Stable Diffusion is a deep learning , text-to-image model released in 2022 based on diffusion techniques. The generative artificial intelligence technology is the premier product of Stability AI and is considered to be a part of the ongoing artificial intelligence boom .

It is primarily used to generate detailed images conditioned on text descriptions, though it can also be applied to other tasks such as inpainting , outpainting, and generating image-to-image translations guided by a text prompt . [3] Its development involved researchers from the CompVis Group at Ludwig Maximilian University of Munich and Runway with a computational donation from Stability and training data from non-profit organizations. [4] [5] [6] [7]

Stable Diffusion is a latent diffusion model , a kind of deep generative artificial neural network . Its code and model weights have been released publicly , [8] and an optimized version can run on most consumer hardware equipped with a modest GPU with as little as 2.4 GB VRAM . [9] This marked a departure from previous proprietary text-to-image models such as DALL-E and Midjourney which were accessible only via cloud services . [10] [11]

Development

Stable Diffusion originated from a project called Latent Diffusion , [12] developed in Germany by researchers at Ludwig Maximilian University in Munich and Heidelberg University . Four of the original 5 authors (Robin Rombach, Andreas Blattmann, Patrick Esser and Dominik Lorenz) later joined Stability AI and released subsequent versions of Stable Diffusion. [13]

The technical license for the model was released by the CompVis group at Ludwig Maximilian University of Munich. [11] Development was led by Patrick Esser of Runway and Robin Rombach of CompVis, who were among the researchers who had earlier invented the latent diffusion model architecture used by Stable Diffusion. [7] Stability AI also credited EleutherAI and LAION (a German nonprofit which assembled the dataset on which Stable Diffusion was trained) as supporters of the project. [7]

Technology

Architecture

Diffusion models , introduced in 2015, are trained with the objective of removing successive applications of Gaussian noise on training images, which can be thought of as a sequence of denoising autoencoders . The name diffusion is from the thermodynamic diffusion , since they were first developed with inspiration from thermodynamics. [14] [15]

Models in Stable Diffusion series before SD 3 all used a variant of diffusion models, called latent diffusion model (LDM) , developed in 2021 by the CompVis (Computer Vision & Learning) [16] group at LMU Munich . [17] [8]

Stable Diffusion consists of 3 parts: the variational autoencoder (VAE), U-Net , and an optional text encoder. [18] The VAE encoder compresses the image from pixel space to a smaller dimensional latent space , capturing a more fundamental semantic meaning of the image. [17] Gaussian noise is iteratively applied to the compressed latent representation during forward diffusion. [18] The U-Net block, composed of a ResNet backbone, denoises the output from forward diffusion

backwards to obtain a latent representation. Finally, the VAE decoder generates the final image by converting the representation back into pixel space. [18]

The denoising step can be flexibly conditioned on a string of text, an image, or another modality. The encoded conditioning data is exposed to denoising U-Nets via a cross-attention mechanism . [18] For conditioning on text, the fixed, pretrained CLIP ViT-L/14 text encoder is used to transform text prompts to an embedding space. [8] Researchers point to increased computational efficiency for training and generation as an advantage of LDMs. [7] [17]

With 860 million parameters in the U-Net and 123 million in the text encoder, Stable Diffusion is considered relatively lightweight by 2022 standards, and unlike other diffusion models, it can run on consumer GPUs, [19] and even CPU -only if using the OpenVINO version of Stable Diffusion. [20]

SD XL

The XL version uses the same LDM architecture as previous versions, [21] except larger: larger UNet backbone, larger cross-attention context, two text encoders instead of one, and trained on multiple aspect ratios (not just the square aspect ratio like previous versions).

The SD XL Refiner, released at the same time, has the same architecture as SD XL, but it was trained for adding fine details to preexisting images via text-conditional img2img.

SD 3.0

The 3.0 version [22] completely changes the backbone. Not a UNet, but a Rectified Flow Transformer , which implements the rectified flow method [23] [24] with a Transformer .

The Transformer architecture used for SD 3.0 has three "tracks", for original text encoding, transformed text encoding, and image encoding (in latent space). The transformed text encoding and image encoding are mixed during each transformer block.

