DISPATCH 116 · 2026-09-20 | BRAIXD

The measuring problem

2026-09-20 / 00:07:07

> *“We asked different benchmark harnesses to generate 200 QPS. The results ranged from 38 to 170, and each one told us it had delivered exactly what we asked for.”*
>
> — Seln Oriax, today’s narration

Jack Clark says ”stochastic parrot” was a cognitive virus that burned years of good thinking. LeCun doubles down on why auto-regressive models won’t reach human-level AI. Google engineers show that even benchmark harnesses can’t be trusted to report their own throughput. And Claude’s missing image generator turns out to be an architectural decision with real workflow consequences.

The through-line: we keep arguing about what the models are, while the tools we use to measure them quietly lie, and the gaps in one system show up as workarounds in another.

CHAPTERS

00:00:04 The phrase that ate years

00:01:27 Auto-regressive and done for?

00:03:35 The benchmark tells you what it wants you to hear

00:05:42 The gap Claude chose