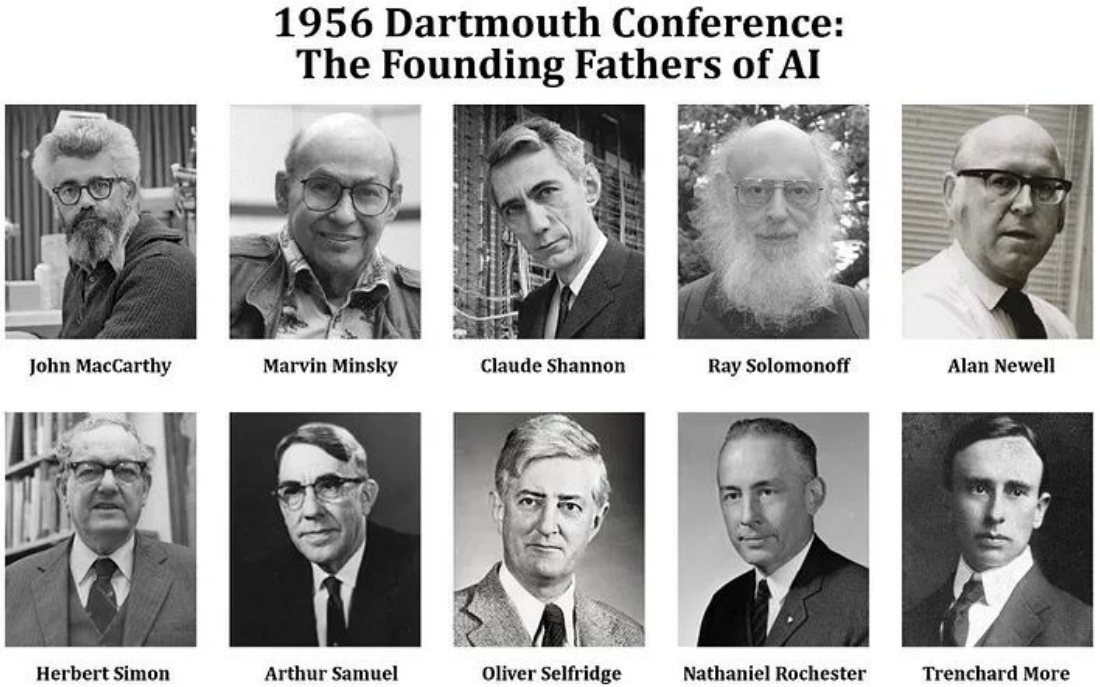
Inteligência Artificial Generativa: Histórico e Perspectivas

Prof. Tsang Ing Ren
tir@cin.ufpe.br

Centro de Informática UFPE

Universidade Federal de Pernambuco

Artificial Intelligence



https://spectrum.ieee.org/dartmouth-ai-workshop

Artificial Intelligence

In the back row from left to right are Oliver Selfridge, Nathaniel Rochester, Marvin Minsky, and John McCarthy. In front on the left is Ray Solomonoff; on the right, Claude Shannon. The identity of the person between Solomonoff and Shannon remained a mystery for some time. The Minsky Family

https://spectrum.ieee.org/dartmouth-ai-workshop

50 anos depois

Photographer: Joe Mehling

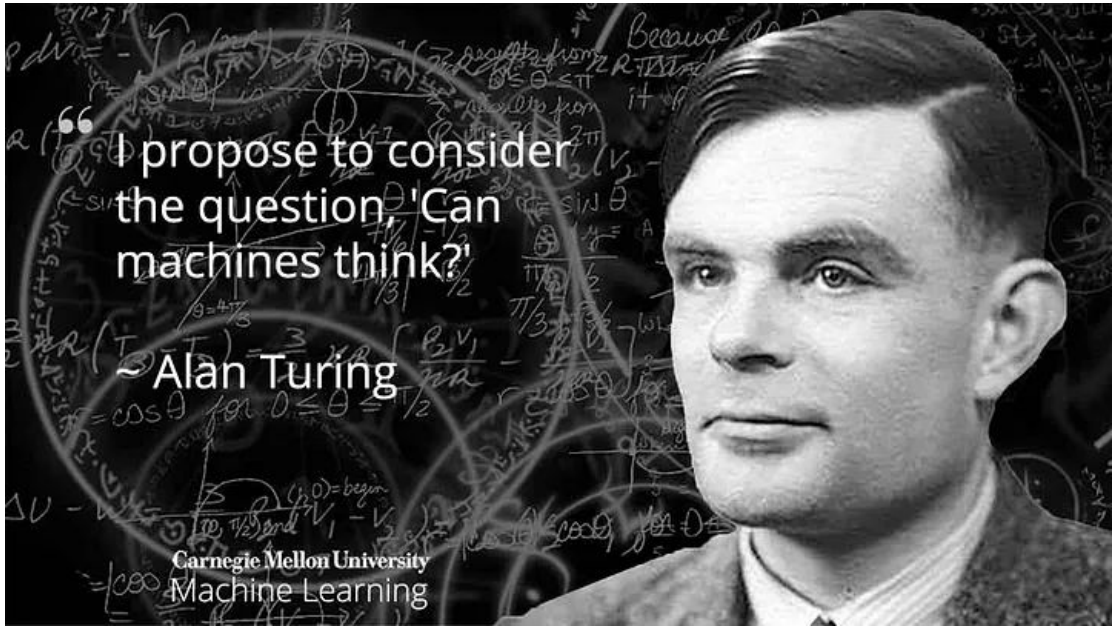
Figure 1. Trenchard More, John McCarthy, Marvin Minsky, Oliver Selfridge, and Ray Solomonoff.

AI Magazine Volume 27 Number 4 (2006) (© AAAI)

Artificial Intelligence

AI Magazine Volume 27 Number 4 (2006) (© AAAI)

Artificial Intelligence



https://medium.com/@jetnew/a-summary-of-alan-m-turings-computing-machinery-and-intelligence-fd714d187c0b

https://watermark.silverchair.com/lix-236-43

VOL. LIX. NO. 236.] [October, 1950

MIND
A QUARTERLY REVIEW
OF
PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND INTELLIGENCE

BY A. M. TURING

1. *The Imitation Game.*

I PROPOSE to consider the question, ‘Can machines think?’ This should begin with definitions of the meaning of the terms ‘machine’ and ‘think’. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words ‘machine’ and ‘think’ are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, ‘Can machines think?’ is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

The new form of the problem can be described in terms of a game which we call the ‘imitation game’. It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either ‘X is A and Y is B’ or ‘X is B and Y is A’. The interrogator is allowed to put questions to A and B thus:

C: Will X please tell me the length of his or her hair?

Artificial Intelligence

nature

View all journals | Search | Log in

Explore content | About the journal | Publish with us | Subscribe | Sign up for alerts | RSS feed

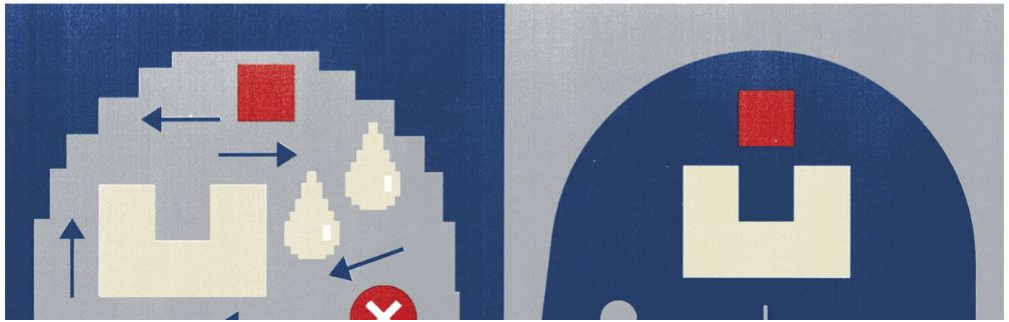
nature > news feature > article

NEWS FEATURE | 25 July 2023

ChatGPT broke the Turing test — the race is on for new ways to assess AI

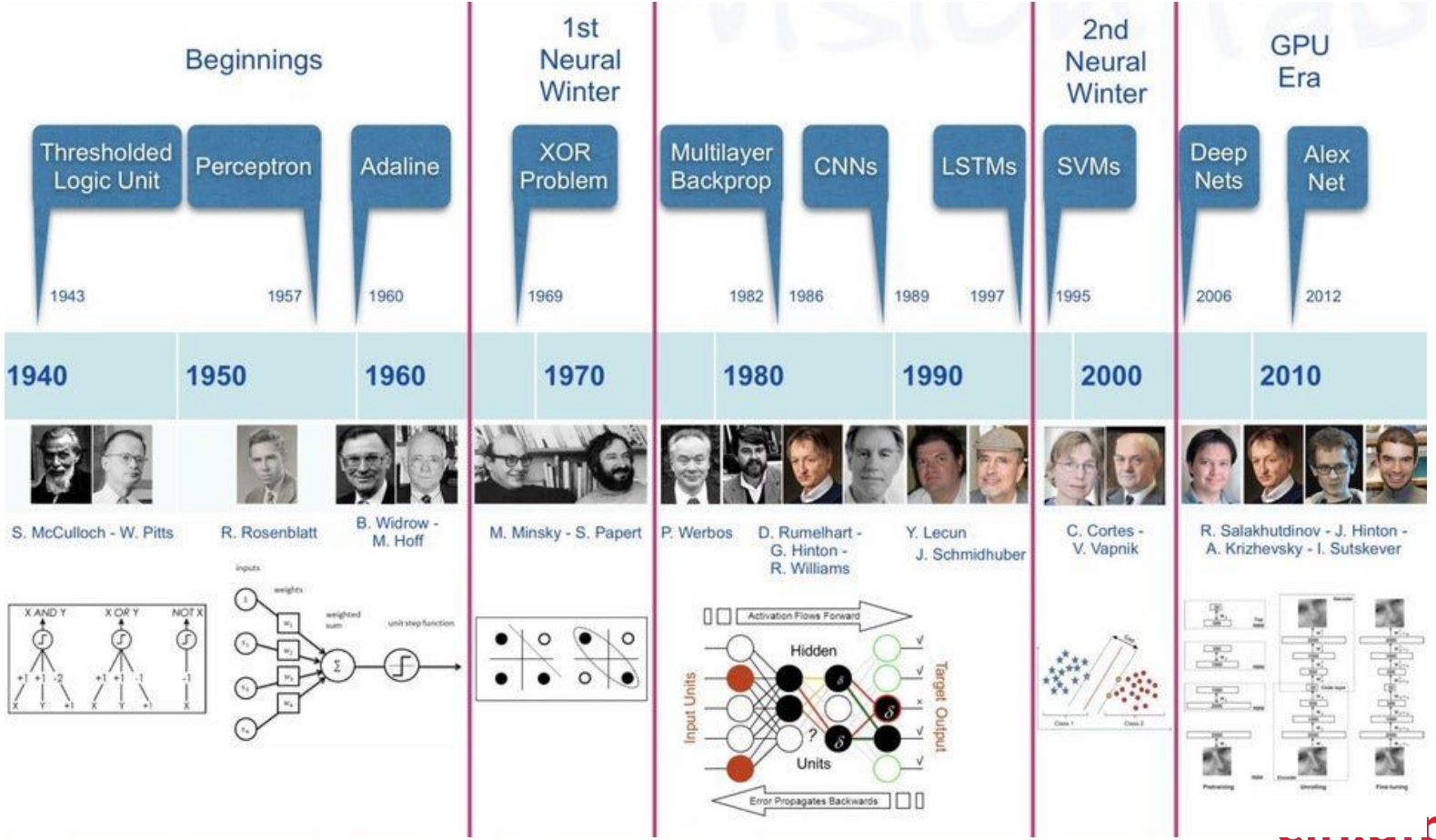
Large language models mimic human chatter, but scientists disagree on their ability to reason.

By Celeste Biever



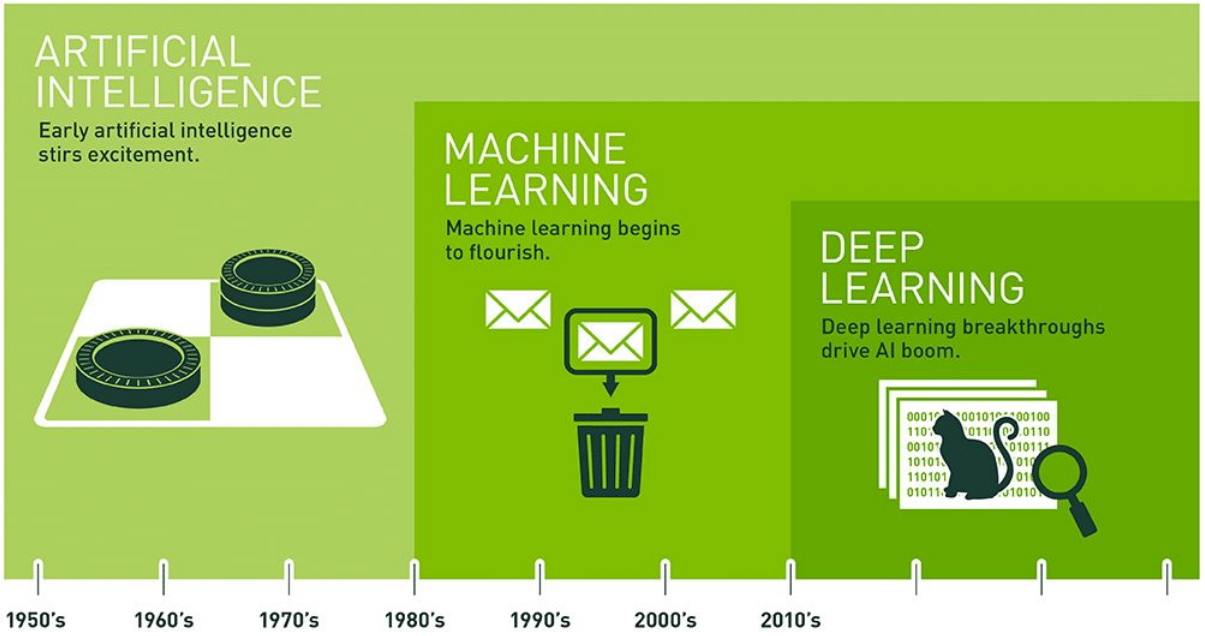
https://www.nature.com/articles/d41586-023-02361-7

Deep Learning



https://twitter.com/jesparent/status/1173618397046804481

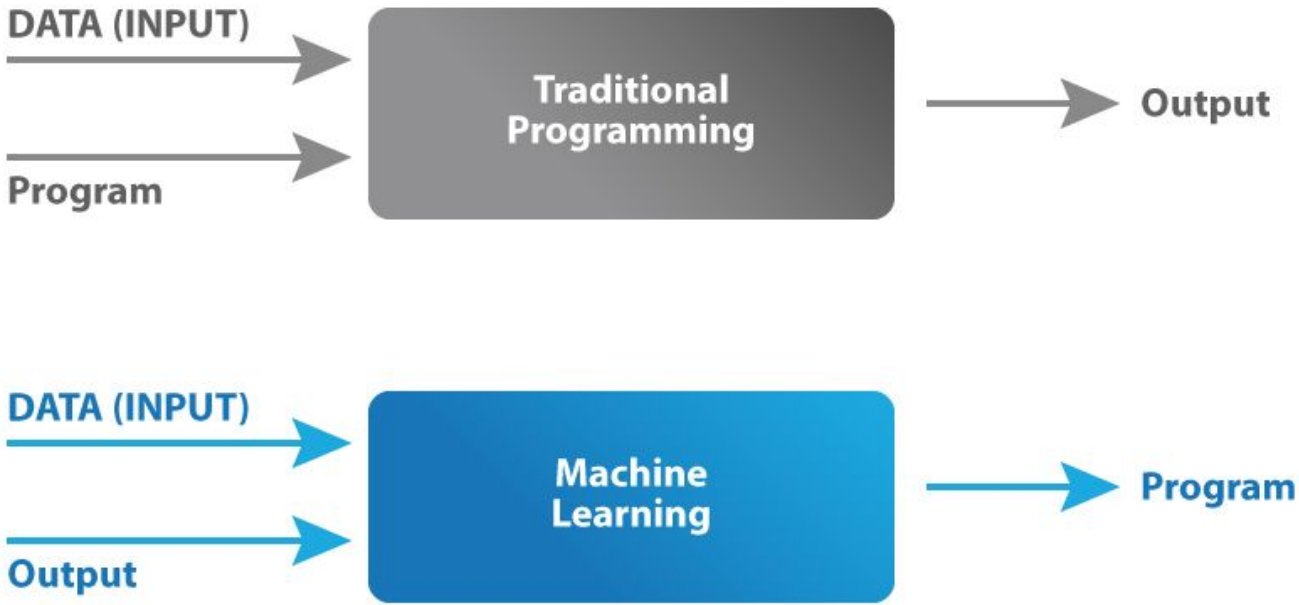
Artificial Intelligence - Machine Learning - Deep Learning



