

Title: OpenAI o3

URL: https://en.wikipedia.org/wiki/OpenAI_o3

PageID: 78665325

Categories: Category:2024 software, Category:2025 in artificial intelligence, Category:ChatGPT, Category:Generative pre-trained transformers, Category:Large language models, Category:OpenAI

Source: Wikipedia (CC BY-SA 4.0).

OpenAI o3 is a reflective generative pre-trained transformer (GPT) model developed by OpenAI as a successor to OpenAI o1 for ChatGPT . It is designed to devote additional deliberation time when addressing questions that require step-by-step logical reasoning. [1] [2] On January 31, 2025, OpenAI released a smaller model, o3-mini, [3] followed on April 16 by o3 and o4-mini . [4]

History

The OpenAI o3 model was announced on December 20, 2024. It was called "o3" rather than "o2" to avoid trademark conflict with the mobile carrier brand named O2 . [1] OpenAI invited safety and security researchers to apply for early access of these models until January 10, 2025. [5] Similarly to o1, there are two different models: o3 and o3-mini. [3]

On January 31, 2025, OpenAI released o3-mini to all ChatGPT users (including free-tier) and some API users. OpenAI describes o3-mini as a "specialized alternative" to o1 for "technical domains requiring precision and speed". [6] o3-mini features three reasoning effort levels: low, medium and high. The free version uses medium. The variant using more compute is called o3-mini-high, and is available to paid subscribers. [3] [7] Subscribers to ChatGPT's Pro tier have unlimited access to both o3-mini and o3-mini-high. [6]

On February 2, OpenAI launched OpenAI Deep Research , a ChatGPT service using a version of o3 that makes comprehensive reports within 5 to 30 minutes, based on web searches . [8]

On February 6, in response to pressure from rivals like DeepSeek , OpenAI announced an update aimed at enhancing the transparency of the thought process in its o3-mini model. [9]

On February 12, OpenAI further increased rate limits for o3-mini-high to 50 requests per day (from 50 requests per week) for ChatGPT Plus subscribers, and implemented file/image upload support. [10]

On April 16, 2025, OpenAI released o3 and o4-mini , a successor of o3-mini. [4]

On June 10, OpenAI released o3-pro, which the company claims is its most capable model yet. [11] OpenAI stated: "We recommend using it for challenging questions where reliability matters more than speed, and waiting a few minutes is worth the tradeoff". [12]

Capabilities

Reinforcement learning was used to teach o3 to "think" before generating answers, using what OpenAI refers to as a "private chain of thought " . [13] This approach enables the model to plan ahead and reason through tasks, performing a series of intermediate reasoning steps to assist in solving the problem, at the cost of additional computing power and increased latency of responses. [14]

o3 demonstrates significantly better performance than o1 on complex tasks, including coding , mathematics , and science . [1] OpenAI reported that o3 achieved a score of 87.7% on the GPQA Diamond benchmark, which contains expert-level science questions not publicly available online. [15]

On SWE-bench Verified, a software engineering benchmark assessing the ability to solve real GitHub issues, o3 scored 71.7%, compared to 48.9% for o1. On Codeforces , o3 reached an Elo score of 2727, whereas o1 scored 1891. [15]

On the Abstraction and Reasoning Corpus for Artificial General Intelligence (ARC-AGI) benchmark, which evaluates an AI's ability to handle new logical and skill acquisition problems, o3 attained three times the accuracy of o1. [1] [16]

See also

List of large language models

References

External links

Introducing OpenAI o3 and o4-mini

O3 is 80% cheaper and introducing o3-pro

v

t

e

ChatGPT in education GPT Store DALL-E ChatGPT Search Sora Whisper

in education

GPT Store

DALL-E

ChatGPT Search

Sora

Whisper

GitHub Copilot

OpenAI Codex

Generative pre-trained transformer GPT-1 GPT-2 GPT-3 GPT-4 GPT-4o o1 o3 GPT-4.5 GPT-4.1
o4-mini GPT-OSS GPT-5

