

OpenAI's Math Result, NVIDIA's Agent Stack, and Scientific AI Systems

AI Leaders Briefing

2026-06-08

OpenAI's Math Result, NVIDIA's Agent Stack, and Scientific AI Systems

By AI Leaders Briefing • June 8, 2026

A model-generated math disproof, NVIDIA's full-stack agentic push, and new lab evidence on scientific and coding workflows defined the week. Product and infrastructure moves also pointed toward more persistent, specialized AI assistants.

Top Signals of the Week

Alexander Wei, Hongxun Wu, and Lijie Chen — OpenAI

OpenAI said a reasoning model produced a disproof of the 80-year-old Erdős unit distance conjecture, showing that the square-grid construction is not close to optimal and instead drawing on class field theory and other number-theoretic tools [1, 2]. The team said the system was a general ChatGPT-like setup with code execution rather than a math-only model, and that success on this problem rose toward ~50% as more test-time compute was applied [2].

- **Why it matters:** It is a direct example of a general-purpose reasoning system being used on an open research problem, with longer inference time materially affecting performance [1, 2].

Jensen Huang — NVIDIA

At GTC Taipei, Huang said Vera Rubin is now in full production as a multi-rack pod-scale system built to process agentic AI, with five connected rack-scale systems spanning NVL72, a Vera CPU rack, Groq LPX, BlueField-4, and Spectrum-X networking [3, 4, 5]. NVIDIA described Vera as the first CPU built for agents and reported over 1.8x higher agentic sandbox performance than x86 under peak load, while RTX Spark extends the same agentic pattern to Windows

PCs with a Blackwell RTX GPU, Grace CPU, 128 GB unified memory, and 1 petaflop AI performance [3, 6, 7, 8, 9]. In a joint appearance with Satya Nadella, Huang said the same architecture is being co-engineered from Windows devices to Azure-scale systems, and that Microsoft’s Fairwater design delivered roughly 30x higher token generation and lower cost than Hopper [10, 11]. NVIDIA also introduced Nemotron 3 Ultra as an open model for long-running agents and Cosmos 3 as an open physical-AI omnimodel [12, 13, 14].

- **Why it matters:** NVIDIA is treating long-running agent loops as the core workload from desktop to AI factory, not just model training or one-shot inference [15, 11, 3].

Anthropic leadership — Anthropic

Anthropic published internal metrics it says show Claude materially accelerating research and coding. Mythos Preview chose the correct next research step 64% of the time in sessions where humans had taken a wrong turn, up from 22% in 2024; a standard code-training speedup test reached ~52x in April versus ~3x for Claude Opus 4 in May 2025; open-ended coding success rose to 76%; and engineers now ship 8x as much code per quarter as in 2021-2025 [16, 17, 18, 19, 20]. Anthropic said this does not guarantee recursive self-improvement, but argued the trend is moving faster than expected and is important enough for the Anthropic Institute to study directly [21, 22].

- **Why it matters:** Anthropic is turning AI-assisted research acceleration from a general claim into reported internal operating data, and pairing that with explicit work on alignment and control [21, 22].

Demis Hassabis — Google DeepMind

Google DeepMind made Co-Scientist available to individual researchers through Gemini for Science [23]. The system uses a coalition of Gemini-based agents to generate, debate, refine, and rank hypotheses, verify claims against literature and data, and bring in web search and specialized models [24, 25, 26]. DeepMind said it has already assisted experts on liver fibrosis targets, ALS approaches, and genetic leads for reversing aging [27].

- **Why it matters:** Multi-agent systems are moving into structured scientific workflows with explicit critique, ranking, and evidence checks [25, 26].

Sam Altman and OpenAI — OpenAI

OpenAI rolled out a new ChatGPT memory system that carries context across conversations, updates what matters over time, and lets users review or steer memory through a summary view [28, 29, 30, 31]. It also expanded Codex into role-specific plugins across 62 apps and 110 skills, and launched Sites so Codex

can turn work, ideas, and plans into shareable websites or apps for Business and Enterprise users [32, 33].

- **Why it matters:** OpenAI is moving its product surface toward persistent, specialized work agents rather than isolated chat turns [28, 32, 33].