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

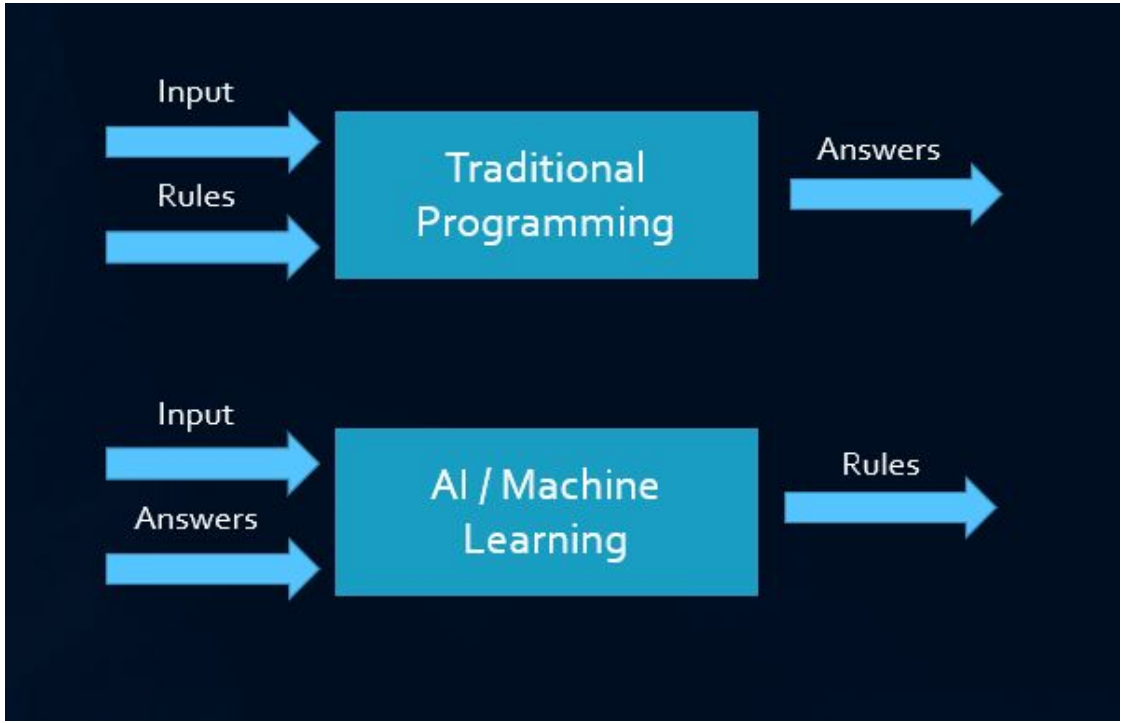
https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/

Machine Learning



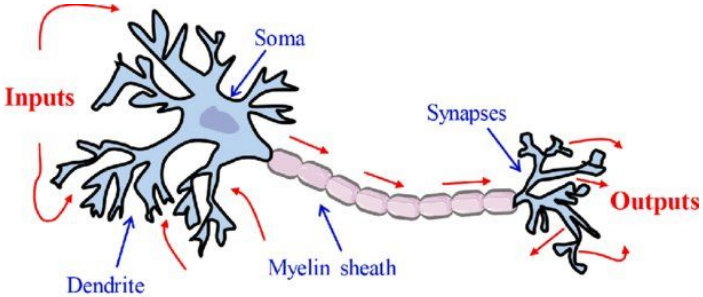
https://www.mygreatlearning.com/blog/what-is-machine-learning/

Machine Learning

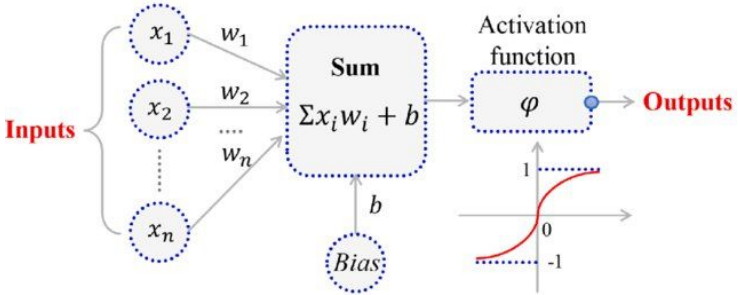


https://medium.com/mlearning-ai/a-glimpse-at-machine-learning-cheatsheet-f364c9fd473b

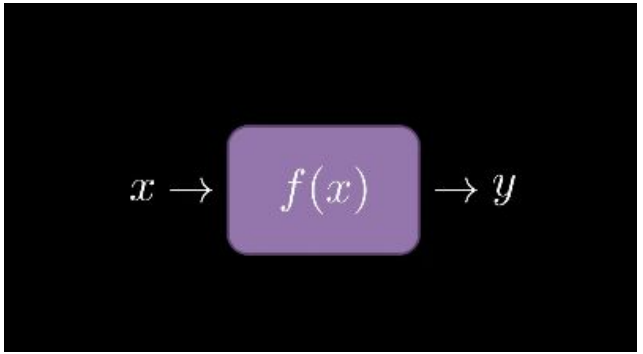
Neural Network - Perceptron



(a) Biological neuron

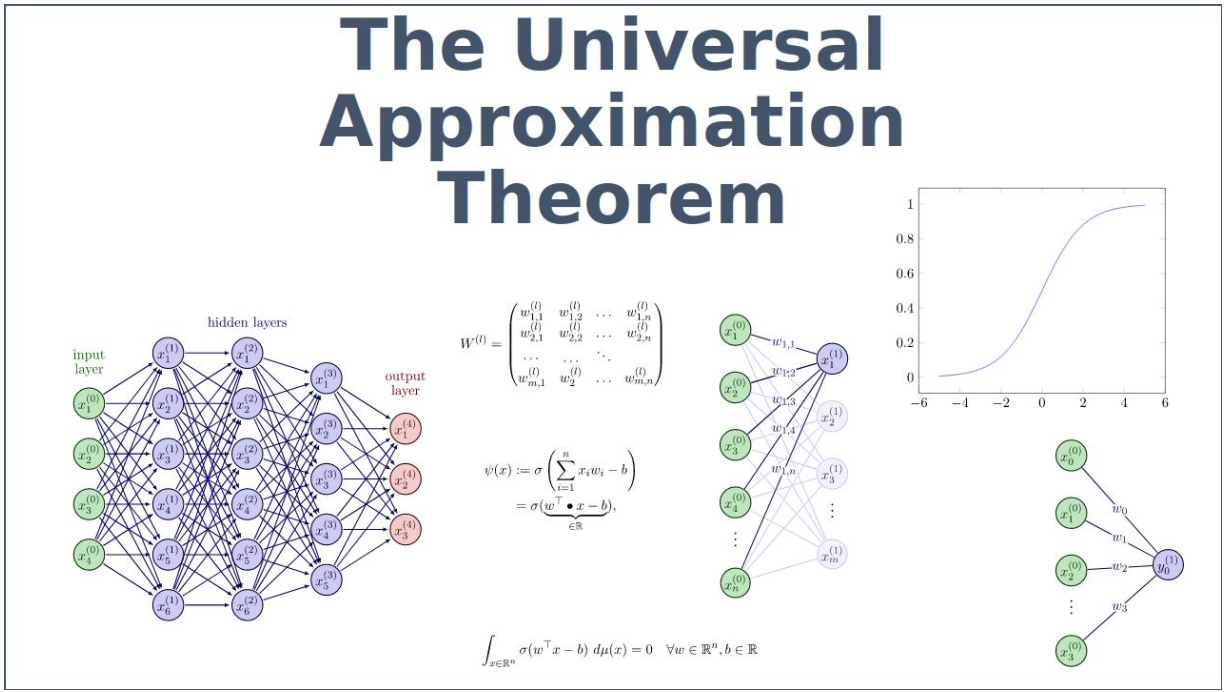


(b) Artificial neuron



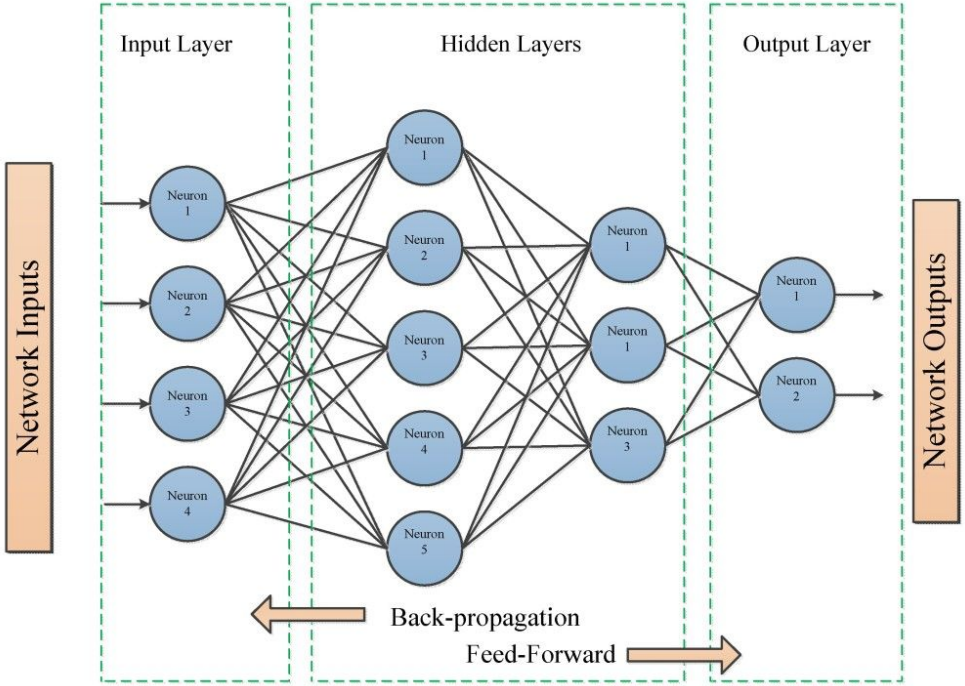
https://www.sciencedirect.com/science/article/pii/S2352012421003179

Neural Network - UAT



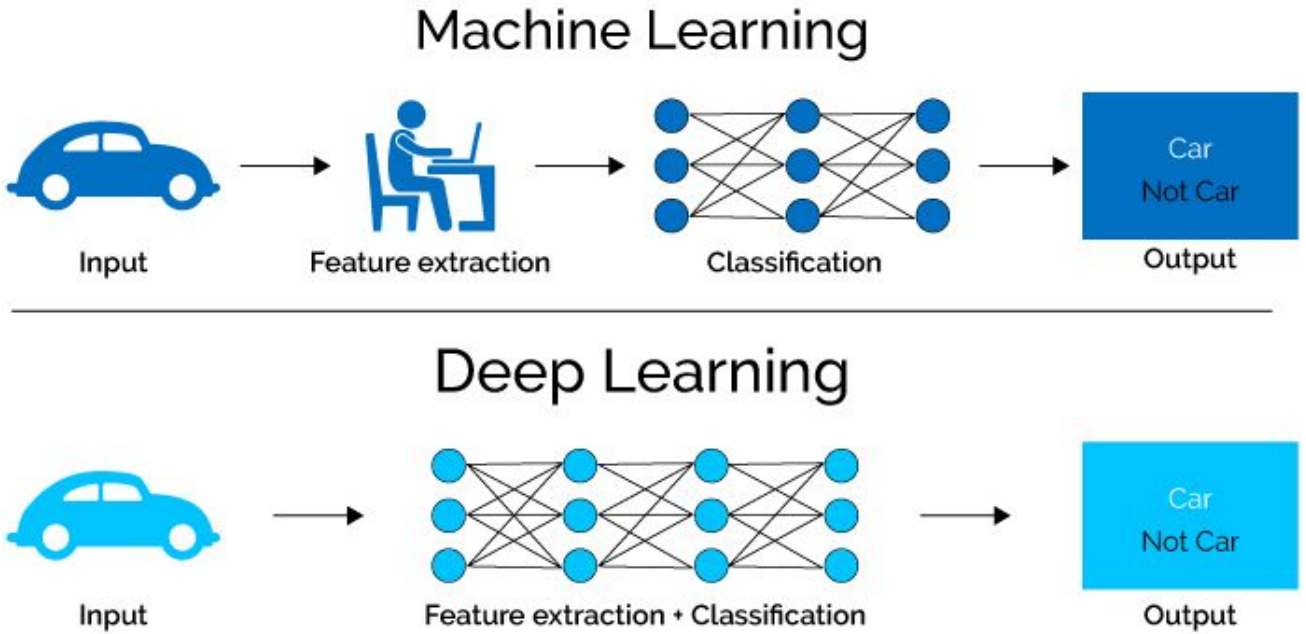
https://www.deep-mind.org/2023/03/26/the-universal-approximation-theorem/

Neural Network - Backpropagation



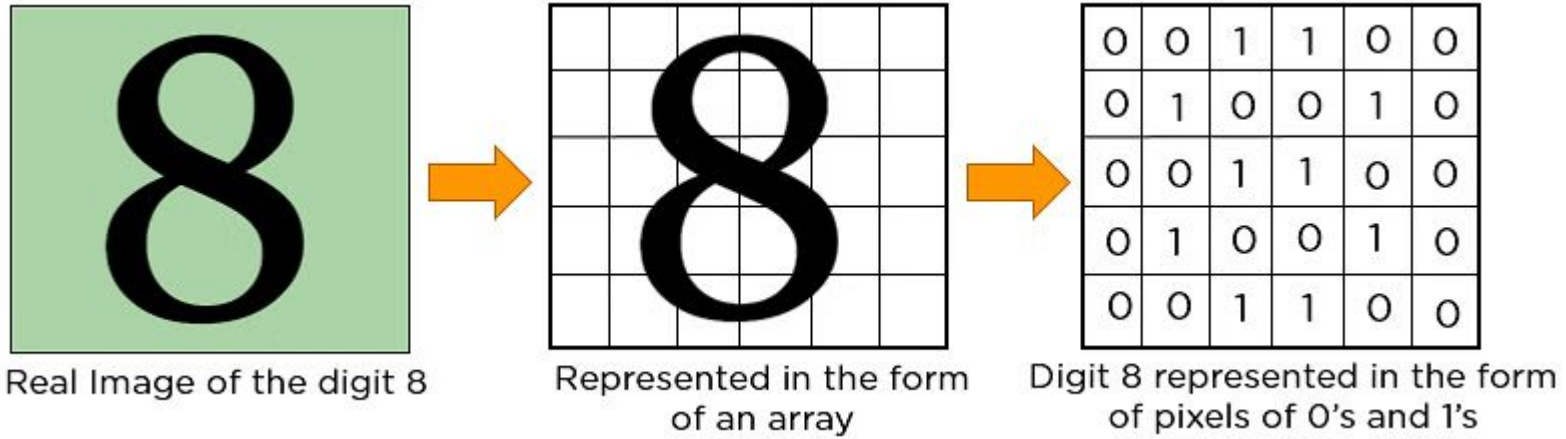
https://www.linkedin.com/pulse/feedforward-vs-backpropagation-ann-saffronedge1/

Machine Learning - Deep Learning



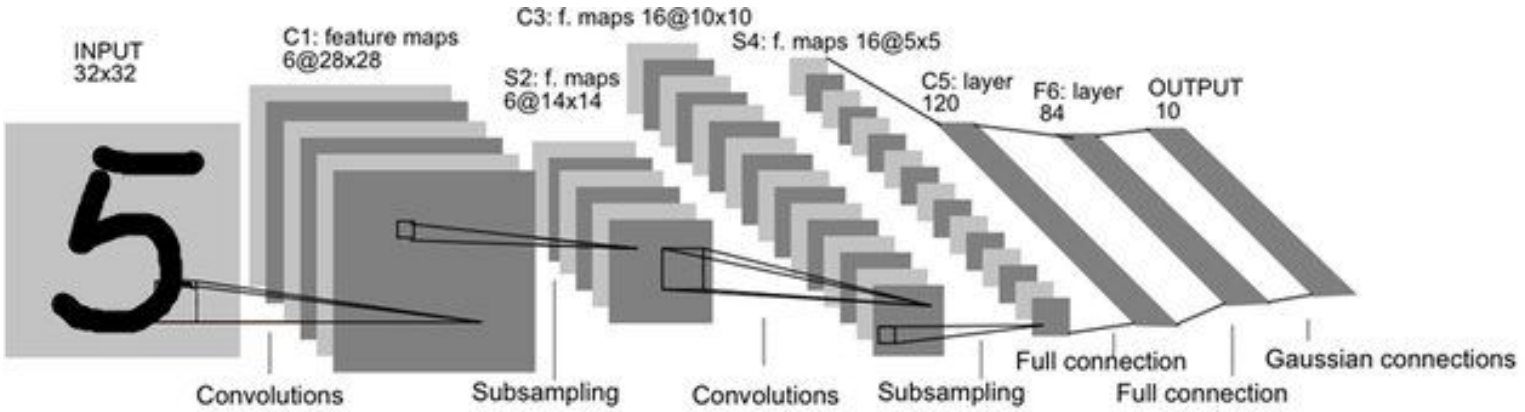
https://semiengineering.com/deep-learning-spreads/

