Grok 4.3 Lands, Codex Broadens Beyond Coding, and GPT-5.5 Closes the Cyber Gap

AI High Signal Digest

2026-05-01

Grok 4.3 Lands, Codex Broadens Beyond Coding, and GPT-5.5 Closes the Cyber Gap

By AI High Signal Digest • May 1, 2026

xAI improved the frontier model price-performance curve with Grok 4.3, OpenAI expanded Codex into general workplace automation, and GPT-5.5 matched Mythos on a key cyber capability threshold. The brief also covers grounded vision research from DeepSeek, DeepMind's medical agent, enterprise security launches, hyperscaler capex, and a notable White House intervention in frontier model access.

Top Stories

Why it matters: The clearest signals today were about where frontier AI competition is moving fastest: price-performance, workplace automation, and offensive/defensive cyber capability.

- **xAI launched Grok 4.3 with a stronger price-performance profile.** Grok 4.3 scored **53** on the Artificial Analysis Intelligence Index, above Muse Spark and Claude Sonnet 4.6 and **4 points** above the latest Grok 4.20, while cutting input prices by about **40%** and output prices by about **60%** versus Grok 4.20 [1]. Artificial Analysis also said it sits on the intelligence-vs-cost Pareto frontier, with a large jump to **1500 Elo** on GDPval-AA, though its AA-Omniscience tradeoff was mixed: higher accuracy, lower non-hallucination than Grok 4.20 [1, 2].
- **OpenAI pushed Codex beyond coding into general computer work.** Sam Altman said a "big upgrade" makes Codex useful for **non-coding computer work** [3]. OpenAI's launch materials position Codex as a work assistant that connects apps like Slack, Google Workspace, and Microsoft 365, summarizes information across apps and docs, drafts work,

plans next steps, and helps with research, slides, spreadsheets, and project plans [4, 5, 6].

- **GPT-5.5 reached the same new cyber threshold as Mythos.** The UK AI Security Institute said GPT-5.5 is the **second model** to complete one of its multi-step cyber-attack simulations end-to-end [7]. OpenAI's Mark Chen said GPT-5.5 performs similarly to Mythos on this long-horizon cyber range eval [8]. One cited evaluation estimated a human expert would need around **20 hours** for the full chain; GPT-5.5 completed it in **2 of 10** attempts, versus **3 of 10** for Mythos Preview [9].

Research & Innovation

Why it matters: The most important research updates targeted grounding, safety, and practical performance in real domains.

- **DeepSeek introduced "Thinking with Visual Primitives."** The method interleaves **points and bounding boxes** directly into reasoning trajectories to anchor language to physical coordinates [10, 11]. DeepSeek highlighted counting, spatial reasoning, and topological reasoning as key tasks, and said the model weights will later be integrated into its foundation model [11, 12].
- **Google DeepMind shared a new multimodal "AI co-clinician."** The research system is designed to support medical decision-making with high-quality evidence and can process live video and audio for cues such as gait, breathing, or rashes [13, 14, 15]. In testing, DeepMind said it made **zero critical errors in 97 of 98** primary-care queries under the adapted NOHARM framework, and matched or outperformed physicians in **68 of 140** assessed areas, while humans still did better on red flags and physical exams [14, 16].
- **Meta researchers proposed "Autodata."** The system frames data creation as an agentic process, with the key idea that more inference compute can be turned into higher-quality training data [17]. Meta said its first implementation, Agentic Self-Instruct, showed strong gains on scientific reasoning tasks versus classical synthetic-data methods [17].

Products & Launches

Why it matters: New launches focused less on demos and more on embedding AI into real developer and enterprise workflows.

- **Anthropic put Claude Security into public beta.** Anthropic said the product is available for **Claude Enterprise** customers and built into Claude Code on the web [18, 19]. It scans repositories for vulnerabilities, validates findings to reduce false positives, and suggests patches for review; commentary around the launch said it is powered by **Opus 4.7** [18, 20].

- **Alibaba released Qwen3.6 open-weight models.** The headline model, **Qwen3.6 27B**, scored **46** on the Intelligence Index, making it the top open-weights model under **150B** parameters; the **35B A3B** variant scored **43** [21]. Both models are Apache 2.0 licensed, support **262K** context, and include native vision input, though Artificial Analysis noted the 27B model is token-hungry and relatively expensive to run at Alibaba Cloud pricing [21].
- **Mistral launched Workflows in public preview.** The product is a Temporal-powered durable execution engine for running **human-in-the-loop** AI processes with data staying inside enterprise infrastructure [22].

Industry Moves

Why it matters: Compute budgets, enterprise deployment, and robotics capital formation are still the clearest structural signals in AI.

- **The hyperscalers' AI spending keeps accelerating.** Meta, Amazon, Alphabet, and Microsoft all beat Q1 2026 expectations, and combined **2026 capex** is on track to exceed **$650B**, with Alphabet guiding **$180-190B**, Microsoft **$190B**, Meta **$125-145B**, and Amazon spending **$44.2B** in Q1 alone [23].
- **Figure AI hit a $39B valuation.** A cited interview summary said the company has raised nearly **$2B** in four years to build general-purpose humanoid robots for real work at scale, and framed the central bottleneck as an **intelligence problem** [24].
- **Cognition highlighted a production Devin deployment in healthcare.** Evinova, AstraZeneca's health-tech subsidiary, is using Devin for regulatory documentation, bug triage, migrations, and test automation; Cognition said regulator documentation is now produced about **8× faster** than the earlier **35-40 hour** process across teams [25].

Policy & Regulation

Why it matters: Access to frontier models is increasingly being shaped by government priority, not just lab policy.

- **A post linking to a Wall Street Journal article said the White House blocked Anthropic from expanding Mythos access** from roughly **50** organizations to about **120**, not because the model was too dangerous, but because officials were concerned wider access would hamper their own use [26, 27].

Quick Takes

Why it matters: A few smaller updates still stood out across security, media generation, robotics, and open models.

- **OpenAI launched Advanced Account Security for ChatGPT**, adding passkeys or physical security keys, disabling password login, tightening recovery, and excluding those conversations from model training [28].
- **Suno V5.5** moved to **#1** on Artificial Analysis' instrumental and vocals leaderboards and added voice cloning, custom models, and personalized recommendations [29].
- **Unitree** launched a **dual-arm humanoid robot** starting at **$4,290**, with binocular vision and voice interaction [30].
- **Gemma 4** has already passed **50 million downloads** with nearly **1,500** community-built models based on it [31].

Sources

1. X post by @ArtificialAnlys
2. X post by @ArtificialAnlys
3. X post by @sama
4. X post by @OpenAI
5. X post by @OpenAI
6. X post by @OpenAI
7. X post by @AISecurityInst
8. X post by @markchen90
9. X post by @cryps1s
10. X post by @teortaxesTex
11. X post by @PKUCXK
12. X post by @teortaxesTex
13. X post by @GoogleDeepMind
14. X post by @GoogleDeepMind
15. X post by @GoogleDeepMind
16. X post by @GoogleDeepMind
17. X post by @jaseweston
18. X post by @claudeai
19. X post by @_catwu
20. X post by @kimmonismus
21. X post by @ArtificialAnlys
22. X post by @dl_weekly
23. X post by @kimmonismus
24. X post by @MollySOShea
25. X post by @cognition
26. X post by @kimmonismus
27. X post by @kimmonismus
28. X post by @cryps1s
29. X post by @ArtificialAnlys
30. X post by @UnitreeRobotics

31. X post by @osanseviero