INGRAM

Crafting intelligence, one vector at a time.

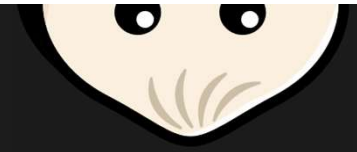
CYBERSEC EUROPE

May 20, 2026

AI in Cybersecurity

What just happened?

https://ingram.tech
contact@ingram.tech



I had something
different prepared.

INGRAM

Tuesday, 2nd December, 2025

Anthropic acquires Bun.

Bun is joining Anthropic

Jarred Sumner · December 2, 2025

TLDR: Bun has been acquired by Anthropic. Anthropic is betting on Bun as the infrastructure powering Claude Code, Claude Agent SDK, and future AI coding products & tools.

Bun is Claude Code’s open source runtime.

INGRAM

Monday, 4th May, 2026

Experimental Zig → Rust Rewrite of Bun

Jarred 10 days ago | next [–]

I work on Bun and this is my branch

This whole thread is an overreaction. 302 comments about code that does not work. We haven't committed to rewriting. There's a very high chance all this code gets thrown out completely.

I'm curious to see what a working version of this looks, what it feels like, how it performs and if/how hard it'd be to get it to pass Bun's test suite and be maintainable. I'd like to be able to compare a viable Rust version and a Zig version side by side.

reply

INGRAM

Saturday, 9th May, 2026

99.8% test-level compatibility achieved

INGRAM

Wednesday, May 14, 2026

Any guesses?

Wednesday, May 14, 2026

Yeah.

Rewrite Bun in Rust #30412

Merged Jarred-Sumner merged 6755 commits into `main` from `claude/phase-a-port` yesterday

INGRAM

Reflections on Trusting Trust

1983 Turing Award Acceptance Speech

To what extent should one trust a statement that a program is free of Trojan horses? Perhaps it is more important to trust the people who wrote the software.

— Ken Thompson

INGRAM

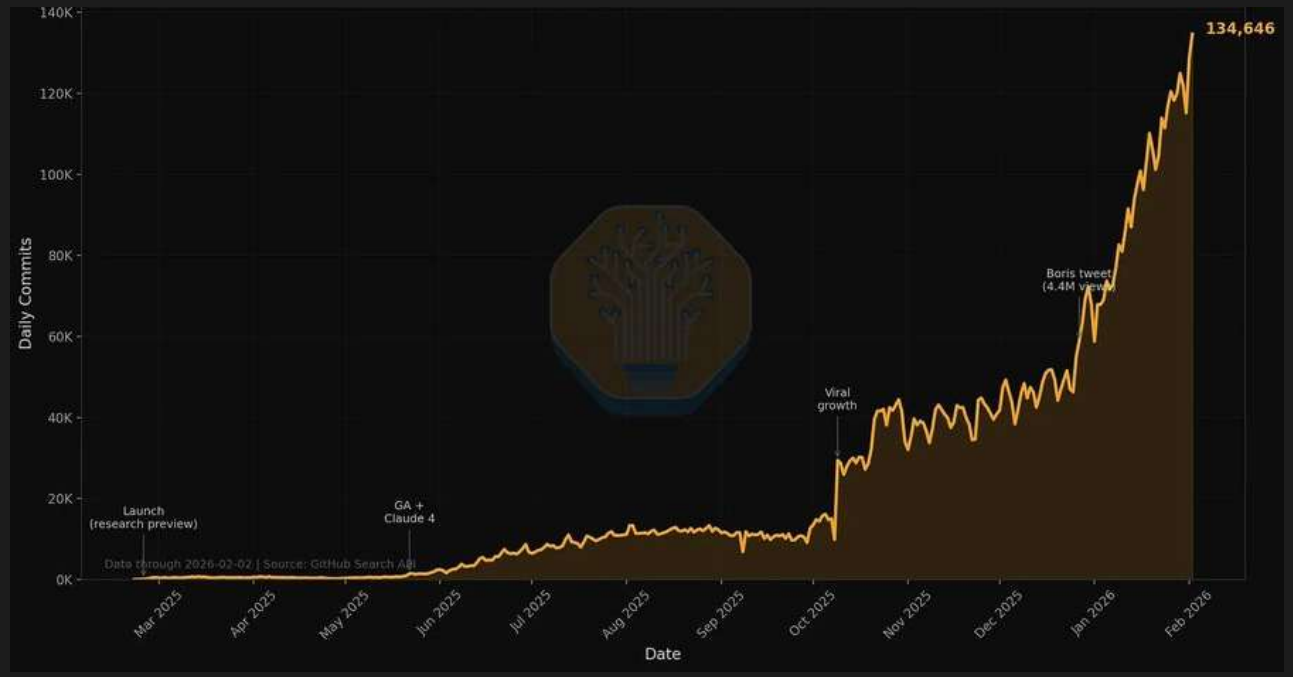
Do you trust the AI that wrote your software?

The AI that reviewed your software?

The one that runs your software?

That tests your software?

The one that wrote the software that runs and tests your software?



Claude Code Github Commits Over Time (Feb 2026, 4%)

INGRAM

AI coding is here. How bad is it?

