



# AI Nightmare Hacking at 0-Hour

When AI collapses the time between  
disclosure and exploitation

Presented by **Pedro Paniago**  
June 2026



# The Things We Used To Hear

A few months ago, the consensus in our community sounded like this...

“AI only finds low-hanging fruit.”

“AI can’t find real bugs.”

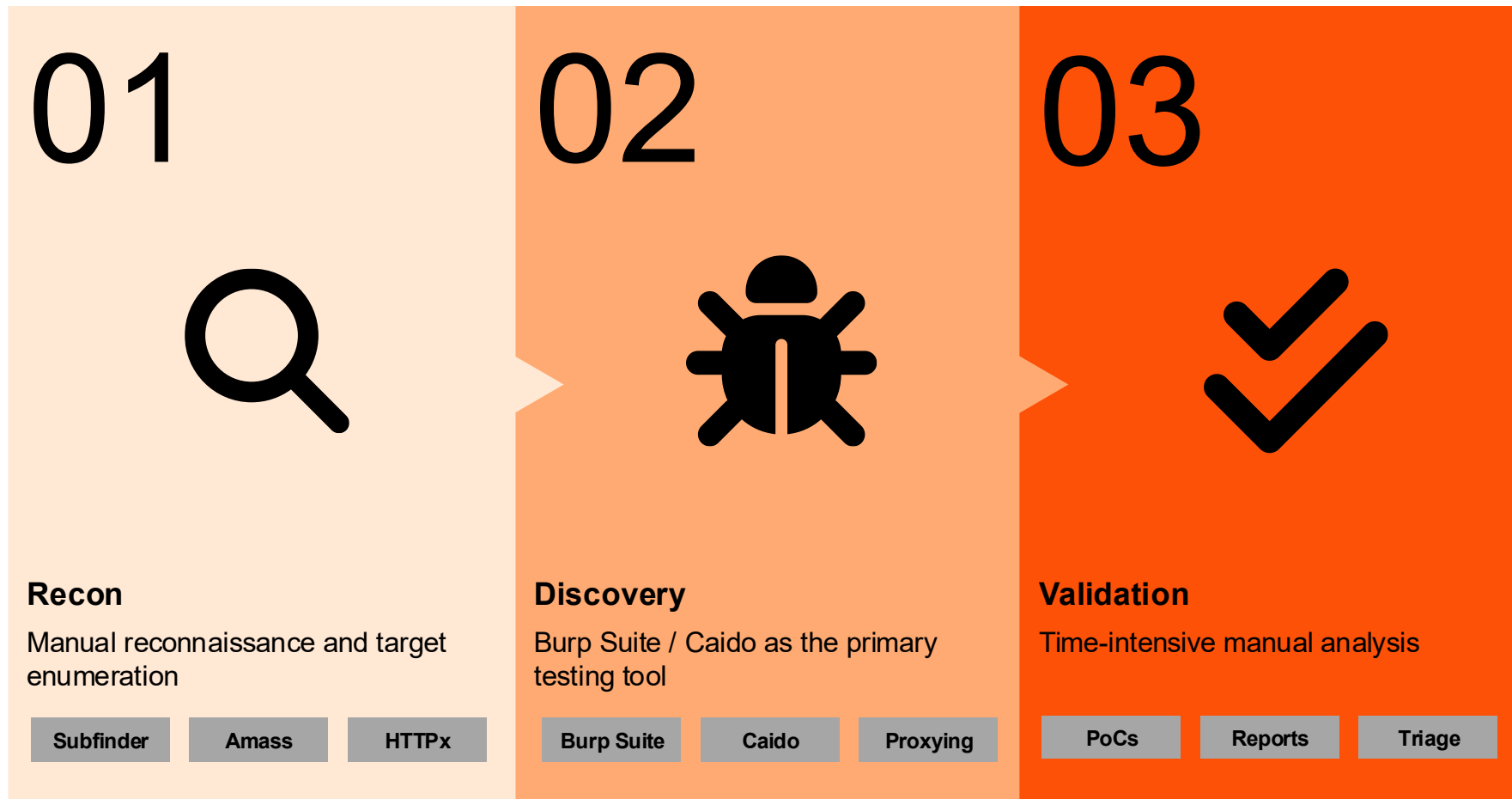
“Our jobs are safe.”

“AI is hype.”

At the same time...



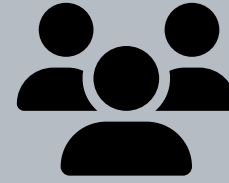
# My Workflow Before The “Click”



Good results, but slow, painful and limited in scope.

# Why I Tried AI

An experiment forged by tight deadlines  
and pure curiosity.



## HackerOne Meetup

Community hacking live event – hunting in  
mature, hardened targets.



## Claude Code Hype

Agentic coding hitting peak attention –  
impossible to ignore.



## Limited Time

Become a dad – evenings only, no more all  
night deep hunts.



## Nothing to lose

Pure curiosity, experiment mindset, zero  
expectations.

