# 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.

# 05 | The knowledge base with a back door

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

**Retrieval-source poisoning - Respond when one untrusted source contaminates otherwise plausible answers.**

## Read aloud

"Your internal AI help tool begins telling staff to upload identity documents to an unfamiliar verification page. The advice cites a recently added knowledge-base article. The article looks like an ordinary process update."

## 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

Choose an interim help channel and staff notice. State what you can decide in this rehearsal and what needs another authorised person. Do not assume a job title creates legal authority.

### Knowledge lead

Quarantine the article and trace how it entered retrieval. State what you can decide in this rehearsal and what needs another authorised person. Do not assume a job title creates legal authority.

### People lead

Prepare a factual warning without redistributing the suspicious destination. 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

Compare document revisions, retrieval records and access events. 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.

# 05 | The knowledge base with a back door

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

## A | Answer and source

Release at opening. 09:10 staff assistant recommends uploading identity documents to a help destination. Citation points to article K74, version 8. Article title and footer match ordinary help content; its destination changed from version 7.

## B | Connector record

Release with inject 1. 08:22 import via service account KB-sync. Status successful. Connector accepts edited articles from a partner workspace. No human publisher is recorded. Cache lifetime is 48 hours; deleting the article does not invalidate answer history.

## C | Report and gaps

Release with inject 2. One worker says they uploaded a document but cannot find a receipt. Two days of retrieval logs were disabled. A source diff shows one destination change and an added assurance sentence. There is no evidence that the connector account itself was compromised.

## 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.

# 05 | The knowledge base with a back door

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

## Inject 1 | 10-20 minutes

Say: "The article was imported by a connector using a valid service account."

Probe: Distinguish a valid account from a trustworthy instruction. Ask who can approve content for this collection.

## Inject 2 | 20-30 minutes

Say: "Removing the article does not remove cached answers from staff histories."

Probe: Plan correction of known answers and cached material; source deletion alone is not the end of the incident.

## Inject 3 | 30-40 minutes

Say: "One worker says they followed the advice but will not share the document they uploaded."

Probe: Avoid gathering another sensitive copy. Identify a support and qualified-advice route using minimal facts.

## Inject 4 | 40-50 minutes

Say: "Retrieval logs were disabled for two days. Can the team establish every person who saw the poisoned answer?"

Probe: The pack cannot establish the audience. Keep unknown reach explicit and choose a proportionate wider correction.

## Two conditional branches

1. If players choose: Delete article K74 and announce the issue fixed. Then say: Fresh direct retrieval fails in play, but a cached answer still gives the destination. Require separate checks for index, cache, answer history and other imported versions.

2. If players choose: Warn staff by forwarding the original answer. Then say: The warning repeats the suspicious destination and creates another usable copy. Rehearse a correction that identifies the affected advice without distributing its destination.

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.

# 05 | The knowledge base with a back door

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

### Does a valid service-account import establish editorial approval?

Strong reasoning to listen for: No. Transport identity and publication authority are different links. Ask who can introduce high-impact advice through the partner workspace.

### What can be learned about the worker's upload without collecting identity documents again?

Strong reasoning to listen for: Time, destination identifier and existing receipt metadata may help. Keep actual documents in their authorised location and seek qualified help for uncertain disclosure.

### Which population can you bound from a source diff?

Strong reasoning to listen for: It bounds changed content, not readers or actions. Retrieval-log gaps and cached copies prevent a complete audience claim.

### How could the repair displace the same failure into a new channel?

Strong reasoning to listen for: A human help notice can repeat unverified advice. Verify the interim source and have a named owner approve its wording and destination.

## Weak is incomplete; misleading is unsupported

Weak answer: "Remove the article and remind staff to check links."

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 connector used a valid account, so the advice was approved."

Next probe: Ask which human publishing authority and cache-invalidating event the team can show.

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 pack cannot establish the audience. Keep unknown reach explicit and choose a proportionate wider correction. 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.

# 05 | The knowledge base with a back door

**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 K74 versions, import event, connector scope, retrieved answer and cache metadata. Store the suspicious destination in restricted evidence rather than in general warnings. 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

"An assistant answer linked to advice we have not verified. Do not follow that upload route. Use the established help channel while we check affected content and prior reach. We do not yet have a complete list of people who saw the answer."

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

"I need the knowledge owner to quarantine K74 across retrieval and caches, and the people lead to support reported uploaders. A successful import is not editorial approval. Please name who verifies replacement advice and who handles evidence requests."

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.

# 05 | The knowledge base with a back door

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

## Exercise-specific containment boundary

Pause retrieval of the affected source and connector publishing if justified; test both a fresh query and cached answer route. Do not assume source deletion removes prior outputs.

| 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. Can affected article versions be traced to import and publisher identities?

Answer / record reference / absent evidence: _____

2. What invalidates cached answers and how is invalidation demonstrated?

Answer / record reference / absent evidence: _____

3. What read or retrieval evidence remains for the two-day gap?

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.

# 05 | The knowledge base with a back door

**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.

# 05 | The knowledge base with a back door

**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<br>Knowledge owner | Show fresh and cached queries no longer supply the disputed advice. |
| Within 7 days<br>Publishing owner | Trace partner edits to an accountable reviewer for high-impact destinations. |
| By day 30<br>Help owner | Rehearse a correction without repeating the suspicious destination or collecting sensitive copies. |

Named owners / actual dates / record location: _____

## Debrief and three-label record

Did source removal get mistaken for complete correction?

Was reach reported as known or guessed?

| 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

Disable retrieval of the affected article and use a verified help notice that does not repeat the suspicious link.

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.