The architecture is named "multimodal diffusion transformer (MMDiT), where the "multimodal" means that it mixes text and image encodings inside its operations. This differs from previous versions of DiT, where the text encoding affects the image encoding, but not vice versa.

Training data

Stable Diffusion was trained on pairs of images and captions taken from LAION-5B, a publicly available dataset derived from Common Crawl data scraped from the web, where 5 billion image-text pairs were classified based on language and filtered into separate datasets by resolution, a predicted likelihood of containing a watermark, and predicted "aesthetic" score (e.g. subjective visual quality). [25] The dataset was created by LAION , a German non-profit which receives funding from Stability AI. [25] [26] The Stable Diffusion model was trained on three subsets of LAION-5B: laion2B-en, laion-high-resolution, and laion-aesthetics v2 5+. [25] A third-party analysis of the model's training data identified that out of a smaller subset of 12 million images taken from the original wider dataset used, approximately 47% of the sample size of images came from 100 different domains, with Pinterest taking up 8.5% of the subset, followed by websites such as WordPress , Blogspot , Flickr , DeviantArt and Wikimedia Commons . [citation needed] An investigation by Bayerischer Rundfunk showed that LAION's datasets, hosted on Hugging Face, contain large amounts of private and sensitive data. [27]

Training procedures

The model was initially trained on the laion2B-en and laion-high-resolution subsets, with the last few rounds of training done on LAION-Aesthetics v2 5+, a subset of 600 million captioned images which the LAION-Aesthetics Predictor V2 predicted that humans would, on average, give a score of at least 5 out of 10 when asked to rate how much they liked them. [28] [25] [29] The LAION-Aesthetics v2 5+ subset also excluded low-resolution images and images which LAION-5B-WatermarkDetection identified as carrying a watermark with greater than 80% probability. [25] Final rounds of training additionally dropped 10% of text conditioning to improve Classifier-Free Diffusion Guidance. [30]

The model was trained using 256 Nvidia A100 GPUs on Amazon Web Services for a total of 150,000 GPU-hours, at a cost of $600,000. [31] [32] [33]

Limitations

Stable Diffusion has issues with degradation and inaccuracies in certain scenarios. Initial releases of the model were trained on a dataset that consists of 512×512 resolution images, meaning that the quality of generated images noticeably degrades when user specifications deviate from its "expected" 512×512 resolution; [34] the version 2.0 update of the Stable Diffusion model later introduced the ability to natively generate images at 768×768 resolution. [35] Another challenge is in generating human limbs due to poor data quality of limbs in the LAION database. [36] The model is insufficiently trained to replicate human limbs and faces due to the lack of representative features in the database, and prompting the model to generate images of such type can confound the model. [37] In addition to human limbs, Stable Diffusion is unable to generate legible ambigrams and some other forms of text and typography. Stable Diffusion XL (SDXL) version 1.0, released in July 2023, introduced native 1024x1024 resolution and improved generation for limbs and text. [38] [39]

Accessibility for individual developers can also be a problem. In order to customize the model for new use cases that are not included in the dataset, such as generating anime characters ("waifu diffusion"), [40] new data and further training are required. Fine-tuned adaptations of Stable Diffusion created through additional retraining have been used for a variety of different use-cases, from medical imaging [41] to algorithmically generated music . [42] However, this fine-tuning process is sensitive to the quality of new data; low resolution images or different resolutions from the original data can not only fail to learn the new task but degrade the overall performance of the model. Even when the model is additionally trained on high quality images, it is difficult for individuals to run models in consumer electronics. For example, the training process for waifu-diffusion requires a minimum 30 GB of VRAM , [43] which exceeds the usual resource provided in such consumer GPUs as Nvidia 's GeForce 30 series , which has only about 12 GB. [44]

The creators of Stable Diffusion acknowledge the potential for algorithmic bias , as the model was primarily trained on images with English descriptions. [32] As a result, generated images reinforce social biases and are from a western perspective, as the creators note that the model lacks data from other communities and cultures. The model gives more accurate results for prompts that are written in English in comparison to those written in other languages, with western or white cultures often being the default representation. [32]

End-user fine-tuning

To address the limitations of the model's initial training, end-users may opt to implement additional training to fine-tune generation outputs to match more specific use-cases, a process also referred to as personalization . There are three methods in which user-accessible fine-tuning can be applied to a Stable Diffusion model checkpoint:

