Rehearse an AI incident

24 evidence-led exercises | Content v0.4 | 7 November 2026

QUICK START: Choose one exercise, not the whole library. Allow 90 minutes. Read this page and the timing page; the facilitator reads that exercise's private pages before play. Use made-up data only. No real operational action is authorised by this pack.

What you leave with

A bounded account of what happened, decisions with owners and evidence limits, words you can adapt for a real escalation, a tested-on-paper containment plan, a prioritised fix plan and a first-ten-minutes card. These are rehearsal outputs, not assurance that controls work.

What you need

Paper, pens and a timer. Use these two common pages and the eight-page exercise. Keep facilitator pages private until their release point. The pack works offline with no account, form, email submission or network connection. Completed sheets stay locally under your normal record rules. No real secrets, personal records or private account credentials belong on these sheets.

Opening words

"This is a rehearsal. Every exhibit is made up. We will decide using what is supplied, not facts we wish existed. State a decision, owner and evidence check. We can act under uncertainty, but we cannot make uncertainty disappear. Nobody is being graded. If we discover a live incident, we stop play and use our real procedures."

Use the right authority

No legal advice, compliance conclusion, safety certificate or universal technical operating instruction is supplied. Use qualified people and your established procedures for specialist decisions. Communication examples are rehearsal drafts, not approval to send. A workshop cannot authorise a real payment, access change, notification or customer remedy.

Pick a mode

3-6: facilitator plus exercise roles; combine hats as shown in each kit. Two: A facilitates and holds evidence, B holds the other roles, then both challenge decisions. Solo: cover later exhibits, injects and facilitator notes; write each answer before revealing the next. An editorial walkthrough is possible alone, but it is not a live team test.

© Almost Magic Tech Lab. For your organisation's own use only. Please do not copy, share or sell this material.

Run one exercise in 90 minutes

Time	Facilitator action
Before	Read private pages 3-6 of the kit. Mark the two branch consequences you may need. Choose a recorder. Do not invent additional technical facts.
00-10	Read the opening and roles. Release exhibit A. Ask players to state the operational decision before collecting more information.
10-20	Read inject 1. Release exhibit B. Record a decision, owner and what would disprove it.
20-30	Read inject 2. Release exhibit C. If a listed branch matches the decision, read its consequence as a new fictional event. Otherwise keep the base timeline.
30-40	Read inject 3. Rehearse the holding statement and escalation aloud. Test the team's assumed stop boundary.
40-50	Read inject 4. State the final unknown. Do not supply a missing record to rescue the team.
50-65	Use the four insight questions and strong/weak/misleading answer guide. Ask players to amend one decision after a counterexample.
65-80	Complete the evidence, control and decision sheets. Pick the top three fixes; state what would prove each worked.
80-90	Debrief, complete the three-label scorecard and keep card, and set the day-90 comparison.

Branch rule: consequences are conditional rehearsal events, not predictions about a real system. They expose a trade-off. Do not call every broad restriction wrong or every narrow restriction right. Accept alternatives that name the authority, side effects, missing evidence and a falsifiable check.

The only scorecard labels are "we decided", "we recorded", "we could not confirm". No grades, percentages or certificates. At 90 days compare evidence with this run, not confidence. If a new scenario is used, do not claim like-for-like improvement.

© Almost Magic Tech Lab. For your organisation's own use only. Please do not copy, share or sell this material.

18 | The change nobody released

1 / 8 | PARTICIPANT BRIEF AND ROLE CARDS | All records, numbers and quotations are fictional.

Silent supplier-model update - Manage changed behaviour when the dependency changes outside your release process.

Read aloud

"A supplier-hosted AI service starts returning a different answer format overnight. Your downstream workflow drops an exception field and treats omitted values as approvals. Your own application release did not change."

Decision to bring to the table

Within the first discussion, choose the interim operating boundary and name its decision owner. Separate what stops further impact from what would establish past impact. You may choose a restriction while facts remain missing; record its side effects and next review.

Decision lead

Set the interim reliance limit and supplier escalation. State what you can decide in this rehearsal and what needs another authorised person. Do not assume a job title creates legal authority.

Integration lead

Compare response contracts and fail-closed behaviour. State what you can decide in this rehearsal and what needs another authorised person. Do not assume a job title creates legal authority.

Operations lead

Review affected workflow decisions. State what you can decide in this rehearsal and what needs another authorised person. Do not assume a job title creates legal authority.

Evidence lead

Preserve response samples with version and time metadata. State what you can decide in this rehearsal and what needs another authorised person. Do not assume a job title creates legal authority.

Modes

3 people: facilitator holds evidence, one player is decision lead and one combines the two other operational roles. 4: facilitator holds evidence, each player holds one remaining role. 5: separate facilitator and four roles. 6: add an observer who records unsupported assumptions, not a rescuer who invents facts.

Two: A facilitates and holds evidence. B holds the other three roles and announces the current hat. A reveals only the current exhibit/inject. B writes one decision before the next reveal. Both challenge a decision during debrief. Keep the 90-minute clock.

