

Title: EleutherAI

URL: <https://en.wikipedia.org/wiki/EleutherAI>

PageID: 73176768

Categories: Category:Applied machine learning, Category:Artificial intelligence laboratories, Category:Deep learning, Category:Language modeling, Category:Open-source artificial intelligence

Source: Wikipedia (CC BY-SA 4.0).

-----

Artificial general intelligence

Intelligent agent

Recursive self-improvement

Planning

Computer vision

General game playing

Knowledge representation

Natural language processing

Robotics

AI safety

Machine learning

Symbolic

Deep learning

Bayesian networks

Evolutionary algorithms

Hybrid intelligent systems

Systems integration

Open-source

Bioinformatics

Deepfake

Earth sciences

Finance

Generative AI Art Audio Music

Art

Audio

Music

Government

Healthcare Mental health

Mental health

Industry

Software development

Translation

Military

Physics

Projects

AI alignment

Artificial consciousness

The bitter lesson

Chinese room

Friendly AI

Ethics

Existential risk

Turing test

Uncanny valley

Timeline

Progress

AI winter

AI boom

AI bubble

Glossary

v

t

e

EleutherAI ( / ˈ ɛ l u θ ɪər / [ 2 ] ) is a grass-roots non-profit artificial intelligence (AI) research group. The group, considered an open-source version of OpenAI , [ 3 ] was formed in a Discord server in July 2020 by Connor Leahy , Sid Black, and Leo Gao [ 4 ] to organize a replication of GPT-3 . In early 2023, it formally incorporated as the EleutherAI Institute, a non-profit research institute. [ 5 ]

History

EleutherAI began as a Discord server on July 7, 2020, under the tentative name "LibreAI" before rebranding to "EleutherAI" later that month, [ 6 ] in reference to eleutheria , the Greek word for liberty . [ 3 ] Its founding members are Connor Leahy, Len Gao, and Sid Black. They co-wrote the code for Eleuther to serve as a collection of open source AI research, creating a machine learning model similar to GPT-3 . [ 7 ]

On December 30, 2020, EleutherAI released The Pile , a curated dataset of diverse text for training large language models . [ 8 ] While the paper referenced the existence of the GPT-Neo models, the models themselves were not released until March 21, 2021. [ 9 ] According to a retrospective written several months later, the authors did not anticipate that "people would care so much about our 'small models. ' " [ 1 ] On June 9, 2021, EleutherAI followed this up with GPT-J-6B , a six billion parameter language model that was again the largest open-source GPT-3-like model in the world. [ 10 ] These language models were released under the Apache 2.0 free software license and are considered to have "fueled an entirely new wave of startups". [ 5 ]

While EleutherAI initially turned down funding offers, preferring to use Google's TPU Research Cloud Program to source their compute, [ 11 ] by early 2021 they had accepted funding from

CoreWeave (a small cloud computing company) and SpellML (a cloud infrastructure company) in the form of access to powerful GPU clusters that are necessary for large scale machine learning research. On Feb 10, 2022, they released GPT-NeoX-20B, a model similar to their prior work but scaled up thanks to the resources CoreWeave provided. [ 12 ]

In 2022, many EleutherAI members participated in the BigScience Research Workshop, working on projects including multitask finetuning, [ 13 ] [ 14 ] training BLOOM , [ 15 ] and designing evaluation libraries. [ 15 ] Engineers at EleutherAI, Stability AI , and NVIDIA joined forces with biologists led by Columbia University and Harvard University [ 16 ] to train OpenFold, an open-source replication of DeepMind's AlphaFold2 . [ 17 ]

In early 2023, EleutherAI incorporated as a non-profit research institute run by Stella Biderman, Curtis Huebner, and Shivanshu Purohit. [ 5 ] [ 18 ] This announcement came with the statement that EleutherAI's shift of focus away from training larger language models was part of a deliberate push towards doing work in interpretability, alignment, and scientific research. [ 18 ] While EleutherAI is still committed to promoting access to AI technologies, they feel that "there is substantially more interest in training and releasing LLMs than there once was," enabling them to focus on other projects. [ 19 ]

In July 2024, an investigation by Proof news found that EleutherAI's The Pile dataset includes subtitles from over 170,000 YouTube videos across more than 48,000 channels. The findings drew criticism and accusations of theft from YouTubers and others who had their work published on the platform. [ 20 ] [ 21 ] In 2025, Stella Biderman served as executive director. Aviya Skowron served as head of policy and ethics. Nora Belrose served as head of interpretability, and Quentin Anthony was head of HPC. [ 22 ]

## Research

According to their website, EleutherAI is a "decentralized grassroots collective of volunteer researchers, engineers, and developers focused on AI alignment , scaling, and open-source AI research". [ 23 ] While they do not sell any of their technologies as products, they publish the results of their research in academic venues, write blog posts detailing their ideas and methodologies, and provide trained models for anyone to use for free. [ citation needed ]

## The Pile

The Pile is an 886 GB dataset designed for training large language models. It was originally developed to train EleutherAI's GPT-Neo models but has become widely used to train other models, including Microsoft 's Megatron-Turing Natural Language Generation, [ 24 ] [ 25 ] Meta AI 's Open

Pre-trained Transformers, [ 26 ] LLaMA , [ 27 ] and Galactica, [ 28 ] Stanford University 's BioMedLM 2.7B, [ 29 ] the Beijing Academy of Artificial Intelligence 's

Chinese-Transformer-XL, [ 30 ] and Yandex 's YaLM 100B. [ 31 ] Compared to other datasets, the Pile's main distinguishing features are that it is a curated selection of data chosen by researchers at EleutherAI to contain information they thought language models should learn and that it is the only such dataset that is thoroughly documented by the researchers who developed it. [ 32 ]

## GPT models

EleutherAI's most prominent research relates to its work to train open-source large language models inspired by OpenAI's GPT-3 . [ 33 ] EleutherAI's "GPT-Neo" model series has released 125 million, 1.3 billion, 2.7 billion, 6 billion, and 20 billion parameter models.

