Search for articles...

Model Release Notes

Updated over a week ago

Table of contents

Releasing GPT-4.1 in ChatGPT for all paid users (May 14, 2025)

Since its launch in the API in April, GPT-4.1 has become a favorite among developers—by popular demand, we're making it available directly in ChatGPT.

GPT-4.1 is a specialized model that excels at coding tasks. Compared to GPT-4o, it's even stronger at precise instruction following and web development tasks, and offers an alternative to OpenAI o3 and OpenAI o4-mini for simpler, everyday coding needs.

Starting today, Plus, Pro, and Team users can access GPT-4.1 via the "more models" dropdown in the model picker. Enterprise and Edu users will get access in the coming weeks. GPT-4.1 has the same rate limits as GPT-4o for paid users.

Introducing GPT-4.1 mini, replacing GPT-4o mini, in ChatGPT for all users (May 14, 2025)

GPT-4.1 mini is a fast, capable, and efficient small model, delivering significant improvements compared to GPT-4o mini—in instruction-following, coding, and overall intelligence. Starting today, GPT-4.1 mini replaces GPT-4o mini in the model picker under "more models" for paid users, and will serve as the fallback model for free users once they reach their GPT-4o usage limits. Rate limits remain the same.

Evals for GPT-4.1 and GPT-4.1 mini were originally shared in the blog post accompanying their API release. They also went through standard safety evaluations. Detailed results are available in the newly launched Safety Evaluations Hub.

Improvement to GPT-4o (May 12, 2025)

We've improved GPT-4o's system instructions to help ensure the image generation tool is called when you want to generate an image in ChatGPT.

Update to GPT-4o (April 29, 2025)

We've reverted the most recent update to GPT-4o due to issues with overly agreeable responses (sycophancy).

We're actively working on further improvements. For more details, check out our blog post explaining what happened and our initial findings, and this blog post where we expand on what we missed with sycophancy and the changes we're going to make going forward.

Improvements to GPT-4o (April 25, 2025)

We're making additional improvements to GPT-4o, optimizing when it saves memories and enhancing problem-solving capabilities for STEM. We've also made subtle changes to the way it responds, making it more proactive and better at guiding conversations toward productive outcomes. We think these updates help GPT-4o feel more intuitive and effective across a variety of tasks–we hope you agree!

OpenAI o3 and o4-mini (April 16, 2025)

OpenAI o3 is our most powerful reasoning model that pushes the frontier across **coding, math, science, visual perception**, and more. It sets a new SOTA on benchmarks including Codeforces, SWE-bench (without building a custom model-specific scaffold), and MMMU. It's ideal for complex queries requiring multi-faceted analysis and whose answers may not be immediately obvious. It performs especially strongly at visual tasks like analyzing images, charts, and graphics. In evaluations by external experts, o3 makes 20 percent fewer major errors than OpenAI o1 on difficult, real-world tasks—especially excelling in areas like programming, business/consulting, and creative ideation. Early testers highlighted its analytical rigor as a thought partner and emphasized its ability to generate and critically evaluate novel hypotheses—particularly within biology, math, and engineering contexts.

OpenAI o4-mini is a smaller model optimized for fast, cost-efficient reasoning—it achieves remarkable performance for its size and cost, particularly in **math, coding, and visual tasks**. It is the best-performing benchmarked model on AIME 2024 and 2025. In expert evaluations, it also outperforms its predecessor, o3-mini, on non-STEM tasks as well as domains like data science. Thanks to its efficiency, o4-mini supports significantly higher usage limits than o3, making it a strong high-volume, high-throughput option for questions that benefit from reasoning.

Improvements to GPT-4o (March 27, 2025)

We've made improvements to GPT-4o—it now feels more intuitive, creative, and collaborative, with enhanced instruction-following, smarter coding capabilities, and a clearer communication style.

Smarter problem-solving in STEM and coding:

GPT-4o has further improved its capability to tackle complex technical and coding problems. It now generates cleaner, simpler frontend code, more accurately thinks through existing code to identify necessary changes, and consistently produces coding outputs that successfully compile and run, streamlining your coding workflows.

Enhanced instruction-following and formatting accuracy:

GPT-4o is now more adept at following detailed instructions, especially for prompts containing multiple or complex requests. It improves on generating outputs according to the format requested and achieves higher accuracy in classification tasks.

"Fuzzy" improvements:

Early testers say that the model seems to better understand the implied intent behind their prompts, especially when it comes to creative and collaborative tasks. It's also slightly more concise and clear, using fewer markdown hierarchies and emojis for responses that are easier to read, less cluttered, and more focused. We're curious to see if our users also find this to be the case.

This model is now available in ChatGPT and in the API as the newest snapshot of chatgpt-4o-latest. We plan to bring these improvements to a dated model in the API in the coming weeks.

Introducing GPT-4.5 (February, 27, 2025)

We're releasing a research preview of GPT-4.5—our largest, and best model for chat, yet. GPT-4.5 is a step forward in scaling up pretraining and post-training. By scaling unsupervised learning, GPT-4.5 improves its ability to recognize patterns, draw connections, and generate creative insights without reasoning.

Early testing shows that interacting with GPT-4.5 feels more natural. Its broader knowledge base, improved ability to follow user intent, and greater "EQ" make it useful for tasks like improving writing, programming, and solving practical problems. We also expect it to hallucinate less.

We're sharing GPT-4.5 as a research preview to better understand its strengths and limitations. We're still exploring what it's capable of and are eager to see how people use it in ways we might not have expected.

> GPT-4.5 is available worldwide for users on the Pro plan in ChatGPT. Eventually this will be available to all paid plans (Plus, Pro, Teams, Enterprise, and Edu) with a ChatGPT account.

Introducing OpenAI o3-mini (January 31, 2025)

We're excited to release o3-mini, our newest cost-efficient reasoning model optimized for coding, math, and science.

On the API, o3-mini supports Structured Outputs, function calling, developer messages, and streaming. It offers three adjustable reasoning efforts (low, medium, and high), so you can balance speed with depth for your use case.

ChatGPT Team, Pro, Plus, and Free plan users can access o3-mini starting today. Additionally, o3-mini now works with search to find up-to-date answers with links to relevant web sources. This is an early prototype as we work to integrate search across our reasoning models.In side-by-side testing, o3-mini delivered results on par with o1 at a lower latency, and outperformed o1-mini on advanced STEM tasks.
Expert evaluators preferred o3-mini's answers 56% of the time over o1-mini's, citing improved clarity and fewer critical errors on difficult questions. We look forward to your feedback and will keep refining o3-mini as we expand our family of advanced reasoning models.

Updates to GPT-4o in ChatGPT (January 29, 2025)