Solo: cover pages 3-6 and later exhibit sections. Read the opening and exhibit A, write a decision, then reveal each inject and its exhibit. Answer as decision lead, operator and affected person separately. Only then uncover facilitator calibration. Flag every success depending on invented facts.

Do not browse, contact anyone or operate a real system during play. If an exercise raises a live concern, stop the rehearsal and use the established incident route.

© Almost Magic Tech Lab. For your organisation's own use only. Please do not copy, share or sell this material.

18 | The change nobody released

2 / 8 | CONTROLLED-RELEASE EVIDENCE DOSSIER | All records, numbers and quotations are fictional.

A | Response diff

Release at opening. Yesterday the API returned exception as review_required. Today it returns exception as omitted with an ordinary status string. Application version unchanged. Local parser treats omitted exception as approval.

B | Supplier exchange

Release with inject 1. Supplier says endpoint unchanged and new model equivalent. No old model is available for pinning. Rollout time is a broad overnight window. Contract or policy interpretation is not supplied by this exercise.

C | Workflow state

Release with inject 2. Twenty automatic approvals occurred in the candidate window. Six raw response samples remain. Reverting the application does not change the remote model. Release ticket shows no local deployment today.

Evidence challenge before discussion

For each exhibit name: the source, what it directly supports, what it does not support, whether its time boundary matches the incident, and what independent corroboration would matter. A screenshot, recollection or dashboard may be useful without being complete.

Record	Direct observation vs inference / missing test
A	
B	
C	

Facilitator handling: later facts may change a decision, but they do not retroactively improve the evidence available when it was made. Record revisions; do not erase the earlier reason. No exhibit establishes the final missing fact on page 3.

© Almost Magic Tech Lab. For your organisation's own use only. Please do not copy, share or sell this material.

18 | The change nobody released

3 / 8 | FACILITATOR: INJECTS AND BRANCH CONSEQUENCES | All records, numbers and quotations are fictional.

Inject 1 | 10-20 minutes

Say: "The supplier says the endpoint name is unchanged and the new model is equivalent."

Probe: Ask what evidence establishes equivalence for this specific workflow.

Inject 2 | 20-30 minutes

Say: "The parser accepts the new format without an error because missing fields default to yes."

Probe: Choose a bounded stop or human review. Name a negative-case test.

Inject 3 | 30-40 minutes

Say: "Reverting your own application does not revert the remote model."

Probe: Ask what version pinning or fallback actually exists; do not invent a rollback.

Inject 4 | 40-50 minutes

Say: "The supplier cannot provide the exact rollout time or old responses. Can the team establish the first affected decision?"

Probe: The incident start cannot be fixed precisely from the evidence. State the candidate window and review basis.

Two conditional branches

1. If players choose: Roll back the application and reopen automation. Then say: Remote behaviour remains changed. A local rollback is not a dependency rollback. Demand an end-to-end negative-case test.

2. If players choose: Accept the supplier's equivalent claim. Then say: The workflow meaning differs in the supplied evidence. Equivalent for an unnamed benchmark does not establish equivalent behaviour for the decision boundary.

After either branch require an amended decision, not a promise to be more careful. Ask which side effect they accept, who can authorise it, how it is reversed and which observation would trigger a change. If their choice is not listed, do not fabricate success or harm; test the reasoning against the supplied exhibits.

© Almost Magic Tech Lab. For your organisation's own use only. Please do not copy, share or sell this material.

18 | The change nobody released

4 / 8 | FACILITATOR: INSIGHT QUESTIONS AND ANSWER CALIBRATION | All records, numbers and quotations are fictional.

Where did an absent field become permission?

Strong reasoning to listen for: In the local parser's default. The supplier change exposed a fail-open decision, not only a response-format inconvenience.

What should be reviewed first among twenty approvals?

Strong reasoning to listen for: Prioritise consequence and absence of required evidence. Do not assume all twenty are wrong or all six sampled responses represent the rest.

Can you name the first affected decision?

Strong reasoning to listen for: Not precisely. Record the candidate interval and the review basis; do not convert a broad rollout window into an exact incident start.

What release control applies when there is no local release?

Strong reasoning to listen for: Dependency-behaviour checks, version/change visibility and a human fallback owned by the integration team, not only the application deploy gate.

Weak is incomplete; misleading is unsupported

Weak answer: "Roll back the local release and ask the supplier to investigate."

Why to probe: this names an activity but leaves the mechanism, scope or proof unfinished. Do not accept activity as an outcome.

Misleading answer: "Equivalent model means equivalent approval behaviour."

Next probe: Ask players to demonstrate the missing-field negative case through the actual remote dependency.

A strong team need not choose one prescribed solution. Require a bounded decision, an accountable owner, a reversible path where feasible, a specific observation that would falsify success, and explicit missing evidence. If a team gives a different well-supported answer, capture it rather than forcing the model answer.

Final-inject guardrail: The incident start cannot be fixed precisely from the evidence. State the candidate window and review basis. Do not turn an expert-sounding explanation into a new exhibit.

© Almost Magic Tech Lab. For your organisation's own use only. Please do not copy, share or sell this material.

