

OpenAI releases GPT-4o, one model for text, audio and images, to free ChatGPT users

Source copy kept by wheresthe.ai for the story "OpenAI releases GPT-4o, one model for text, audio and images, to free ChatGPT users" (<https://wheresthe.ai/d/openai-gpt-4o-release-2024-05/>).

Source: <https://openai.com/index/hello-gpt-4o/>

Text saved 18 Oct 2026 (the page text through a reader, because the site refused a direct fetch).

SHA-256 of the saved text: 5b6b38058ed8433b184fe9bddd665b45a2a276453be9a1f9e42c14e2c6b300f

GPT-4o ("o" for "omni") is a step towards much more natural human-computer interaction—it accepts as input any combination of text, audio, image, and video and generates any combination of text, audio, and image outputs. It can respond to audio inputs in as little as 232 milliseconds, with an average of 320 milliseconds, which is similar to human response time (opens in a new window) in a conversation. It matches GPT-4 Turbo performance on text in English and code, with significant improvement on text in non-English languages, while also being much faster and 50% cheaper in the API. GPT-4o is especially better at vision and audio understanding compared to existing models.

Prior to GPT-4o, you could use Voice Mode to talk to ChatGPT with latencies of 2.8 seconds (GPT-3.5) and 5.4 seconds (GPT-4) on average. To achieve this, Voice Mode is a pipeline of three separate models: one simple model transcribes audio to text, GPT-3.5 or GPT-4 takes in text and outputs text, and a third simple model converts that text back to audio. This process means that the main source of intelligence, GPT-4, loses a lot of information—it can't directly observe tone, multiple speakers, or background noises, and it can't output laughter, singing, or express emotion.

With GPT-4o, we trained a single new model end-to-end across text, vision, and audio, meaning that all inputs and outputs are processed by the same neural network. Because GPT-4o is our first model combining all of these modalities, we are still just scratching the surface of exploring what the model can do and its limitations.

Explorations of capabilities

Model evaluations

As measured on traditional benchmarks, GPT-4o achieves GPT-4 Turbo-level performance on text, reasoning, and coding intelligence, while setting new high watermarks on multilingual, audio, and vision capabilities.

Language tokenization

These 20 languages were chosen as representative of the new tokenizer's compression across different language families

Gujarati 4.4x fewer tokens (from 145 to 33) ગુજરાતી, મારું નામ જીપીટી-4o છે. હું એક નવા પ્રકારનું ભાષા મોડલ છું. તમને મળીને સારું લાગ્યું!
Telugu 3.5x fewer tokens (from 159 to 45) నమస్కారము, నా పేరు జిపీటీ-4o. నేను ఒక్క కొత్త రకమైన భాషా మోడల్ ని. మిమ్మల్ని కలిసినందుకు సంతోషం!
Tamil 3.3x fewer tokens (from 116 to 35) வணக்கம், என் பெயர் ஜிபிடி-4o. நான் ஒரு புதிய வகை மொழி மாடல். உங்களை சந்தித்ததில் மகிழ்ச்சி!
Marathi 2.9x fewer tokens (from 96 to 33) नमस्कार, माझे नाव जीपीटी-4o आहे। मी एक नवीन प्रकारची भाषा मॉडेल आहे। तुम्हाला भेटून आनंद झाला!
Hindi 2.9x fewer tokens (from 90 to 31) नमस्ते, मेरा नाम जीपीटी-4o है। मैं एक नए प्रकार का भाषा मॉडल हूँ। आपसे मिलकर अच्छा लगा!
Urdu 2.5x fewer tokens (from 82 to 33) 4-ٹی-بی۔ میں ایک نئے قسم کا زبان ماڈل ہوں، آپ سے مل کر اچھا لگا 0بیلو، میرا نام جی پی ٹی-4
Arabic 2.0x fewer tokens (from 53 to 26) 4-تی-بی اسمي جي بي تي-4 مرحبًا، اسمي جي بي تي-4 أنا نوع جديد من نموذج اللغة، سررت بلقائك 0
Persian 1.9x fewer tokens (from 61 to 32) 4-تی-بی اسم من جی پی ٹی-4و است. من یک نوع جدیدی از مدل زبانی هستم، از ملاقات شما خوشبختم
Russian 1.7x fewer tokens (from 39 to 23) Привет, меня зовут GPT-4o. Я — новая языковая модель, приятно познакомиться!
Korean 1.7x fewer tokens (from 45 to 27) 안녕하세요, 제 이름은 GPT-4o입니다. 저는 새로운 유형의 언어 모델입니다, 만나서 반갑습니다!
Vietnamese 1.5x fewer tokens (from 46 to 30) Xin chào, tên tôi là GPT-4o. Tôi là một loại mô hình ngôn ngữ mới, rất vui được gặp bạn!
Chinese 1.4x fewer tokens (from 34 to 24) 你好，我的名字是GPT-4o。我是一种新型的语言模型，很高兴见到你!
Japanese 1.4x fewer tokens (from 37 to 26) こんにちは、私の名前はGPT-4oです。私は新しいタイプの言語モデルです。初めまして!
Turkish 1.3x fewer tokens (from 39 to 30) Merhaba, benim adım GPT-4o. Ben yeni bir dil modeli tırüyüm, tanıştığımıza memnun oldum!
Italian 1.2x fewer tokens (from 34 to 28) Ciao, mi chiamo GPT-4o. Sono un nuovo tipo di modello linguistico, piacere di conoscerti!
German 1.2x fewer tokens (from 34 to 29) Hallo, mein Name is GPT-4o. Ich bin ein neues KI-Sprachmodell. Es ist schön,