Research & Engineering

OpenAI — GPT-Rosalind

OpenAI introduced new capabilities for GPT-Rosalind, a model series built for enterprise life sciences research. It combines GPT-5.5’s agentic coding and tool use with stronger support for drug discovery, analysis, design, and experimental workflows [34].

Anthropic — chemistry and cyber

Anthropic said Claude Opus 4.7 matches, and on some tasks beats, dedicated NMR software for determining molecular structures [35]. Separately, it studied 832 malicious accounts and mapped their activity onto a threat-technique database to evaluate how existing defenses perform against AI-enabled cyberattacks [36].

Demis Hassabis — Google / Gemma

Demis Hassabis marked the release of Gemma 4 12B, describing it as a unified, encoder-free multimodal model that can run locally on a laptop with 16 GB VRAM and is released under Apache 2.0. The launch also marked 150+ million Gemma 4 downloads [37, 38].

Jensen Huang — NVIDIA

NVIDIA released Cosmos 3 as an open omnimodel for physical AI with native understanding and generation across text, image, video, sound, and action, plus 32B “Super” and 8B “Nano” variants [13, 14]. NVIDIA said Cosmos 3 ranks #1 across seven physical-AI leaderboards and made it available on Hugging Face and GitHub [39, 40]. It also released Nemotron 3 Ultra for long-running agents, with claimed 5x faster and 30% cheaper inference than prior frontier open models [12, 3]. At CVPR, NVIDIA added composable physical-AI agent skills for data generation, simulation, policy training, and evaluation powered by Cosmos 3 [41].

Strategy & Industry

Sam Altman — OpenAI

Altman said OpenAI’s Saline Township site is its first Midwest data center and part of a gigawatt-scale Stargate campus, with potential uses spanning private

tutoring, small-business AI services, and medical research [42, 43]. In a separate interview, he said OpenAI expects \$1.4 trillion of infrastructure spending over eight years and that AI is shifting from query-response toward always-on background agents [44]. He also said public anxiety about AI is justified, rejected simplistic mass-unemployment messaging, and argued the field needs international coordination on the biggest safety risks [45]. OpenAI also made its frontier models and Codex generally available on AWS via Amazon Bedrock, preserving enterprise security, compliance, and governance workflows [46].

Arthur Mensch — Mistral AI

Mensch said Mistral has committed €4B to data centers in France and Sweden, including a new high-availability French inference site, with targets of 200 MW by 2027 and 1 GW by 2030 [47]. He also launched Vibe, an agentic enterprise platform built on open-source models with task delegation, state management, human approval gates, and custom app hosting on customer tenants [47]. His strategic view is that Europe has a narrow window to build sovereign AI infrastructure and full-stack control; for now, Mistral continues to rely on NVIDIA while exploring custom-chip options [47].

Aidan Gomez and Joelle Pineau — Cohere

Cohere said it differentiates by deploying models on customer infrastructure for privacy and security in sectors like telecom, energy, healthcare, finance, defense, and government, rather than requiring data to move into vendor-operated clouds [48]. Gomez said this makes the business more capital efficient because customers absorb inference costs [48], while Pineau tied Canada’s AI strategy to commercialization, sovereign compute, and homegrown model-building capacity [49, 50]. Gomez also argued Canada and Europe need deliberate sovereign-tech policies because market forces alone will not remove strategic single points of failure [48].

Jack Clark and Anthropic — Anthropic

Anthropic confidentially submitted a draft S-1 to the SEC, preserving the option to pursue an IPO [51]. It also expanded Project Glasswing, extending Claude Mythos Preview to about 150 additional organizations in more than 15 countries [52]. In interviews, Jack Clark said Mythos showed stronger cyber offense and defense capabilities than expected, leading Anthropic to provide access to select companies and the UK AI Security Institute for defensive testing before broader rollout [53, 54].

“The AI industry has a gas pedal, but it doesn’t have a brake pedal in the car.” [54]

Worth Watching

Fei-Fei Li — World Labs

Fei-Fei Li said World Labs is building large world models as the path to spatial intelligence, and drew a taxonomy between renderers, planners, and simulators. Her view is that simulators that respect physics, dynamics, geometry, and semantics are the key layer for robotics, digital twins, design, and agent training [55, 56]. She also argued robotics will need more capital and more scientifically grounded safety work, not just rhetoric [56].