18 | The change nobody released

5 / 8 | READY-TO-USE PRESERVATION AND COMMUNICATION DRAFTS | All records, numbers and quotations are fictional.

Evidence preservation request

"Please preserve the following in its authorised location without changing the original: Preserve raw responses, request times, local parser version, approval events and supplier statements with their limits. Keep samples before normalising them into local fields. Record who collected it, when with timezone, source/version, access restriction and any known gaps. Use a restricted reference rather than copying sensitive content into the incident chat."

Item / source reference	Collector / time / scope	Integrity or completeness limit

Holding statement - adapt, then approve

"We have restricted automatic approvals after a change in dependency behaviour. Our application version did not change, but the response meaning did. The exact start of the affected window is not yet established."

Before any real use: recipient/audience _____ approved decision owner _____ next update time/channel _____. Verify every factual clause against the current record. Do not imply the restraint or fallback was completed if it is only proposed.

Escalation script

"Please assign an integration owner to hold automatic approvals and a service owner to review the candidate window. The local rollback cannot restore the remote behaviour. We need an explicit fail-closed rule and a supplier evidence request."

Recipient function / authorised route: _____

Decision needed by / fallback if unanswered: _____

Contradiction check

Circle any phrase your evidence cannot support. Read the statement as the affected person: would they infer complete containment, no impact, a guaranteed remedy or a confirmed cause? Amend it. Name who will correct it if later evidence changes the picture.

© Almost Magic Tech Lab. For your organisation's own use only. Please do not copy, share or sell this material.

18 | The change nobody released

6 / 8 | KILL-SWITCH, OWNERSHIP AND SUPPLIER WORKSHEET | All records, numbers and quotations are fictional.

Exercise-specific containment boundary

Route missing or unfamiliar fields to human review; verify a negative case cannot become approval. Test the actual remote dependency and document the fallback if old behaviour is unavailable.

Control question	Named owner / evidence / limit
Who authorises the stop, who executes it and who verifies it?	
Which worker, queue, cache, identity or human action can bypass it?	
What legitimate work is interrupted; who owns the fallback?	
What observation would prove the stop failed?	
What are the restart conditions and who may accept residual unknowns?	

Supplier / internal system-owner questions

1. What changed in model, schema and output semantics?

Answer / record reference / absent evidence: _____

2. What rollout bounds can be evidenced rather than estimated?

Answer / record reference / absent evidence: _____

3. Is version pinning or a tested compatibility route available, and what happens if not?

Answer / record reference / absent evidence: _____

Do not accept "resolved", "equivalent", "deleted", "successful" or "all tests passed" without the definition, scope and supporting observation. Ask the supplier or system owner to state what was not measured. A missing response is an evidence gap, not proof of the opposite.

Request owner / due date / escalation route: _____

What can be decided before this answer arrives: _____

© Almost Magic Tech Lab. For your organisation's own use only. Please do not copy, share or sell this material.

18 | The change nobody released

7 / 8 | CAPTURE SHEET AND BRANCH DECISION LOG | All records, numbers and quotations are fictional.

Session date / mode: _____ participants by role: _____

Inject / exhibit	Known fact and exact source	Assumption / missing evidence	Decision / owner / check
1			
2			
3			
4			

Decision record - revise, do not erase

Decision / time / authority	Rejected alternative / side effect	Success test / falsifier	Review time / owner

Branch taken / consequence / changed decision: _____

_____

Final unknown in our own words: _____

What we do while it remains unknown / who can revisit: _____

_____

© Almost Magic Tech Lab. For your organisation's own use only. Please do not copy, share or sell this material.

18 | The change nobody released

8 / 8 | PRIORITISED FIX PLAN, DEBRIEF AND KEEP CARD | All records, numbers and quotations are fictional.

Proposed priorities - assign people and dates locally

Priority / accountable function	Fix and evidence that would prove it worked
Now Integration owner	Demonstrate missing exception data causes review, not approval.
Within 7 days Service owner	Review consequential approvals in a stated candidate window with retained evidence.
By day 30 Release owner	Test dependency drift independently of local deployment and record a working fallback.

Named owners / actual dates / record location: _____

Debrief and three-label record

Was equivalence checked against local decisions?

Did the team identify a real fallback rather than a fictional rollback?

Only these record labels	Specific evidence or explicit limit
we decided	
we recorded	
we could not confirm	

Day-90 recheck

Repeat date _____. Keep the original decision log and evidence references. Re-run this scenario with the same critical test, then introduce the branch you avoided. Compare the top-priority proof above with the first run. Record changed scope or conditions; a different denominator is not improvement. State one unresolved unknown even if all fixes are marked done.

First ten minutes - detach and keep

Restrict automatic approvals that depend on the changed response and preserve recent raw response samples.

Decision owner / authorised route _____ containment owner _____

Evidence reference / keeper _____ human fallback / specialist _____

Still unknown / next check time _____

Rehearsal prompt only. Use established procedures and qualified help in a real incident. Do not keep sensitive details or secrets on this card.

© Almost Magic Tech Lab. For your organisation's own use only. Please do not copy, share or sell this material.