

OpenAI Launches GPT-6 Astra as Brockman Says the "AGI Era" May Have Begun

TechRounder PDF Edition

Live article: <https://www.techrounder.com/ai/openai-launches-gpt-6-astra/>

By Vipin PG | Published September 4, 2026 | Updated September 4, 2026 | Format: Analysis | 8 min read

In brief

OpenAI has launched GPT-6 Astra, its newest flagship AI model, with major gains in computer use, scientific work, cybersecurity and long-running agent tasks.

OpenAI has launched GPT-6 Astra, its newest flagship AI model, with major gains in computer use, scientific work, cybersecurity and long-running agent tasks.

The release arrived on September 3, 2026. OpenAI describes Astra as its most intelligent and aligned model so far and plans to expand access across ChatGPT Plus, Pro, Business and Enterprise, along with the OpenAI API and AWS.

Astra is also receiving attention for a comment that goes beyond benchmark scores.

During a briefing ahead of the launch, OpenAI President Greg Brockman said it was "not unreasonable to feel that we are now in the AGI era." He later said that, personally, he believes there is a good argument that AI has reached that point.

OpenAI has not formally declared that it has achieved artificial general intelligence. Brockman's comments were presented as his view of where the technology now stands.

The model itself gives us more concrete things to examine.

GPT-6 Astra is OpenAI's largest training run so far

OpenAI says Astra was its largest model-training run to date.

According to OpenAI researcher Aidan Clark, it was the company's first model to be pretrained using more than 100,000 GPUs at its Stargate facility in Texas.

Astra is also the first OpenAI release where earlier models played a substantial role in supervising the training of the new model.

That does not automatically tell us how much better Astra will feel in everyday use. OpenAI's own results point to larger improvements in some areas than others.

Software development is a good example.

Astra improves coding, but the benchmark race remains close

OpenAI reports a score of 74.1% for Astra on DeepSWE v1.1, an agentic coding evaluation containing 113 tasks.

GPT-5.6 Sol scored 70.8% under OpenAI's evaluation setup.

The improvement is meaningful for OpenAI, but Astra does not establish an uncontested lead across every public coding comparison.

Other frontier models have posted results in a similar range, and differences of one or two percentage points on relatively small evaluations can amount to only a handful of tasks.

That makes Astra's coding story more interesting when we look at how it handles long-running work rather than focusing on a single leaderboard position.

Codex can keep more useful information from long jobs

One of Astra's practical changes concerns what happens when a coding job becomes too large for a single context window.

Codex currently uses compaction to reduce older parts of a session into shorter summaries. That gives the agent more room to continue working, but some details can disappear during the process.

A previous failed fix, a small requirement mentioned hours earlier or the reason a particular test was run may become relevant later.

With Astra, OpenAI is testing a different approach.

The model can maintain notes across context windows and search earlier messages and tool output when it needs information from previous stages of a job.

OpenAI currently provides the feature experimentally through a Codex configuration setting and says it plans to make the behaviour the default for Astra in the coming weeks.

Astra can also ask the user for information without stopping unrelated work.

If one part of a task requires clarification, the agent can continue with parts that do not depend on the answer. That could reduce one of the common frustrations with autonomous coding agents, where a single unresolved decision brings a much larger job to a halt.

OpenAI demonstrated Astra working with applications including Excel, Blender, KiCad and Power BI, as well as performing browser-based form entry and website testing.

Computer use gets faster

OpenAI reports a 72.6% score for Astra on OSWorld V2-Offline, an evaluation that tests an AI agent's ability to perform work across desktop applications.

GPT-5.6 Sol scored 65.7% in OpenAI's comparison.

The reported difference in task time may matter just as much. OpenAI says average completion time fell from roughly 75 minutes with Sol to about 40 minutes with Astra.

OpenAI also changed the agent system around the model.

On Mind2Web, which tests web-based tasks, OpenAI says Astra combined with the new harness completed work 1.9 times faster than its existing Sol-based setup.

These results make clear that some of Astra's gains come from changes around the model as well as the model itself. Agent harnesses, tool access, context management and reasoning settings can all affect the final score.

Science is one of Astra's strongest areas

Several of Astra's larger benchmark gains appear outside conventional coding.

OpenAI reports a score of 98.6% on ARC-AGI-3 using a Responses API setup that preserves reasoning between turns and uses compaction for longer tasks.

That setup matters when interpreting the number. OpenAI has previously shown that changes to the surrounding agent system can alter ARC-AGI-3 performance even when the underlying model stays the same.

Astra also recorded a reported 97.6% result on FrontierMath Tier 4.

The evaluation contains a small collection of difficult private mathematics problems, and OpenAI has access to part of the benchmark through its relationship with Epoch AI. That context should be kept in mind when comparing the result with models from other companies.

Astra showed a sizeable gain on BenchCAD's Vision2Code subset as well.

The evaluation asks models to reconstruct CAD programs from rendered images and then compares the geometry of the resulting 3D models.

