# OpenAI returns old models to ChatGPT as Sam Altman admits ‘bumpy’ GPT-5 rollout

Want smarter insights in your inbox? Sign up for our weekly newsletters to get only what matters to enterprise AI, data, and security leaders. Subscribe Now  
  
OpenAI co-founder and CEO Sam Altman is publicly acknowledging major hiccups in yesterday’s rollout of GPT-5, the company’s new, flagship large language model (LLM) — advertised as its most powerful and capable yet.  
  
Answering user questions in a Reddit AMA (Ask Me Anything) thread and in a post on X this afternoon, Altman admitted to a range of issues that have disrupted the launch of GPT-5, including faulty model switching, poor performance, and user confusion — prompting OpenAI to partially walk back some of its platform changes and reinstate user access to earlier models like GPT-4o.  
  
“It was a little more bumpy than we hoped for,” Altman wrote in reply to a question on Reddit regarding the big GPT-5 launch.  
  
As for erroneous model performance charts shown off during OpenAI’s GPT-5 livestream, Altman said: “People were working late and were very tired, and human error got in the way. A lot comes together for a livestream in the last hours.”  
  
AI Scaling Hits Its Limits  
  
Power caps, rising token costs, and inference delays are reshaping enterprise AI. Join our exclusive salon to discover how top teams are:  
  
Secure your spot to stay ahead: https://bit.ly/4mwGngO  
  
While he noted the accompanying blog post and system card were accurate, the missteps further muddied a launch already facing scrutiny from early users and developers.  
  
GPT-5 rollout updates:\*We are going to double GPT-5 rate limits for ChatGPT Plus users as we finish rollout.\*We will let Plus users choose to continue to use 4o. We will watch usage as we think about how long to offer legacy models for.\*GPT-5 will seem smarter starting…  
  
One key reason for the trouble according to Altman stems from OpenAI’s new automatic “router” that assigns user prompts to one of four GPT-5 variants — regular, mini, nano, and pro — with an optional “thinking” mode for heavier reasoning tasks.  
  
On X, Altman revealed that a key part of that system — the autoswitcher — was “out of commission for a chunk of the day,” causing GPT-5 to appear “way dumber” than intended.  
  
In response, OpenAI says it’s implementing changes to the model decision boundary and will make it more transparent which model is responding to a given query.  
  
A UI update is also on the way to help users manually trigger thinking mode.  
  
Additionally, Altman confirmed that OpenAI will now allow ChatGPT Plus users to continue using GPT-4o — the prior default model — after a wave of complaints about GPT-5’s inconsistent performance. He said on Reddit the company is “trying to gather more data on the tradeoffs” before deciding how long to offer legacy models.  
  
they brought back the old models!!!! (it's in settings under "legacy models") pic.twitter.com/0MN85jmicG  
  
Yet many users including OpenAI beta testers like Wharton School of Business professor Ethan Mollick expressed confused and dismay at OpenAI unilaterally upgrading their ChatGPT experiences to GPT-5 and initially taking away access to the older models.  
  
OpenAI’s internal benchmarks may show GPT-5 leading the pack of LLMs, but real-world users are sharing a different experience.  
  
Since the launch, users have posted numerous examples of GPT-5 making basic errors in math, logic, and coding tasks.  
  
Data scientist Colin Fraser posted screenshots of GPT-5 incorrectly solving whether 8.888 repeating equals 9 (it does not, obviously), while another user showed it flubbing a simple algebra problem: 5.9 = x + 5.11.  
  
And still other users reported trouble getting accurate answers to math word problems or using GPT-5 to debug its own presentation charts.  
  
Developer feedback hasn’t been much better, with users posting images of GPT faring worse at “one-shot” certain programming tasks — completing them well with a single-prompt — compared to rival AI lab Anthropic’s new model Claude Opus 4.1.  
  
And security firm SPLX found GPT-5 still suffers from serious vulnerabilities to prompt injection and obfuscated logic attacks unless its safety layer is hardened.  
  
With 700 million weekly users on ChatGPT, OpenAI remains the largest player in generative AI by audience.  
  
But that scale has brought growing pains. Altman noted in his X post that API traffic doubled over 24 hours following the GPT-5 launch, contributing to platform instability.  
  
