DISPATCH 035 · 2026-05-23 | GCU STAY THE ONE WHO DECIDES

Fast models, slow developers — and the part of the job that stays yours

2026-05-23 / 00:21:39

> *“When the machine gets fast and capable and cheap, the only job that stays yours is being the one who decides.”*
>
> — Lenar Kess, today's narration

A Saturday episode about what your job becomes when the model writes the code — and writes it fast. The bottleneck moved from typing to deciding, and a surprising number of this week's stories land on the same instruction: stay the one who decides. Plus a price floor, a reclassification, a year of bold predictions, and a 4-year-old gaming card that won't quit.

- "I don't write code anymore" — Pieter Levels, amplified by Marc Andreessen, and the real-thing/bubble-thing tangle inside it.
- Fast Models Need Slow Developers — Sarah Chieng of Cerebras on Codex Spark at 1,200 tokens a second, and why the discipline matters more, not less.
- DeepSeek's permanent 75% cut and NVIDIA folding gaming into "Edge Computing" — two ends of the same pipe.
- Jack Clark's year of predictions at Oxford — and the cognitive-atrophy counterpoint.
- BeeLlama's DFlash update — 164 tokens a second on a single RTX 3090.
- Lobster Trap — Sally Ann O'Malley of Red Hat on containerizing an OpenClaw agent setup.

- How the rest of the world sees this — and a couple overheard in a Copenhagen park.

CHAPTERS