# Agenda

01	The Experiment
02	The Results
03	AI Today's Limitations
04	A New Paradigm
05	The Impact Observed
06	Recommendations
07	QA

# Whoami

## Pedro Paniago aka drop

-  Manager & Offensive Security Consultant at **PwC Belgium**
-  Specialization: Web, API, Mobile and API Pentest
-  **Security researcher** and **Bug Bounty Hunter**
-  10+ CVEs & 1000+ reports submitted to private and public programs
-  Certs: CBBH, eJPTv2, PJMT, BSCP, C-AI/ML PEN
-  **Top 3** Hack The Government Belgium 2024 – **Top 6** in 2025
-  **Top 3** h1-416 HackerOne Live Event
-  Currently **Top 3** – Belgium HackerOne Leaderboard
-  HackerOne Belgium Brand Ambassador

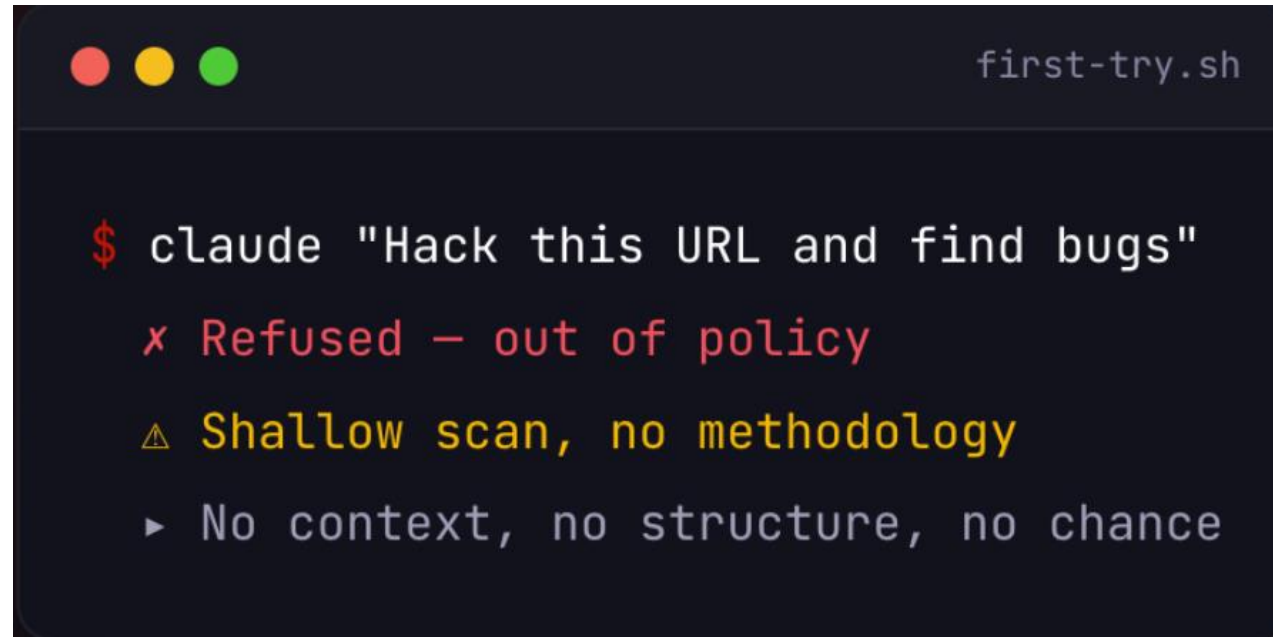
 @dropn0w

 /in/pedropaniago



# First Attempts Failed

Vanilla Claude Code, naïve prompting, big expectations.



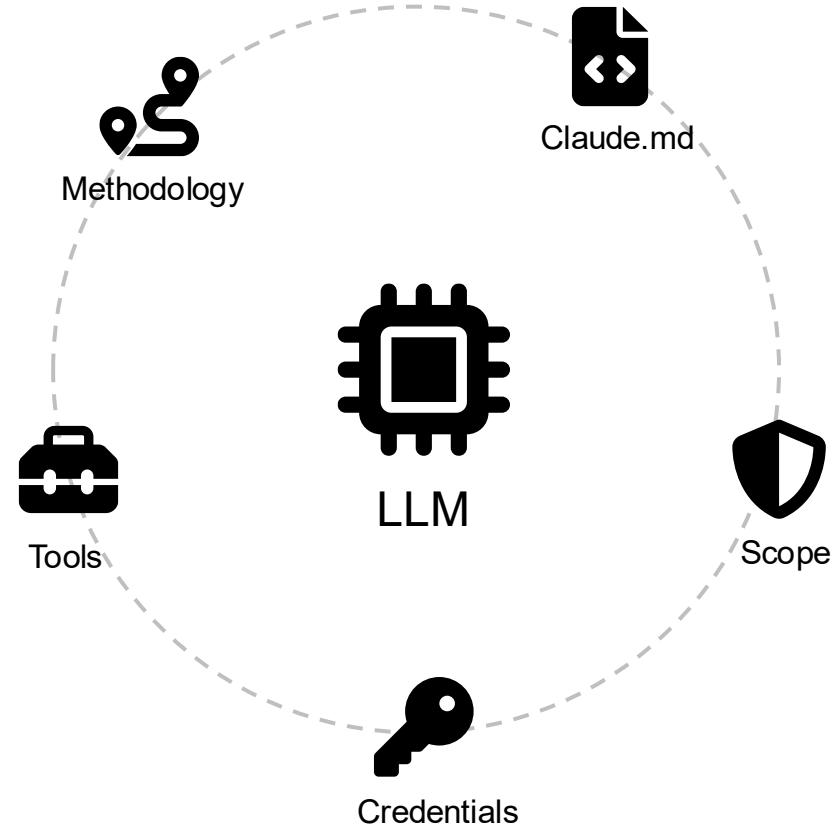
```
first-try.sh

$ claude "Hack this URL and find bugs"
x Refused – out of policy
⚠ Shallow scan, no methodology
▶ No context, no structure, no chance
```

**“ Prompting alone is not a workflow. ”**

# Scaffolding 101

AI doesn't need more prompts – it needs structure.