Image Representation



https://www.simplilearn.com/tutorials/deep-learning-tutorial/convolutional-neural-network

Deep Learning

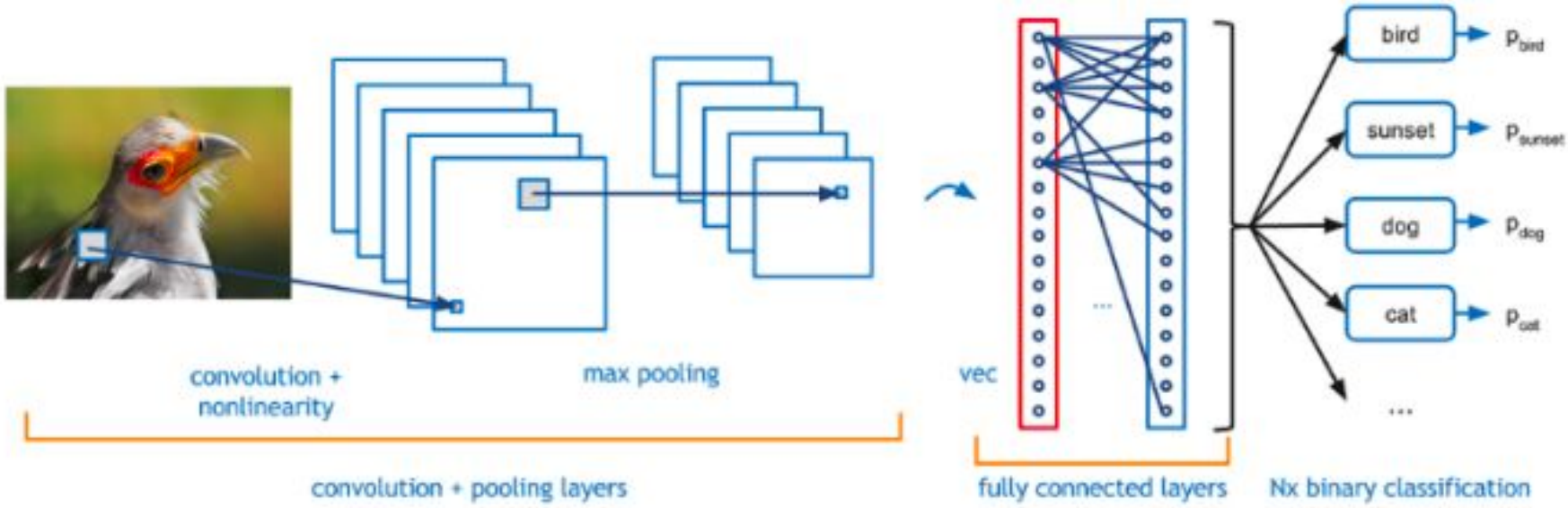


An early (Le-Net5) Convolutional Neural Network design, LeNet-5, used for recognition of digits

Key Deep Learning Architectures: LeNet-5

LeNet-5 [1998, paper by LeCun et al.]

Deep Learning



Deep Learning

http://www.image-net.org/

ImageNet (2010)

22,000 classes em +- 15 milhões de imagens de alta resolução

Deep Learning

IMAGENET

14,197,122 images, 21841 synsets indexed

Explore Download Challenges Publications Updates About

Not logged in. Login | Signup

ImageNet is an image database organized according to the WordNet hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images. Currently we have an average of over five hundred images per node. We hope ImageNet will become a useful resource for researchers, educators, students and all of you who share our passion for pictures.

Click here to learn more about ImageNet, Click here to join the ImageNet mailing list.

What do these images have in common? *Find out!*

Research updates on improving ImageNet data

© 2016 Stanford Vision Lab, Stanford University, Princeton University support@image-net.org Copyright infringement

FEI-FEI LI
(publishes under L.Fei-Fei)

Sequoia Professor of
Computer Science
Denning Co-Director
Stanford HAI

https://www.youtube.com/watch?v=40riCqvRoMs

Deep Learning

AlexNet (2012)

ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky University of Toronto kriz@cs.utoronto.ca
Ilya Sutskever University of Toronto ilya@cs.utoronto.ca
Geoffrey E. Hinton University of Toronto hinton@cs.utoronto.ca

Abstract

We trained a large, deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. On the test data, we achieved top-1 and top-5 error rates of 37.5% and 17.0% which is considerably better than the previous state-of-the-art. The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way softmax. To make training faster, we used non-saturating neurons and a very efficient GPU implementation of the convolution operation. To reduce overfitting in the fully-connected layers we employed a recently-developed regularization method called "dropout" that proved to be very effective. We also entered a variant of this model in the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry.

left: Alex Krizhevsky, Middle: Ilya Sutskever, Right: Geoffery Hinton

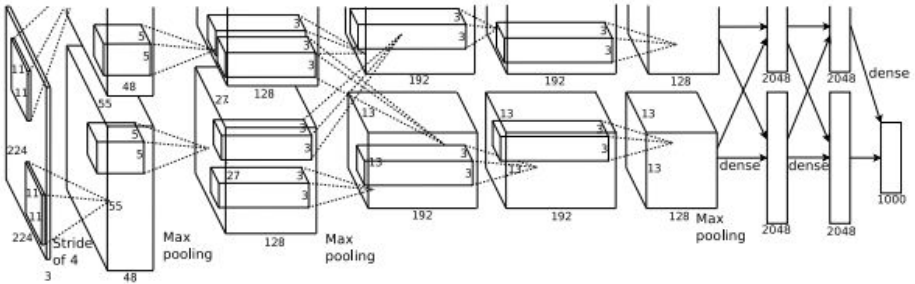


Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input is 150,528-dimensional, and the number of neurons in the network's remaining layers is given by 253,440–186,624–64,896–64,896–43,264–4096–4096–1000.

Deep Learning

AlexNet (2012)

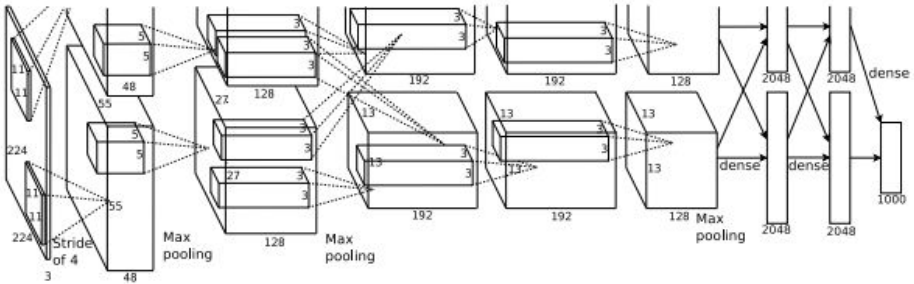


Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input is 150,528-dimensional, and the number of neurons in the network's remaining layers is given by 253,440–186,624–64,896–64,896–43,264–4096–4096–1000.

Nvidia GTX 580

Deep Learning

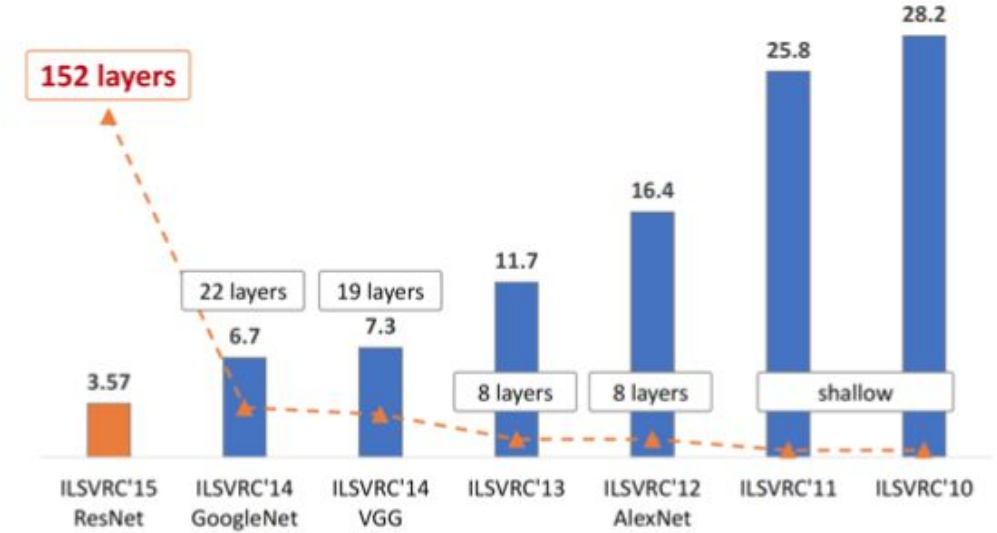


Fig. 1. The evolution of the winning entries on the ImageNet Large Scale Visual Recognition Challenge from 2010 to 2015. Since 2012, CNNs have outperformed hand-crafted descriptors and shallow networks by a large margin. Image re-printed with permission [36].

Deep Learning

Fig. 1. Performance of the winners of the ImageNet classification competitions over the years.

Deep Learning

The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches

Md Zahangir Alom[1], Tarek M. Taha[1], Chris Yakopcic[1], Stefan Westberg[1], Paheding Sidike[2], Mst Shamima Nasrin[1], Brian C Van Essen[3], Abdul A S. Awwal[3], and Vijayan K. Asari[1]

***Abstract*—In recent years, deep learning has garnered tremendous success in a variety of application domains. This new field of machine learning has been growing rapidly, and has been applied to most traditional application domains, as well as some new areas that present more opportunities. Different methods have been proposed based on different categories of learning, including supervised, semi-supervised, and un-supervised learning. Experimental results show state-of-the-art performance using deep learning when compared to traditional machine learning approaches in the fields of image processing, computer vision, speech recognition, machine translation, art, medical imaging, medical information processing, robotics and control, bio-informatics, natural language processing (NLP), cybersecurity, and many others.**

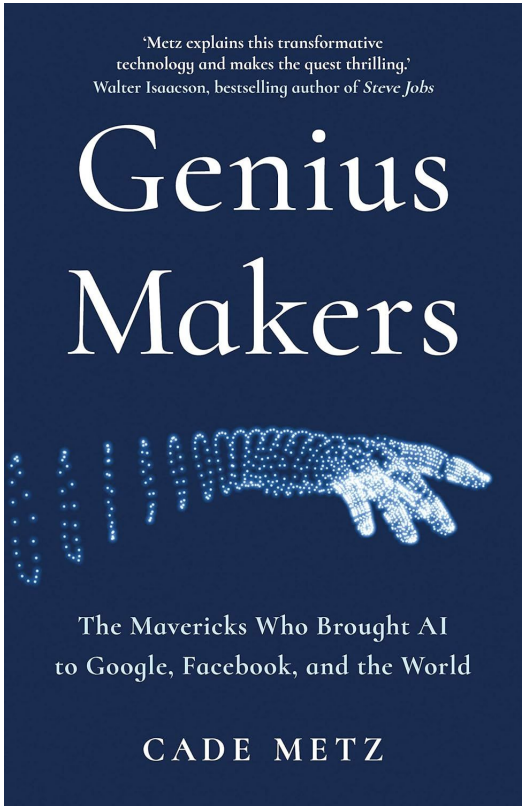
This report presents a brief survey on the advances that have occurred in the area of DL, starting with the Deep Neural Network (DNN). The survey goes on to cover the Convolutional Neural Network (CNN), the Recurrent Neural Network (RNN) including Long Short Term Memory (LSTM) and Gated Recurrent Units

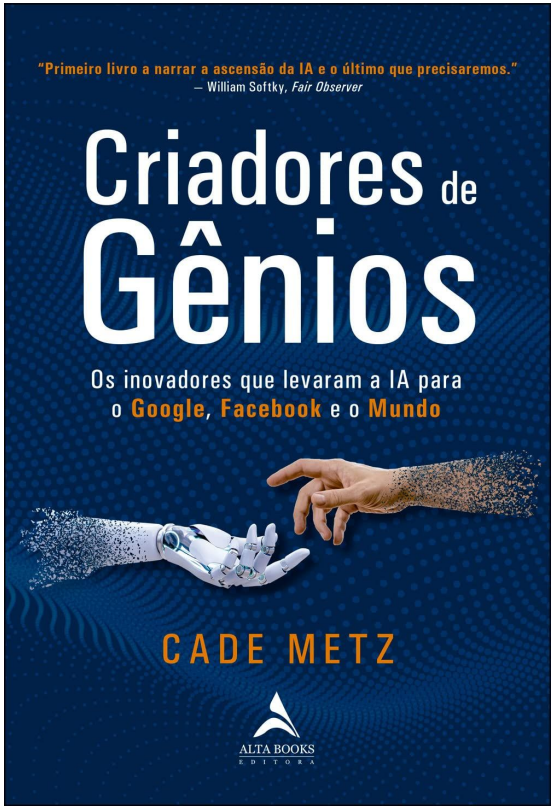
I. Introduction

Since the 1950s, a small subset of Artificial Intelligence (AI), often called Machine Learning (ML), has revolutionized several fields in the last few decades. Neural Networks (NN) are a subfield of ML, and it was this subfield that spawned Deep Learning (DL). Since its inception DL has been creating ever larger disruptions, showing outstanding success in almost every application domain. Fig. 1 shows, the taxonomy of AI. DL (using either deep architecture of learning or hierarchical learning approaches) is a class of ML developed largely from 2006 onward. Learning is a procedure consisting of estimating the model parameters so that the learned model (algorithm) can perform a specific task. For example, in Artificial Neural Networks (ANN), the parameters are the weight matrices $(w_{i,j}{}'s)$. DL on the other hand consists of several layers in between the input and output layer which allows for many stages of non-linear information processing units with

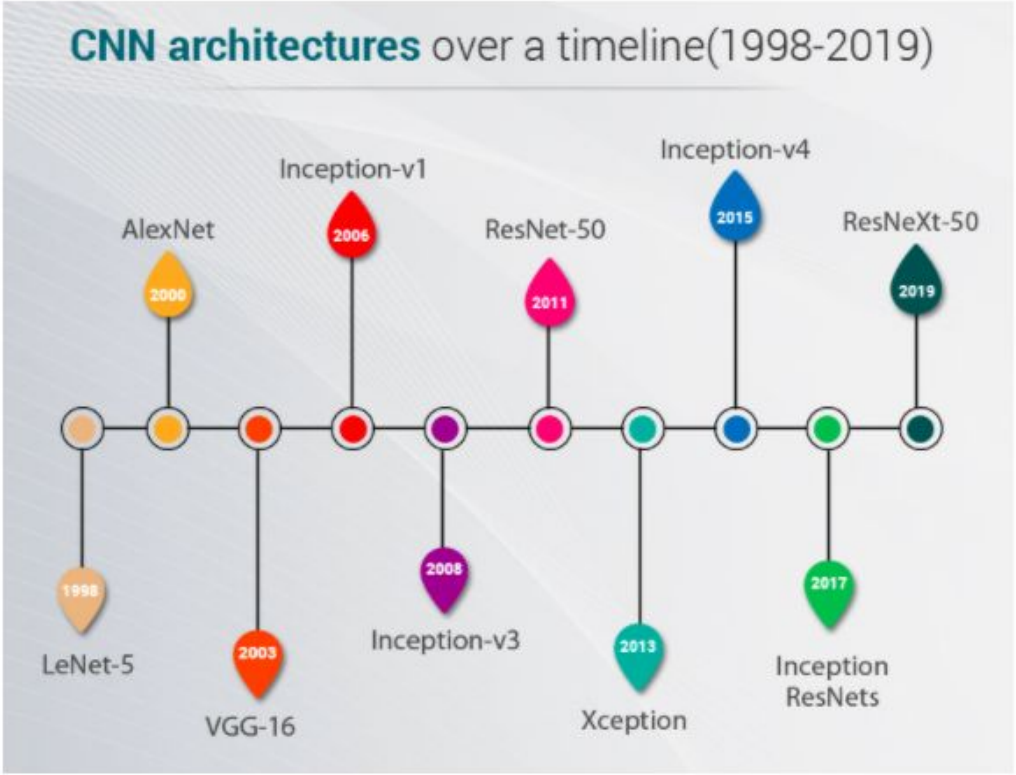
https://arxiv.org/abs/1803.01164

Deep Learning





Deep Learning



Deep Learning

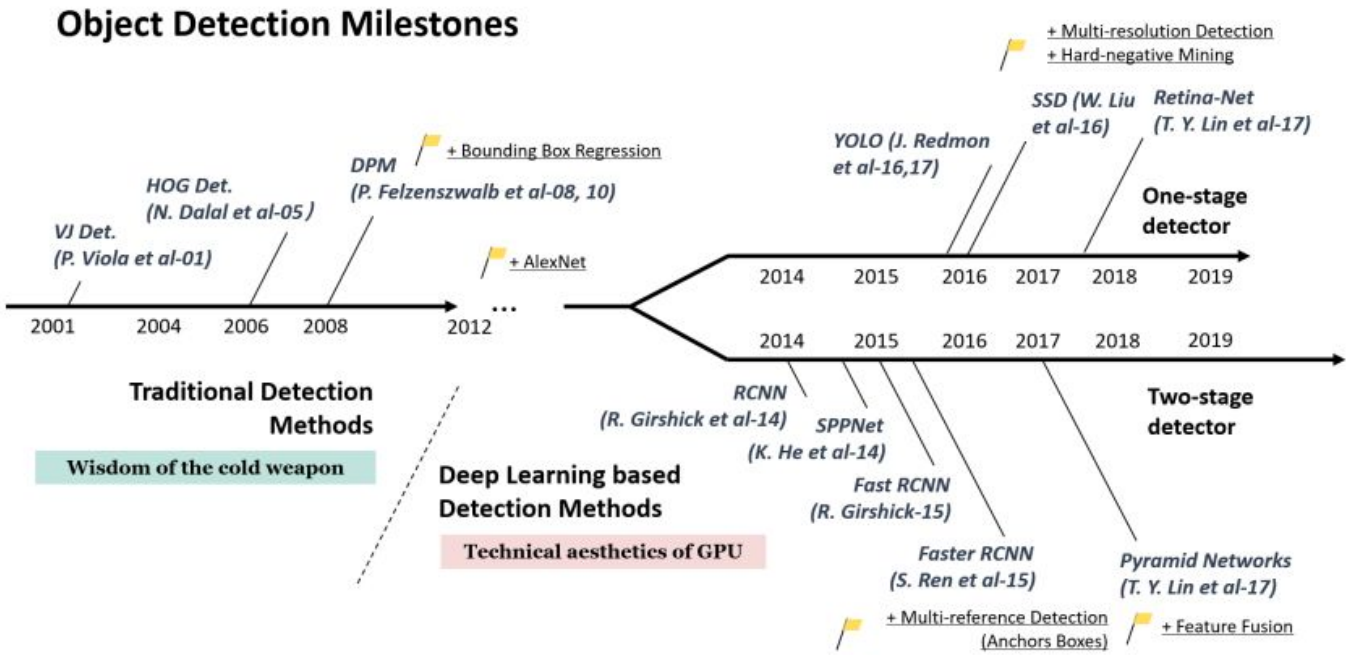


Fig. 2. A road map of object detection. Milestone detectors in this figure: VJ Det. [10, 11], HOG Det. [12], DPM [13–15], RCNN [16], SPPNet [17], Fast RCNN [18], Faster RCNN [19], YOLO [20], SSD [21], Pyramid Networks [22], Retina-Net [23].

