

Title: IBM Granite

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

PageID: 76093849

Categories: Category:2023 in artificial intelligence, Category:2023 software, Category:Artificial neural networks, Category:Free software, Category:Generative pre-trained transformers, Category:IBM products, Category:IBM software, Category:Large language models, Category:Open-source artificial intelligence

Source: Wikipedia (CC BY-SA 4.0).

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
Naïve 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

IBM Granite is a series of decoder-only AI foundation models created by IBM . [3] It was announced on September 7, 2023, [4] [5] and an initial paper was published 4 days later. [6] Initially intended for use in the IBM's cloud-based data and generative AI platform Watsonx along with other models, [7] IBM opened the source code of some code models. [8] [9] Granite models are trained on datasets curated from Internet , academic publishings , code datasets, legal and finance documents. [10] [11] [1]

Foundation models

A foundation model is an AI model trained on broad data at scale such that it can be adapted to a wide range of downstream tasks. [12]

Granite's first foundation models were Granite.13b.instruct and Granite.13b.chat. The "13b" in their name comes from 13 billion, the amount of parameters they have as models, lesser than most of the larger models of the time. Later models vary from 3 to 34 billion parameters. [4] [13]

On May 6, 2024, IBM released the source code of four variations of Granite Code Models under Apache 2 , an open source permissive license that allows completely free use, modification and sharing of the software, and put them on Hugging Face for public use. [14] [15] According to IBM's own report, Granite 8b outperforms Llama 3 on several coding related tasks within similar range of parameters. [16] [17]

See also

Mistral AI , a company that also provides open source models

GPT

LLaMA

Cyc

Gemini

References

External links

[1]

GitHub page

IBM Granite Playground

v

t

e

History World War II

World War II

Mergers and acquisitions PC business acquisition by Lenovo

PC business acquisition by Lenovo

Mainframe IBM Z

IBM Z

Power microprocessors

Power Systems

Storage FlashSystem DS8000

FlashSystem

DS8000

Quantum Q System One Q System Two Eagle Osprey Heron Condor

Q System One

Q System Two

Eagle

Osprey

Heron
Condor
Blue Gene
Cell microprocessors
PowerPC
Midrange computer
Personal Computer
Selectric
ThinkPad
Carbon Design System
Cloud Cloudant
Cloudant
Cognos Analytics
Connections
Criminal Reduction Utilising Statistical History
Fortran
ILOG
Information Management Software
Lotus Software
Mainframe operating systems
Mashup Center
Planning Analytics
PureQuery
Quantum Platform Qiskit OpenQASM
Qiskit
OpenQASM
Rational Software
SPSS
Tivoli Software Service Automation Manager
Service Automation Manager
Watson
Watsonx Granite
Granite
WebSphere
Apptio
Center for The Business of Government
Consulting Promontory
Promontory

Kenexa

International subsidiaries India

India

Press

Red Hat

Research

AdStar

AIM alliance Kaleida Labs Taligent

Kaleida Labs

Taligent

Ambra Computer

Cognos

EduQuest

Kyndryl

Lexmark

Merative

Microelectronics

Product Center

Science Research Associates

Service Bureau

The Weather Company (Weather Underground)

Towers 1250 René-Lévesque , Montreal, QC One Atlantic Center , Atlanta, GA

1250 René-Lévesque , Montreal, QC

One Atlantic Center , Atlanta, GA

Software Labs Rome Software Lab Toronto Software Lab

Rome Software Lab

Toronto Software Lab

IBM Buildings Chicago Honolulu New York Seattle

Chicago

Honolulu

New York

Seattle

Facilities Thomas J. Watson Research Center Hakoziaki Facility Yamato Facility

Thomas J. Watson Research Center

Hakoziaki Facility

Yamato Facility

Cambridge Scientific Center

IBM Hursley

Canada Head Office Building
IBM Rochester
Academy of Technology
Deep Thunder Develothon
Develothon
Fellow
The Great Mind Challenge
Linux Technology Center
SkillsBuild
Smarter Planet
Virtual Universe Community
World Community Grid
Think conference
Automated teller machine
Cynefin framework
DRAM
Electronic keypunch
Floppy disk
Hard disk drive
Magnetic stripe card
Relational model
Sabre airline reservation system
Scanning tunneling microscope
Financial swaps
Universal Product Code
Big Blue
Commercial Processing Workload
Customer engineer
Globally integrated enterprise
e-business
Think slogan
Thomas J. Watson (1914–1956)
Thomas Watson Jr. (1956–1971)
T. Vincent Learson (1971–1973)
Frank T. Cary (1973–1981)
John R. Opel (1981–1985)
John Fellows Akers (1985–1993)
Louis V. Gerstner Jr. (1993–2002)

Samuel J. Palmisano (2002–2011)

Ginni Rometty (2012–2020)

Arvind Krishna (since 2020)

Thomas Buberl

David Farr

Alex Gorsky

Michelle J. Howard

Arvind Krishna

Andrew Liveris

Martha E. Pollack

Joseph R. Swedish

Peter R. Voser

A Boy and His Atom

Big Blue sports teams American football Rugby union

American football

Rugby union

Common Public License / IBM Public License Wallace v. International Business Machines Corp.

Wallace v. International Business Machines Corp.

Deep Blue

Deep Thought

Dynamic infrastructure

GlobalFoundries

GUIDE International

IBM and the Holocaust

International chess tournament

Lucifer cipher

Mathematica

IBM Plex

SHARE computing

ScicomP

Unions

Commons

Category

Navigational boxes FOSS Midrange computers Operating systems Personal computers
System/360 System/370 Typewriters Vacuum tube computers

FOSS

Midrange computers

Operating systems

Personal computers
System/360
System/370
Typewriters
Vacuum tube computers
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