Title: GPT-2

URL: https://en.wikipedia.org/wiki/GPT-2

PageID: 66045029

Categories: Category:2019 in artificial intelligence, Category:2019 software, Category:Generative pre-trained transformers, Category:Large language models, Category:OpenAI, Category:Software using the MIT license

---

Supervised learning

Unsupervised learning

Semi-supervised learning

Self-supervised learning

Reinforcement learning

Meta-learning

Online learning

Batch learning

Curriculum learning

Rule-based learning

Neuro-symbolic AI

Neuromorphic engineering

Quantum machine learning

Classification

Generative modeling

Regression

Clustering

Dimensionality reduction

Density estimation

Anomaly detection

Data cleaning

AutoML

Association rules

Semantic analysis

Structured prediction

Feature engineering

Feature learning

Learning to rank

Grammar induction

Ontology learning

Multimodal learning

Apprenticeship learning

Decision trees

Ensembles Bagging Boosting Random forest

Bagging

Boosting

Random forest

k -NN

Linear regression

Naive Bayes

Artificial neural networks

Logistic regression

Perceptron

Relevance vector machine (RVM)

Support vector machine (SVM)

BIRCH

CURE

Hierarchical

k -means

Fuzzy

Expectation–maximization (EM)

DBSCAN

OPTICS

Mean shift

Factor analysis

CCA

ICA

LDA

NMF

PCA

PGD

t-SNE

SDL

Graphical models Bayes net Conditional random field Hidden Markov

Bayes net

Conditional random field

Hidden Markov

RANSAC

k -NN

Local outlier factor

Isolation forest

Autoencoder

Deep learning

Feedforward neural network

Recurrent neural network LSTM GRU ESN reservoir computing

LSTM

GRU

ESN

reservoir computing

Boltzmann machine Restricted

Restricted

GAN

Diffusion model

SOM

Convolutional neural network U-Net LeNet AlexNet DeepDream

U-Net

LeNet

AlexNet

DeepDream

Neural field Neural radiance field Physics-informed neural networks

Neural radiance field

Physics-informed neural networks

Transformer Vision

Vision

Mamba

Spiking neural network

Memtransistor

Electrochemical RAM (ECRAM)

Q-learning

Policy gradient

SARSA

Temporal difference (TD)

Multi-agent Self-play

Self-play

Active learning

Crowdsourcing

Human-in-the-loop

Mechanistic interpretability

RLHF

Coefficient of determination

Confusion matrix

Learning curve

ROC curve

Kernel machines

Bias–variance tradeoff

Computational learning theory

Empirical risk minimization

Occam learning

PAC learning

Statistical learning

VC theory

Topological deep learning

AAAI

ECML PKDD

NeurIPS

ICML

ICLR

IJCAI

ML

JMLR

Glossary of artificial intelligence

List of datasets for machine-learning research List of datasets in computer vision and image processing

List of datasets in computer vision and image processing

Outline of machine learning

v

t

e

Generative Pre-trained Transformer 2 (GPT-2) is a large language model by OpenAI and the second in their foundational series of GPT models. GPT-2 was pre-trained on a dataset of 8 million web pages. [2] It was partially released in February 2019, followed by full release of the 1.5-billion-parameter model on November 5, 2019. [3] [4] [5]

GPT-2 was created as a "direct scale-up" of GPT-1 [6] with a ten-fold increase in both its parameter count and the size of its training dataset. [5] It is a general-purpose learner and its ability to perform the various tasks was a consequence of its general ability to accurately predict the next item in a sequence, [2] [7] which enabled it to translate texts, answer questions about a topic from a text, summarize passages from a larger text, [7] and generate text output on a level

sometimes indistinguishable from that of humans ; however, it could become repetitive or nonsensical when generating long passages. [8] It was superseded by the GPT-3 and GPT-4 models, which are no longer open source.

GPT-2 has, like its predecessor GPT-1 and its successors GPT-3 and GPT-4, a generative pre-trained transformer architecture, implementing a deep neural network , specifically a transformer model, [6] which uses attention instead of older recurrence- and convolution-based architectures. [9] [10] Attention mechanisms allow the model to selectively focus on segments of input text it predicts to be the most relevant. [11] [12] This model allows for greatly increased parallelization , and outperforms previous benchmarks for RNN/CNN/LSTM-based models. [6]