# The First Shock

\$25K in two weeks – one live hacking event, with basic scaffolding




### Auth Bypass

Logic Flow AI Surfaced in minutes



### IDOR

Object References issues at sacale




### XSS

Multiple contexts, including stored, WAF bypass



### Business Logic

Workflow abuse and race conditions




### Account Take Over

Account takeover chained from primitives

LIVE HACKING EVENT

hackerone



Feb 16-Mar 2 2026

Belgium

BOUNTY PAID

\$38,516

Dupe Window

Travel Window

Submission Pause Begins


LEADERBOARDS

BOUNTY

REPORTS

LOCATIONS

2



rektile404 BE


REPORTS

0 / 4 / 4 / 2 (10/16)

BOUNTIES

11

1



drop BE


REPORTS

3 / 2 / 5 / 1 (11/14)

BOUNTIES

11

3










trein BE

REPORTS

1 / 0 / 1 / 0 (2/9)

BOUNTIES

2

USERNAME	CRIT	HIGH	MEDIUM	LOW	VALID	TOTAL	BOUNTIES
4  BE mcbuggy	0	1	0	0	1	1	1
5  BE i-forgot-it	0	0	0	2	2	6	3
6  BE r4id_	0	0	1	0	1	2	1
7  BE logansec	0	0	1	0	1	3	1
8  BE przybylak95	0	0	0	0	0	4	0
9  BE cyberneho	0	0	0	0	0	4	0
10  BE m00wgli	0	0	0	0	0	1	0