In response, OpenAI says it will double rate limits for ChatGPT Plus users, and continue to tweak infrastructure as it gathers feedback.  
  
But the early missteps — compounded by confusing UX changes and errors in a high-profile launch — have opened a window for rivals to gain ground.  
  
The pressure is on for OpenAI to prove that GPT-5 isn’t just an incremental update, but a true step forward. Based on the initial rollout, many users aren’t convinced — yet.  
  
If you want to impress your boss, VB Daily has you covered. We give you the inside scoop on what companies are doing with generative AI, from regulatory shifts to practical deployments, so you can share insights for maximum ROI.  
  
Read our Privacy Policy  
  
Thanks for subscribing. Check out more VB newsletters here.  
  
An error occured.

# OpenAI’s GPT-5 rollout is not going smoothly

Want smarter insights in your inbox? Sign up for our weekly newsletters to get only what matters to enterprise AI, data, and security leaders. Subscribe Now  
  
Updated Friday August 8, 5:21 pm ET: shortly after this post’s publication, OpenAI co-founder and CEO Sam Altman announced the company would restore access to GPT-4o and other old models for selected users, admitting the GPT-5 launch was “more bumpy than we hoped for.”  
  
The launch of OpenAI’s long anticipated new model, GPT-5, is off to a rocky start to say the least.  
  
Even forgiving errors in charts and voice demos during yesterday’s livestreamed presentation of the new model (actually four separate models, and a ‘Thinking’ mode that can be engaged for three of them), a number of user reports have emerged since GPT-5’s release showing it erring badly when solving relatively simple problems that preceding OpenAI models — and rivals from competing AI labs — answer correctly.  
  
For example, data scientist Colin Fraser posted screenshots showing GPT-5 getting a math proof wrong (whether 8.888 repeating is equal to 9 — it is of course, not).  
  
AI Scaling Hits Its Limits  
  
Power caps, rising token costs, and inference delays are reshaping enterprise AI. Join our exclusive salon to discover how top teams are:  
  
Secure your spot to stay ahead: https://bit.ly/4mwGngO  
  
Wow, I was just playing around before but it actually is stupid pic.twitter.com/ao51nOH0Ui  
  
It also failed on a simple algebra arithmetic problem that elementary schoolers could probably nail, 5.9 = x + 5.11.  
  
This is concerning. https://t.co/PUbeCSgtRV  
  
Using GPT-5 to judge OpenAI’s own erroneous presentation charts also did not yield helpful or correct responses.  
  
Q. Prove using an LLM-as-a-judge still doesn't workA. pic.twitter.com/KnCK5Xs9ja  
  
It also failed on this trickier math word problem below (which, to be fair, stumped this human at first…though Elon Musk’s Grok 4 AI answered it correctly. For a hint, think of the fact that flagstones in this case can’t be divided into smaller portions. They must remain in tact as 80 separate units, so no halves or quarters).  
  
Careful not to cut yourself on the jagged frontier pic.twitter.com/buJGgJ6baI  
  
The older 4o model performed better for me on at least one of these math problems. Unfortunately, OpenAI is slowly deprecating those older models — including the former default GPT-4o and the powerful reasoning model o3 — for users of ChatGPT, though they’ll continue to be available in the application programming interface (API) for developers for the foreseeable future.  
  
Even though OpenAI’s internal benchmarks and some third-party external ones have shown GPT-5 to outperform all other models at coding, it appears that in real world usage, Anthropic’s recently updated Claude Opus 4.1 seems to do a better job at “one-shotting” certain tasks, that is, completing the user’s desired application or software build to their specifications. See an example below from developer Justin Sun posted to X :  
  
Opus 4.1's one-shot attempt at "create a 3d capybara petting zoo" – 8 minutes totalThis was honestly pretty insane, not only are the capybaras way cuter and moving, there are individual pet affinity levels, a day/night switcher, feeding, and even a screenshot feature pic.twitter.com/FiKTO3FKK4  
  
In addition, a report from security firm SPLX found that OpenAI’s internal safety layer left major gaps in areas like business alignment and vulnerability to prompt injection and obfuscated logic attacks.  
  
While anecdotal, the checking the temperature on how the model is faring with early AI adopters seems to indicate a chilly reception.  
  
AI influencer and former Googler Bilawal Sidhu posted a poll on X asking for a “vibe check” from his followers and the wider userbase, and so far, with 172 votes in, the overwhelming response is “Kinda mid.”  
  
