Title: Chinchilla (language model)

URL: https://en.wikipedia.org/wiki/Chinchilla_(language_model)

PageID: 72752788

Categories: Category:2022 in artificial intelligence, Category:Chatbots, Category:Generative pre-trained transformers, Category:Google DeepMind, Category:Large language models

-----

Chinchilla is a family of large language models (LLMs) developed by the research team at Google DeepMind , presented in March 2022. [1]

Models

It is named " chinchilla " because it is a further development over a previous model family named Gopher. Both model families were trained in order to investigate the scaling laws of large language models . [2]

It claimed to outperform GPT-3 . It considerably simplifies downstream utilization because it requires much less computer power for inference and fine-tuning. Based on the training of previously employed language models, it has been determined that if one doubles the model size, one must also have twice the number of training tokens. This hypothesis has been used to train Chinchilla by DeepMind. Similar to Gopher in terms of cost, Chinchilla has 70B parameters and four times as much data. [3]

Chinchilla has an average accuracy of 67.5% on the Measuring Massive Multitask Language Understanding (MMLU) benchmark, which is 7% higher than Gopher's performance. Chinchilla was still in the testing phase as of January 12, 2023. [4]

Chinchilla contributes to developing an effective training paradigm for large autoregressive language models with limited compute resources. The Chinchilla team recommends that the number of training tokens is twice for every model size doubling, meaning that using larger, higher-quality training datasets can lead to better results on downstream tasks. [5] [6]

It has been used for the Flamingo vision-language model . [7]

Architecture

Both the Gopher family and Chinchilla family are families of transformer models .

In particular, they are essentially the same as GPT-2 , with different sizes and minor modifications. Gopher family uses RMSNorm instead of LayerNorm ; relative positional encoding rather than absolute positional encoding. The Chinchilla family is the same as the Gopher family, but trained with AdamW instead of Adam optimizer .

The Gopher family contains six models of increasing size, from 44 million parameters to 280 billion parameters. They refer to the largest one as "Gopher" by default. Similar naming conventions apply for the Chinchilla family.

Table 1 of [2] shows the entire Gopher family:

Table 4 of [1] compares the 70-billion-parameter Chinchilla with Gopher 280B.

See also

LaMDA

References

v

t

e

Google

Google Brain

Google DeepMind

AlphaGo (2015)

Master (2016)

AlphaGo Zero (2017)

AlphaZero (2017)

MuZero (2019)

Fan Hui (2015)

Lee Sedol (2016)

Ke Jie (2017)

AlphaGo (2017)

The MANIAC (2023)

AlphaFold (2018)

AlphaStar (2019)

AlphaDev (2023)

AlphaGeometry (2024)

AlphaGenome (2025)

Inception (2014)

WaveNet (2016)

MobileNet (2017)

Transformer (2017)

EfficientNet (2019)

Gato (2022)

Quantum Artificial Intelligence Lab

TensorFlow

Tensor Processing Unit

Assistant (2016)

Sparrow (2022)

Gemini (2023)

BERT (2018)

XLNet (2019)

T5 (2019)

LaMDA (2021)

Chinchilla (2022)

PaLM (2022)

Imagen (2023)

Gemini (2023)

VideoPoet (2024)

Gemma (2024)

Veo (2024)

DreamBooth (2022)

NotebookLM (2023)

Vids (2024)

Gemini Robotics (2025)

" Attention Is All You Need "

Future of Go Summit

Generative pre-trained transformer

Google Labs

Google Pixel

Google Workspace

Robot Constitution

Category

Commons

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