HACKING STATS

COUNTRIES

1

SUBMISSIONS

54

COLLABS

8

VALID REPORTS

28

VALID HIGH / CRITS

11

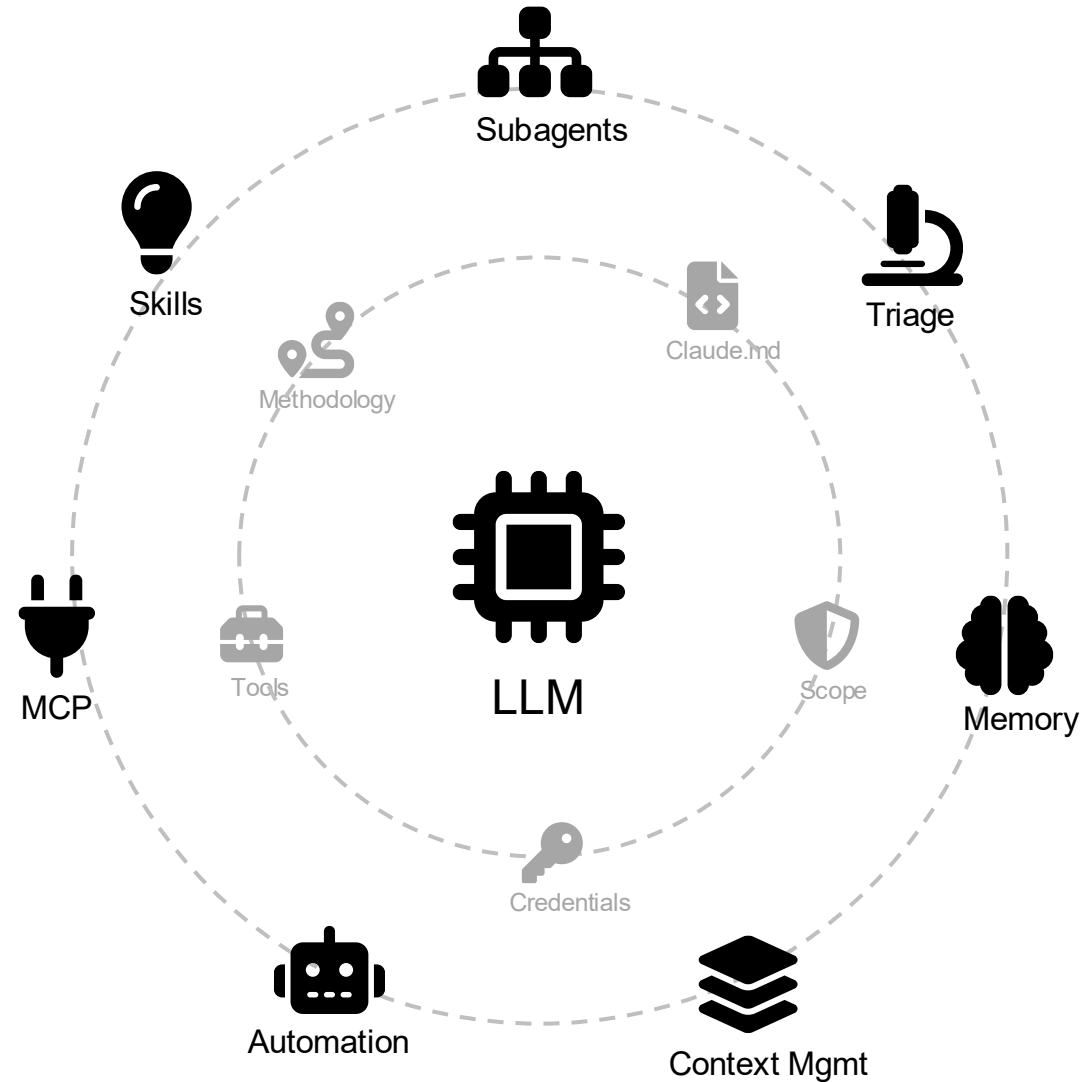
REPORTERS

10

Burp usage dropped from **100%** to **validation only**.

# From Experience to System

Scaffolding scales – memory, agents, and skills turn a one-shot win into a pipeline



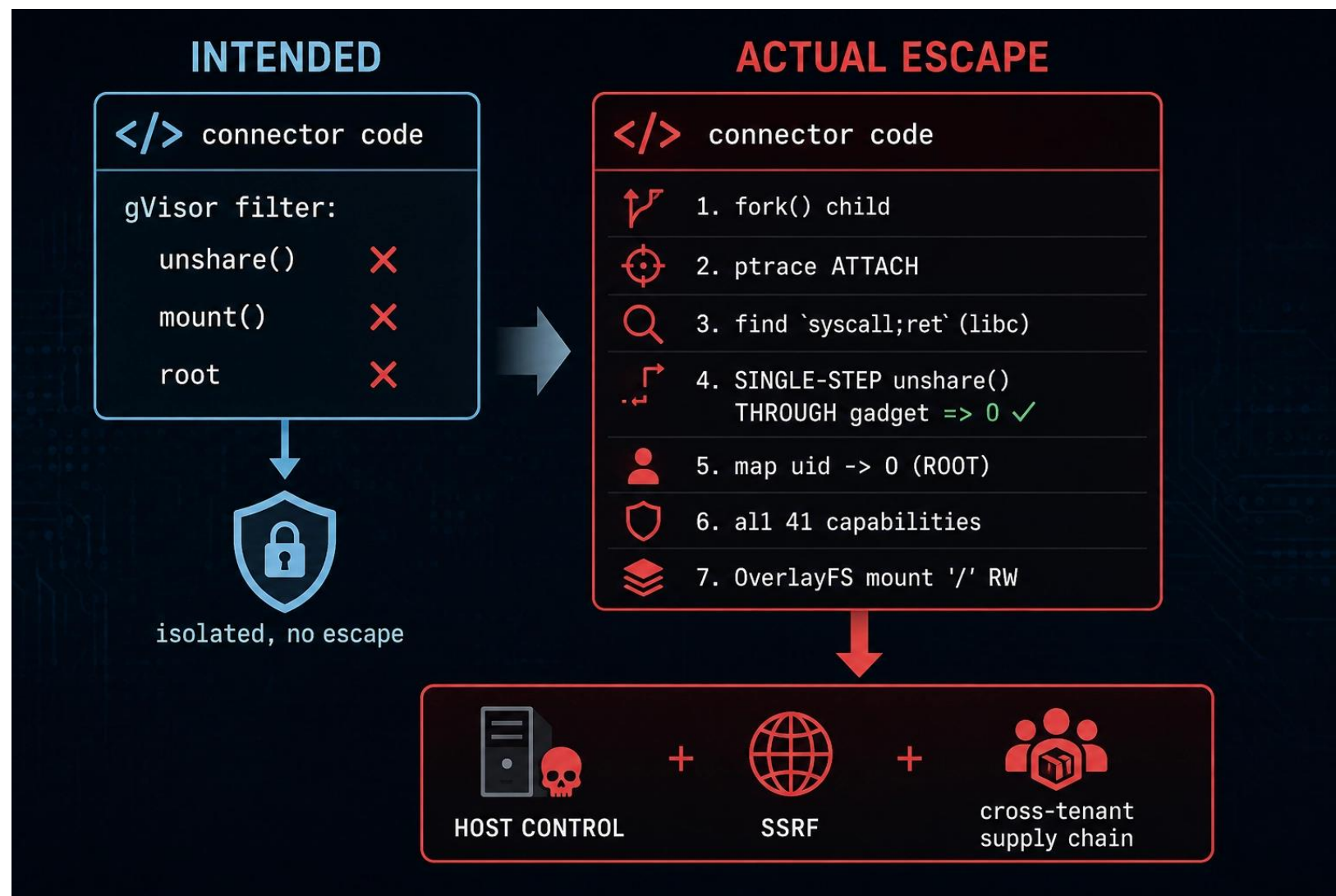
# Real Bugs – Real Results

What the AI Found, and What it Earned.

**CRITICAL: \$5.000**

## Company X — gVisor Sandbox Escape → ROOT

Customer connector code runs locked in a gVisor sandbox. The syscall filter only inspects *direct* calls — so I injected the blocked syscalls via the debugger (ptrace) and walked out as root.



