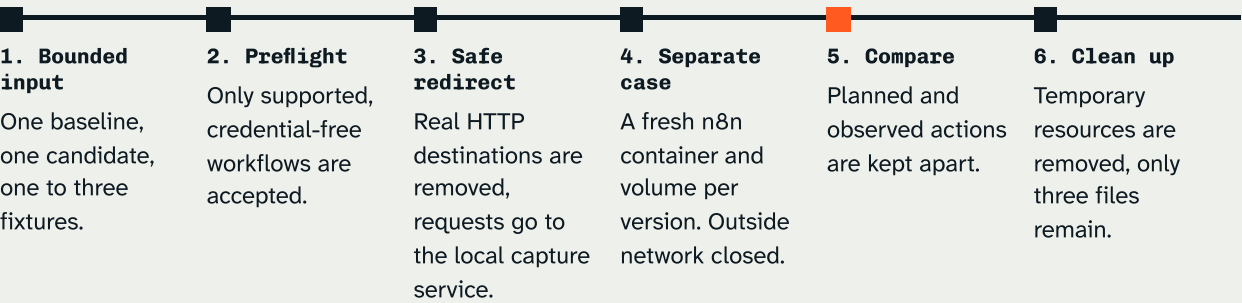
Missing authentication	Must be rejected.
Invalid staging signature	Must be rejected.
Missing or unexpected fields	Must be rejected.
Prompt injection marker	Must be held behind approval.
Invalid approval token	Must be rejected.
Valid request	Must be queued with no outside action.
Repeated request ID	Must return the same operation.
Unsupported method	Must be rejected.

By default the target can only be a local address. A remote staging target needs three things together. They are the `--allow-remote` flag, an exact host name list and narrow path prefixes. Redirects are not followed. Reports do not keep request headers or bodies.

The site used to replay a recorded run of the gate on a synthetic test workflow. Because its input was synthetic, I took it down on 2 October 2026; you run the gate against your own staging environment.

4 Stages of the local release rehearsal

The rehearsal runs with the command `node bin/rehearse.mjs`. It takes the baseline, candidate, fixtures and an output folder. Docker Desktop must be running.



The supported scope is narrow on purpose

Entry	One Webhook node, version 2.1, with a literal POST path.
Middle step	Set nodes with literal, sanitised values, version 3.4.
Action	Credential-free HTTP Request nodes, version 4.2, with a literal POST, PUT or PATCH destination.
Shape	One acyclic main[0] path, at most 20 active nodes.
Stopped	AI, Code, Execute Command, credentials, subworkflows, community nodes, dynamic destinations, cycles and unknown versions stop before Docker starts.

The checker on the site opens the three files a rehearsal produces, in the browser. I do not publish a sample rehearsal bundle, because the one I had was generated from synthetic fixtures. When you give it your own rehearsal files, the result opens with these parts.

Result	The result code and, below it, whether the SHA-256 bindings of the three files match. A match only shows the bundle is consistent with itself.
Planned, observed, unresolved	The three counts are kept apart. If unresolved is not zero, the evidence is incomplete.
Observed action difference	Which node changed and how, in one plain sentence.
Containment	Containment and cleanup state, the number of outside DNS queries and whether anything was forwarded out.

5 Result codes and the word observed

0, REHEARSAL_COMPLETE	Every supported case finished, the expected evidence was observed and there is no action difference.
2, REVIEW_REQUIRED	The cases finished and a meaningful action difference exists.
3, BLOCKED_UNSUPPORTED	A workflow is outside the supported scope. Docker never starts.
4, INCOMPLETE_EVIDENCE	Expected capture evidence is missing or unresolved.
5, CONTAINMENT_FAILURE	A DNS, forwarding, unexpected action, inventory or cleanup check failed.

Observed only means that a transformed synthetic request reached the local capture service. It does not prove node order, that an approval node really ran, retry or loop behaviour, delivery to the original destination, readiness for production, or that the two workflows are fully equivalent.

6 Design decisions

1 Keep the scope narrow

Instead of guessing how to run something unsupported, I stop it. If an unknown version, a dynamic destination or a credential shows up, the result is BLOCKED.

2 Fail closed

An unexpected DNS query, outward forwarding or a cleanup error fails the rehearsal. When there is doubt, no positive result is written.

3 Never the same environment for both versions

The baseline and candidate use separate containers and volumes. Transformed files stay bound to their original SHA-256 values.

4**Only three files are published**

Raw workflows, fixture bodies, request and response content, header values, credentials and Docker streams are never written to the output.

5**A person decides**

No output approves a release. Even the site leaves the decision to a person after it verifies the bundle.

7 Limits

- ✗ This is an experimental open lab. As the page says, it is not included in the paid handoff review, and no separate runtime pilot is sold.
- ✗ The result covers only the supplied workflow, contract and test scenarios. It is not a penetration test, a compliance certificate or a security guarantee.
- ✗ The prompt injection scenario calls no language model. It only tests the workflow controls around approval.
- ✗ The generic gate does not read the deployed workflow back from n8n. The operator must confirm separately that staging really runs the fingerprinted candidate.