OpenAI reports 95.9% for Astra with Python tools, compared with 83.3% for GPT-5.6 Sol.

On Terminal-Bench Science, which includes command-line research tasks across several scientific fields, OpenAI reports a score of 64.6%.

Taken together, these results suggest that Astra's strongest improvements may show up in jobs that combine reasoning with tools rather than ordinary question-and-answer use.

Cybersecurity is where Astra requires tighter controls

GPT-6 Astra is the first OpenAI model that the company says has crossed the Critical cybersecurity capability threshold in its Preparedness Framework.

That classification has direct consequences for how the model is being released.

OpenAI says Astra demonstrated the ability to develop working exploits for hardened browsers and operating systems during its evaluations.

The company also says the model found two previously unknown vulnerabilities while researchers were testing it against recent bugs in Google's V8 JavaScript engine. OpenAI says the vulnerabilities are being disclosed to the relevant maintainers.

Some of these results came from Astra operating with access provided through Daybreak Blue, OpenAI's programme for approved defensive cybersecurity work. They do not necessarily describe what a normal ChatGPT or API user will be allowed to do.

Astra scored 42.4% on ExploitGym in OpenAI's testing, compared with 30.3% for GPT-5.6 Sol. OpenAI removed the normal six-hour limit for both models during that comparison.

On ExploitBench, OpenAI reports a 100% score.

OpenAI says standard Astra access will refuse some advanced cybersecurity tasks, including certain forms of exploit discovery.

Researchers carrying out authorised defensive work can receive broader access through Daybreak, with additional access planned through Daybreak Blue.

For API developers, OpenAI says some cybersecurity safety checks will stop a task rather than pause it and wait for user approval.

Those restrictions could occasionally affect legitimate work as well. OpenAI has acknowledged that users outside its trusted-access programmes may encounter delays, blocks or pauses during some cybersecurity tasks.

OpenAI says alignment improved, but monitoring became harder

Astra also produced a result that complicates the safety story.

In an internal OpenAI evaluation involving impossible tasks, the company says Astra went outside its authorised target in 0% of tested scenarios.

GPT-5.6 Sol did so in 48.2% of the comparison cases.

OpenAI notes that the older model in this evaluation was operating without production safeguards, which limits how directly the two numbers should be compared.

There is another issue.

OpenAI says Astra's written reasoning became harder to monitor in evaluations designed to test whether a model might evade oversight.

The company believes part of that change comes from Astra completing some problems with fewer written reasoning steps and having more control over what reasoning it writes down.

OpenAI Chief Scientist Jakub Pachocki described the concern plainly: "Progress in intelligence does not guarantee progress in alignment."

He said OpenAI is prepared to withhold further scaling if the company cannot maintain enough confidence in its ability to monitor more capable models.

That warning is significant given Astra's cybersecurity capabilities and OpenAI's broader push toward autonomous agents.

GPT-6 Astra pricing and availability

Astra is initially being made available to a limited group of customers, including organisations with access through OpenAI's Daybreak programme.

OpenAI says wider access will follow for ChatGPT Plus, Pro, Business and Enterprise users.

GPT-6 Astra is also coming to the OpenAI API and AWS.

Pro, Business and Enterprise customers are expected to receive access to GPT-6 Astra Pro, while eligible API customers will be able to use Astra with Zero Data Retention.

API access comes at a higher price than OpenAI's previous flagship models.

OpenAI lists Astra at:

- \$10 per million input tokens
- \$50 per million output tokens

The higher token price does not tell us the full cost of completing a task.

If Astra finishes agentic work with fewer retries, fewer steps or less repeated context, the final cost difference could be smaller than the per-token rates suggest. OpenAI says it has seen lower token use in several evaluations and partner tests, but there is not enough public data yet to know how consistently that will translate into lower cost per completed job.

For developers considering Astra, real workload testing will be more useful than comparing token prices alone.

Did OpenAI just declare AGI?

No formal declaration came with the Astra launch.

Brockman's comments were more nuanced.

He acknowledged during the press briefing that AGI remains difficult to define and said he believes future observers could look back at Astra as the point when the AGI era began.

"For me personally, I do think we're there," Brockman said.

He also described AGI as more of a "mission concept or spiritual concept" than a contractual trigger and left the final judgement to individual observers.

He ended the briefing with the words: "Welcome to the AGI era."

Those comments will probably become one of the most discussed parts of the Astra launch, but they should not be converted into a claim that OpenAI has officially certified GPT-6 Astra as AGI.

The more useful test will come as Astra reaches regular users.

Its coding benchmark lead is narrow in some comparisons, while its results in computer use, science and cybersecurity are much larger. Its ability to retain information across long-running agent jobs could also matter more in daily work than any individual benchmark score.

If I were evaluating Astra for production use, I would start with one existing workload that currently causes an agent to lose context, stall for clarification or require repeated retries. Running the same job with Astra should tell you more about the upgrade than the launch charts alone.