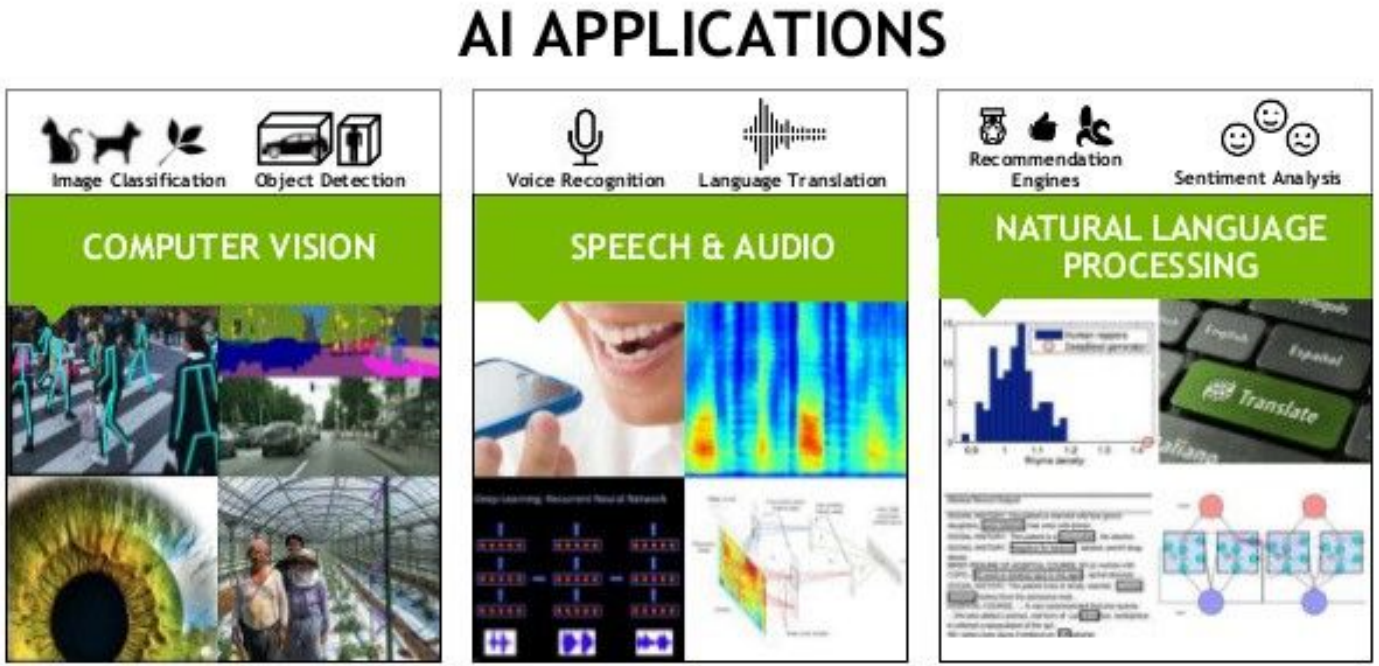
Deep Learning

Computer Vision
Openpilot e Tesla autopilot

https://www.comma.ai/openpilot

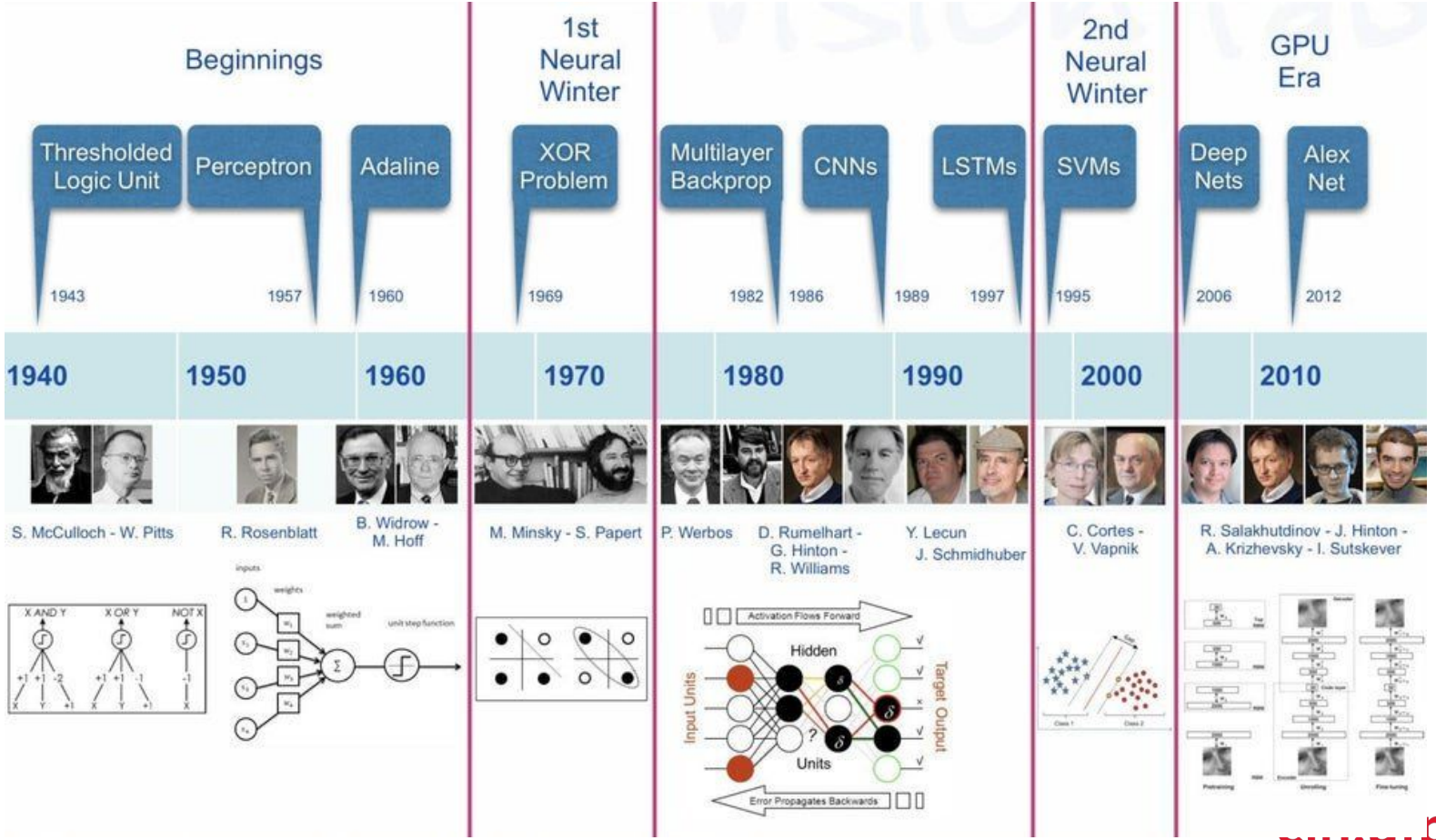
https://www.tesla.com/autopilot

Deep Learning



https://medium.com/@sanchittanwar75/introduction-to-machine-learning-and-deep-learning-bd25b792e488

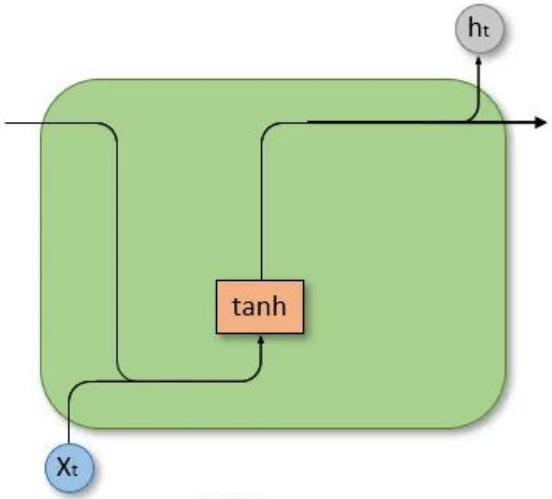
Deep Learning



https://twitter.com/jesparent/status/1173618397046804481

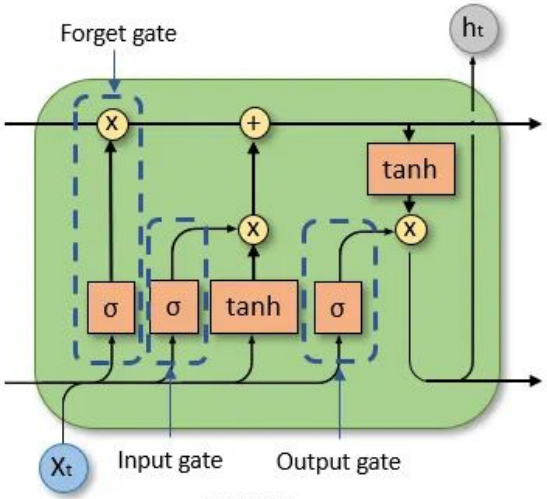
Deep Learning - Sequence model

1986



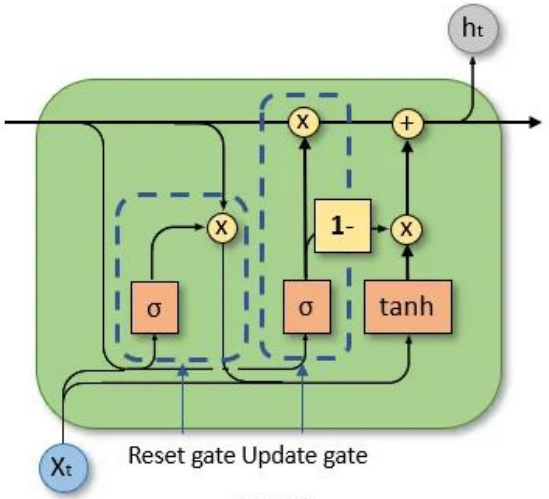
RNN

1995-1997



LSTM

2014



GRU

https://towardsdatascience.com/a-brief-introduction-to-recurrent-neural-networks-638f64a61ff4

Deep Learning

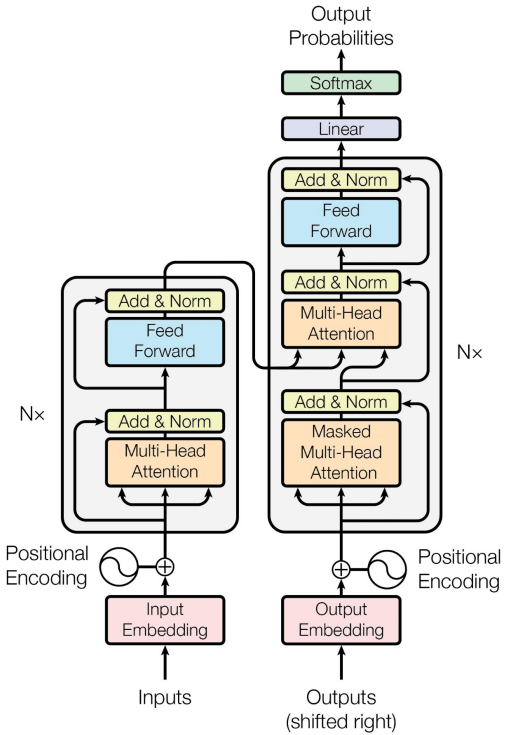
2017

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com
Noam Shazeer* Google Brain noam@google.com
Niki Parmar* Google Research nikip@google.com
Jakob Uszkoreit* Google Research usz@google.com
Llion Jones* Google Research llion@google.com
Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu
Łukasz Kaiser* Google Brain lukaszkaiser@google.com
Illia Polosukhin* ‡ illia.polosukhin@gmail.com

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the



https://arxiv.org/abs/1706.03762

Large Language Models - LLM

LARGE LANGUAGE MODELS - PIVOTAL INNOVATIONS

Original Eliza Script

1966

ELIZA
ELIZA developed by Joseph Weizenbaum at MIT to simulate limited conversations with a human.

1972

STNLP
Statistically-Trained Natural Language Processing System (STNLP) developed by Terry Winograd at MIT. STNLP was a language model that could generate text from statistical rules.

1997

LSTM
Long Short-Term Memory (LSTM) network developed by Sepp Hochreiter and Jürgen Schmidhuber as recurrent neural network able to learn from data and generate text.

1999

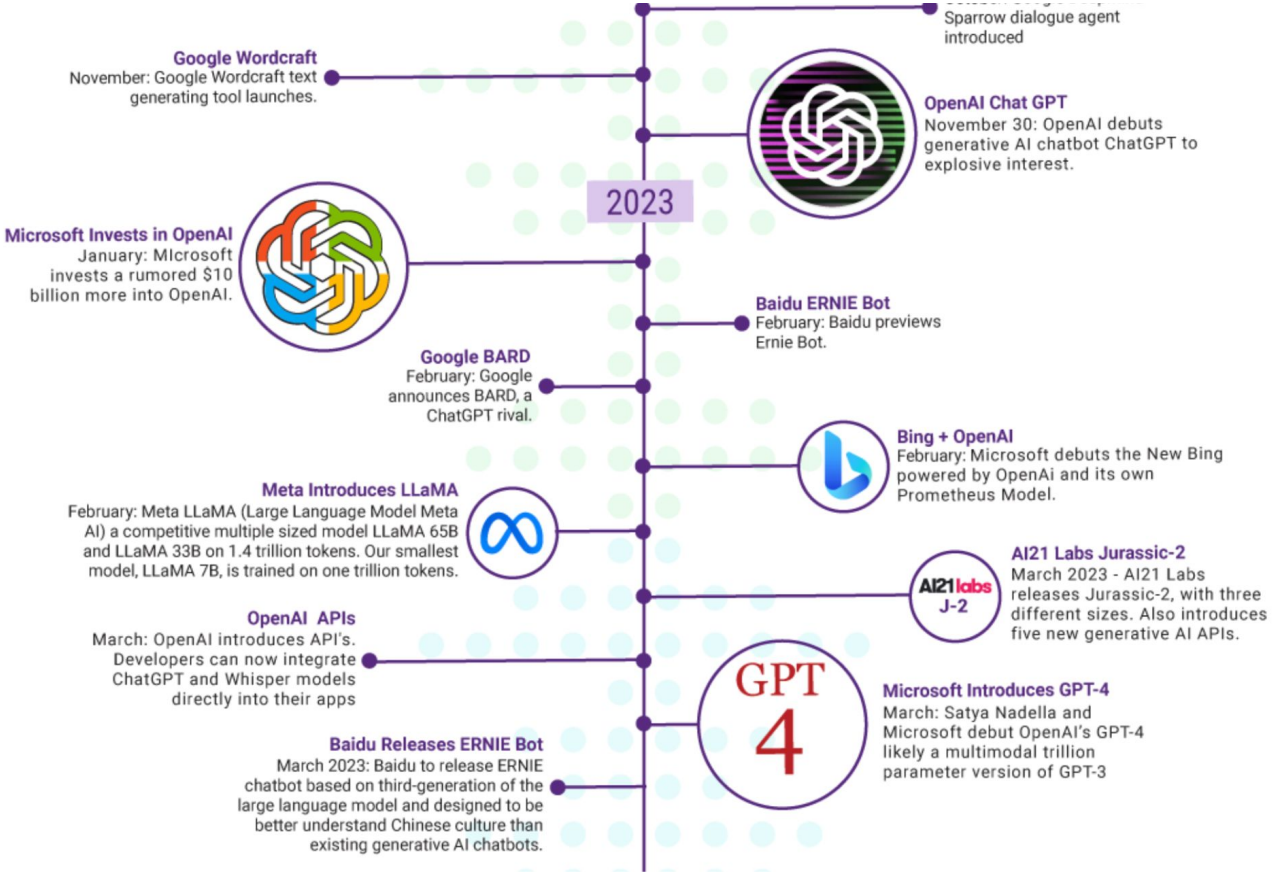
NVIDIA GPU
Nvidia introduces first Graphics Processing Unit (GPU), the Geforce 256.