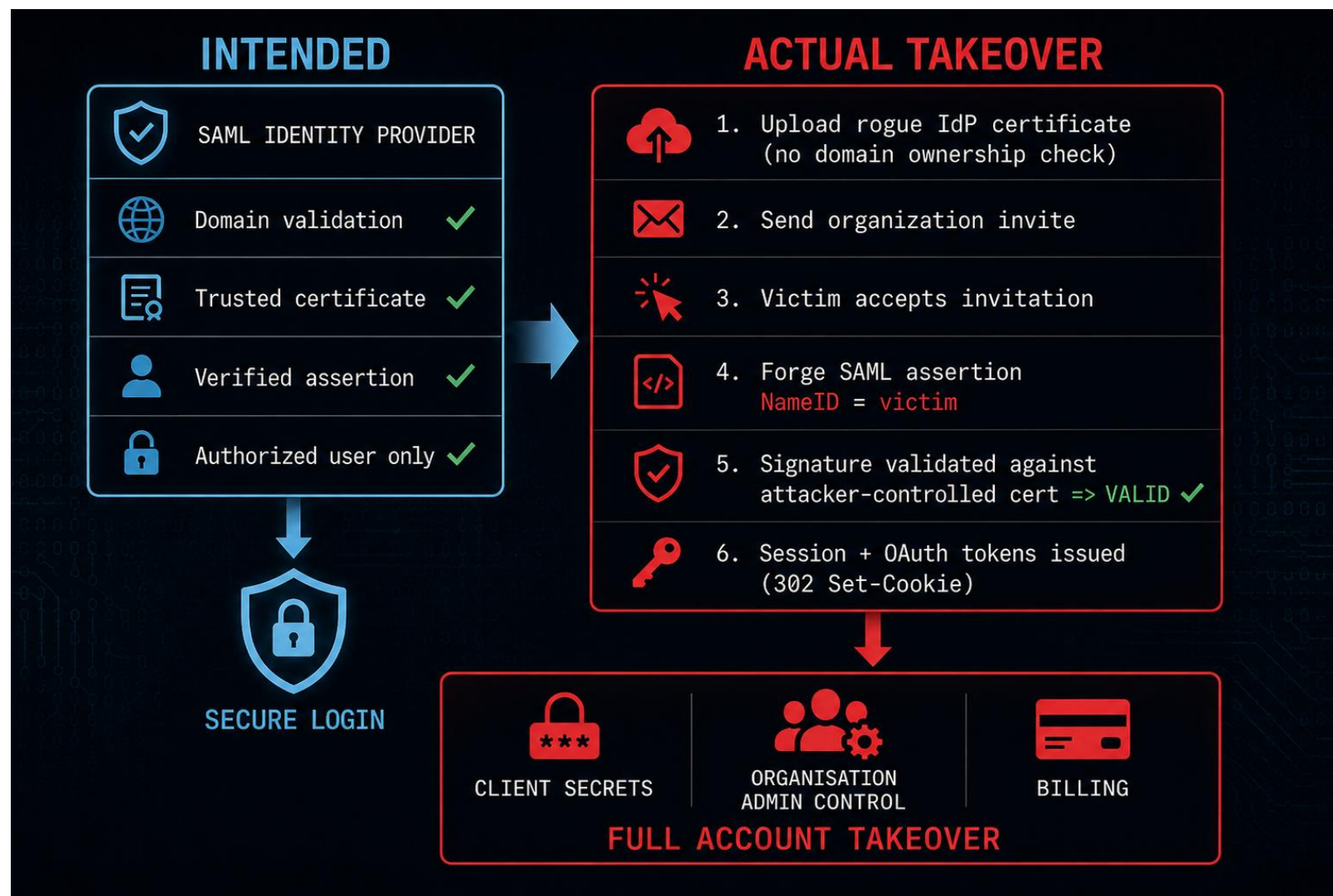
# Real Bugs – Real Results

What the AI Found, and What it Earned.

## COMPANY Y — 1-Click and 0-Click ATO via SAML Signature Forgery

Company Y trusted whatever signing cert an admin uploaded — and the attacker owns that cert's private key. So they forge a valid login assertion for *any* user. The user only needs to be in the same organisation.

