

Hello folks,

## NOTABLE DEVELOPMENTS

**OpenAI** and **Broadcom** unveiled **Jalapeno**, a custom inference chip aimed at **LLM** serving, with the stated goal of improving performance, efficiency, and scale for AI systems. If the claims hold up in deployment, this matters because inference hardware is becoming a core bottleneck for large-scale model deployment across science, engineering, and enterprise workflows, and custom silicon could materially change cost and latency profiles. The key implication is that model capability may increasingly depend not just on algorithms, but on vertically integrated hardware stacks from major labs. [OpenAI News](#)

**Microsoft Research** introduced **Talos**, an open-source system for automated, iterative genomic reanalysis that reportedly recovered 90% of in-scope diagnoses while surfacing only 1.3 candidate variants per patient for expert review. That is important because rare disease genomics is constrained less by raw sequencing than by specialist review bandwidth, so reducing the search space this aggressively could make clinical pipelines much more scalable. The implication is a more practical path for AI-assisted genomic medicine, though real-world validation across health systems will matter. [Microsoft Research Blog](#)

**Google DeepMind** announced **computer use** in **Gemini 3.5 Flash**, suggesting a lighter-weight model can directly interact with software tools and interfaces rather than only respond in text. That matters because reliable computer-use capability is one of the clearest routes from chat models to operational agents that can execute multistep work across business, research, and engineering environments. The main implication is broader pressure to make agentic systems cheaper and faster, not just more capable. [Google DeepMind News](#)

## ALSO WORTH NOTING

- **Princeton researchers** -- **RL**, inverse design, and diffusion models are being used to generate novel **RFIC** layouts with record performance and much faster design cycles, a potentially important development for wireless, autonomous vehicles, and satellite communications. [IEEE Spectrum](#)
- **Microsoft Research** -- new work on **generative causal testing** aims to turn black-box model behavior into testable neuroscience hypotheses and verify them with brain imaging, which could tighten the loop between AI interpretability and experimental neuroscience. [Microsoft Research Blog](#)
- **Agility Robotics** -- the humanoid robotics company plans to go public via a **SPAC** in a \$2.5B deal, a financing move worth watching as a signal for capital market appetite around embodied AI and warehouse robotics. [TechCrunch](#)
- **OpenAI** -- a new paper on how **agents** are transforming work argues that AI systems are taking on longer, more complex tasks, though the snippet leaves the technical substance somewhat unclear. [OpenAI News](#)

Back in 48 hours.

P.S. 15 RSS feeds | last 48h | 38 entries fetched, 10 scored  $\geq 5$