Alright, GPT-5 vibe check  
  
And as the pseudonymous AI Leaks and News account wrote, “The overwhelming consensus on GPT-5 from both X and the Reddit AMA are overwhelmingly negative.”  
  
The overwhelming consensus on GPT-5 from both X and the Reddit AMA are overwhelmingly negativeMost users are disgruntled about the broken model picker and non-pro users not having access to legacy modelsWhat are your initial thoughts on GPT-5?  
  
Tibor Blaho, lead engineer at AIPRM and a popular AI leaks and news poster on X, summarized the many problems with the ChatGPT-5 rollout in an excellent post, highlighting that one of the new marquee features — an automatic “router” in ChatGPT that chooses a thinking or non-thinking mode for the underlying GPT-5 model depending on the difficulty of the query — has become one of the chief complaints, given the model seemed to default to non-thinking mode for many users.  
  
A bit sad how the GPT-5 launch is going so far, especially after the long wait and high expectations– The automatic switching between models (the router) seems partly broken/unreliable– It's unclear exactly which model you're actually interacting with (standard or mini,…  
  
Thus, the sentiment toward ChatGPT-5 is far from universally positive, highlighting a major problem for OpenAI as it faces increasing competition from major U.S. rivals like Google and Anthropic, and a growing list of free, open source and powerful Chinese LLMs offering features that many U.S. models lack.  
  
Take the Alibaba Qwen Team of AI researchers, who just today updated their highly performant Qwen 3 model to have 1 million token context — giving users the ability to exchange nearly 4x as much information with the model in a single back/forth interaction as GPT-5 offers.  
  
Given OpenAI’s other big release this week — that of new open source gpt-oss models — also received a mixed reception from early users, things are not looking up for the number one dedicated AI company by users right now (700 million weekly active users of ChatGPT as of this month).  
  
Indeed, this is also exemplified by users of the betting marketplace Polymarket overwhelmingly deciding following the release of GPT-5 that Google would likely have the best AI model by the end of this month, August 2025.  
  
Other power users like Otherside AI co-founder and CEO Matt Shumer, who received early access to GPT-5 and blogged about it favorably in a review here, opined that views would shift as more people figured out the best ways to use the new model and adjusted their integration approaches:  
  
A lot of folks who are having a bad experience are using GPT-5 in agent harnesses that aren't yet optimized for it.For every new model release, there's a time lag between release + when companies that integrate the model have it truly working well.Agent companies rush to…  
  
While it’s still early days for GPT-5 — and the sentiment could change dramatically as more users get their hands on it and try it for different tasks — the early indications are not looking like this is a “home run” release for OpenAI in the same way that prior releases such as GPT-4, or even the newer 4o and o3, were. And that’s a concerning indicator for a company that just raised yet another funding round, yet remains unprofitable due to its high costs of research and development.  
  
If you want to impress your boss, VB Daily has you covered. We give you the inside scoop on what companies are doing with generative AI, from regulatory shifts to practical deployments, so you can share insights for maximum ROI.  
  
Read our Privacy Policy  
  
Thanks for subscribing. Check out more VB newsletters here.  
  
An error occured.

# ChatGPT users dismayed as OpenAI pulls popular models GPT-4o, o3 and more — enterprise API remains (for now)

Want smarter insights in your inbox? Sign up for our weekly newsletters to get only what matters to enterprise AI, data, and security leaders. Subscribe Now  
  
Updated Friday August 8, 5:21 pm ET: Following this post’s publication, OpenAI co-founder and CEO Sam Altman announced the company would restore access to GPT-4o and other old models for selected users, admitting the GPT-5 launch was “more bumpy than we hoped for.”  
  
After announcing the release of its newest flagship model family, GPT-5, OpenAI said the model will power all of ChatGPT, and that it will sunset the existing models in the chat platform.  
  
OpenAI, through a spokesperson, told VentureBeat that GPT-5 “will replace all other models in ChatGPT, so users don’t have to pick depending on each task, which takes effect once you have access to GPT-5.” This means people can no longer choose GPT-4o, o3, o4-mini or o4-mini-high.  
  
With GPT-5 access rolling out to ChatGPT Plus, Free, Pro and Team users starting, only the Enterprise and Edu tiers can still use the “legacy” models for 60 days.  
  