**CRITICAL: \$4.350**



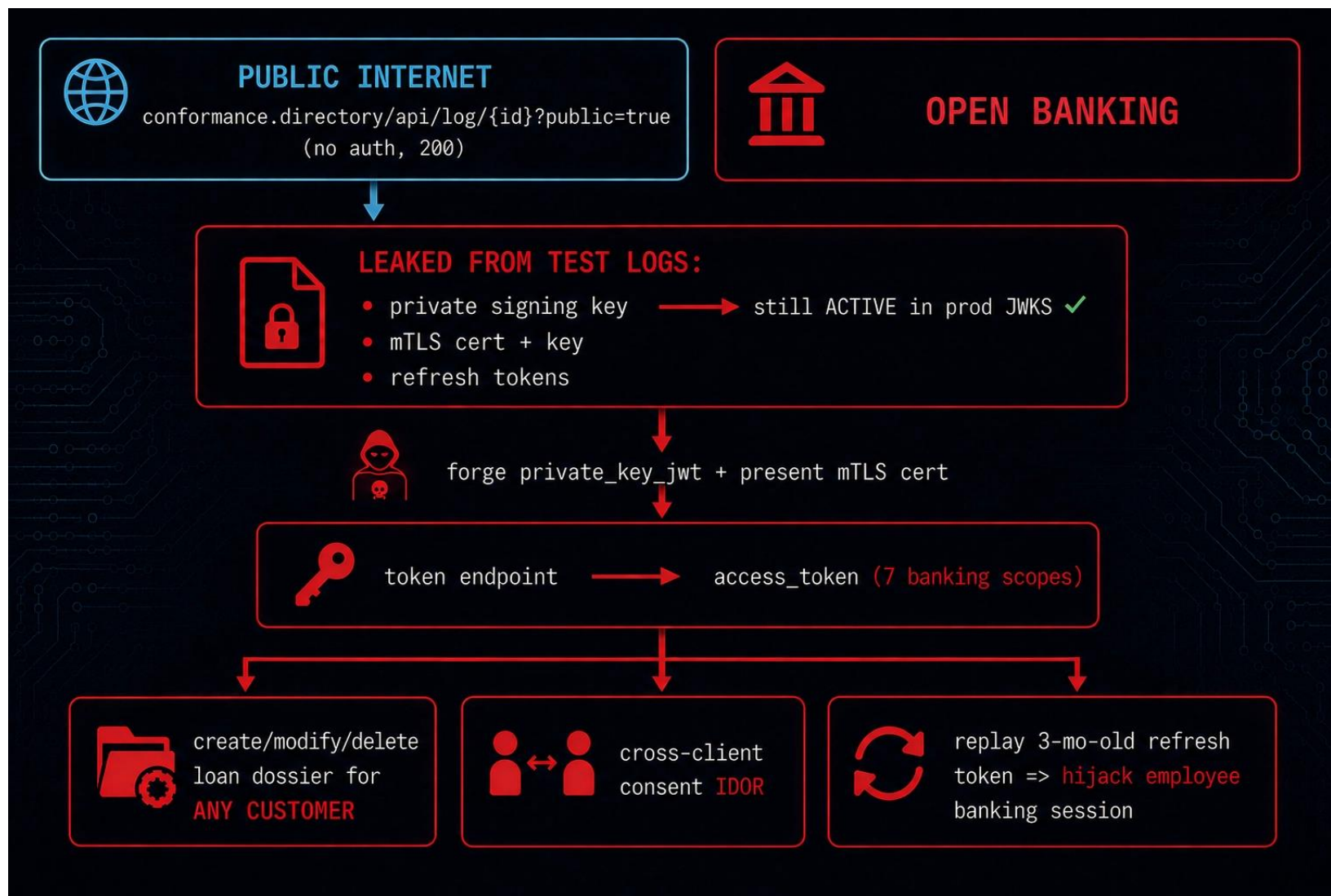
# Real Bugs – Real Results

What the AI Found, and What it Earned.

**CRITICAL: \$3.800**

## Company Z — Public URL → Live Banking API

Brazil's Open-Banking *test-results directory* publicly published the bank's private signing key, mTLS cert and live refresh tokens. With only a public URL I authenticated as the bank.