GPT-1

GPT-2

GPT-3

GPT-4

GPT-4o

o1

o3

GPT-4.5

GPT-4.1

o4-mini

GPT-OSS

GPT-5

ChatGPT Deep Research

Operator

Sam Altman removal

removal

Greg Brockman
Sarah Friar
Jakub Pachocki
Scott Schools
Mira Murati
Emmett Shear
Sam Altman
Adam D'Angelo
Sue Desmond-Hellmann
Zico Kolter
Paul Nakasone
Adebayo Ogunesi
Nicole Seligman
Fidji Simo
Lawrence Summers
Bret Taylor (chair)
Greg Brockman (2017–2023)
Reid Hoffman (2019–2023)
Will Hurd (2021–2023)
Holden Karnofsky (2017–2021)
Elon Musk (2015–2018)
Ilya Sutskever (2017–2023)
Helen Toner (2021–2023)
Shivon Zilis (2019–2023)
Stargate LLC
Apple Intelligence
AI Dungeon
AutoGPT
Contrastive Language-Image Pre-training
" Deep Learning "
LangChain
Microsoft Copilot
OpenAI Five
Transformer
Category
v
t
e

Autoencoder
Deep learning
Fine-tuning
Foundation model
Generative adversarial network
Generative pre-trained transformer
Large language model
Model Context Protocol
Neural network
Prompt engineering
Reinforcement learning from human feedback
Retrieval-augmented generation
Self-supervised learning
Stochastic parrot
Synthetic data
Top-p sampling
Transformer
Variational autoencoder
Vibe coding
Vision transformer
Waluigi effect
Word embedding
Character.ai
ChatGPT
DeepSeek
Ernie
Gemini
Grok
Copilot
Claude
Gemini
Gemma

GPT 1 2 3 J 4 4o 4.5 4.1 OSS 5

1

2

3

J

4

4o
4.5
4.1
OSS
5
Llama
o1
o3
o4-mini
Qwen
Base44
Claude Code
Cursor
Devstral
GitHub Copilot
Kimi-Dev
Qwen3-Coder
Replit
Xcode
Aurora
Firefly
Flux
GPT Image 1
Ideogram
Imagen
Midjourney
Qwen-Image
Recraft
Seedream
Stable Diffusion
Dream Machine
Hailuo AI
Kling
Midjourney Video
Runway Gen
Seedance
Sora
Veo

Wan
15.ai
Eleven
MiniMax Speech 2.5
WaveNet
Eleven Music
Endel
Lyria
Riffusion
Suno AI
Udio
Agentforce
AutoGLM
AutoGPT
ChatGPT Agent
Devin AI
Manus
OpenAI Codex
Operator
Replit Agent
01.AI
Aleph Alpha
Anthropic
Baichuan
Canva
Cognition AI
Cohere
Contextual AI
DeepSeek
ElevenLabs
Google DeepMind
HeyGen
Hugging Face
Inflection AI
Krikey AI
Kuaishou
Luma Labs
Meta AI

MiniMax
Mistral AI
Moonshot AI
OpenAI
Perplexity AI
Runway
Safe Superintelligence
Salesforce
Scale AI
SoundHound
Stability AI
Synthesia
Thinking Machines Lab
Upstage
xAI
Z.ai
Category
v
t
e
History timeline
timeline
Companies
Projects
Parameter Hyperparameter
Hyperparameter
Loss functions
Regression Bias–variance tradeoff Double descent Overfitting
Bias–variance tradeoff
Double descent
Overfitting
Clustering
Gradient descent SGD Quasi-Newton method Conjugate gradient method
SGD
Quasi-Newton method
Conjugate gradient method
Backpropagation
Attention

