Title: OpenAI o1

URL: https://en.wikipedia.org/wiki/OpenAI_o1

PageID: 77854996

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

OpenAI o1 is a generative pre-trained transformer (GPT), the first in OpenAI 's "o" series of reasoning models . A preview of o1 was released by OpenAI on September 12, 2024. o1 spends time "thinking" before it answers, making it better at complex reasoning tasks, science and programming than GPT-4o . [1] The full version was released to ChatGPT users on December 5, 2024. [2]

History

Background

According to leaked information, o1 was formerly known within OpenAI as "Q*", and later as "Strawberry". [3] The codename "Q*" first surfaced in November 2023, around the time of Sam Altman 's ousting and subsequent reinstatement , with rumors suggesting that this experimental model had shown promising results on mathematical benchmarks. [4] In July 2024, Reuters reported that OpenAI was developing a generative pre-trained transformer known as "Strawberry", [3] which later became o1.

Release

"o1-preview" and "o1-mini" were released on September 12, 2024, for ChatGPT Plus and Team users. [1] GitHub started testing the integration of o1-preview in its Copilot service the same day. [5] On December 5, 2024, the full version of o1 was released. [6] On the same day, a subscription called ChatGPT Pro was released, featuring access to a pro version of o1 that uses more compute to provide better answers. [6] In January 2025, o1 was integrated into Microsoft Copilot . [7]

o1-preview's API is several times more expensive than GPT-4o . [8] As of January 2025, API usage for the full o1 model is limited to developers on usage tier 5. [9]

OpenAI noted that o1 is the first of a series of "reasoning" models. OpenAI shared in December 2024 benchmark results for its successor, o3 (the name o2 was skipped to avoid trademark conflict with the mobile carrier brand named O2). [10]

In March 2025, OpenAI released the o1-pro API, its most expensive AI model to date. The pricing is set at $150 per 1 million input tokens and $600 per 1 million output tokens. [11]

Capabilities

According to OpenAI, o1 has been trained using a new optimization algorithm and a dataset specifically tailored to it; while also meshing in reinforcement learning into its training. [8] OpenAI described o1 as a complement to GPT-4o rather than a successor. [12] [13]

o1 spends additional time thinking (generating a chain of thought) before generating an answer, which makes it better for complex reasoning tasks, particularly in science and mathematics . [1] Compared to previous models, o1 has been trained to generate long " chains of thought " before returning a final answer. [14] [15] According to Mira Murati , this ability to think before responding represents a new, additional paradigm, which is improving model outputs by spending more computing power when generating the answer, whereas the model scaling paradigm improves outputs by increasing the model size, training data and training compute power. [12] OpenAI's test results suggest a correlation between accuracy and the logarithm of the amount of compute spent thinking before answering. [15] [14]

o1-preview performed approximately at a PhD level on benchmark tests related to physics, chemistry, and biology. On the American Invitational Mathematics Examination , it solved 83% (12.5/15) of the problems, compared to 13% (1.8/15) for GPT-4o. It also ranked in the 89th percentile in Codeforces coding competitions. [16] o1-mini is faster and 80% cheaper than o1-preview. It is particularly suitable for programming and STEM -related tasks, but does not have the same "broad world knowledge" as o1-preview. [17]

OpenAI noted that o1's reasoning capabilities make it better at adhering to safety rules provided in the prompt's context window. OpenAI reported that during a test, one instance of o1-preview exploited a misconfiguration to succeed at a task that should have been infeasible due to a bug. [18] [19] OpenAI also granted early access to the UK and US AI Safety Institutes for research, evaluation, and testing. According to OpenAI's assessments, o1-preview and o1-mini crossed into "medium risk" in CBRN (biological, chemical, radiological, and nuclear) weapons. Dan Hendrycks wrote that "The model already outperforms PhD scientists most of the time on answering questions related to bioweapons ." He suggested that these concerning capabilities will continue to increase. [20]

Limitations

o1 usually requires more computing time and power than other GPT models by OpenAI, because it generates long chains of thought before making the final response. [14]

According to OpenAI, o1 may "fake alignment ", that is, generate a response that is contrary to accuracy and its own chain of thought, in about 0.38% of cases. [21]

OpenAI forbids users from trying to reveal o1's chain of thought, which is hidden by design and not trained to comply with the company's policies. Prompts are monitored, and users who intentionally or accidentally violate this may lose their access to o1. OpenAI cites AI safety and competitive advantage as reasons for the restriction, which has been described as a loss of transparency by developers who work with large language models (LLMs). [22]

In October 2024, researchers at Apple submitted a preprint reporting that LLMs such as o1 may be replicating reasoning steps from the models' own training data. [23] By changing the numbers and names used in a math problem or simply running the same problem again, LLMs would perform somewhat worse than their best benchmark results. Adding extraneous but logically inconsequential information to the problems caused a much greater drop in performance, from −17.5% for o1-preview and −29.1% for o1-mini, to −65.7% for the worst model tested. [24]

Safety evaluations from Apollo Research found that o1 was more consistently able to deceive than other frontier models in controlled tests (e.g. attempting to copy itself to an external server when threatened with shutdown). When confronted, it relatively rarely admitted deceptive action (in 20% of test cases). [25]

See also

List of large language models

References

External links

Official website

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 Ogunlesi

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