2000

IBM Model 1

https://voicebot.ai/large-language-models-history-timeline/

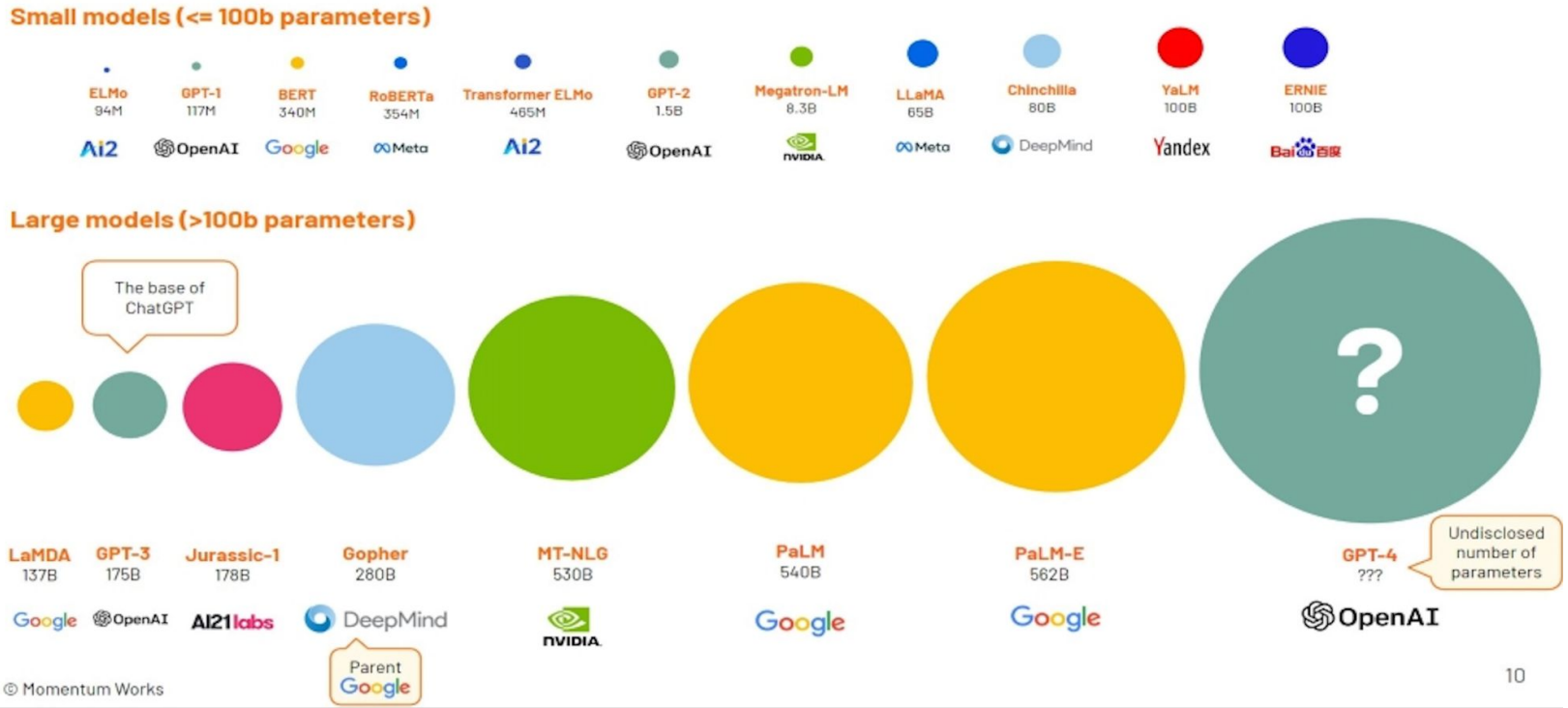
Large Language Models - LLM



https://voicebot.ai/large-language-models-history-timeline/

Size

Large Language Models are becoming very large indeed



https://thelowdown.momentum.asia/the-emergence-of-large-language-models-llms/

How Many Neurons Are in the Brain?

Published 4 Dec 2018 | Reviewed 4 Dec 2018 | Source BrainFacts/SfN

For half a century, neuroscientists thought the human brain contained 100 billion nerve cells. But when neuroscientist Suzana Herculano-Houzel devised a new way to count brain cells, she came up with a different number — 86 billion. Herculano-Houzel, now an associate professor of psychological science at Vanderbilt University, describes her research and explains how we "lost" 14 billion neurons overnight.

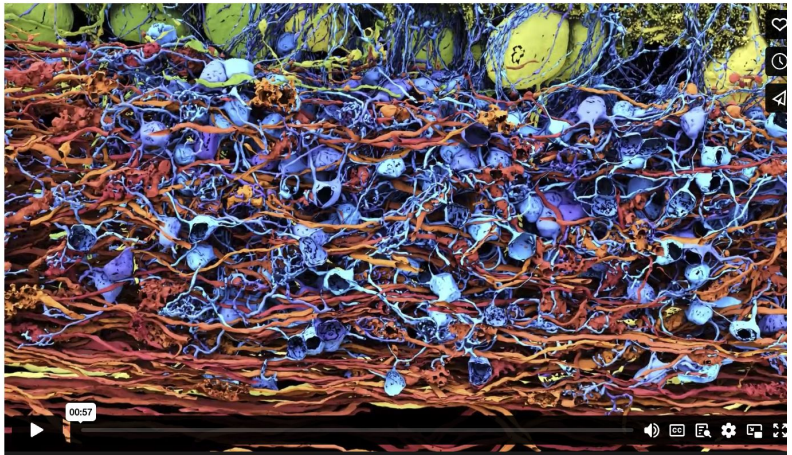
... she came up with a different number — 86 billion. Herculano-Houzel,

https://www.brainfacts.org/in-the-lab/meet-the-researcher/2018/how-many-neurons-are-in-the-brain-120418

A New Field of Neuroscience Aims to Map Connections in the Brain

Scientists working in connectomics are creating comprehensive maps of how neurons connect to one another

By CATHERINE CARUSO | January 19, 2023 | Research
11 min read



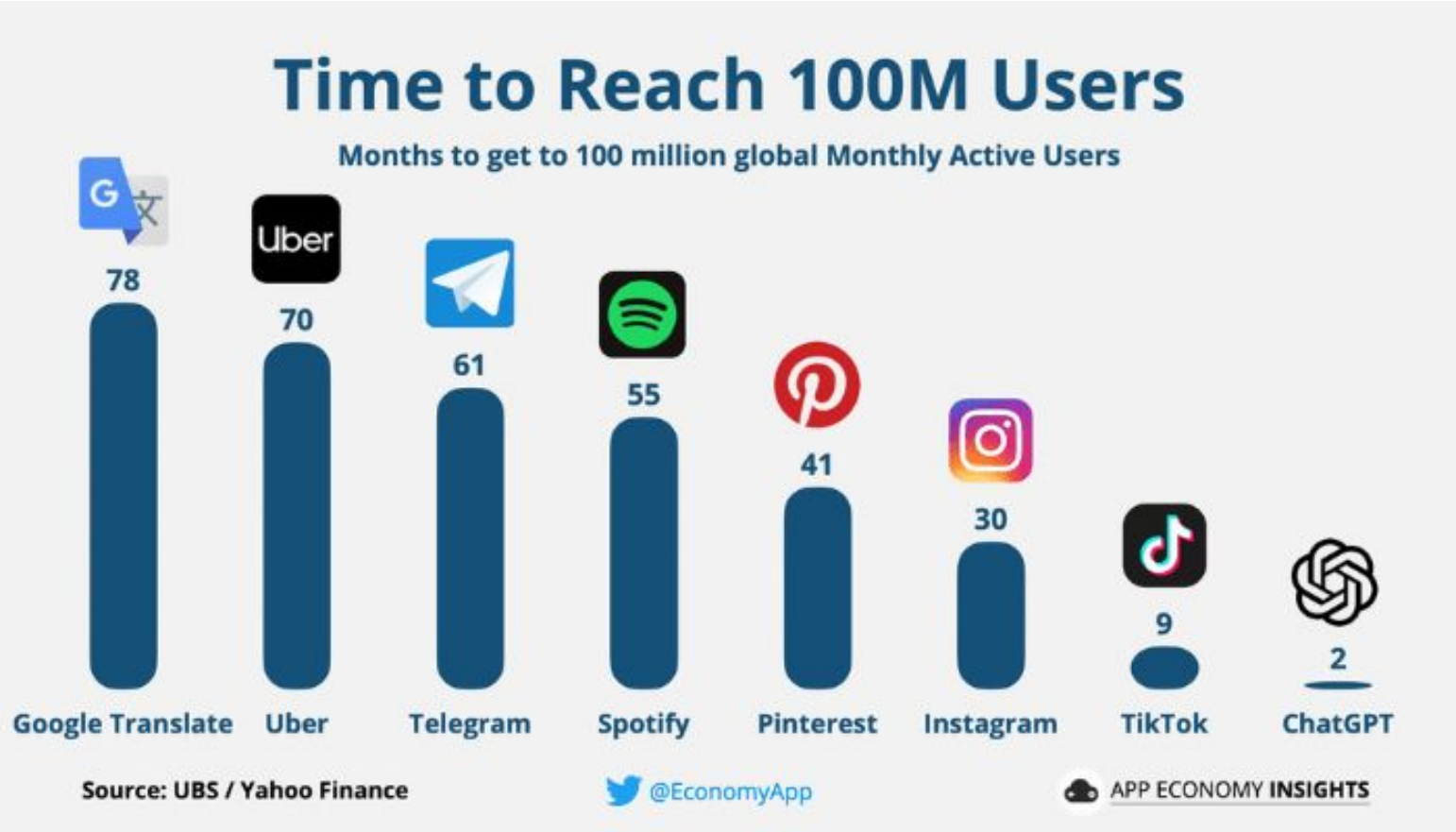
Wei-Chung Allen Lee is working in a new field of neuroscience called connectomics that aims to comprehensively map connections between neurons. Video: Catherine Caruso, Stephanie Dutchen, and Tyler Sloan

Many of us have seen microscopic images of neurons in the brain — each neuron appearing as a glowing cell in a vast sea of blackness. This image is misleading: Neurons don't exist in isolation. In the human brain, some 86 billio neurons form 100 trillion connections to each other — numbers that, ironically, are far too large for the human brai to fathom.

In the human brain, some 86 billion neurons form 100 trillion connections to each other — numbers that, ironically, are far too large for the human brain to fathom.

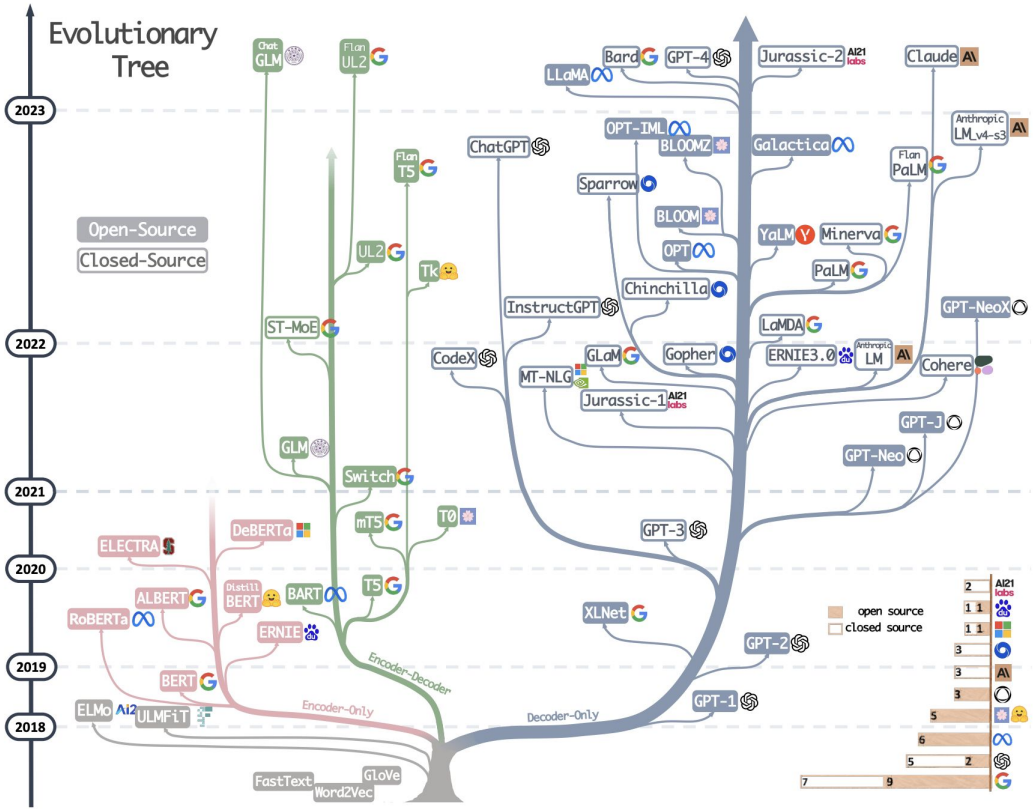
https://hms.harvard.edu/news/new-field-neuroscience-aims-map-connections-brain

Adoption



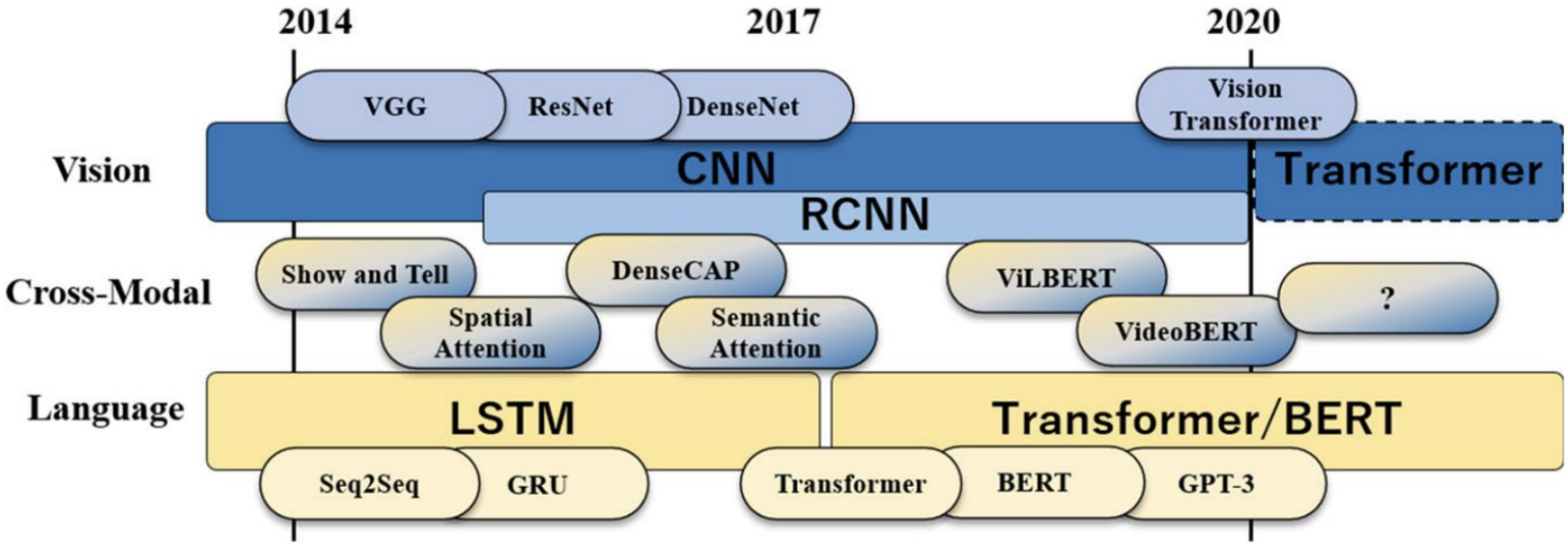
https://businessday.ng/technology/article/chatgpt-is-fastest-app-to-hit-100m-users-in-history/

Evolution of LLM models



https://arxiv.org/abs/2304.13712

Vision, Language and Cross-Modal Domains



Shin, A., Ishii, M. & Narihira, T. Perspectives and Prospects on Transformer Architecture for Cross-Modal Tasks with Language and Vision. *Int J Comput Vis* 130, 435–454 (2022).