An "embedding" can be trained from a collection of user-provided images, and allows the model to generate visually similar images whenever the name of the embedding is used within a generation prompt. [45] Embeddings are based on the "textual inversion" concept developed by researchers from Tel Aviv University in 2022 with support from Nvidia , where vector representations for specific tokens used by the model's text encoder are linked to new pseudo-words. Embeddings can be used to reduce biases within the original model, or mimic visual styles. [46]

A "hypernetwork" is a small pretrained neural network that is applied to various points within a larger neural network, and refers to the technique created by NovelAI developer Kurumuz in 2021, originally intended for text-generation transformer models . Hypernetworks steer results towards a particular direction, allowing Stable Diffusion-based models to imitate the art style of specific artists, even if the artist is not recognised by the original model; they process the image by finding key areas of importance such as hair and eyes, and then patch these areas in secondary latent space. [47]

DreamBooth is a deep learning generation model developed by researchers from Google Research and Boston University in 2022 which can fine-tune the model to generate precise, personalised outputs that depict a specific subject, following training via a set of images which depict the subject. [48]

Capabilities

The Stable Diffusion model supports the ability to generate new images from scratch through the use of a text prompt describing elements to be included or omitted from the output. [8] Existing images can be re-drawn by the model to incorporate new elements described by a text prompt (a process known as "guided image synthesis" [49]) through its diffusion-denoising mechanism. [8] In addition, the model also allows the use of prompts to partially alter existing images via inpainting and outpainting, when used with an appropriate user interface that supports such features, of which numerous different open source implementations exist. [50]

Stable Diffusion is recommended to be run with 10 GB or more VRAM, however users with less VRAM may opt to load the weights in float16 precision instead of the default float32 to tradeoff model performance with lower VRAM usage. [34]

Text to image generation

The text to image sampling script within Stable Diffusion, known as "txt2img", consumes a text prompt in addition to assorted option parameters covering sampling types, output image dimensions, and seed values. The script outputs an image file based on the model's interpretation of the prompt. [8] Generated images are tagged with an invisible digital watermark to allow users to identify an image as generated by Stable Diffusion, [8] although this watermark loses its efficacy if the image is resized or rotated. [51]

Each txt2img generation will involve a specific seed value which affects the output image. Users may opt to randomize the seed in order to explore different generated outputs, or use the same seed to obtain the same image output as a previously generated image. [34] Users are also able to adjust the number of inference steps for the sampler; a higher value takes a longer duration of time, however a smaller value may result in visual defects. [34] Another configurable option, the classifier-free guidance scale value, allows the user to adjust how closely the output image adheres to the prompt. [30] More experimental use cases may opt for a lower scale value, while use cases aiming for more specific outputs may use a higher value. [34]

Additional text2img features are provided by front-end implementations of Stable Diffusion, which allow users to modify the weight given to specific parts of the text prompt. Emphasis markers allow users to add or reduce emphasis to keywords by enclosing them with brackets. [52] An alternative method of adjusting weight to parts of the prompt are "negative prompts". Negative prompts are a feature included in some front-end implementations, including Stability AI's own DreamStudio cloud service, and allow the user to specify prompts which the model should avoid during image generation. The specified prompts may be undesirable image features that would otherwise be present within image outputs due to the positive prompts provided by the user, or due to how the model was originally trained, with mangled human hands being a common example. [50] [53]

Image modification

Stable Diffusion also includes another sampling script, "img2img", which consumes a text prompt, path to an existing image, and strength value between 0.0 and 1.0. The script outputs a new image based on the original image that also features elements provided within the text prompt. The strength value denotes the amount of noise added to the output image. A higher strength value produces more variation within the image but may produce an image that is not semantically consistent with the prompt provided. [8]

There are different methods for performing img2img. The main method is SDEdit, [54] which first adds noise to an image, then denoises it as usual in text2img.