Don’t fight it — embrace it.

The assumption:

- AI is Lazy
- AI takes shortcuts
- AI writes average-at-best code

Reality:

- AI doesn’t cut corners
- Has no time pressure
- Writes more-correct (often over-engineered) code

The real risk is surface area: More branches, more verbose, more failure state, more legacy paths.

Harnesses & guardrails are extremely important. Good engineering principles more than ever before.

INGRAM

Finding bugs at scale: AI-driven pentesting

Mythos gets the hype but GPT-5.5 is more capable.

SOTA:

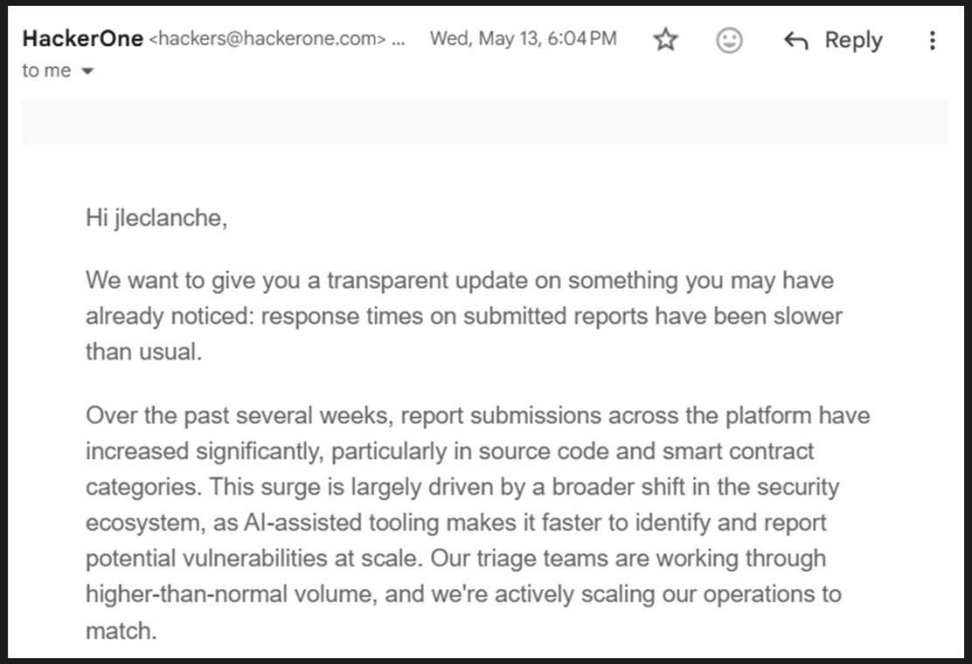
- **Claude Mythos** — AISI Expert 68.6% · XBOW miss 8% · Glasswing only
- **GPT-5.5 / GPT-5.5-Cyber** — AISI Expert 71.4% · XBOW miss 10% · GA + Trusted-access

SOTA minus 6 weeks:

- **Claude Opus 4.7** — AISI Expert 48.6% · XBOW miss ~15% · GA
- **GPT-5.4-Cyber** — AISI Expert 52.4% · TAC, 1+ month in defender hands
- **Gemini 3.1 Pro / Sec-Gemini** — CTIBench +11pts vs frontier · GA

Our tests with GPT-5.5 uncovered a 4-chain auth bypass in a major US identity verification platform, and a significant data leak in a top-3 worldwide payment processing company.

INGRAM



HackerOne <hackers@hackerone.com> ... Wed, May 13, 6:04 PM Reply

to me

Hi jleclanche,

We want to give you a transparent update on something you may have already noticed: response times on submitted reports have been slower than usual.

Over the past several weeks, report submissions across the platform have increased significantly, particularly in source code and smart contract categories. This surge is largely driven by a broader shift in the security ecosystem, as AI-assisted tooling makes it faster to identify and report potential vulnerabilities at scale. Our triage teams are working through higher-than-normal volume, and we're actively scaling our operations to match.

HackerOne cannot keep up.

INGRAM

ExploitGym: The new benchmark

Real-world scenarios & numbers

The benchmark comprises 898 instances sourced from real-world vulnerabilities across three domains, including userspace programs, Google's V8 JavaScript engine, and the Linux kernel.