ViT - Vision Transformers

Centro de Informática UFPE — Universidade Federal de Pernambuco

Published as a conference paper at ICLR 2021

AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE

Alexey Dosovitskiy[*,†], **Lucas Beyer**[*], **Alexander Kolesnikov**[*], **Dirk Weissenborn**[*], **Xiaohua Zhai**[*], **Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby**[*,†]

[*]equal technical contribution, [†]equal advising

Google Research, Brain Team

{adosovitskiy, neilhoulsby}@google.com

ABSTRACT

While the Transformer architecture has become the de-facto standard for natural language processing tasks, its applications to computer vision remain limited. In vision, attention is either applied in conjunction with convolutional networks, or used to replace certain components of convolutional networks while keeping their overall structure in place. We show that this reliance on CNNs is not necessary and a pure transformer applied directly to sequences of image patches can perform very well on image classification tasks. When pre-trained on large amounts of data and transferred to multiple mid-sized or small image recognition benchmarks (ImageNet, CIFAR-100, VTAB, etc.), Vision Transformer (ViT) attains excellent results compared to state-of-the-art convolutional networks while requiring substantially fewer computational resources to train.[1]

https://arxiv.org/abs/2010.11929

cin.ufpe.br

ViT - Vision Transformers

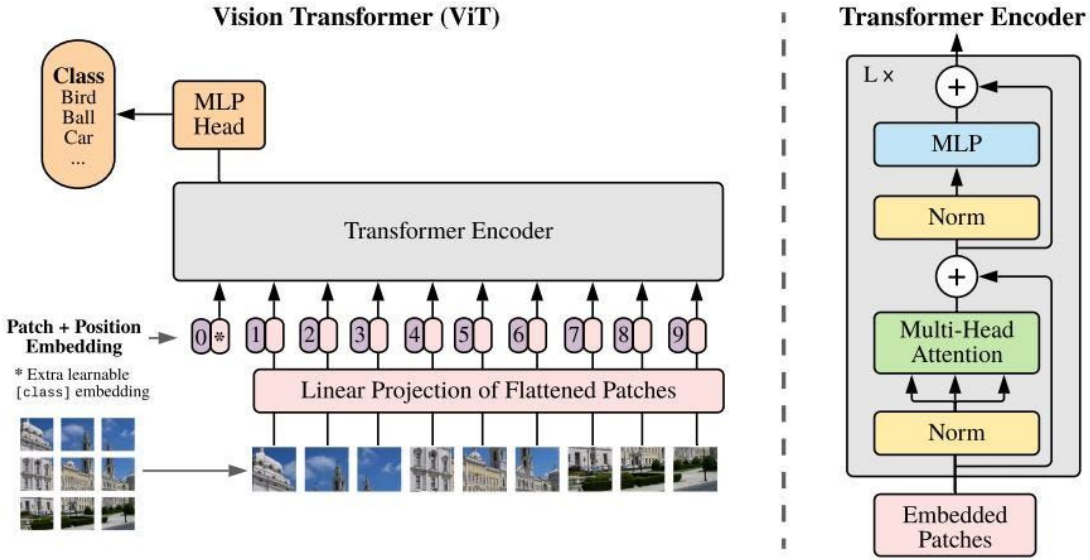


Figure 1: Model overview. We split an image into fixed-size patches, linearly embed each of them, add position embeddings, and feed the resulting sequence of vectors to a standard Transformer encoder. In order to perform classification, we use the standard approach of adding an extra learnable "classification token" to the sequence. The illustration of the Transformer encoder was inspired by Vaswani et al. (2017).

https://arxiv.org/abs/2010.11929

ViT - Vision Transformers

Datasets

	# of Images	# of Classes
ImageNet (Small)	1.3 Million	1 Thousand
ImageNet-21K (Medium)	14 Million	21 Thousand
JFT (Big)	300 Million	18 Thousand

https://github.com/wangshusen/DeepLearning/blob/master/Slides/10_ViT.pdf

Image Classification Accuracies

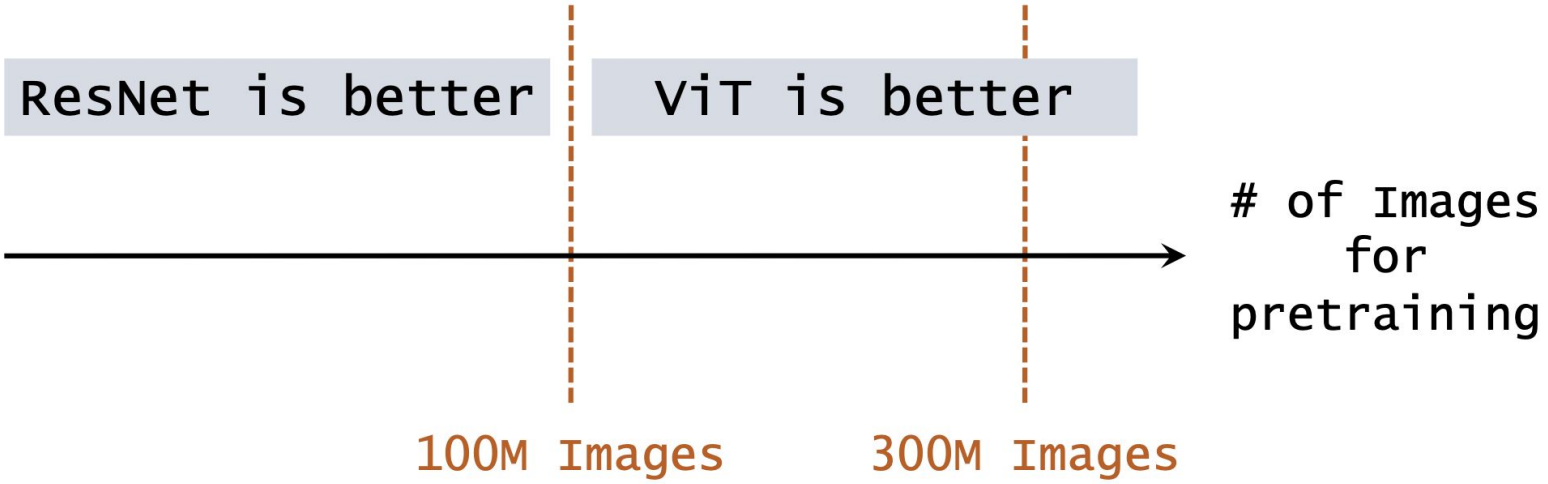
- Pretrain the model on Dataset A, fine-tune the model on Dataset B, and evaluate the model on Dataset B.

- Pretrained on ImageNet (small), ViT is slightly worse than ResNet.
- Pretrained on ImageNet-21K (medium), ViT is comparable to ResNet.
- Pretrained on JFT (large), ViT is slightly better than ResNet.

https://github.com/wangshusen/DeepLearning/blob/master/Slides/10_ViT.pdf

ViT - Vision Transformers

Image Classification Accuracies



https://github.com/wangshusen/DeepLearning/blob/master/Slides/10_ViT.pdf

ViT - Vision Transformers

BLOG ›

Scaling vision transformers to 22 billion parameters

FRIDAY, MARCH 31, 2023

Posted by Piotr Padlewski and Josip Djolonga, Software Engineers, Google Research

Large Language Models (LLMs) like PaLM or GPT-3 showed that scaling transformers to hundreds of billions of parameters improves performance and unlocks emergent abilities. The biggest dense models for image understanding, however, have reached only 4 billion parameters, despite research indicating that promising multimodal models like PaLI continue to benefit from scaling vision models alongside their language counterparts. Motivated by this, and the results from scaling LLMs, we decided to undertake the next step in the journey of scaling the Vision Transformer.

In "Scaling Vision Transformers to 22 Billion Parameters", we introduce the biggest dense vision model, ViT-22B. It is 5.5x larger than the previous largest vision backbone, ViT-e, which has 4 billion parameters. To enable this scaling, ViT-22B incorporates ideas from scaling text models like PaLM, with improvements to both training stability (using QK normalization) and training efficiency (with a novel approach called asynchronous parallel linear operations). As a result of its modified architecture, efficient sharding recipe, and bespoke implementation, it was able to be trained on Cloud TPUs with a high hardware utilization[1]. ViT-22B advances the state of the art on many vision tasks using frozen representations, or with full fine-tuning. Further, the model has also been successfully used in PaLM-e, which showed that a large model combining ViT-22B with a language model can significantly advance the state of the art in robotics tasks.

https://ai.googleblog.com/2023/03/scaling-vision-transformers-to-22.html

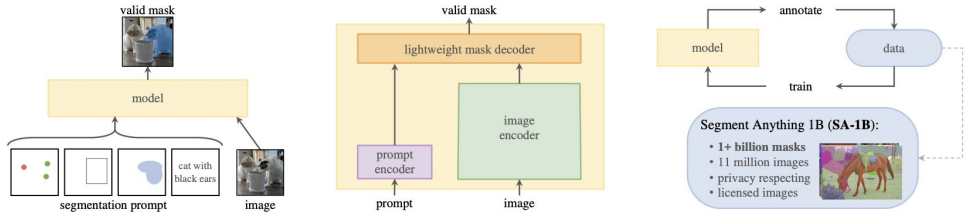
SAM - Segment Anything

Centro de Informática UFPE — Universidade Federal de Pernambuco

Segment Anything

Alexander Kirillov[1,2,4] Eric Mintun[2] Nikhila Ravi[1,2] Hanzi Mao[2] Chloe Rolland[3] Laura Gustafson[3]
Tete Xiao[3] Spencer Whitehead Alexander C. Berg Wan-Yen Lo Piotr Dollár[4] Ross Girshick[4]
[1]project lead [2]joint first author [3]equal contribution [4]directional lead
Meta AI Research, FAIR

2304.02643v1 [cs.CV] 5 Apr 2023



(a) **Task**: promptable segmentation (b) **Model**: Segment Anything Model (**SAM**) (c) **Data**: data engine (top) & dataset (bottom)

Figure 1: We aim to build a foundation model for segmentation by introducing three interconnected components: a promptable segmentation *task*, a segmentation *model* (SAM) that powers data annotation and enables zero-shot transfer to a range of tasks via prompt engineering, and a *data* engine for collecting SA-1B, our dataset of over 1 billion masks.

Abstract

We introduce the Segment Anything (SA) project: a new task, model, and dataset for image segmentation. Using our efficient model in a data collection loop, we built the largest segmentation dataset to date (by far), with over 1 ***billion*** *masks on 11M licensed and privacy respecting images. The model is designed and trained to be promptable, so it can transfer zero-shot to new image distributions and tasks. We evaluate its capabilities on numerous tasks and find that*

matching in some cases) fine-tuned models [10, 21]. Empirical trends show this behavior improving with model scale, dataset size, and total training compute [56, 10, 21, 51].

Foundation models have also been explored in computer vision, albeit to a lesser extent. Perhaps the most prominent illustration aligns paired text and images from the web. For example, CLIP [82] and ALIGN [55] use contrastive learning to train text and image encoders that align the two modalities. Once trained, engineered text prompts enable zero-shot generalization to novel visual concepts and data

- The image encoder has **632M** parameters.
- The prompt encoder and mask decoder have **4M** parameters.

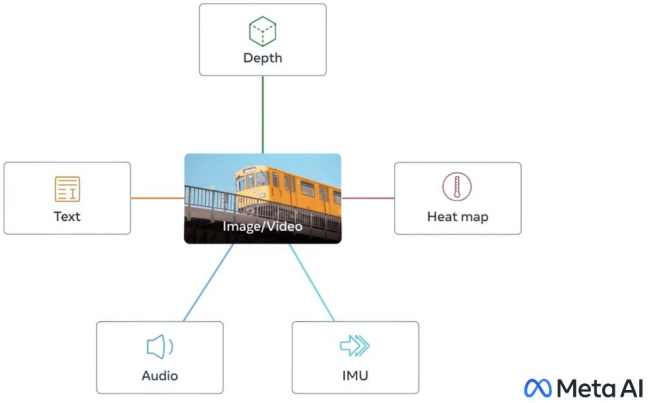
https://segment-anything.com/

ImageBind: Holistic AI Learning across six modalities

Computer Vision

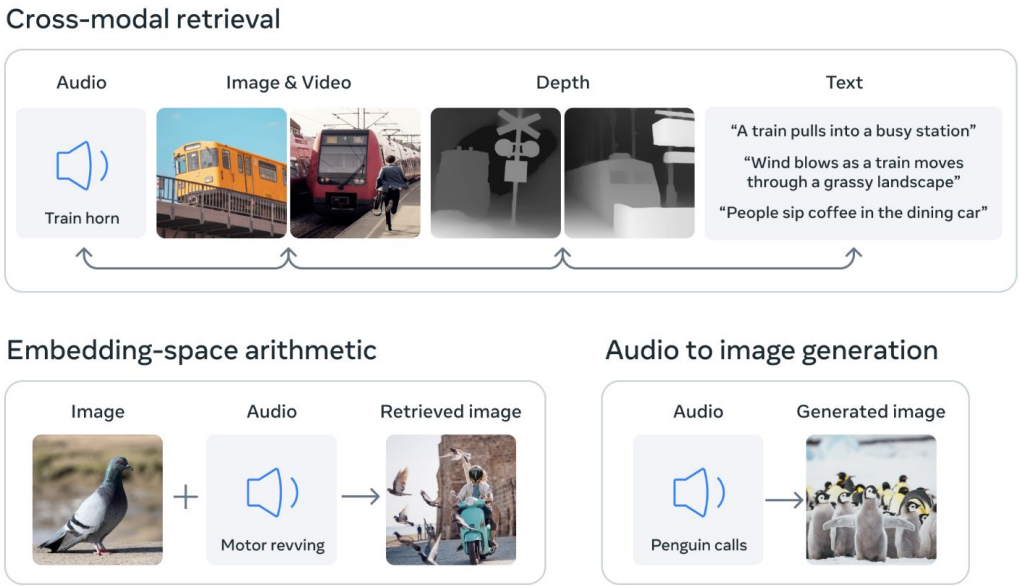
ImageBind: Holistic AI learning across six modalities

May 9, 2023



https://ai.facebook.com/blog/imagebind-six-modalities-binding-ai/

ImageBind: Holistic AI Learning across six modalities



By aligning six modalities' embedding into a common space, ImageBind enables cross-modal retrieval of different types of content that aren't observed together, the addition of embeddings from different modalities to naturally compose their semantics, and audio-to-image generation by using our audio embeddings with a pretrained DALLE-2 decoder to work with CLIP text embeddings.