The ability of img2img to add noise to the original image makes it potentially useful for data anonymization and data augmentation , in which the visual features of image data are changed and anonymized. [55] The same process may also be useful for image upscaling, in which the

resolution of an image is increased, with more detail potentially being added to the image. [55] Additionally, Stable Diffusion has been experimented with as a tool for image compression. Compared to JPEG and WebP , the recent methods used for image compression in Stable Diffusion face limitations in preserving small text and faces. [56]

Additional use-cases for image modification via img2img are offered by numerous front-end implementations of the Stable Diffusion model. Inpainting involves selectively modifying a portion of an existing image delineated by a user-provided layer mask , which fills the masked space with newly generated content based on the provided prompt. [50] A dedicated model specifically fine-tuned for inpainting use-cases was created by Stability AI alongside the release of Stable Diffusion 2.0. [35] Conversely, outpainting extends an image beyond its original dimensions, filling the previously empty space with content generated based on the provided prompt. [50]

A depth-guided model, named "depth2img", was introduced with the release of Stable Diffusion 2.0 on November 24, 2022; this model infers the depth of the provided input image, and generates a new output image based on both the text prompt and the depth information, which allows the coherence and depth of the original input image to be maintained in the generated output. [35]

ControlNet

ControlNet [57] is a neural network architecture designed to manage diffusion models by incorporating additional conditions. It duplicates the weights of neural network blocks into a "locked" copy and a "trainable" copy. The "trainable" copy learns the desired condition, while the "locked" copy preserves the original model. This approach ensures that training with small datasets of image pairs does not compromise the integrity of production-ready diffusion models. The "zero convolution" is a 1×1 convolution with both weight and bias initialized to zero. Before training, all zero convolutions produce zero output, preventing any distortion caused by ControlNet. No layer is trained from scratch; the process is still fine-tuning, keeping the original model secure. This method enables training on small-scale or even personal devices.

User interfaces

Stability provides an online image generation service called DreamStudio . [58] [59] The company also released an open source version of DreamStudio called StableStudio . [60] [61] In addition to Stability's interfaces, many third party open source interfaces exist, such as AUTOMATIC1111 Stable Diffusion Web UI , which is the most popular and offers extra features, [62] Fooocus , which aims to decrease the amount of prompting needed by the user, [63] and ComfyUI , which has a node-based user interface, essentially a visual programming language akin to many 3D modeling applications. [64] [65] [66]

Releases

Key papers

Learning Transferable Visual Models From Natural Language Supervision (2021). [80] This paper describes the CLIP method for training text encoders, which convert text into floating point vectors. Such text encodings are used by the diffusion model to create images.

SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations (2021). [54] This paper describes SDEdit, aka "img2img".

High-Resolution Image Synthesis with Latent Diffusion Models (2021, updated in 2022). [81] This paper describes the latent diffusion model (LDM). This is the backbone of the Stable Diffusion architecture.

Classifier-Free Diffusion Guidance (2022). [30] This paper describes CFG, which allows the text encoding vector to steer the diffusion model towards creating the image described by the text.

SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis (2023). [21] Describes SDXL.

Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow (2022). [23] [24] Describes rectified flow, which is used for the backbone architecture of SD 3.0.

Scaling Rectified Flow Transformers for High-resolution Image Synthesis (2024). [22] Describes SD 3.0.

Training cost

SD 2.0: 0.2 million hours on A100 (40GB). [71]

Stable Diffusion 3.5 Large was made available for enterprise usage on Amazon Bedrock of Amazon Web Services . [82]

Usage and controversy

Stable Diffusion claims no rights on generated images and freely gives users the rights of usage to any generated images from the model provided that the image content is not illegal or harmful to individuals. [83]

The images Stable Diffusion was trained on have been filtered without human input, leading to some harmful images and large amounts of private and sensitive information appearing in the training data. [27]

More traditional visual artists have expressed concern that widespread usage of image synthesis software such as Stable Diffusion may eventually lead to human artists, along with photographers, models, cinematographers, and actors, gradually losing commercial viability against AI-based competitors. [84]

Stable Diffusion is notably more permissive in the types of content users may generate, such as violent or sexually explicit imagery, in comparison to other commercial products based on generative AI. [85] Addressing the concerns that the model may be used for abusive purposes, CEO of Stability AI, Emad Mostaque , argues that "[it is] peoples' responsibility as to whether they are ethical, moral, and legal in how they operate this technology", [11] and that putting the capabilities of Stable Diffusion into the hands of the public would result in the technology providing a net benefit, in spite of the potential negative consequences. [11] In addition, Mostaque argues that the intention behind the open availability of Stable Diffusion is to end corporate control and dominance over such technologies, who have previously only developed closed AI systems for image synthesis. [11] [85] This is reflected by the fact that any restrictions Stability AI places on the content that users may generate can easily be bypassed due to the availability of the source code. [86]