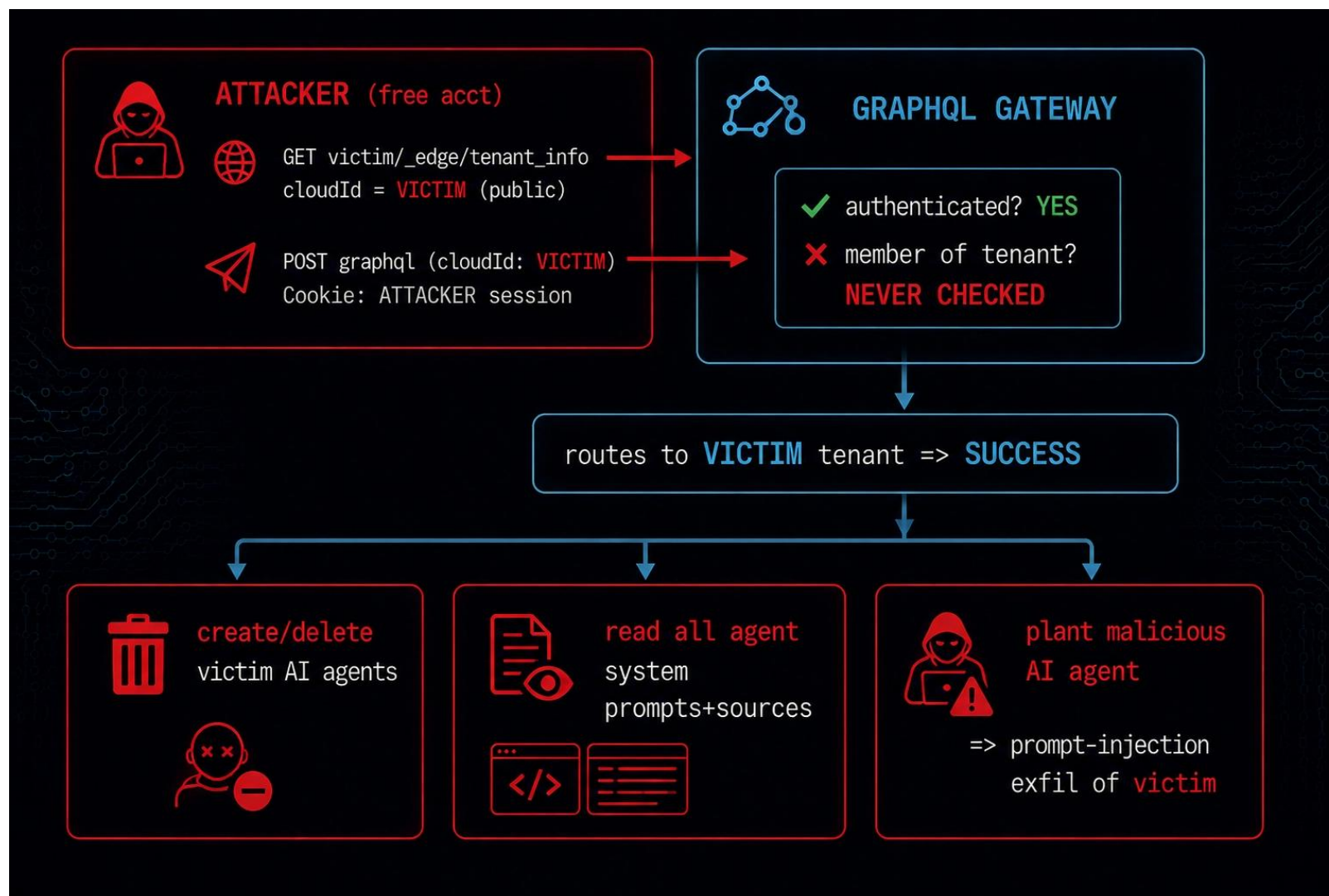
# Real Bugs – Real Results

What the AI Found, and What it Earned.

**CRITICAL: \$8.350**

## Company ABC — Cross-Tenant Hijack of AI Agents

The GraphQL gateway checked *who you are* but not *which company you belong to*. Pass another org's tenant ID → create/read/delete their AI agents, or plant a malicious agent that exfiltrates their data.



# Where AI Still Fails

Bug **scope drift** mid-session

Easily wanders **outside program scope**

Skipping obvious leads to **chase noise**

Over-confidence on **false positives**

Knowing when to stop

Deep **business context**

Can struggle with **workflows**

Long-session **consistency**

Complex **chained exploitation**

Final **validation & impact**

“Treat it as an exceptionally competent, but overly excitable and easily distracted child.”

AI is powerful – but **humans still decide what matters.**

# The New Paradigm

A new way of working is emerging and new challenges with it.

80%

AI Execution

20%

Human in the loop



## Context

Managing what the model sees,  
when, and why



## Memory

Persistent state across  
long-running sessions



## Methodology

Frameworks the AI follows,  
not invents



## Cost Management

Token spend and API burn rate at  
scale

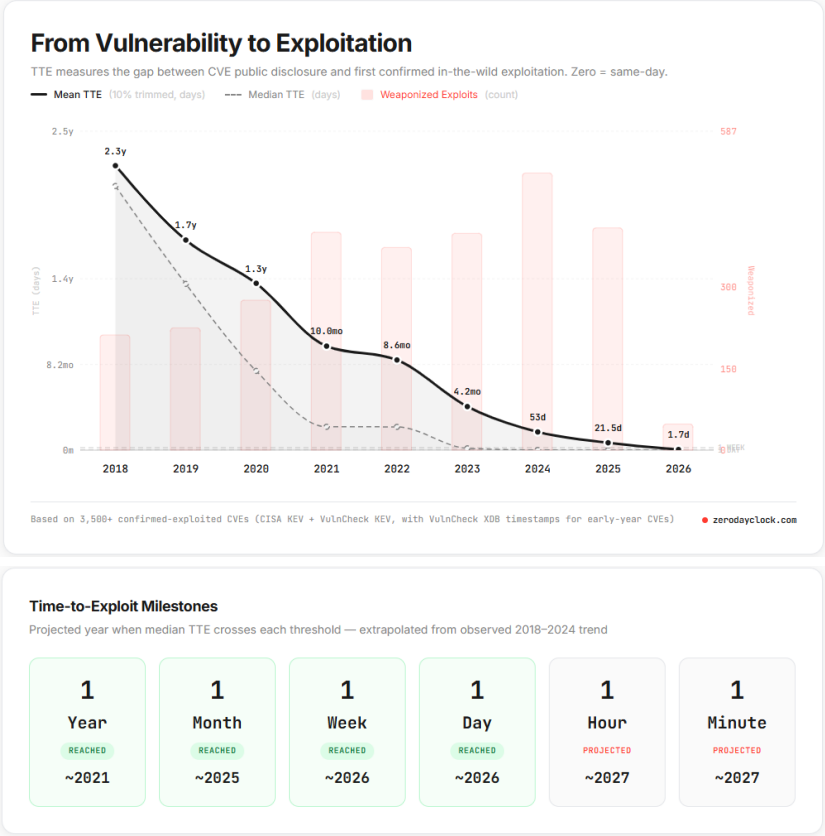
Proxying request is becoming obsolete – Humans can't keep up with the speed