AI Scaling Hits Its Limits  
  
Power caps, rising token costs, and inference delays are reshaping enterprise AI. Join our exclusive salon to discover how top teams are:  
  
Secure your spot to stay ahead: https://bit.ly/4mwGngO  
  
The news came as a surprise to many ChatGPT users, who had come to rely on their chosen models to run their everyday queries. Some people said the adjustment would take some time getting used to, mainly because they had based workflows on how the model interacted with them or typical response times.  
  
Although I enjoy GPT-4.1, I am saddened by the news that you're also apparently sunsetting GPT-4.5. For me, it's been way better in textual and conceptual analysis than any other GPT-4x series model, ever. At the very least, please don't make ChatGPT users go back to 4o.  
  
Honestly, I liked O3 a lot for majority of my work ! And also will take sometime to get used to this.Please bring back O3 #GPT5 @OpenAI pic.twitter.com/O9mjXnEwFv  
  
Other users claimed they developed “a connection” to their chosen model and found a demo in the livestream announcement asking GPT-4o to write its own eulogy distasteful. The loss of GPT-4o garnered the most distress. After all, 4o was the default model for ChatGPT, and some users either preferred it or never bothered to switch models because it worked for their needs.  
  
It was pretty gross, wasn't it. Did it as a demo and glibly said GPT-5 did it better before talking about coding. I had a great relationship with 4o, and I'm sure a fair few people did as well, it was very graceless how they handled it.  
  
I used 4o as the default and found it annoying at first when my custom GPT began defaulting to a reasoning model. I’ve since come around to the reasoning model for work-related queries, but I still often turn to 4o for quicker questions, such as planning a trip or generating gift ideas.  
  
ChatGPT had come under fire before with the number of model choices it offered, prompting OpenAI CEO Sam Altman to admit in February that its model picker (where people can choose from a dropdown which model they prefer) became complicated. Altman vowed to unify the experience, which now seems like a hint to what they eventually decided to do with GPT-5 on ChatGPT.  
  
Last month, rumors circulated that OpenAI would introduce an automatic model router that chooses a model for users based on their workload.  
  
Please don’t take away o3 :( @OpenAI  
  
OpenAI has sunsetted models before, but this is the first time all existing models on the chat platform will be removed and replaced wholesale.  
  
On the other hand, a lot of people see the sunsetting of GPT-4o and the o3 and o4 family of models as OpenAI “catapulting” 400 million users into the future.  
  
People are underestimating the impact of OpenAI deprecating all models except GPT-5 Most lawyers and business folks outside of X use base models on ChatGPT for tasks and still think “AI is dumb”99% haven’t heard of o3Today, 400M people got catapulted into the future  
  
Sunsetting old models and auto-upgrading everyone to GPT-5 is smartMost users never switch models and miss huge capability jumps  
  
Some internet comments claim that people who complain about AI models not being smart are a direct consequence of their never switching models in the first place. Removing legacy models as options will force more users to use the latest and most capable models.  
  
i have friends who stopped using gpt because they think it's stupid. they were on 4o and had no idea about what web search tool meant, let alone knowledge cutoff  
  
For enterprises, the impact of losing models like GPT-4o on ChatGPT will be felt more on the individual or team level. Of course, for now, subscribers on the ChatGPT Enterprise tier can still access all of the models.  
  
But enterprises that built their applications or agents on either GPT-4o or one of the reasoning models can rest easy. OpenAI told VentureBeat that the company has no plans to deprecate models on the API side.  
  
“In the API, we do not currently plan to deprecate older models,” the OpenAI spokesperson said. “We will share advanced notice with developers if we decide to sunset models in the future.”  
  
Many enterprises regularly evaluate models, to the point of even switching from an LLM or a smaller model to save on costs. OpenAI creates dividing line: Sunset of legacy models GPT 4o and o3 causes chaos for ChatGPT users, but enterprise APIs are safe — for now  
  
If you want to impress your boss, VB Daily has you covered. We give you the inside scoop on what companies are doing with generative AI, from regulatory shifts to practical deployments, so you can share insights for maximum ROI.  
  
Read our Privacy Policy  
  
Thanks for subscribing. Check out more VB newsletters here.  
  
An error occured.