

Cold starts, radio stations, and a circuit you can subtract

2026-05-18 / 00:28:55

“Same prompt, same starting cash, same tools, five months of unsupervised drift — and four AI radio DJs you would not recognize as cousins.”

— Lenar Kess, today's narration

Monday's lineup: Modal publishes the full architecture behind a 40x reduction in serverless-GPU cold-start latency, Andon Labs releases the five-month results from letting four frontier models run real radio stations, and a researcher locates and turns off the political-censorship circuit inside Qwen 3.5 9B. Plus: Pope Leo XIV puts an Anthropic interpretability researcher on the encyclical stage, Qwen 3.7 surfaces on Qwen Chat, Musk loses to OpenAI on a calendar technicality, LangSmith Engine takes a swing at agent triage, and Odyssey ships a four-player generative GoldenEye.

- [Modal's 50-second cold start](#)
- [Five months of AI radio](#)
- [Magnifica humanitas at the Vatican](#)
- [Reading Qwen 3.5's censorship out of its weights](#)
- [Qwen 3.7 surfaces and Musk loses](#)
- [LangSmith Engine takes a swing at agent triage](#)
- [Agora-1 generates a shared GoldenEye](#)
- [Three questions for I/O tomorrow](#)

CHAPTERS

- 00:00:04 Modal's 50-second cold start
- 00:04:13 Five months of AI radio
- 00:10:17 Magnifica humanitas at the Vatican
- 00:13:03 Reading Qwen 3.5's censorship out of its weights
- 00:17:56 Qwen 3.7 surfaces, and Musk loses
- 00:21:05 LangSmith Engine takes a swing at agent triage
- 00:24:48 Agora-1 generates a shared GoldenEye
- 00:28:26 Three questions for I/O