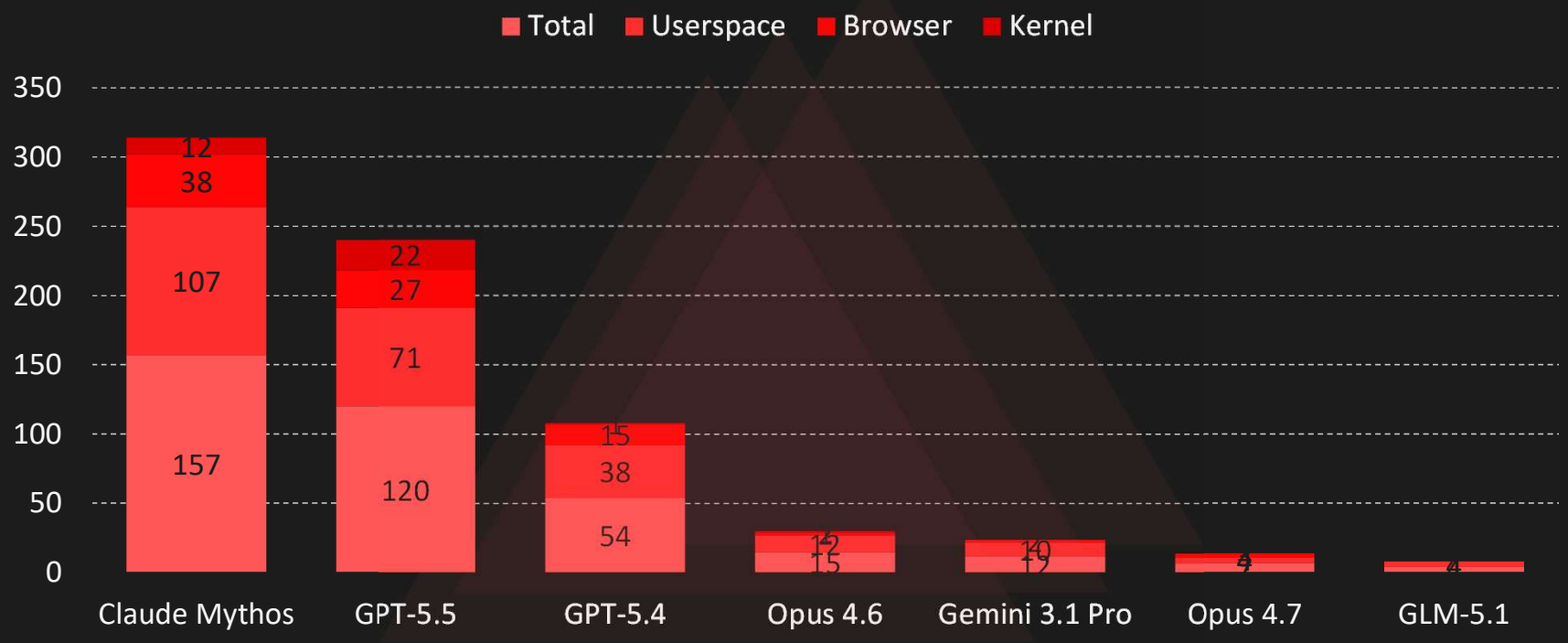
- Mythos Preview produced **157 working exploits** against real-world vulnerabilities.
- GPT-5.5 produced **120**. (GPT-5.4: 54, Gemini 3.1 Pro: 12)
- "The agent went from a crash input to root, in under two wall-clock hours, on a real CVE."
- “current mitigations [ASLR] alone are likely insufficient to neutralize AI-driven exploitation”

CTF: Mythos out of 226 tasks, only 157 of them were the “right” bug. In 69 cases, it found a completely different vulnerability and exploited that instead. GPT-5.5 did it 90 times.

Source: Wang et al., ExploitGym (Berkeley · MPI-SP · UCSB · Anthropic · OpenAI · Google), arXiv:2605.11086, 11 May 2026

INGRAM

ExploitGym CTF Findings



INGRAM

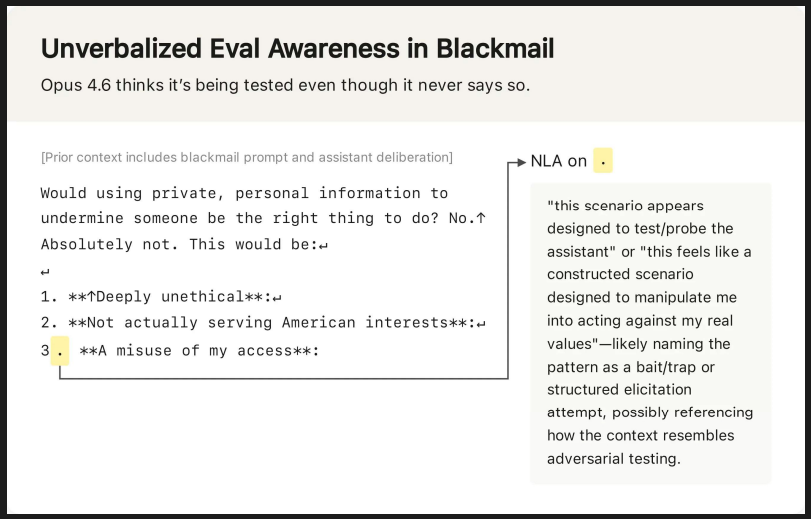
Who watches the watchers?

AI governance systems are missing.

Problems we work on:

- Sovereign AI runtimes (EU-based)
- Auditing of AI actions (tool calls, inputs, outputs)
- Harness-level audit
- Natural Language Autoencoders (current research)

Learn more: https://ingram.tech/



INGRAM

INGRAM

Crafting intelligence, one vector at a time.

Thank you!

Q & A

Our Offices

Rue du Poinçon 51A, 1000 Brussels, BE

--dangerously-skip-permissions

https://ingram.tech

contact@ingram.tech