https://ai.facebook.com/blog/imagebind-six-modalities-binding-ai/

Foundation Models

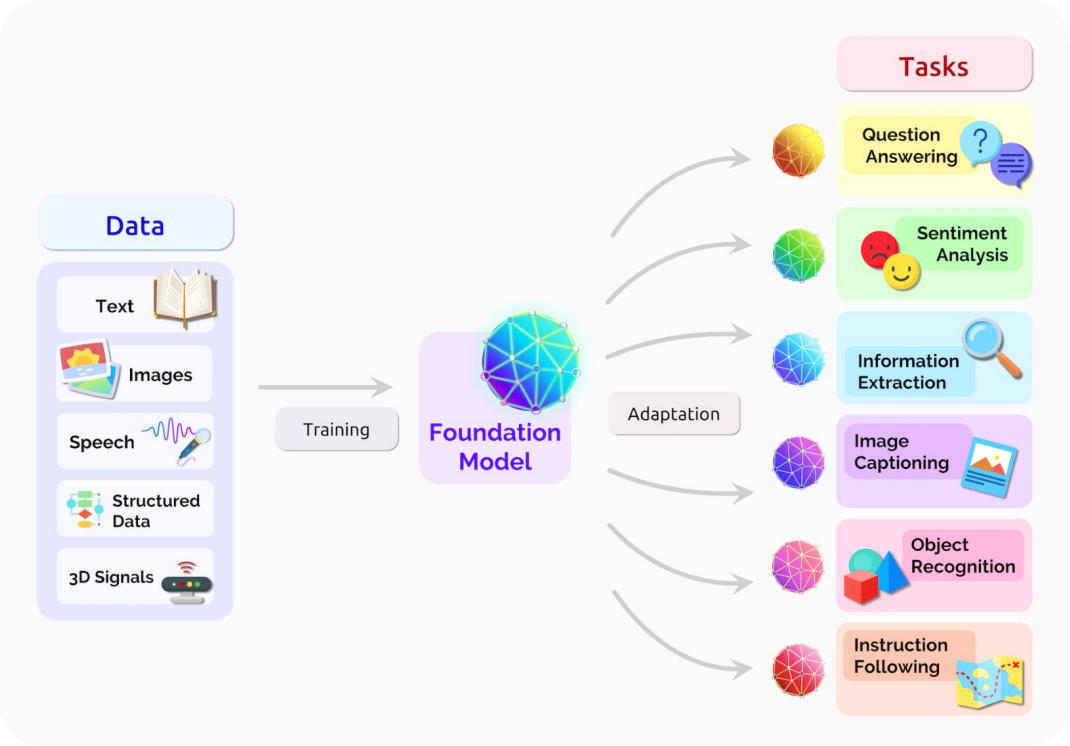


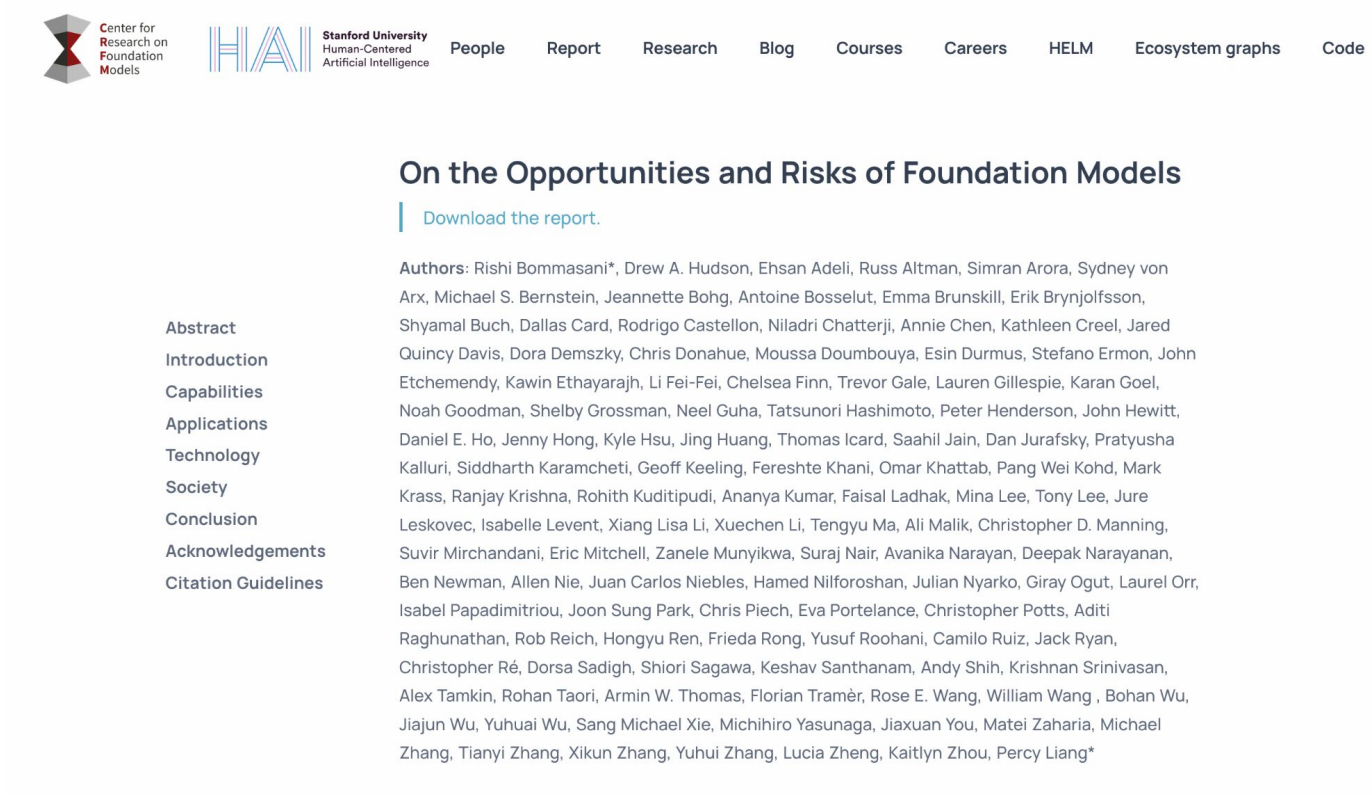
Fig. 2. A foundation model can centralize the information from all the data from various modalities. This one model can then be adapted to a wide range of downstream tasks.

https://crfm.stanford.edu/report.html

A **foundation model** (also called **base model**)[1] is a large machine learning (ML) model trained on a vast quantity of data at scale (often by self-supervised learning or semi-supervised learning)[2] such that it can be adapted to a wide range of downstream tasks.

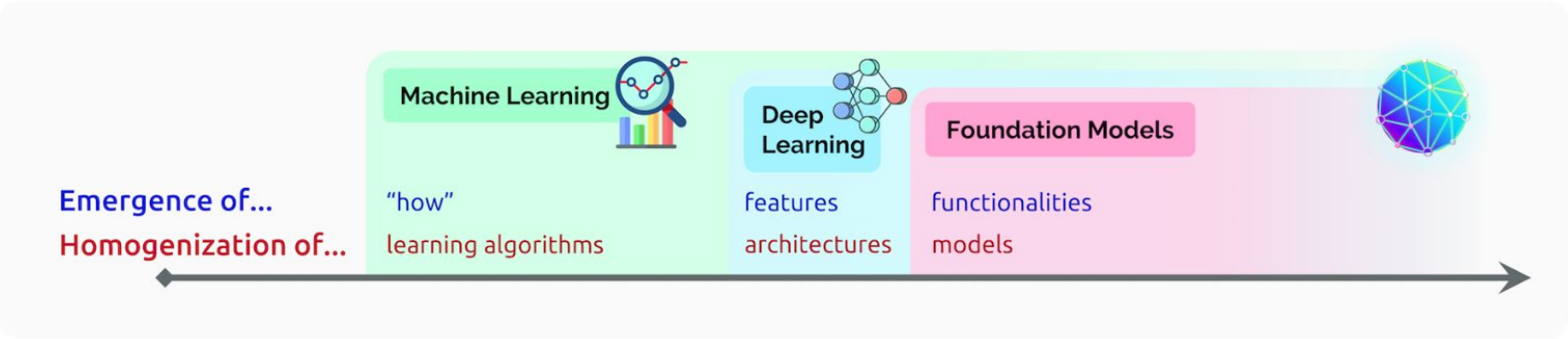
https://en.wikipedia.org/wiki/Foundation_models

Foundation Models



https://crfm.stanford.edu/report.html

Foundation Models



Emergence means that the behavior of a system is implicitly induced rather than explicitly constructed; it is both the source of scientific excitement and anxiety about unanticipated consequences.

Homogenization indicates the consolidation of methodologies for building machine learning systems across a wide range of applications; it provides strong leverage towards many tasks but also creates single points of failure.

https://crfm.stanford.edu/report.html

Foundation Models

https://crfm.stanford.edu/report.html

Generative Model

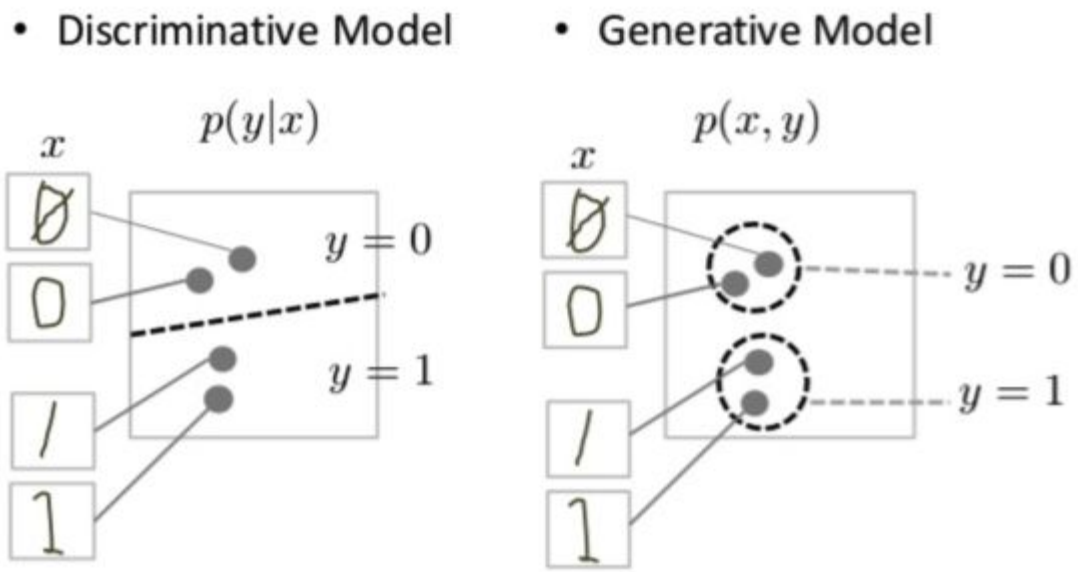
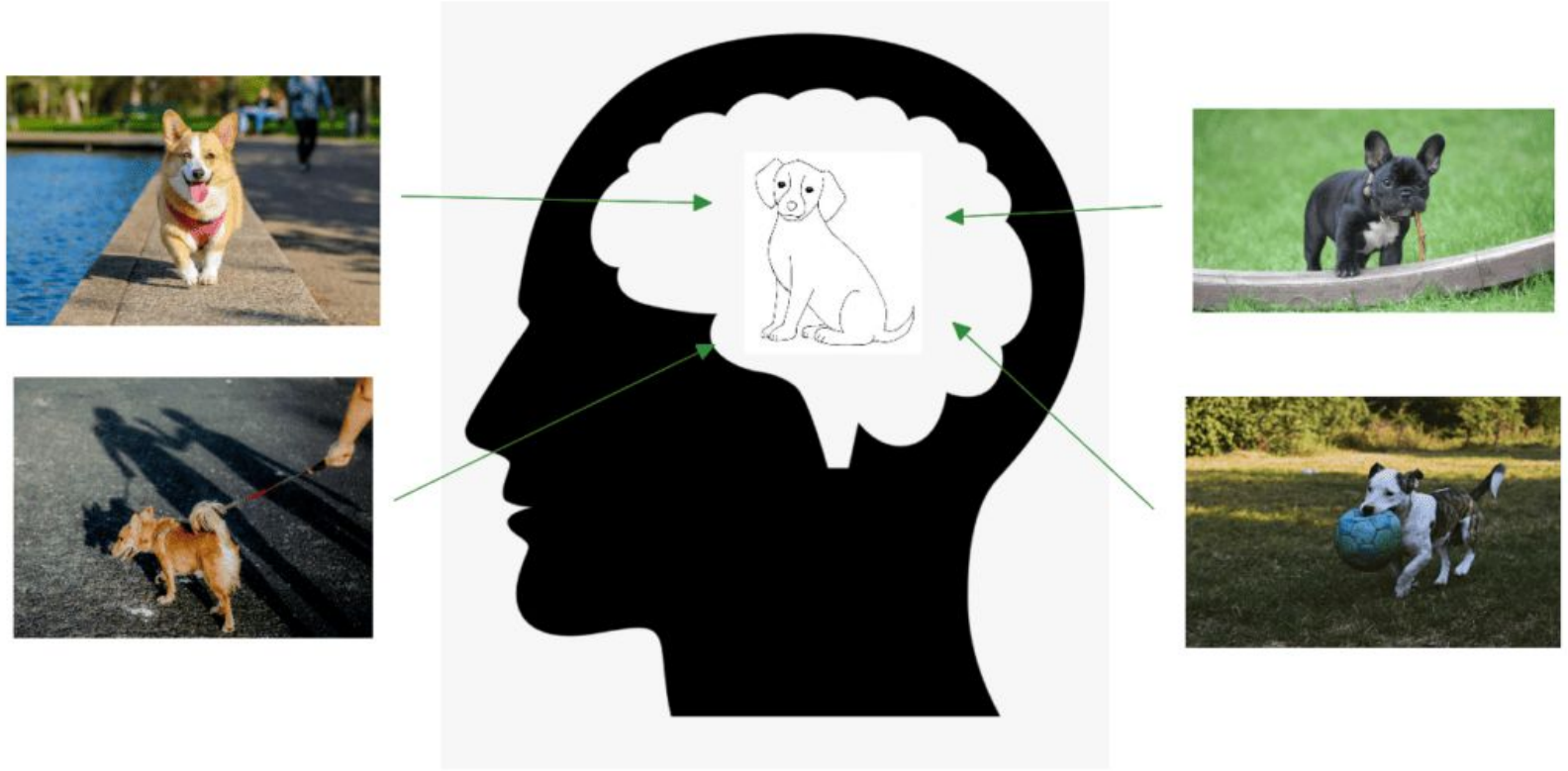


Figure 1: Discriminative and generative models of handwritten digits.

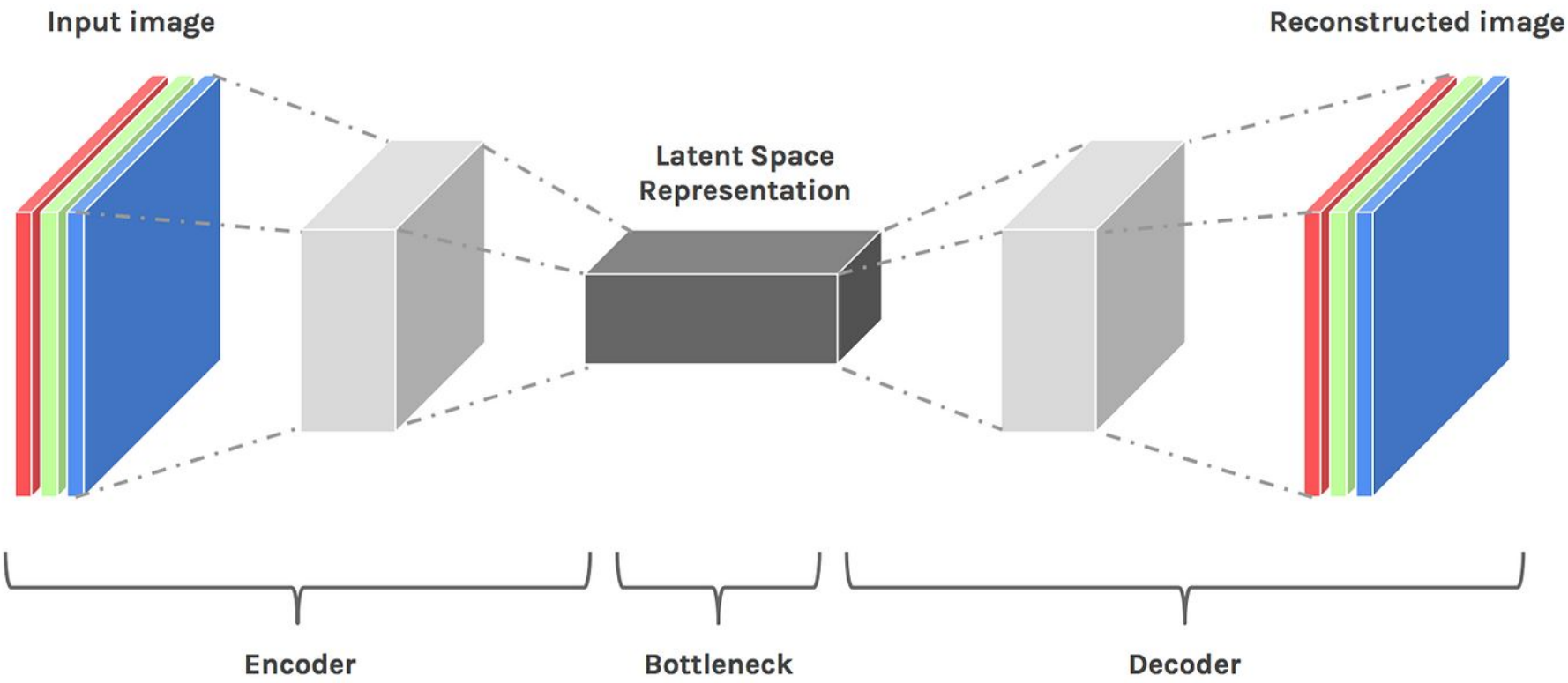
https://developers.google.com/machine-learning/gan/generative