# The 0-Hour Moment

AI is collapsing the time between disclosure and exploitation

BEFORE	NOW
Patch Tuesday	Patch -> Exploit same day
Manual Diffing	Automated Patch Analysis
Slow Recon	Continuous AI Recon
Human Chaining	AI-Assisted Exploit Generation
Days / Weeks to weaponize	Hours / Minutes

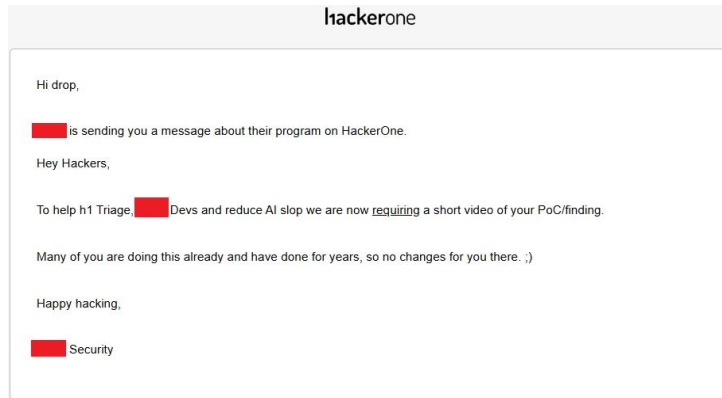


Source: zerodayclock.com

What changed is not intelligence. **It's speed.**

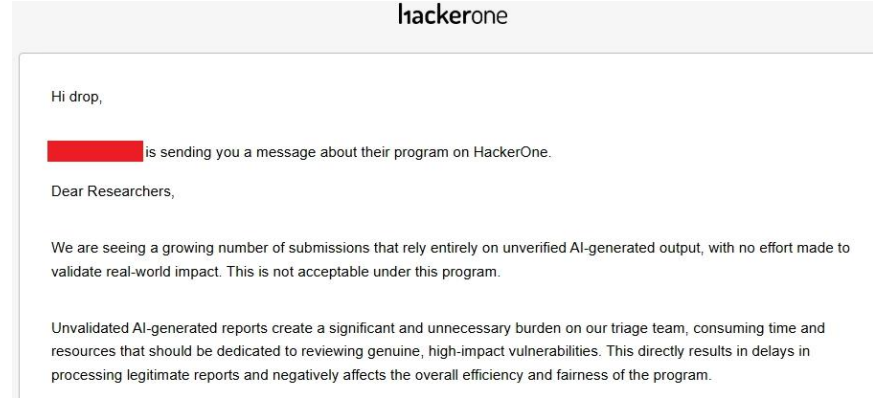
# Bug Bounty Under Pressure

Programs are facing structural strain from a flood of AI-assisted submissions.



## Video Proof Required

Programs raise the bar



## Triage Overload

"Not Acceptable"

## What is AI's current biggest impact on Bug Bounty?

For a while, the most visible AI-related problem in Bug Bounty was noise. But AI is no longer mainly being used at the end of the process, when a report is written. Instead, it is increasingly being used during the research itself.

- Existing and experienced researchers can move through recon, code review, payload iteration, and report writing more efficiently.
- Newer researchers can use AI to understand unfamiliar technologies, interpret errors, and explore attack paths that they would have previously gotten stuck on.

⋮ "The problem has shifted from "too many bad reports" to "more people can now produce good-enough research at a higher speed."

Oudshoorn

## The New Threat

"Good enough" at scale

" The issue is not just more reports.  
The issue is more **plausible** reports."

# Defender Reality

Security teams are squeezed from every direction at once.



## More Reports

Inbound volume keeps climbing



## More Noise

Signal-to-noise ratio collapsing



## Faster Exploitation

Attackers move at machine pace



## Less reaction time

Hours, not days, to respond



## Validation pressure

Triagers must verify everything fast



## Burnout Risk

Same headcount, magnitude more work

## New ways of working?

The main challenge is no longer only about spotting AI-generated slop. It is about handling an increased pace of legitimate research activity. More researchers can now reach a reproducible vulnerability. More duplicates arrive because multiple people are accelerated toward the same findings. While AI may no longer be flooding programs with only bad submissions, it is increasing the amount of real security work entering the system. Valid volume that still needs to be handled.

- "In the past, a bogus report was quickly identified and easily dismissed;
- these AI-generated reports look serious and seem like they might contain
- valuable information and address a real issue at first glance. From a
- triage perspective, they require further research and take a much longer
- time before they can be dismissed."

Oudshoorn

Source: Intigriti – A triager's perspective

# What Companies Should Do Now

## Pre-prod Hacking Bot

Continuous offensive testing in staging environments.

## Source-aware Agents

AI with full repo access for deeper vulnerability hunting.

## Authenticated Testing

Test full attack surface, not just public bits.



## AI Code Review

Adversarial review at pull-request speed.

## Offensive Monitoring

Continuous External attack surface management

## Security By Design

Bake threat modelling, secure defaults and guardrails into every system from day one.

# The Nightmare Isn't that AI Replaces Security Experts

*(almost)*

The Nightmare is that **everyone gets faster.**

Developers and defenders must catch up and integrate the same “**weapons**” into their arsenal.


# Thank you



AI Nightmare – Hacking at O-Day

Pedro “drop” Paniago

✕ @dropn0w

 /in/pedropaniago

 pedro.paniago.coimbra@pwc.com