Convolution
Normalization Batchnorm
Batchnorm
Activation Softmax Sigmoid Rectifier
Softmax
Sigmoid
Rectifier
Gating
Weight initialization
Regularization
Datasets Augmentation
Augmentation
Prompt engineering
Reinforcement learning Q-learning SARSA Imitation Policy gradient
Q-learning
SARSA
Imitation
Policy gradient
Diffusion
Latent diffusion model
Autoregression
Adversary
RAG
Uncanny valley
RLHF
Self-supervised learning
Reflection
Recursive self-improvement
Hallucination
Word embedding
Vibe coding
Machine learning In-context learning
In-context learning
Artificial neural network Deep learning
Deep learning
Language model Large language model NMT
Large language model
NMT

Reasoning language model

Model Context Protocol

Intelligent agent

Artificial human companion

Humanity's Last Exam

Artificial general intelligence (AGI)

AlexNet

WaveNet

Human image synthesis

HWR

OCR

Computer vision

Speech synthesis 15.ai ElevenLabs

15.ai

ElevenLabs

Speech recognition Whisper

Whisper

Facial recognition

AlphaFold

Text-to-image models Aurora DALL-E Firefly Flux Ideogram Imagen Midjourney Recraft Stable Diffusion

Aurora

DALL-E

Firefly

Flux

Ideogram

Imagen

Midjourney

Recraft

Stable Diffusion

Text-to-video models Dream Machine Runway Gen Hailuo AI Kling Sora Veo

Dream Machine

Runway Gen

Hailuo AI

Kling

Sora

Veo

Music generation Riffusion Suno AI Udio

Riffusion
Suno AI
Udio
Word2vec
Seq2seq
GloVe
BERT
T5
Llama
Chinchilla AI
PaLM
GPT 1 2 3 J ChatGPT 4 4o o1 o3 4.5 4.1 o4-mini 5
1
2
3
J
ChatGPT
4
4o
o1
o3
4.5
4.1
o4-mini
5
Claude
Gemini Gemini (language model) Gemma
Gemini (language model)
Gemma
Grok
LaMDA
BLOOM
DBRX
Project Debater
IBM Watson
IBM Watsonx
Granite
PanGu- Σ

DeepSeek
Qwen
AlphaGo
AlphaZero
OpenAI Five
Self-driving car
MuZero
Action selection AutoGPT
AutoGPT
Robot control
Alan Turing
Warren Sturgis McCulloch
Walter Pitts
John von Neumann
Claude Shannon
Shun'ichi Amari
Kunihiko Fukushima
Takeo Kanade
Marvin Minsky
John McCarthy
Nathaniel Rochester
Allen Newell
Cliff Shaw
Herbert A. Simon
Oliver Selfridge
Frank Rosenblatt
Bernard Widrow
Joseph Weizenbaum
Seymour Papert
Seppo Linnainmaa
Paul Werbos
Geoffrey Hinton
John Hopfield
Jürgen Schmidhuber
Yann LeCun
Yoshua Bengio
Lotfi A. Zadeh
Stephen Grossberg

Alex Graves
James Goodnight
Andrew Ng
Fei-Fei Li
Alex Krizhevsky
Ilya Sutskever
Oriol Vinyals
Quoc V. Le
Ian Goodfellow
Demis Hassabis
David Silver
Andrej Karpathy
Ashish Vaswani
Noam Shazeer
Aidan Gomez
John Schulman
Mustafa Suleyman
Jan Leike
Daniel Kokotajlo
François Chollet
Neural Turing machine
Differentiable neural computer
Transformer Vision transformer (ViT)
Vision transformer (ViT)
Recurrent neural network (RNN)
Long short-term memory (LSTM)
Gated recurrent unit (GRU)
Echo state network
Multilayer perceptron (MLP)
Convolutional neural network (CNN)
Residual neural network (RNN)
Highway network
Mamba
Autoencoder
Variational autoencoder (VAE)
Generative adversarial network (GAN)
Graph neural network (GNN)
Category