

AI-for-Science Claims Split as Agent Workflows Move Toward Production

AI News Digest

2026-06-12

AI-for-Science Claims Split as Agent Workflows Move Toward Production

By AI News Digest • June 12, 2026

A bold automated-research announcement landed beside benchmark evidence showing current limits in scientific synthesis. Meanwhile, OpenAI, Perplexity, and BBVA all pointed to the same quieter trend: AI systems are being packaged for longer-running, more governed, production use.

What stood out today

Automated discovery claims got stronger, but so did the evidence on current limits

Recursive unveiled what Richard Socher called a v0.1 “Eureka Machine”—an automated open-ended discovery system positioned as an early milestone toward recursive self-improving superintelligence—and said it reached state-of-the-art results on NanoGPT speedrun, NanoChat, and NVIDIA’s Sol-ExecBench, with the code and ideas behind those results invented by the AI itself and open-sourced for community investigation [1]. A new preprint pointed the other way: SciConBench introduces 9.11k scientific questions derived from Cochrane Systematic Reviews and reports that frontier AI agents cannot synthesize scientific conclusions well [2]. The contrast matters because DeepMind is explicitly building science-focused systems: Demis Hassabis described Gemini for science as a Gemini variant with tools for citations, literature lookup, and graph reading, and pointed to AlphaFold’s release of roughly 200 million protein structures, now used by more than 3 million researchers across 190 countries, as an example of “science at digital speed” [3, 4].

“science at digital speed” [4]

Agent workflows are getting more production-oriented

OpenAI reaches for secure background execution with Ona

OpenAI said it has reached an agreement to acquire Ona, whose secure cloud execution technology is meant to help Codex take on longer-running work even when laptops are closed and help more organizations deploy agents securely in production; after closing, Ona will join the Codex team [5]. OpenAI's description of the deal centered on secure execution and production deployment rather than a model release [5].

Perplexity folds Deep Research into its Computer agent

Perplexity said Deep Research is now a native skill inside its Computer agent harness and that the system is built on a new "Search as Code" architecture [6, 7]. The company says the model writes code to assemble searches, runs thousands of retrieval steps in parallel tailored to each question, and outperforms legacy Deep Research on every benchmark [6].

A concrete enterprise deployment example

BBVA lays out a bank-wide AI operating model across 120,000 employees

In an OpenAI event, BBVA described a top-down AI agenda organized around six specialized "robots" covering retail customer experience, banker advisory, risk, back-office work, software development with Codex, and general-purpose employee agents, alongside two pillars: data readiness and agent orchestration [8]. The bank said it has rolled out ChatGPT Enterprise to 120,000 employees worldwide and backed the rollout with dedicated adoption teams, executive dashboards, and training across regions [8]. BBVA also said bottom-up experimentation has produced more than 100 GPTs used by thousands of employees, with 70-80% time savings in many cases, and that its OpenAI partnership helped it make major course corrections along the way [8].

Worth watching

Google DeepMind launches a \$10M fund on collective AI behavior

Google DeepMind, together with Schmidt Sciences, Cooperative AI, and ARIA Research and with support from Google.org, launched a \$10 million fund to study the collective behaviors that can emerge when millions of AI agents interact [9]. The stated goal is to understand how AI systems behave as a group, not just one model at a time [9].

Gemini Omni Flash is being positioned for developers, not just demos

Logan Kilpatrick said Google DeepMind’s Gemini Omni Flash is state-of-the-art on image-to-video, text-to-video, and video editing, pointed developers to a public benchmarks page, and said API access is coming soon [10, 11]. The announcement emphasized both benchmark claims and near-term developer distribution [11, 10].

Sources

1. X post by @RichardSocher
2. X post by @manoelribeiro
3. The AI Breakthrough That Will Change Everything (Google DeepMind CEO Interview)
4. AI and science with Demis Hassabis | The Royal Society x Nobel Prize
5. X post by @OpenAINewsroom
6. X post by @perplexity__ai
7. X post by @AravSrinivas
8. Customer Ignite Talk: Antonio Bravo Acin (Global Head of AI Transformation, BBVA) & OpenAI
9. X post by @GoogleDeepMind
10. X post by @OfficialLoganK
11. X post by @OfficialLoganK