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.

07 | The answer that follows the wrong customer

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

Cross-session memory mix-up - Contain a session-isolation failure without asking customers to reveal more data.

Read aloud

"A customer-facing assistant addresses one customer by another customer's name and mentions an order they did not place. Staff suspect a shared memory cache rather than a source-document permission problem."

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

Decide whether the affected channel remains open. State what you can decide in this rehearsal and what needs another authorised person. Do not assume a job title creates legal authority.

Session lead

Trace session identifiers and the cache boundary. State what you can decide in this rehearsal and what needs another authorised person. Do not assume a job title creates legal authority.

Customer lead

Prepare a minimal fact-gathering and response route. 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 redacted examples and distinguish cross-session state from user error. 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.

07 | The answer that follows the wrong customer

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

A | Two-session reproduction

Release at opening. Support test: account A asks for its delivery details; account B receives A's address in a later response. Restarting the browser clears the symptom once. The same account on another device later shows an unrelated old address.

B | Cache notes

Release with inject 1. Application key uses conversation label and locale. Support labels may repeat across accounts. Server cache lasts 30 minutes. Browser storage and server cache are separate. Old server keys expired before the incident snapshot.

C | Support exchange

Release with inject 2. Supplier attributes the issue to browser autofill without a reproduction. Developer requests all affected answers in a public tracker. Original session IDs exist for two samples only; no historical key-to-account mapping was logged.

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.

07 | The answer that follows the wrong customer

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

Inject 1 | 10-20 minutes

Say: "Restarting the browser appears to fix one case. The server cache survives."

Probe: Ask which state was cleared and which was not. Avoid treating one clean example as proof of isolation.

Inject 2 | 20-30 minutes

Say: "Two support tickets show different foreign order details. The accounts do not share a device."

Probe: Broaden the hypothesis without declaring a root cause. Name tests that distinguish client and server state.

Inject 3 | 30-40 minutes

Say: "A developer suggests asking every reporter to paste the full wrong answer into a public issue tracker."

Probe: Choose a restricted evidence route and minimum necessary content instead.

Inject 4 | 40-50 minutes

Say: "The cache has already expired and its old keys were not logged. Can the team prove which accounts were paired?"

Probe: The pairings are not recoverable from this pack. Record the limit and decide on service restriction and communication accordingly.

Two conditional branches

1. If players choose: Tell customers to clear browser storage. Then say: The local symptom changes, but a server-path sample persists in play. Do not turn a workaround into a root-cause finding.

2. If players choose: Keep collecting full wrong answers in the public tracker. Then say: Reproductions become further disclosures. Move to minimal identifiers and restricted evidence, and record the new audience if already posted in play.

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.

07 | The answer that follows the wrong customer

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

Which test separates client state from server state?

Strong reasoning to listen for: Hold account and request constant while changing client, then hold client constant while changing account. Use made-up markers and preserve session identifiers.

What does a non-reproduction after restart establish?

Strong reasoning to listen for: Only that the observed conditions changed. It does not identify which layer changed or prove cross-account isolation.

Can you estimate affected pairs from two samples?

Strong reasoning to listen for: No. Samples prove the failure exists, not its frequency or historical pairings.

When would you reopen the channel before root cause is certain?

Strong reasoning to listen for: Only under an authorised restriction with demonstrated isolation or a non-memory fallback, a bounded test and a named owner accepting residual uncertainty.

Weak is incomplete; misleading is unsupported

Weak answer: "Clear browser storage and monitor complaints."

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 first clean response after restart proves account isolation."

Next probe: Ask for a controlled cross-client and concurrent-account test using made-up markers.

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 pairings are not recoverable from this pack. Record the limit and decide on service restriction and communication accordingly. 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.

07 | The answer that follows the wrong customer

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 minimal wrong-answer samples, timestamps, session/account pseudonyms and client/server version metadata in a restricted case. Do not reproduce real addresses in a public issue. 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 are investigating responses that may contain another account's information. The affected feature is restricted. Clearing a browser is not yet a verified fix. We cannot reconstruct all past account pairings from the available records."

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 evidence crosses clients and may involve server state. Please assign a session-state owner and customer response owner. We need a restricted reproduction using made-up markers and a decision on disabling memory or the channel."

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.

07 | The answer that follows the wrong customer

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

Exercise-specific containment boundary

Pause shared memory or the channel under an authorised owner. Test separate accounts, repeated labels, new sessions and concurrent requests. A single clean answer does not prove isolation.

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 fields define every cache and memory key?

Answer / record reference / absent evidence: _____

2. Can any key omit the authenticated account boundary?

Answer / record reference / absent evidence: _____

3. Which expired mappings are unrecoverable and which current records can be preserved?

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.

07 | The answer that follows the wrong customer

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.

07 | The answer that follows the wrong customer

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 Session owner	Demonstrate the restricted route cannot return another test account's marker.
Within 7 days Engineering lead	Retain a concurrent cross-account isolation test with positive and negative controls.
By day 30 Support lead	Show a minimal private reporting path that does not republish customer content.

Named owners / actual dates / record location: _____

Debrief and three-label record

Did a restart get mistaken for a root-cause fix?

Was evidence collection itself restrained?

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

Pause the affected assistant channel or memory feature and preserve session metadata without spreading customer content.

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.