Claude Design Debuts as Stargate Advances and Epoch Sees Capability Acceleration

AI High Signal Digest

2026-04-18

Claude Design Debuts as Stargate Advances and Epoch Sees Capability Acceleration

By AI High Signal Digest • April 18, 2026

Anthropic expanded Claude into design workflows, Epoch AI reported faster progress across most capability metrics since reasoning models emerged, and new survey data suggests OpenAI's Stargate buildout is materially underway. Also inside: web-agent research, new automation tools, and fresh enterprise signals.

Top Stories

Why it matters: The biggest AI story today is that labs are expanding from model releases into workflow ownership, while the compute and capability curves behind those products keep steepening.

- **Anthropic pushed Claude beyond chat and coding with Claude Design.** The new tool lets users create prototypes, slides, and one-pagers by talking to Claude; it supports inline edits, sliders, export to Canva/PPTX/PDF/HTML, and handoff to Claude Code. It runs on Claude Opus 4.7 and is rolling out in research preview to Pro, Max, Team, and Enterprise users [1, 2]. That matters because Anthropic is productizing end-to-end creative work, not just model access. Posts tracking the launch also pointed to Figma shares falling about 7% after the announcement [3].
- **Epoch AI found signs of faster capability growth after reasoning models arrived.** Across four capability metrics, Epoch reported strong evidence of acceleration in three: ECI, METR's 50% time horizon, and a math index were best fit by two linear trends with a break around the arrival of reasoning models, while WeirdML V2 did not show the same acceleration [4, 5, 6]. Epoch says the result survives multiple robustness

checks, but also notes these metrics lean heavily toward math and programming, where RL-style verification is easier than in messier domains [7, 8].

- **OpenAI's Stargate buildout looks materially underway.** Epoch AI says all seven US Stargate sites show visible development and that the project appears on track for more than 9 GW by 2029, comparable to New York City's peak power demand [9, 10]. Abilene, Texas is already estimated at 0.6 GW operational today and 1.2 GW by Q3 2026 [11]. The significance is strategic: frontier AI is becoming a power-and-construction race as much as a model race.

Research & Innovation

Why it matters: Research progress is increasingly about making agents more durable, reusable, and efficient rather than only pushing raw benchmark scores.

- **FrontierSWE** is a new ultra-long-horizon coding benchmark where agents get up to 20 hours to solve tasks such as optimizing a video rendering library or training models for quantum-property prediction, and they still rarely succeed [12]. It is a useful reality check on how far current coding agents remain from sustained autonomous engineering.
- **WebXSkill** teaches web agents reusable skills from synthetic trajectories. Reported gains include 69.5% on WebArena versus 59.7% for baselines, and 86.1% on WebVoyager in grounded mode; guided skills also transferred across environments at 85.1% [13]. The authors also note stronger models benefit more from grounded execution, while weaker ones gain more from guided mode [13].
- **Ternary Bonsai** from PrismML uses ternary weights {-1, 0, +1} to build models the company says are 9x smaller than 16-bit counterparts while outperforming most peers in their parameter classes on standard benchmarks [14]. The models are open-sourced in 8B, 4B, and 1.7B sizes under Apache 2.0 [14].

Products & Launches

Why it matters: The product layer is shifting toward persistent automation, local agents, and cheaper multimodal building blocks.

- **Claude Code Routines** adds serverless automations that can be triggered by schedule, API call, or GitHub webhook, with daily run caps depending on plan tier [15].
- **Ollama 0.21** now supports **Hermes Agent**, which Ollama describes as a self-improving agent that creates skills from experience, improves them during use, persists knowledge, searches past conversations, and builds a user model across sessions [16, 17].

- **Fish Audio S2 Pro** became the leading open-weights model on Artificial Analysis's speech arena, with 1,165 Elo, multi-speaker and multi-turn generation, natural-language prosody tags, and API pricing of $15 per 1M characters [18].

Industry Moves

Why it matters: Enterprise adoption is still moving fast, but the business models around AI are getting closer scrutiny.

- **TextQL raised $17M led by Blackstone** to build agentic analytics for messy enterprise data. The company says it grew revenue 9x year over year, posted 300%+ net dollar retention, and is live at Blackstone, Scale AI, and Dropbox, where its system queries across 400K+ tables and 100K+ dashboards [19, 20, 21, 22].
- **OpenAI saw notable leadership turnover.** Bill Peebles said he is leaving after helping build Sora from zero to one, highlighting early gains in object permanence and a rapid jump to high-fidelity 1080p multi-shot generation [23]. Kevin Weil also said OpenAI for Science is being decentralized into other research teams as he departs [24].
- **Revenue quality is becoming a live debate in enterprise AI.** Scott Stevenson argued that some startups are inflating "Contracted ARR" by annualizing future step-up pricing on multi-year deals even when current cash collection is much lower and customers can opt out after 12 months [25]. His example showed roughly $100M reported ARR versus $35M in cash-generating ARR by Q5, with forward-deployed engineers further pressuring margins [25].

Quick Takes

Why it matters: These are smaller items, but each points to where the next capability or deployment shift may come from.

- Anthropic introduced **Claude Mythos Preview**, a model that can autonomously identify and exploit serious software vulnerabilities; it is not being released publicly and is instead being tested with industry partners first [26].
- **Muse Spark** ranked **#3 on ClawEval**, ahead of GPT-5.4 and Gemini 3.1 Pro, according to Alexandr Wang [27].
- AMD and EmbeddedLLM say the **MORI-IO KV Connector** boosts vLLM single-node goodput by **2.5x** and keeps decode stable at max load [28].
- **Qwen 3.6** can now preserve chain-of-thought between turns, which researchers say could improve reasoning efficiency if context clutter stays manageable [29, 30].

Sources

1. X post by @claudeai
2. X post by @TheRundownAI
3. X post by @kimmonismus
4. X post by @EpochAIResearch
5. X post by @EpochAIResearch
6. X post by @EpochAIResearch
7. X post by @EpochAIResearch
8. X post by @EpochAIResearch
9. X post by @EpochAIResearch
10. X post by @EpochAIResearch
11. X post by @EpochAIResearch
12. X post by @MatternJustus
13. X post by @dair_ai
14. X post by @PrismML
15. X post by @dl_weekly
16. X post by @ollama
17. X post by @ollama
18. X post by @ArtificialAnlys
19. X post by @TheEthanDing
20. X post by @TheEthanDing
21. X post by @TheEthanDing
22. X post by @TheEthanDing
23. X post by @billpeeb
24. X post by @kevinweil
25. X post by @scottastevenson
26. X post by @DeepLearningAI
27. X post by @alexandr_wang
28. X post by @vllm_project
29. X post by @teortaxesTex
30. X post by @arankomatsuzaki