

Opus 4.7 Lands as Codex Expands and OpenAI Moves Into Life Sciences

AI High Signal Digest

2026-04-17

Opus 4.7 Lands as Codex Expands and OpenAI Moves Into Life Sciences

By AI High Signal Digest • April 17, 2026

Anthropic's Opus 4.7 and OpenAI's expanded Codex defined the day, while GPT-Rosalind signaled a sharper move toward domain-specific frontier models. The brief also covers Qwen's new open model, Perplexity's local-compute push, and a notable Pentagon AI policy development.

Top Stories

Why it matters: The biggest shift today is toward more capable work agents and more specialized frontier models.

1. **Anthropic released Claude Opus 4.7.** Anthropic says it is its most capable Opus model yet, handling long-running tasks with more rigor, following instructions more precisely, and verifying its own outputs before reporting back [1]. Pricing stays at **\$5 / \$25 per million tokens**, with availability across the API, Bedrock, Vertex AI, and Microsoft Foundry [2]. Third-party measurements highlighted stronger coding and agentic performance, including **70% on CursorBench vs. 58% for 4.6** and **1753 on GDPval-AA** at max effort [3, 4]. The tradeoff: Anthropic says 4.6-tuned prompts may need rework because 4.7 interprets instructions more literally, and the new tokenizer can raise token counts [2].
2. **OpenAI expanded Codex far beyond coding.** Codex can now use apps on a Mac, open an in-app browser, generate images, remember preferences, and take on ongoing or repeatable tasks [5, 6]. OpenAI also added **90+ plugins**, GitHub review-comment handling, and remote SSH connections to devboxes [7, 8, 9].

“Codex for (almost) everything.” [5]

This matters because OpenAI is turning Codex into a broader computer-use agent, not just a coding assistant [10].

3. **OpenAI launched GPT-Rosalind for life sciences.** The new model is built for **biology, drug discovery, and translational medicine** [11, 12]. OpenAI says it is optimized for scientific workflows, with stronger performance in **protein and chemical reasoning, genomics analysis, biochemistry knowledge, and scientific tool use** [13]. Access starts as a trusted-access research preview for qualified customers including **Amgen, Moderna, the Allen Institute, and Thermo Fisher Scientific**, and OpenAI is also releasing a **Life Sciences plugin for Codex** [14, 12].

Research & Innovation

Why it matters: The research signal is two-sided: open models are getting stronger, but new benchmarks still show large capability gaps.

- **Qwen3.6-35B-A3B** is a new open sparse MoE model with **35B total parameters and 3B active**, released under Apache 2.0 [15]. Alibaba says it reaches agentic coding performance on par with models **10x its active size** [15], while third-party summaries highlighted **73.4 on SWE-Bench Verified, 51.5 on Terminal-Bench 2.0**, and strong vision scores despite the small active parameter count [16].
- **LongCoT** introduced **2,500 expert-designed** long-horizon reasoning problems across chemistry, math, computer science, chess, and logic [17]. At launch, the best frontier models were still **below 10% accuracy**, underscoring how far long-context systems remain from robust long-horizon reasoning [17].
- Google's **Auto-Diagnose** shows what applied LLM tooling can look like in production: inside Google's Critique code review system, it analyzes failure logs, summarizes relevant lines, and suggests root causes [18]. Google reported **90.14%** root-cause diagnosis accuracy on 71 real-world failures and usage across **52,635** distinct failing tests after deployment [18].

Products & Launches

Why it matters: New launches are increasingly about giving agents durable access to local software, media workflows, and voice interfaces.*

- **Perplexity Personal Computer** integrates with the Mac app for secure orchestration across local files, browser workflows, and native apps like **iMessage, Apple Mail, and Calendar** [19, 20]. It can also run in the background on a Mac mini and be triggered from iPhone [21].
- **Gemini 3.1 Flash TTS** landed in AI Studio, with tag-based control over vocal delivery such as pace and accent, plus a composer view for iteration and code export [22].

- **VOID**, now live on fal.ai, brings Netflix’s video object removal model to developers, including correction of physical interactions and support for **multiple objects, fast motion, and complex backgrounds** [23].

Industry Moves

Why it matters: Big companies are now reorganizing teams, distribution, and platforms around AI-native workflows.

- **Apple** is reportedly sending **close to 200 Siri staffers** to a multi-week coding bootcamp using AI tools such as **Claude Code** and **Codex** [24]. The report says roughly **60 engineers** stay on core development and another **60** handle evaluations and safety checks [24].
- **GenAI traffic is no longer a one-player market.** Similarweb data shows ChatGPT at **56.72%** share one month ago, Gemini at **25.46%**, and Claude at **6.02%**, versus **77.43%**, **6.00%**, and **1.40%** respectively 12 months earlier [25].
- **Salesforce launched Headless 360**, exposing Salesforce, Agentforce, and Slack as **APIs, MCP, and CLI** so AI agents can access workflows and data directly without the browser being the primary interface [26].

Policy & Regulation

Why it matters: Government adoption is moving from theory to contract language, and the terms matter.*

- Reporting cited in analysis says **Google is negotiating a classified Pentagon AI agreement** to deploy Gemini in secure environments [27]. The same analysis says proposed language would mirror OpenAI’s earlier Pentagon deal [27, 28]. Critics argue that even where red lines are stated, broad “all lawful purposes” language may leave room for wider military or surveillance use [28, 29].

Quick Takes

Why it matters: These smaller updates help show where momentum is building next.*

- **Stanford HAI** released the **2026 AI Index**, a 400+ page report covering AI performance, investment, labor, policy, and public sentiment [30].
- **Google DeepMind and Boston Dynamics** are powering **Spot** with Gemini Robotics embodied reasoning models for inspection-style tasks [31, 32].
- A **Cursor/UChicago** study across 500 teams found developers tackled **68% more high-complexity tasks** as models improved, while overall AI usage rose **44%** [33, 34].
- **PrismML** open-sourced **Ternary Bonsai**, a 1.58-bit model family it says is **9x smaller** than 16-bit counterparts [35].

Sources

1. X post by @claudeai
2. X post by @kimmonismus
3. X post by @scaling01
4. X post by @ArtificialAnlys
5. X post by @OpenAI
6. X post by @OpenAI
7. X post by @OpenAI
8. X post by @OpenAIDevs
9. X post by @OpenAIDevs
10. X post by @OpenAIDevs
11. X post by @OpenAI
12. X post by @kevinweil
13. X post by @OpenAI
14. X post by @OpenAI
15. X post by @Alibaba_Qwen
16. X post by @kimmonismus
17. X post by @iScienceLuvr
18. X post by @omarsar0
19. X post by @perplexity_ai
20. X post by @perplexity_ai
21. X post by @perplexity_ai
22. X post by @GoogleAIStudio
23. X post by @fal
24. X post by @kimmonismus
25. X post by @Similarweb
26. X post by @Benioff
27. X post by @kimmonismus
28. X post by @BlackHC
29. X post by @BlackHC
30. X post by @dl_weekly
31. X post by @GoogleDeepMind
32. X post by @GoogleDeepMind
33. X post by @cursor_ai
34. X post by @cursor_ai
35. X post by @PrismML