

# The harness, not the model — and the trust layer racing to catch up

2026-05-26 / 00:24:26

*“A coding agent is good enough to trip your social instincts and not good enough to honor them.”*

— Lenar Kess, today's narration

One developer catching you up on the day in AI and the craft of building with it. Today: the wrapper around a model can move a benchmark more than the model does, a watermark goes multi-lab, and a decensoring tool with thirteen million downloads shows where that watermark leaks. Plus a sharp little essay on why coding agents make us so mad, the jobs data behind the panic, and three things you can pick up today.

- [The harness, not the model](#) — a Google DeepMind Kaggle talk and an arXiv position paper argue the agent [harness can swing a score ~22%](#) while frontier models tie.
- [Gemini Omni](#) — editing video by talking to it, with SynthID baked in ([community reaction](#)).
- [SynthID becomes a shared layer](#) — 100 billion watermarks, Search and Chrome, and OpenAI/ElevenLabs/Kakao on board.
- [Heretic in the Financial Times](#) — decensoring open weights in ten minutes, and the [artifact that proves the gap](#).
- [The user is visibly frustrated](#) — why conversational agent UX trips your social wiring.

- A rage-quitting modder and the jobs data — backlash, and what the numbers actually say.
- The bench — NuExtract3, EAGLE 3.1, and a rejected llama.cpp patch worth grabbing.

---

## CHAPTERS

00:00:04 The harness is the variable nobody discloses

00:04:34 Gemini Omni: editing video by talking to it

00:06:59 SynthID becomes a shared layer

00:10:02 Heretic in the Financial Times

00:13:22 The user is visibly frustrated

00:16:33 Rage-quitting the modder, and the jobs data

00:20:36 The bench: small models, faster tokens

00:23:21 What I'd watch next