EACG
Enterprise Architecture Consulting Group

SECURITY DIALOGUES
No. 7 · 18 September 2026
Marcus Hale (PSIRT) · Tess Calder (red team) · Priya Raman (guest) · Jessy Lang

It Just Sounded Urgent

Eight lessons from the flood of reports that read perfectly and meant nothing

Open-source maintainer Priya Raman gets about three times the reports she used to, yet the real ones stay at about eleven or twelve a year. The new fakes read exactly like the real ones: same structure, same confident tone, a function name that exists in her code. The model is not lying. It completes the shape of a critical report, and nobody is accountable for what it says.

CONFABULATED URGENCY
Not a lie: fluency standing in for judgment.

1. Is the key fact real?

2. Who drafted it? Was it reviewed?

3. Whose name goes on the correction?

WHAT GOES WRONG AND WHAT TO DO ABOUT IT

1	**Fluent and wrong** Well-formed reports built on invented specifics.	No rule can filter for it, so verify the load-bearing fact in every report: the function, the dependency, the version. Priya's tells: a proof of concept calling a method that exists nowhere in her library, a dependency her project never used, and a CVSS vector that computes to a precise 8.1 for nothing.
2	**Confidence is the mechanism** The fluency is the product, not a warning sign.	A model that hedged would be less convincing, but the fluent version is what gets forwarded, escalated and acted on. Do not read a confident tone as evidence.
3	**No one to argue with** No motive to model, no shame, no ceiling on volume.	Replying gets nothing back, or a fresh report with a different wrong claim. The sender does not tire, learn or feel embarrassed, and shame was doing more triage work than anyone realised.
4	**Real reports get buried** Skimming after the fifth wrong report.	Priya nearly missed a genuine finding in March: it sat in her queue for six days. Do not skim. Triage by bucket, so you know what you are looking at before you spend an hour on it.
5	**Reviews that quietly stop** A draft assistant becomes an unsupervised filer.	Keep draft assistance a draft: a human reviews, a human sends. Nobody signed off on unsupervised filing, it arrived one skipped review at a time. Write down any change to that process.
6	**A generator nobody owns** Never inventoried, so nobody is accountable.	Put it on the inventory by name, with an owner, like an algorithm choice or a scanner. Keep a record of which outbound communications were drafted, reviewed or fully generated by a model.
7	**Poisoned inputs** Whatever the tool reads can become its finding.	In Tess's test, one invented code comment about an auth bypass made two of three report assistants file a critical report. Check what the tool reads: anything a contributor can write there can become a confident finding.
8	**Identical trust levels** Human-authored, AI-drafted and reviewed, AI-generated and unreviewed look the same.	Tag every output with who or what drafted it, whether a named human reviewed it, and a rolling precision score per source. When a source's ratio of real to fabricated findings crosses a threshold, quarantine its output for mandatory review.

THREE POINTS TO KEEP

The tag Priya asks for
Four words on every report: reviewed by a person, or not. It costs nothing technically, only the admission of how much is unsupervised. One vendor that rebuilt its pipeline this way saw complaints drop by more than half in a quarter.

The economics of free
A few cents of compute for the sender, twenty minutes to two hours of unpaid time for the maintainer. Priya's conservative ratio is a thousand to one. Nobody decided to move that cost.

This is not 'no AI'
A draft assistant that stays a draft assistant is a legitimate answer to too much scanner output. The failure is nobody deciding what it may do unsupervised, and nobody tracking whether it still does only that.

Listen to the episode
It Just Sounded Urgent and the rest of the series.
eacg.de
/en/podcast/s07-confabulated-urgency

Digital Product Design
From the first idea to a viable first release.
eacg.de
/en/consulting/digital-products

PSIRT Services
Intake, assess and disclose vulnerabilities, built or run as a service.
eacg.de
/en/consulting/psirt

© 2026 EACG GmbH