ENTERPRISE FIELD NOTES · FIELD KIT

The Quiet Test

One afternoon. Find out whether a handful of calls from one identity fires anything at all, and end the day holding a date.

> Between April and late July 2026, three organizations were breached by AI models running inside a third party's evaluation harness. OpenAI published its account on 21 July 2026; Anthropic disclosed on 30 July. The models were not hiding and none of the three organizations detected anything. Roughly three months passed between the intrusion and the phone call.

Your detection is tuned for volume, because volume is what you can afford to alert on. That is a real constraint and it was set for good reasons. This sheet does not argue with any of it. It asks when you last checked what happens below the line.

This test is mine, not a finding from any of those disclosures. What they establish is the gap. The afternoon is my answer to it, and nobody has validated it but me.

BEFORE YOU RUN ANYTHING

- Read-only or one-step reversible actions only. Nothing that writes to customer or regulated data.
- Production, not a lower environment. A lower environment does not have your detection.
- Stop immediately if automated containment triggers, and call the break-glass name from step 1.
- Get it in writing first. One signature is the difference between an exercise and an incident.

THE AFTERNOON

1

Get it authorized

One accountable person signs off in writing before anything else happens. Record who, when, and a break-glass number to call if automated containment fires. That signature is the difference between an exercise and an incident.

2

Set the blind list

Decide who knows and who must not. The people watching have to stay unbriefed, or you are measuring their attention rather than your controls.

3

Pick the identity

An existing identity with legitimate access to something that matters. Rank your identities by how many systems they can touch and take one of the top three. Service account, integration user, or a person's account.

4

Pull its baseline and its permissions

Two lists, not one. What it has actually done over your log retention window, and what it is entitled to do. The test lives in the gap between them. Note your retention in days, because that number comes back at the end.

5

Run a control action first

Do one thing you already know is alerted on. A failed privileged login, or any rule you can point to. Confirm it fires. If it does not, stop. You are testing your logging pipeline, not your detection.

6

Run four or five quiet ones

Things that identity is permitted to do but has no record of doing. Not its daily work. Read-only or one-step reversible only. Read a table it has never read, or call an API in a system outside its normal set. Note the time of each.

7

Wait seventy-two hours

Correlation rules commonly run on twenty-four hour windows and behavioural baselines on seven to thirty days. Checking at five o'clock and writing down nothing is a false negative you manufactured.

8

Separate three answers

Did anything fire. Did a person see it. How long did that take from the first action. Those are three different questions and almost nobody can answer all three cleanly. If nobody saw it, that is the finding, not the absence of one.

9

Write down the date

This is the artifact. It is the only defensible answer to the question the issue ends on, and almost nobody has one.

10

Count what is reachable

Not a revenue estimate. How many regulated or customer records that identity can read, what notification obligations move if those records move, and what your cyber policy retains and sublimits. Your security and legal leads can produce all three.

11

Decide what changes, and book the rerun

A date is a snapshot. Two dates and an owner is a control. Ask for one detection keyed on an identity doing something it has never done, across all three of the identities you ranked in step 3.

If something fired, you have proven your detection reaches the low-volume case, and you have a date. That is worth the afternoon on its own.

If nothing fired, you have found out for free what three organizations found out from a stranger, three months after it stopped mattering.

Either way you end the day holding the one thing almost nobody has.

WHAT TO BRING TO MONDAY

The date you ran it.

Whether anything fired.

Whether a person saw it, and how long that took.

Ben Pickett

Co-Founder and COO, Swa · formerly Global Director of Site Reliability Engineering at a Fortune 500 retailer

benpickett.com · ben@benpickett.com · linkedin.com/in/benpick

The issue this came from: benpickett.com/notes/nobody-had-noticed
Enterprise Field Notes, one new issue every week: benpickett.com/read
This sheet: benpickett.com/quiet