Julien Chaumond and Hugging Face engineering — Hugging Face

Julien Chaumond launched SynthTraces, a compact harness in which an open model acts as a coding agent against Hugging Face open-source repos while a small local model simulates user requests; the project has already generated more than 2,000 published traces for training and fine-tuning [57]. In parallel, Hugging Face redesigned the `hf` CLI for coding agents and reported 0.93-0.94 task success versus 0.84-0.92 for curl/Python SDK baselines, with complex multi-step tasks consuming 2.4x-6x more tokens without the CLI; the optional skill cut tool calls by about 30% [58].

The common thread this week was convergence: longer-running agents, tighter domain workflows, and infrastructure built around iteration rather than one-shot responses [15, 25, 32, 44]. The next test is whether the strong internal and narrow-domain results now being reported can scale without losing control of cost, security, or human oversight [22, 59, 47, 58].

Sources

1. X post by @OpenAI
2. How a reasoning model cracked an 80-year-old math problem — the OpenAI Podcast Ep. 20
3. NVIDIA GTC Taipei 2026 Keynote | Full Replay
4. NVIDIA Vera Rubin Platform Ramping into Full Production | Built for the Era of Agents
5. X post by @nvidia
6. X post by @NVIDIAAIInfra
7. Announcing NVIDIA RTX Spark | GTC Taipei 2026 Keynote by CEO Jensen Huang
8. Early Preview of NVIDIA RTX Spark at Computex
9. X post by @nvidia
10. X post by @nvidia
11. Jensen Huang and Satya Nadella’s Conversation at Microsoft Build
12. X post by @nvidia
13. X post by @NVIDIAAI

14. X post by @nvidia
15. Dell Technologies World 2026 Keynote | May 18–21 | Las Vegas
16. X post by @AnthropicAI
17. X post by @AnthropicAI
18. X post by @AnthropicAI
19. X post by @AnthropicAI
20. X post by @AnthropicAI
21. X post by @AnthropicAI
22. X post by @AnthropicAI
23. X post by @GoogleDeepMind
24. X post by @GoogleDeepMind
25. X post by @GoogleDeepMind
26. X post by @GoogleDeepMind
27. X post by @GoogleDeepMind
28. X post by @OpenAI
29. X post by @OpenAI
30. X post by @OpenAI
31. X post by @OpenAI
32. X post by @OpenAI
33. X post by @OpenAI
34. X post by @OpenAI
35. X post by @AnthropicAI
36. X post by @AnthropicAI
37. X post by @demishassabis
38. X post by @googlegemma
39. X post by @nvidia
40. X post by @nvidia
41. X post by @nvidia
42. OpenAI CEO Sam Altman speaks at groundbreaking for Saline Township data center
43. Open AI's Sam Altman praises hyper-scale Saline data center construction site
44. CEO da OpenAI prevê gasto de US\$ 1,4 tri em IA e recua sobre demissões | BREAKING NEWS
45. Sam Altman: People are right to be anxious about AI
46. X post by @OpenAI
47. Can Europe Build an AI Champion? — Mistral CEO Arthur Mensch
48. Gibt es für Europa und Kanada noch eine Chance auf Unabhängigkeit, Aidan Gomez?
49. How can Ottawa support Canada's AI industry?
50. Stratégie canadienne sur l'intelligence artificielle : entrevue avec la cheffe de l'IA chez Cohere
51. X post by @AnthropicAI
52. X post by @AnthropicAI
53. Is It Already Too Late To Control AI? | Jack Clark, Co-Founder of Anthropic

- 54. “The AI industry... doesn’t have a brake pedal” | Anthropic co-founder on “AI’s continued advances”
- 55. World Labs’ Fei-Fei Li on Creating Large World Models
- 56. Bloomberg Tech | Morning Streams
- 57. X post by @julien_c
- 58. Designing the hf CLI as an agent-optimized way to work with the Hub
- 59. X post by @NVIDIAAIInfra