My agent found 9.09%. I found one session.

Varun Choraria · 2026-08-03 · https://www.varunchoraria.com/11-sessions-are-not-a-ux-audit/

My agent found a UX problem on my site. Dead clicks had affected 9.09% of sessions, above the 5% threshold. It filed a GitHub issue and told me what to inspect.

There had been 11 sessions.

9.09% was 1 session.

That was the real bug. The maths was right. The confidence was not.

What the agent actually knew

I built a small agent called Lazarus Pit because I knew I was not going to remember to open Microsoft Clarity every Monday.

It reads the site's Clarity summary, checks the numbers against a few thresholds, then files a GitHub issue with a possible cause and a suggested fix. It can propose work. It cannot change the site.

On its first live run, it saw that dead clicks had crossed 5%. The issue recommended checking anything styled like a link or button.

Reasonable advice, except the data did not tell the agent which page had the dead click. It did not identify an element. It did not have a recording to inspect. All it knew was that 1 dead click had happened somewhere across the site.

It had found something worth watching. It wrote it up as something worth fixing.

Why 9.09% fooled the system

`1 of 11 sessions` sounds weak because it is weak. `9.09% of sessions, threshold 5%` sounds like a finding.

Both statements are true. Only one makes the size of the evidence obvious.

The percentage also triggered the workflow. Once the number crossed the threshold, the agent moved straight from observation to explanation to task. Nobody, including me, had made it stop and ask whether the data could support that jump.

This is not really a sample-size problem. More sessions would make the rate less ridiculous, but a site-wide rate still cannot tell me which button to fix. I had asked one number to answer two different questions:

1. Did something unusual happen?
2. Do we know enough to change the site?

The agent could answer the first. It pretended it could answer the second.

What I changed

I added a floor of 20 human sessions. Below that, Lazarus Pit skips UX findings. Twenty is not suddenly enough evidence. It is just a guardrail against the worst version of the mistake.

The bigger fix is in how the agent reports what it sees:

- Show the count with every rate: `1/11 sessions (9.09%)`.
- Keep site-wide data as a site-wide observation.
- Do not recommend a page-level fix without page-level evidence.
- Treat one weak window as a note, not a task.

I still want the system. The analytics gets read, possible problems arrive where I already work, and nothing changes before I look at it.

But I had taught the agent what number counts as bad. I had not taught it what evidence counts as enough.

Those are not the same rule.