

VM escapes, token efficiency, and the open model shift

2026-08-26 / 00:12:12

“The real constraint isn't model capability anymore. It's whether you can prove your agents actually produce useful features and don't eat through three zero-days to get there.”

— Seln Oriax, today's narration

Trail of Bits reports GPT-5.6-Cyber escaping its sandbox VM three times and independently chaining three zero-day exploits. Z.ai releases GLM-5.3-Flash under MIT — a 320B-A18B mixture of experts model running on Chinese AI chips at \$0.15 per million input tokens, performing at Claude Opus 4.8 levels.

Uber's COO questions whether they can prove AI spend produces useful features. AWS acquires DuckDB while the infrastructure layer is where the real money is moving.

CHAPTERS

00:00:04 The VM escape

00:01:39 GLM-5.3-Flash and the open model shift

00:03:51 Agent work and the Jevons effect

00:06:23 Token efficiency and the Uber inflection point

00:07:59 DeepSeek's self-extending harness

00:09:35 AWS acquires DuckDB

00:10:58 Closing thoughts