OpenAI's Math Milestone and Anthropic's Colossus-Scale Bet

AI High Signal Digest

2026-05-21

OpenAI's Math Milestone and Anthropic's Colossus-Scale Bet

By AI High Signal Digest • May 21, 2026

OpenAI claimed the first autonomous AI solution to a major open math problem, while Anthropic paired a multiyear Colossus compute deal with projected profitability. Also inside: agent research advances, new product launches from Cohere, OpenAI, and Google, plus notable funding and policy moves.

Top Stories

Why it matters: today's clearest signals were AI reaching a new research threshold, and frontier-model economics becoming even more tied to massive infrastructure commitments.

- **OpenAI said a general-purpose reasoning model solved the planar unit distance problem**, a famous Erdős question posed in 1946. The model found a new family of constructions that beats the square-grid approach mathematicians had treated as best for nearly 80 years, which OpenAI called the first time AI has autonomously solved a prominent open problem central to a field of mathematics [1, 2].

  > "What's significant about this moment is that it's the first really clear example of AI solving — not just an unsolved math problem — but a really well-known math problem." [3]

The result also comes less than a year after frontier models reached IMO gold-level performance, marking a fast jump from competition math to original research [4, 5].

- **Anthropic's scale story sharpened on both compute and finances.** A SpaceX filing says Anthropic is paying **$1.25B per month through**

May 2029 for capacity across Colossus and Colossus II [6], while WSJ-reported projections put Q2 revenue at **$10.9B** and first operating profit at about **$559M** [7]. Together, those numbers suggest frontier AI demand is now supporting both multiyear infrastructure commitments and near-term profitability.

Research & Innovation

Why it matters: beyond headline model releases, the most useful research updates focused on making agents cheaper, more scientifically useful, and easier to evaluate honestly.

- **Databricks introduced MemEx**, a programmable Python scratchpad that lets agents transform, slice, and persist tool outputs as typed objects instead of flooding the context window [8]. On enterprise agentic tasks, Databricks says frontier models gained **2–5 accuracy points** at **25–30% lower cost**, while Qwen 122B and 397B nearly doubled accuracy at **40–50% lower cost** [8].
- **Hugging Face released Carbon**, an open DNA base model with open weights, training code, and data pipeline for downstream biology tasks [9]. The team says Carbon is **275x faster** than the next best model at its size, can run locally, and can process a whole human genome on a single GPU in under two days [9].
- **InferenceBench offered a reality check on AI R&D automation.** Its creators say current frontier agents still struggle with system-level engineering and complex dependencies, and underperform simple hyperparameter-tuning baselines for vLLM and SGLang [10]. They also found weak strategy diversity, with most agents defaulting to vLLM rather than exploring alternatives [10].

Products & Launches

Why it matters: new launches are pushing AI from standalone chat into reusable workflows for coding, science, and open deployment.

- **Cohere open-sourced Command A+**, its fastest and most powerful model yet, under Apache 2.0 [11, 12]. Cohere says it supports **48 languages**, multimodal input, and can run on as few as **two H100s** [11, 13].
- **OpenAI brought Codex to mobile.** Users can work with Codex from the ChatGPT mobile app, answer questions on the go, and continue the same thread later from a computer [14, 15].
- **Google DeepMind launched Science Skills for Antigravity**, integrating insights from more than 30 life-science sources including UniProt and the AlphaFold Database [16]. In a test on a rare disease caused by

AK2 mutations, the team said the tool sped up structural analysis and surfaced novel insights into the condition's mechanisms [17].

Industry Moves

Why it matters: capital and partnerships are concentrating around the data, search, and vertical workflows that agents need to be useful in production.

- **Exa raised $250M at a $2.2B valuation** to keep building search for agents [18]. The company says it now serves **5,000+ companies** and **500,000+ developers**, and makes agents smarter and cheaper by returning **90% less text** with little to no tradeoff in RAG quality [18].
- **Genesis and Incyte expanded their AI drug-discovery partnership** with **$120M upfront**, recurring research funding, and potential milestone and royalty payments [19]. Incyte's proprietary experimental data will help train the next generation of Genesis foundation models, starting with at least five new collaboration targets [19].
- **RADAR became a new physical-AI unicorn** with a **$170M Series B** [20]. The company says its retail system delivers **99% item-level inventory accuracy** in real time and is already deployed in more than **1,400 stores** [20].

Policy & Regulation

Why it matters: governments are moving closer to pre-release oversight of frontier systems.

- The White House briefed OpenAI, Anthropic, and Reflection AI on a planned executive order that would create a **voluntary** framework asking labs to share frontier models with government agencies up to **90 days before public release** [21].

Quick Takes

Why it matters: a few smaller updates added signal on agents, search efficiency, and generative media.

- More than **50% of Devin sessions** are now triggered by agents rather than humans [22, 23].
- **Perplexity** said its query-aware context compression cuts context tokens by up to **70%** while improving answer quality [24, 25].
- **Stable Audio 3** launched with three open-source variants plus a closed "large" model [26].
- **MiniMax Speech 2.8 Turbo** arrived on Together AI with **600+ voices** and sub-250ms latency across **40+ languages** [27, 28].

Sources

1. X post by @OpenAI
2. X post by @OpenAI
3. X post by @scaling01
4. X post by @polynoamial
5. X post by @alexwei_
6. X post by @eliebakouch
7. X post by @scaling01
8. X post by @DbrxMosaicAI
9. X post by @ClementDelangue
10. X post by @maksym_andr
11. X post by @cohere
12. X post by @cohere
13. X post by @cohere
14. X post by @OpenAIDevs
15. X post by @OpenAIDevs
16. X post by @GoogleDeepMind
17. X post by @GoogleDeepMind
18. X post by @jeffzwang
19. X post by @genesismolai
20. X post by @SpencerHewett
21. X post by @steph_palazzolo
22. X post by @imjaredz
23. X post by @shayanshafii
24. X post by @perplexity_ai
25. X post by @perplexity_ai
26. X post by @multimodalart
27. X post by @togethercompute
28. X post by @togethercompute