GPT-Neo (125M, 1.3B, 2.7B): [ 34 ] released in March 2021, it was the largest open-source GPT-3-style language model in the world at the time of release.

GPT-J (6B): [ 35 ] released in March 2021, it was the largest open-source GPT-3-style language model in the world at the time of release. [ 36 ]

GPT-NeoX (20B): [ 37 ] released in February 2022, it was the largest open-source language model in the world at the time of release.

Pythia (13B): [ 38 ] While prior models focused on scaling larger to close the gap with closed-sourced models like GPT-3, the Pythia model suite goes in another direction. The Pythia suite was designed to facilitate scientific research on the capabilities of and learning processes in large language models. [ 38 ] Featuring 154 partially trained model checkpoints, fully public training data, and the ability to reproduce the exact training order, Pythia enables research on verifiable training, [ 39 ] social biases, [ 38 ] memorization, [ 40 ] and more. [ 41 ]

## VQGAN-CLIP

Following the release of DALL-E by OpenAI in January 2021, EleutherAI started working on text-to-image synthesis models. When OpenAI did not release DALL-E publicly, EleutherAI's Katherine Crowson and digital artist Ryan Murdock developed a technique for using CLIP (another model developed by OpenAI) to convert regular image generation models into text-to-image synthesis ones. [ 44 ] [ 45 ] [ 46 ] [ 47 ] Building on ideas dating back to Google's DeepDream , [ 48 ] they found their first major success combining CLIP with another publicly available model called VQGAN and the resulting model is called VQGAN-CLIP. [ 49 ] Crowson released the technology by tweeting notebooks demonstrating the technique that people could run for free without any special equipment. [ 50 ] [ 51 ] [ 52 ] This work was credited by Stability AI CEO Emad Mostaque as motivating the founding of Stability AI. [ 53 ]

## Public reception

### Praise

EleutherAI's work to democratize GPT-3 won the UNESCO Netexplo Global Innovation Award in 2021, [ 54 ] InfoWorld's Best of Open Source Software Award in 2021 [ 55 ] and 2022, [ 56 ] was nominated for VentureBeat's AI Innovation Award in 2021. [ 57 ]

Gary Marcus , a cognitive scientist and noted critic of deep learning companies such as OpenAI and DeepMind, [ 58 ] has repeatedly [ 59 ] [ 60 ] praised EleutherAI's dedication to open-source and transparent research.

Maximilian Gahntz, a senior policy researcher at the Mozilla Foundation , applauded EleutherAI's efforts to give more researchers the ability to audit and assess AI technology. "If models are open and if data sets are open, that'll enable much more of the critical research that's pointed out many of the flaws and harms associated with generative AI and that's often far too difficult to conduct." [ 61 ]

### Criticism

Technology journalist Kyle Wiggers has raised concerns about whether EleutherAI is as independent as it claims, or "whether the involvement of commercially motivated ventures like Stability AI and Hugging Face —both of which are backed by substantial venture capital—might influence EleutherAI's research." [ 62 ]

### See also

List of artificial intelligence companies

## References

v

t

e

AGI

AI alignment

AI boom

AI capability control

AI safety

AI takeover  
Consequentialism  
Effective accelerationism  
Ethics of artificial intelligence  
Existential risk from artificial intelligence  
Friendly artificial intelligence  
Instrumental convergence  
Vulnerable world hypothesis  
Intelligence explosion  
Longtermism  
Machine ethics  
Suffering risks  
Superintelligence  
Technological singularity  
Alignment Research Center  
Center for AI Safety  
Center for Applied Rationality  
Center for Human-Compatible Artificial Intelligence  
Centre for the Study of Existential Risk  
EleutherAI  
Future of Humanity Institute  
Future of Life Institute  
Google DeepMind  
Humanity+  
Institute for Ethics and Emerging Technologies  
Leverhulme Centre for the Future of Intelligence  
Machine Intelligence Research Institute  
OpenAI  
Safe Superintelligence  
Scott Alexander  
Sam Altman  
Yoshua Bengio  
Nick Bostrom  
Paul Christiano  
Eric Drexler  
Sam Harris  
Stephen Hawking  
Dan Hendrycks

Geoffrey Hinton

Bill Joy

Shane Legg

Elon Musk

Steve Omohundro

Huw Price

Martin Rees

Stuart J. Russell

Ilya Sutskever

Jaan Tallinn

Max Tegmark

Frank Wilczek

Roman Yampolskiy

Eliezer Yudkowsky

Artificial Intelligence Act

Do You Trust This Computer?

Human Compatible

Open letter on artificial intelligence (2015)

Our Final Invention

Roko's basilisk

Statement on AI risk of extinction

Superintelligence: Paths, Dangers, Strategies

The Precipice

If Anyone Builds It, Everyone Dies