Training

Since the transformer architecture enabled massive parallelization , GPT models could be trained on larger corpora than previous NLP (natural language processing) models. While the GPT-1 model demonstrated that the approach was viable, GPT-2 would further explore the emergent properties of networks trained on extremely large corpora. CommonCrawl , a large corpus produced by web crawling and previously used in training NLP systems, [13] was considered due to its large size, but was rejected after further review revealed large amounts of unintelligible content. [2] [13] Instead, OpenAI developed a new corpus, known as WebText ; rather than scraping content indiscriminately from the World Wide Web , WebText was generated by scraping only pages linked to by Reddit posts that had received at least 3 karma prior to December 2017. The corpus was subsequently cleaned; HTML documents were parsed into plain text, duplicate pages were eliminated, and Wikipedia pages were removed (since their presence in many other datasets could have induced overfitting). [2]

While the cost of training GPT-2 is known to have been $256 per hour, [14] [15] the amount of hours it took to complete training is unknown; therefore, the overall training cost cannot be estimated accurately. [16] However, comparable large language models using transformer architectures have had their costs documented in more detail; the training processes for BERT and XLNet consumed, respectively, $6,912 and $245,000 of resources. [15]

Release

GPT-2 was first announced on 14 February 2019. A February 2019 article in The Verge by James Vincent said that, while "[the] writing it produces is usually easily identifiable as non-human", it remained "one of the most exciting examples yet" of language generation programs: [17]

Give it a fake headline, and it'll write the rest of the article, complete with fake quotations and statistics. Feed it the first line of a short story, and it'll tell you what happens to your character next. It can even write fan fiction, given the right prompt. [17]

The Guardian described this output as "plausible newspaper prose"; [8] Kelsey Piper of Vox said "one of the coolest AI systems I've ever seen may also be the one that will kick me out of my job". [18] GPT-2's flexibility was described as "impressive" by The Verge ; specifically, its ability to translate text between languages, summarize long articles, and answer trivia questions were noted. [17]

A study by the University of Amsterdam employing a modified Turing test found that at least in some scenarios, participants were unable to distinguish poems generated by GPT-2 from those written by humans. [19]

The GPT-2 series contained 4 models, reported in the paper. They were not released all at once, but in stages.

Restrictions and partial release

While previous OpenAI models had been made immediately available to the public, OpenAI initially refused to make a public release of GPT-2's source code when announcing it in February, citing the risk of malicious use; [8] limited access to the model (i.e. an interface that allowed input and provided output, not the source code itself) was allowed for selected press outlets on announcement. [8] One commonly-cited justification was that, since generated text was usually completely novel, it could be used by spammers to evade automated filters ; OpenAI demonstrated

a version of GPT-2 fine-tuned to "generate infinite positive – or negative – reviews of products". [8]

Another justification was that GPT-2 could be used to generate text that was obscene or racist . Researchers such as Jeremy Howard warned of "the technology to totally fill Twitter, email, and the web up with reasonable-sounding, context-appropriate prose, which would drown out all other speech and be impossible to filter". [17] The Allen Institute for Artificial Intelligence , in response to GPT-2, announced a tool to detect "neural fake news". [20]

However, opinion was divided. A February 2019 article in The Verge argued that the threat posed by GPT-2 had been exaggerated; [21] Anima Anandkumar , a professor at Caltech and director of machine learning research at Nvidia , said that there was no evidence that GPT-2 had the capabilities to pose the threats described by OpenAI, and that what they did was the "opposite of open", characterizing their refusal to release the full model as "malicious BS ". [21] The Gradient published an open letter to OpenAI requesting that they release the model publicly, comparing the threat posed by text-generation AI to the threat posed by the printing press , and giving Photoshop as an example of "a technology that has (thankfully) not destroyed modern society despite its potential for chaos": [22]

Thirty years later, society has emerged relatively unscathed despite Photoshop being simple enough for high school students to use and ubiquitous enough to commandeer its own verb. Why? Precisely because everyone knows about Photoshop. [22]

774M release

While OpenAI did not release the fully-trained model or the corpora it was trained on, description of their methods in prior publications (and the free availability of underlying technology) made it possible for GPT-2 to be replicated by others as free software ; one such replication, OpenGPT-2, was released in August 2019, in conjunction with a freely licensed version of WebText called OpenWebText. The cloud compute costs for OpenGPT-2 were given as approximately $50,000. [23]

On August 20, 2019, OpenAI released a partial version of GPT-2, with 774 million parameters (roughly half the size of the full 1.5 billion parameter model). [24]