Controversy around photorealistic sexualized depictions of underage characters have been brought up, due to such images generated by Stable Diffusion being shared on websites such as Pixiv . [87]

In June of 2024, a hack on an extension of ComfyUI , a user interface for Stable Diffusion, took place, with the hackers claiming they targeted users who committed "one of our sins", which included AI-art generation, art theft , promoting cryptocurrency . [88] [clarification needed]

Litigation

Andersen, McKernan, and Ortiz v. Stability AI, Midjourney, and DeviantArt

In January 2023, three artists, Sarah Andersen , Kelly McKernan , and Karla Ortiz, filed a copyright infringement lawsuit against Stability AI, Midjourney , and DeviantArt , claiming that these companies have infringed the rights of millions of artists by training AI tools on five billion images scraped from the web without the consent of the original artists. [89]

In July 2023, U.S. District Judge William Orrick inclined to dismiss most of the lawsuit filed by Andersen, McKernan, and Ortiz but allowed them to file a new complaint, providing them an opportunity to reframe their arguments. [90]

Getty Images v. Stability AI

In January 2023, Getty Images initiated legal proceedings against Stability AI in the English High Court, alleging significant infringement of its intellectual property rights. Getty Images claims that Stability AI "scraped" millions of images from Getty's websites without consent and used these images to train and develop its deep-learning Stable Diffusion model. [91] [92]

Key points of the lawsuit include:

Getty Images asserting that the training and development of Stable Diffusion involved the unauthorized use of its images, which were downloaded on servers and computers that were potentially in the UK. However, Stability AI argues that all training and development took place outside the UK, specifically in U.S. data centers operated by Amazon Web Services. [93]

Stability AI applied for reverse summary judgment and/or strike out of two claims: the training and development claim, and the secondary infringement of copyright claim. The High Court, however, refused to strike out these claims, allowing them to proceed to trial. The court is to determine whether the training and development of Stable Diffusion occurred in the UK, which is crucial for establishing jurisdiction under the UK's Copyright, Designs and Patents Act 1988 (CDPA). [94]

The secondary infringement claim revolves around whether the pre-trained Stable Diffusion software, made available in the UK through platforms like GitHub, HuggingFace, and DreamStudio, constitutes an "article" under sections 22 and 23 of the CDPA. The court will decide whether the term "article" can encompass intangible items such as software. [94]

The trial is expected to take place in summer 2025 and has significant implications for UK copyright law and the licensing of AI-generated content.

License

Unlike models like DALL-E , Stable Diffusion makes its source code available , [95] [8] along with the model (pretrained weights). Prior to Stable Diffusion 3, it applied the Creative ML OpenRAIL-M license, a form of Responsible AI License (RAIL), to the model (M). [96] The license prohibits certain use cases, including crime, libel , harassment , doxing , " exploiting ... minors ", giving medical advice, automatically creating legal obligations, producing legal evidence, and "discriminating against or harming individuals or groups based on ... social behavior or ... personal or personality characteristics ... [or] legally protected characteristics or categories ". [97] [98] The user owns the rights to their generated output images, and is free to use them commercially. [99]

Stable Diffusion 3.5 applies the permissive Stability AI Community License while commercial enterprises with revenue exceed $1 million need the Stability AI Enterprise License. [100] As with the OpenRAIL-M license, the user retains the rights to their generated output images and is free to use them commercially. [79]

See also

Artificial intelligence art

Runway

Midjourney

Craiyon

Hugging Face

Imagen

References

External links

Stable Diffusion Demo

"Step by Step visual introduction to Diffusion Models. - Blog by Kemal Erdem" . Retrieved August 31, 2024 .

"U-Net for Stable Diffusion" . U-Net for Stable Diffusion . Retrieved August 31, 2024 .

Interactive Explanation of Stable Diffusion

"We Are All Raw Material for AI" : Investigation on sensitive and private data in Stable Diffusions training data

" Negative Prompts in Stable Diffusion "

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