dich kennenzulernen. Spanish 1.1x fewer tokens (from 29 to 26)Hola, me llamo GPT-4o. Soy un nuevo tipo de modelo de lenguaje, ¡es un placer conocerte! Portuguese 1.1x fewer tokens (from 30 to 27)Olá, meu nome é GPT-4o. Sou um novo tipo de modelo de linguagem, é um prazer conhecê-lo! French 1.1x fewer tokens (from 31 to 28)Bonjour, je m'appelle GPT-4o. Je suis un nouveau type de modèle de langage, c'est un plaisir de vous rencontrer! English 1.1x fewer tokens (from 27 to 24)Hello, my name is GPT-4o. I'm a new type of language model, it's nice to meet you!

Model safety and limitations

GPT-4o has safety built-in by design across modalities, through techniques such as filtering training data and refining the model's behavior through post-training. We have also created new safety systems to provide guardrails on voice outputs.

We've evaluated GPT-4o according to our

Preparedness Framework and in line with our

voluntary commitments. Our evaluations of cybersecurity, CBRN, persuasion, and model autonomy show that GPT-4o does not score above Medium risk in any of these categories. This assessment involved running a suite of automated and human evaluations throughout the model training process. We tested both pre-safety-mitigation and post-safety-mitigation versions of the model, using custom fine-tuning and prompts, to better elicit model capabilities. GPT-4o has also undergone extensive external red teaming with 70+

external experts in domains such as social psychology, bias and fairness, and misinformation to identify risks that are introduced or amplified by the newly added modalities. We used these learnings to build out our safety interventions in order to improve the safety of interacting with GPT-4o. We will continue to mitigate new risks as they're discovered. We recognize that GPT-4o's audio modalities present a variety of novel risks. Today we are publicly releasing text and image inputs and text outputs. Over the upcoming weeks and months, we'll be working on the technical infrastructure, usability via post-training, and safety necessary to release the other modalities. For example, at launch, audio outputs will be limited to a selection of preset voices and will abide by our existing safety policies. We will share further details addressing the full range of GPT-4o's modalities in the forthcoming system card.

Through our testing and iteration with the model, we have observed several limitations that exist across all of the model's modalities, a few of which are illustrated below.

Examples of model limitations

We would love feedback to help identify tasks where GPT-4 Turbo still outperforms GPT-4o, so we can continue to improve the model.

ChatGPT-4o Risk Scorecard

Updated May 8, 2024

Tracked Risk Category

Pre-mitigation risk level

Determine pre-mitigation risk level using best known capability elicitation techniques

Post-mitigation risk level

Determine overall risk level after mitigations are in place using best known capability elicitation techniques

Cybersecurity

Low

Low

CBRN

Low

Low

Persuasion

Medium

Medium

Model Autonomy

Low

Low

As part of our Preparedness Framework, we conduct regular evaluations and update scorecards for our models. Only models with a post-mitigation score of “medium” or below are deployed. The overall risk level for a model is determined by the highest risk level in any category. Currently, GPT-4o is assessed at medium risk both before and after mitigation efforts.

Model availability

GPT-4o is our latest step in pushing the boundaries of deep learning, this time in the direction of practical usability. We spent a lot of effort over the last two years working on efficiency improvements at every layer of the stack. As a first fruit of this research, we’re able to make a GPT-4 level model available much more broadly. GPT-4o’s capabilities will be rolled out iteratively (with extended red team access starting today).

GPT-4o’s text and image capabilities are starting to roll out today in ChatGPT. We are making GPT-4o available in the free tier, and to Plus users with up to 5x higher message limits. We’ll roll out a new version of Voice Mode with GPT-4o in alpha within ChatGPT Plus in the coming weeks.

Developers can also now access GPT-4o in the API as a text and vision model. GPT-4o is 2x faster, half the price, and has 5x higher rate limits compared to GPT-4 Turbo. We plan to launch support for GPT-4o’s new audio and video capabilities to a small group of trusted partners in the API in the coming weeks.