Full 1.5B release

Initial concerns that GPT-2 would lend itself to widespread misuse did not come to pass; The Verge said that "there are reasons to be skeptical about claims that AI technology will usher in some sort of ‘infopocalypse.’ For a start, we already have programs that can generate plausible text at high volume for little cost: humans." [25] By November 2019, OpenAI said that they had "seen no strong evidence of misuse so far", and the full version, with 1.5 billion parameters trained with forty gigabytes of data, "about eight thousand times larger than the collected works of Shakespeare", [26] was released on November 5, 2019. [3] [4]

Small and Medium Releases

Two other smaller releases of GPT-2 are available, including the small version of 124M parameters and the medium size of 355M parameters. Both are available to download from Huggingface. [27] [28]

Limitations

While GPT-2's ability to generate plausible passages of natural language text were generally remarked on positively, its shortcomings were noted as well, especially when generating texts longer than a couple paragraphs; Vox said "the prose is pretty rough, there’s the occasional non-sequitur, and the articles get less coherent the longer they get". [18] The Verge similarly noted that longer samples of GPT-2 writing tended to "stray off topic" and lack overall coherence; [17] The Register opined that "a human reading it should, after a short while, realize something's up", and noted that "GPT-2 doesn't answer questions as well as other systems that rely on algorithms to extract and retrieve information." [14]

GPT-2 deployment is resource-intensive; the full version of the model is larger than five gigabytes, making it difficult to embed locally into applications, and consumes large amounts of RAM. In

addition, performing a single prediction "can occupy a CPU at 100% utilization for several minutes", and even with GPU processing, "a single prediction can take seconds". To alleviate these issues, the company Hugging Face created DistilGPT2 , using knowledge distillation to produce a smaller model that "scores a few points lower on some quality benchmarks", but is "33% smaller and twice as fast". [citation needed]

Application and subsequent research

Even before the release of the full version, GPT-2 was used for a variety of applications and services, as well as for entertainment. In June 2019, a subreddit named r/SubSimulatorGPT2 was created in which a variety of GPT-2 instances trained on different subreddits made posts and replied to each other's comments, creating a situation where one could observe "an AI personification of r/Bitcoin argue with the machine learning-derived spirit of r/ShittyFoodPorn"; [25] by July of that year, a GPT-2-based software program released to autocomplete lines of code in a variety of programming languages was described by users as a "game-changer". [29]

In 2019, AI Dungeon was launched, which used GPT-2 to generate dynamic text adventures based on user input. [30] AI Dungeon now offers access to the largest release of GPT-3 API as an optional paid upgrade, the free version of the site uses the 2nd largest release of GPT-3. [31] Latitude, the company formed around AI Dungeon, raised $3.3 million in seed funding in 2021. [32] Several websites host interactive demonstrations of different instances of GPT-2 and other transformer models. [33] [34] [35]

In February 2021, a crisis center for troubled teens announced that they would begin using a GPT-2-derived chatbot to help train counselors by allowing them to have conversations with simulated teens (this use was purely for internal purposes, and did not involve having GPT-2 communicate with the teens themselves). [36]

On May 9, 2023, OpenAI released a mapped version of GPT-2. OpenAI used successor model, GPT-4 , to map each neuron of GPT-2 to determine their functions. [37]

Performance and evaluation

GPT-2 became capable of performing a variety of tasks beyond simple text production due to the breadth of its dataset and technique: answering questions, summarizing, and even translating between languages in a variety of specific domains , without being instructed in anything beyond how to predict the next word in a sequence. [17] [18]

One example of generalized learning is GPT-2's ability to perform machine translation between French and English, for which task GPT-2's performance was assessed using WMT-14 translation tasks. GPT-2's training corpus included virtually no French text; non-English text was deliberately removed while cleaning the dataset prior to training, and as a consequence, only 10MB of French of the remaining 40,000MB was available for the model to learn from (mostly from foreign-language quotations in English posts and articles). [2]

Despite this, GPT-2 achieved 5 BLEU on the WMT-14 English-to-French test set (slightly below the score of a translation via word-for-word substitution). It was also able to outperform several contemporary (2017) unsupervised machine translation baselines on the French-to-English test set, where GPT-2 achieved 11.5 BLEU. This remained below the highest-performing contemporary unsupervised approach (2019), which had achieved 33.5 BLEU. [2] However, other models used large amounts of French text to achieve these results; GPT-2 was estimated to have used a monolingual French corpus approximately 1/500 the size of comparable approaches. [2]

GPT-2 was to be followed by the 175-billion-parameter GPT-3 , [39] revealed to the public in 2020 [40] (whose source code has never been made available). Access to GPT-3 is provided exclusively through APIs offered by OpenAI and Microsoft . [41] That was then later followed by GPT-4 .

See also

List of large language models

References

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