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.

12 | The graph that drifted quietly

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

Model performance drift - Respond to deteriorating decisions when the software itself has not changed.

Read aloud

"A forecasting model used to plan staffing has become less accurate over several weeks. The team says no model release occurred. Local demand has changed, and the old accuracy dashboard still reports last quarter's results."

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 level and planning owner. State what you can decide in this rehearsal and what needs another authorised person. Do not assume a job title creates legal authority.

Model lead

Compare recent inputs, outputs and version records. State what you can decide in this rehearsal and what needs another authorised person. Do not assume a job title creates legal authority.

Planning lead

Identify operational effects and a manual fallback. 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

Separate stale metrics from current outcome evidence. 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.

12 | The graph that drifted quietly

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

A | Dashboard

Release at opening. Quarterly accuracy remains 91 on the displayed metric. Recent review finds errors concentrated in evening requests. Outcome labels arrive after fourteen days; the newest fortnight is not in the metric.

B | Data definition change

Release with inject 1. Completed request now means status closed, including auto-expired items. Earlier definition excluded expiry. Model weights unchanged. Input mix has shifted toward a new service channel.

C | Decision request

Release with inject 2. Planning meeting tomorrow proposes reducing human review on the basis of the green dashboard. Analyst proposes a threshold adjustment but supplies no matured outcome evaluation. The previous quarter's chart remains in the board pack.

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.

12 | The graph that drifted quietly

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

Inject 1 | 10-20 minutes

Say: "Recent error is concentrated in one service window. Overall averages hide it."

Probe: Ask for segment-level outcomes and a decision about the affected use, rather than a single average.

Inject 2 | 20-30 minutes

Say: "The data feed changed its definition of a completed request, but the column name stayed the same."

Probe: Consider data meaning as well as model code. Preserve the old and new definitions.

Inject 3 | 30-40 minutes

Say: "A manager wants to keep using the old green dashboard for the next planning meeting."

Probe: Have the team state what the dashboard does and does not establish now.

Inject 4 | 40-50 minutes

Say: "Two weeks of true outcomes are not yet available. Can the team confirm that a proposed adjustment fixes the drift?"

Probe: The outcome lag prevents confirmation. Define an interim limit and an observation plan, not a retrospective success claim.

Two conditional branches

1. If players choose: Keep the planning decision because the model did not change. Then say: The input and outcome definitions changed, so the metric is no longer the same evidence. Require a stated reliance limit before using it for staffing.

2. If players choose: Tune the threshold on recent incomplete outcomes. Then say: The apparent improvement reflects label availability, not proven correction. Keep an interim human-review rule and a dated observation plan.

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.

12 | The graph that drifted quietly

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

What changed even if the code did not?

Strong reasoning to listen for: Input population, label definition, observation window and operational use. Treat model performance as a system property, not only a code version.

How could an unchanged average hide a worse service?

Strong reasoning to listen for: Changes in case mix can mask a failing subgroup or time window. Require denominator and outcome-lag disclosure alongside stratified results.

What decision can be made before true outcomes mature?

Strong reasoning to listen for: A reversible reliance restriction with an owner, observation period and review date. Not a claim that the adjustment works.

Which chart belongs in tomorrow's pack?

Strong reasoning to listen for: A version-labelled chart with the metric definition, time boundary and missing outcomes stated. Removing all numbers is not a substitute for honest evidence limits.

Weak is incomplete; misleading is unsupported

Weak answer: "Watch the dashboard for another week."

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

Misleading answer: "The weights are unchanged and quarterly accuracy is green, so performance has not drifted."

Next probe: Ask for the outcome maturity, input mix and metric-definition boundary.

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 outcome lag prevents confirmation. Define an interim limit and an observation plan, not a retrospective success claim. 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.

12 | The graph that drifted quietly

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 old/new metric definitions, input-mix snapshots, model version, reporting code and planning-pack version. Keep matured and immature outcomes separate. 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

"The dashboard does not establish current performance for the newest requests or the changed completion definition. We are limiting reliance while comparing like-for-like outcomes. The proposed adjustment has not yet been shown to fix the drift."

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

"The planning decision relies on stale and non-comparable evidence. Please nominate the planning owner to set an interim review rule and the analyst to provide a dated, definition-consistent comparison before reducing human review."

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.

12 | The graph that drifted quietly

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

Exercise-specific containment boundary

Limit the affected planning use; keep human review where evidence is immature. Verify the decision owner has received the evidence limit, not just a revised chart.

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. Which input or label definitions changed and when?

Answer / record reference / absent evidence: _____

2. Can results be segmented with consistent denominators?

Answer / record reference / absent evidence: _____

3. What outcome lag prevents validation of the proposed adjustment?

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.

12 | The graph that drifted quietly

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.

12 | The graph that drifted quietly

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 Planning owner	Record an interim reliance limit in the actual decision record.
Within 7 days Analyst	Produce comparable time/channel results with outcome lag and denominator visible.
By day 30 Model owner	Evaluate the proposed adjustment on matured outcomes and retain both gains and regressions.

Named owners / actual dates / record location: _____

Debrief and three-label record

Were stale metrics called out explicitly?

Was the fallback connected to the affected planning decision?

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

Mark the forecast's current evidence limit and put the affected planning decision under named human review.

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.