Latent Space



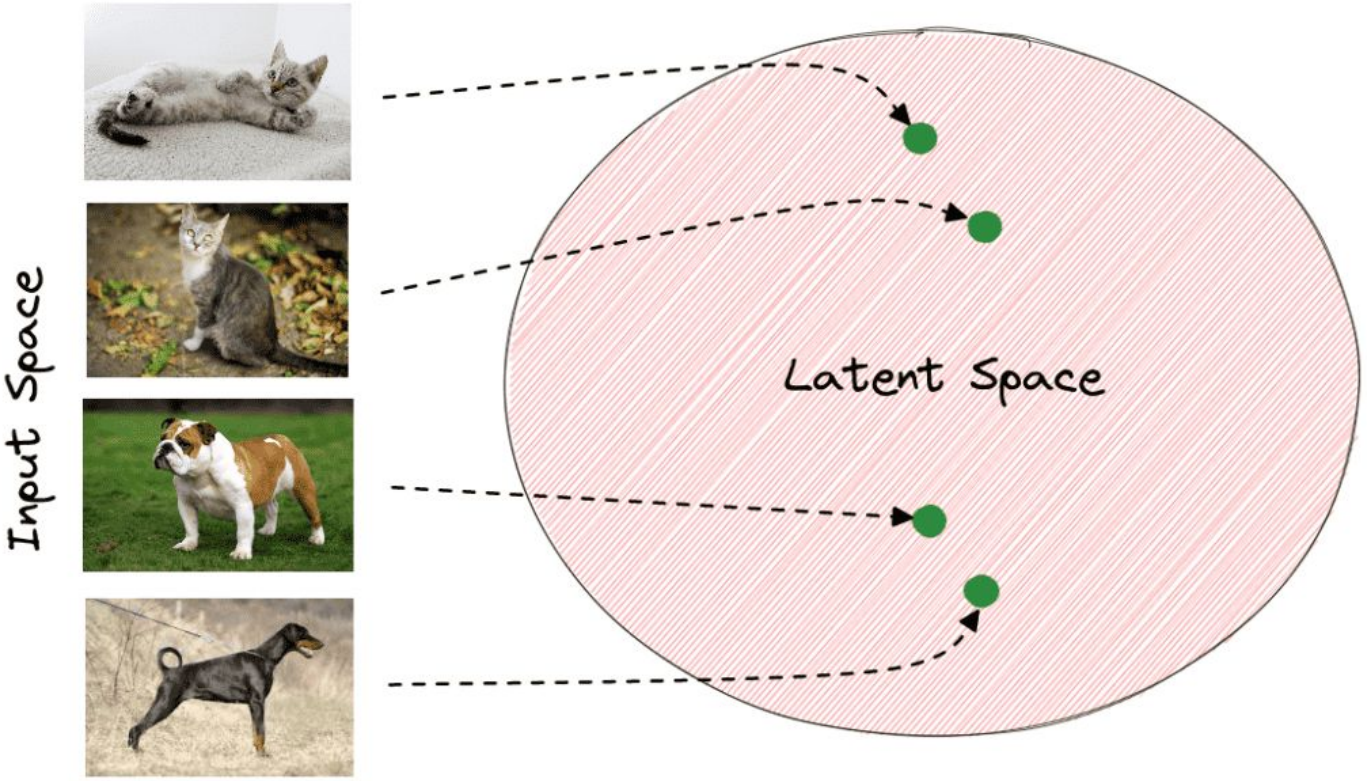
https://www.baeldung.com/cs/dl-latent-space

Latent Space



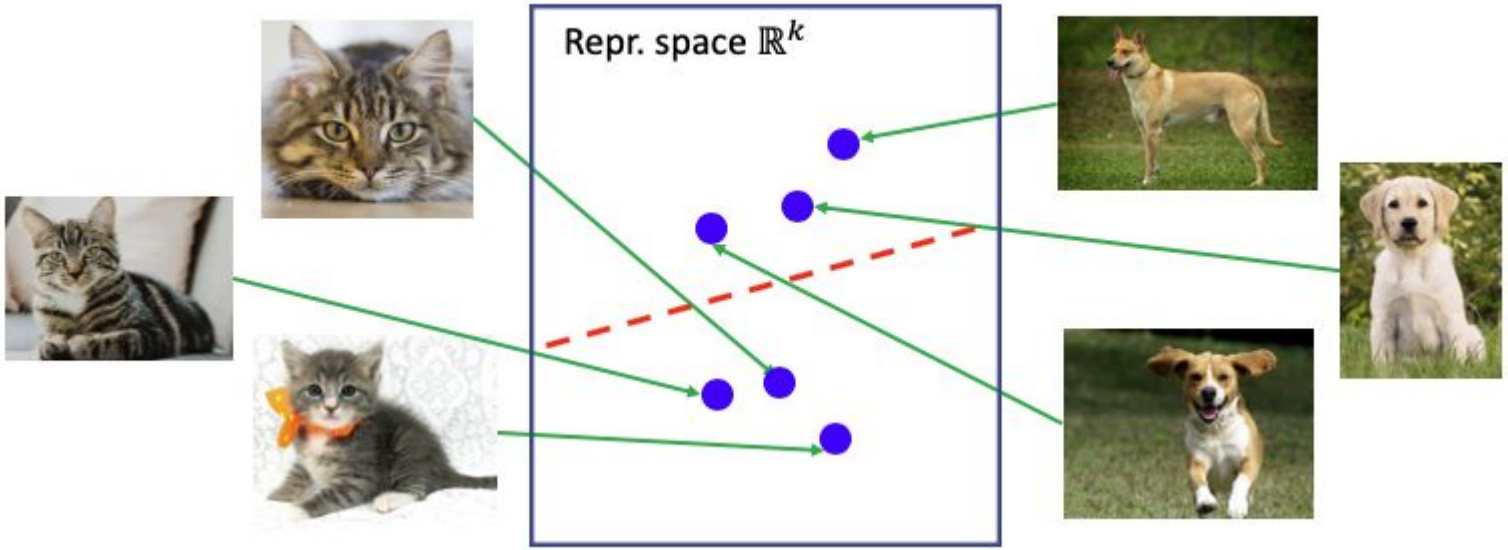
https://towardsdatascience.com/understanding-latent-space-in-machine-learning-de5a7c687d8d

Latent Space



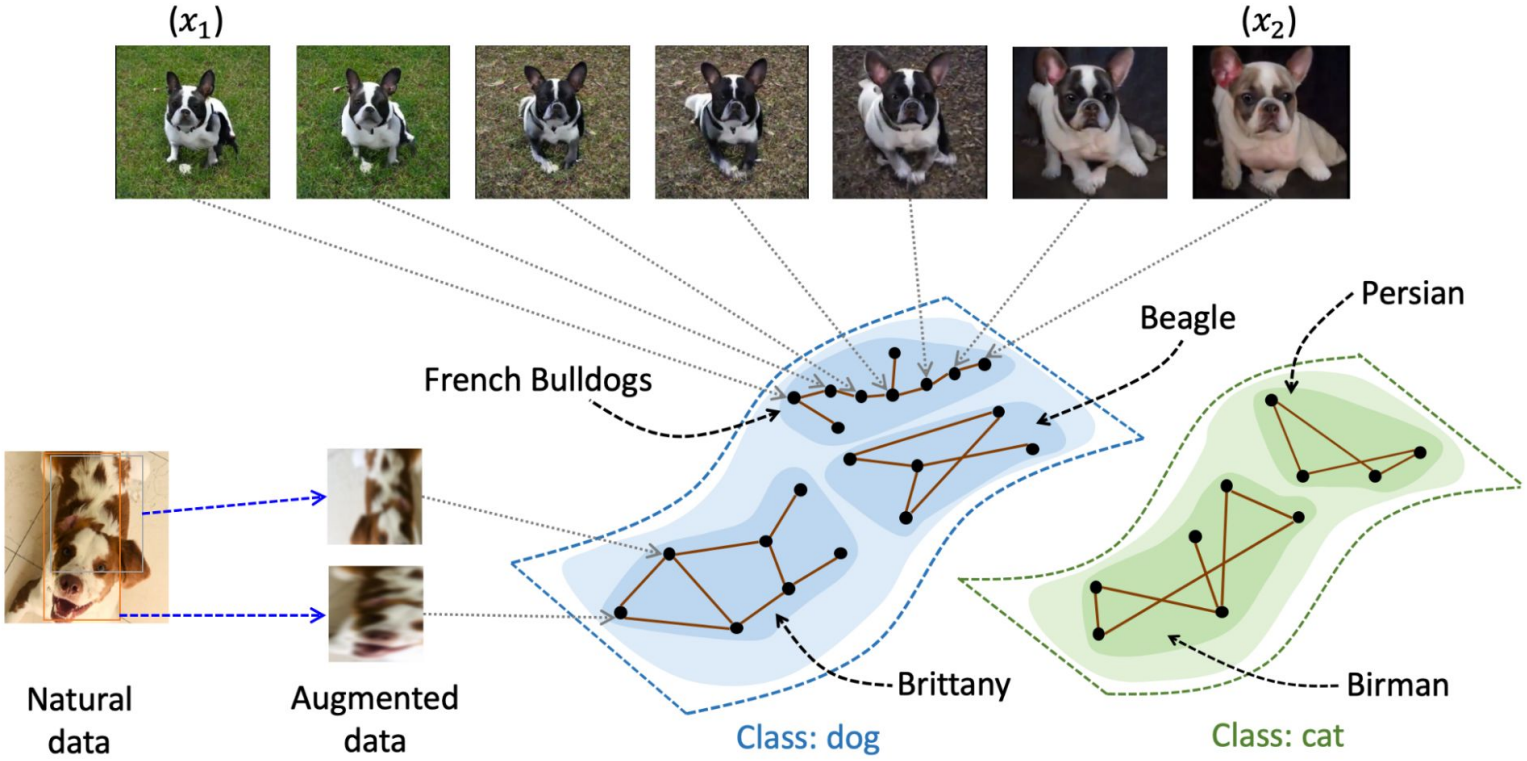
https://www.baeldung.com/cs/dl-latent-space

Latent Space



https://ai.stanford.edu/blog/understanding-contrastive-learning/

Latent Space



https://ai.stanford.edu/blog/understanding-contrastive-learning/

GAN - Generative Adversarial Network

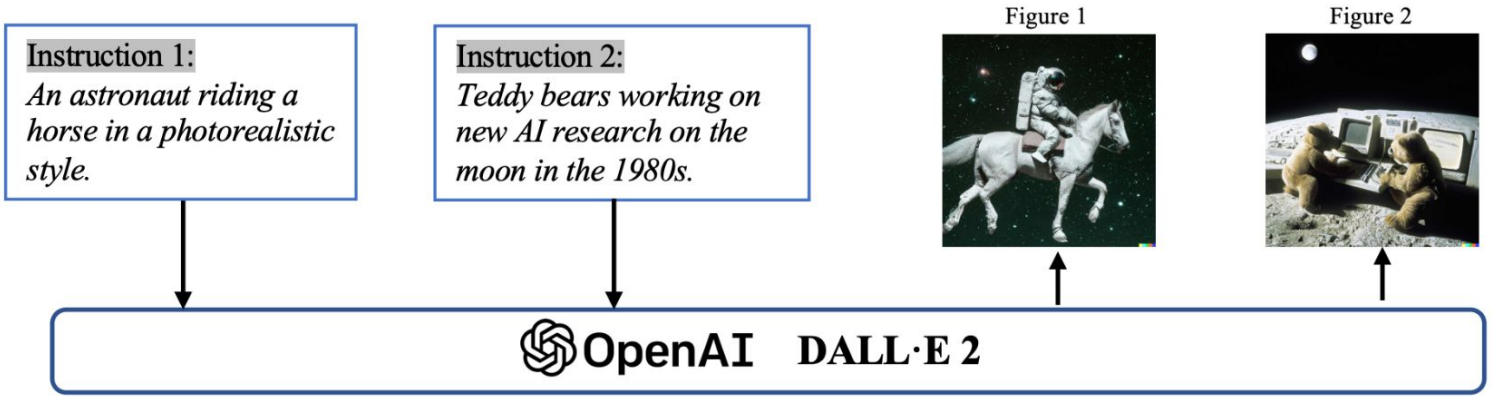


Fig. 1. Examples of AIGC in image generation. Text instructions are given to OpenAI DALL-E-2 model, and it generates two images according to the instructions.

https://arxiv.org/abs/2303.04226

GAN - Generative Adversarial Network

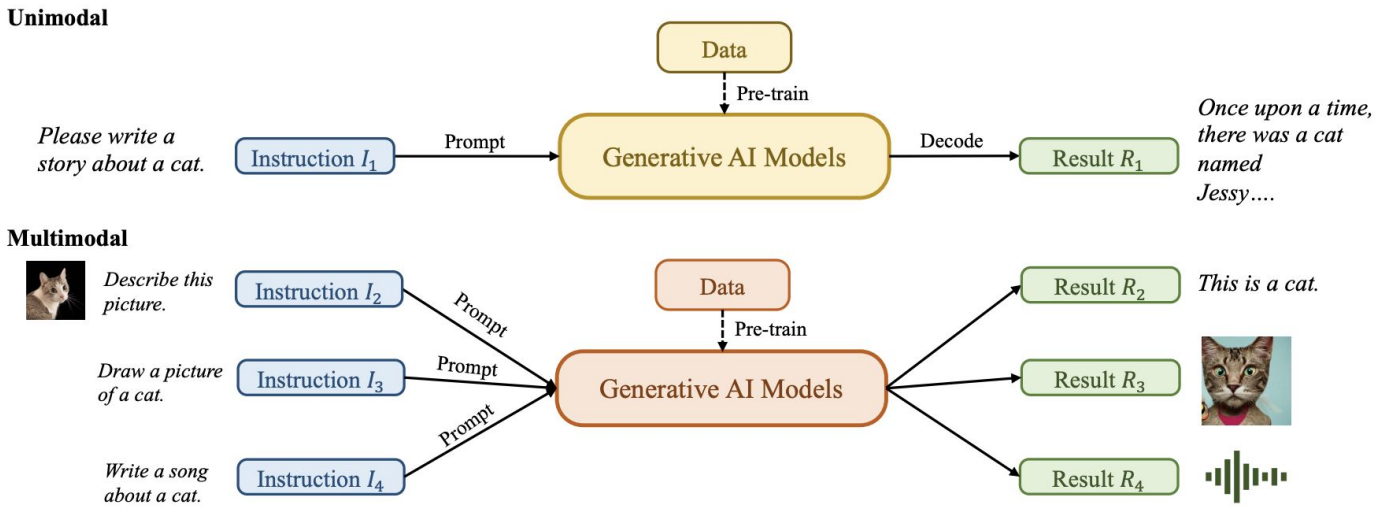


Fig. 2. Overview of AIGC. Generally, GAI models can be categorized into two types: unimodal models and multimodal models. Unimodal models receive instructions from the same modality as the generated content modality, whereas multimodal models accept cross-modal instructions and produce results of different modalities.

https://arxiv.org/abs/2303.04226

GAN - Generative Adversarial Network

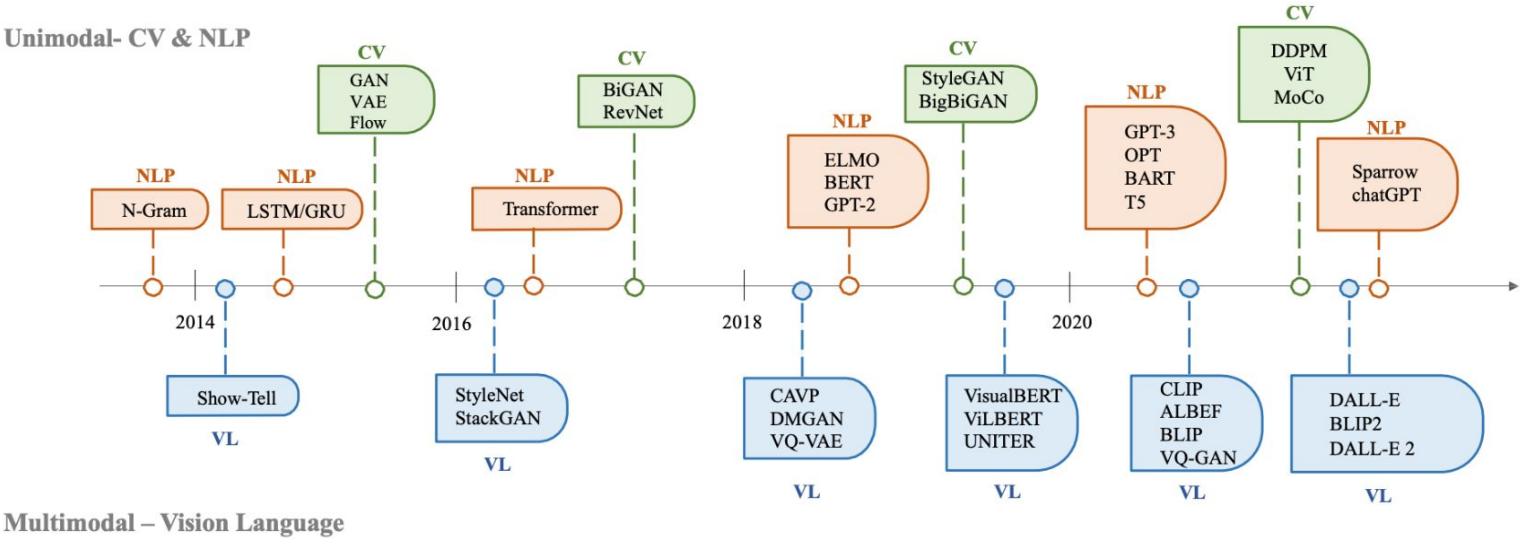


Fig. 3. The history of Generative AI in CV, NLP and VL.

https://arxiv.org/abs/2303.04226

GAN - Generative Adversarial Network

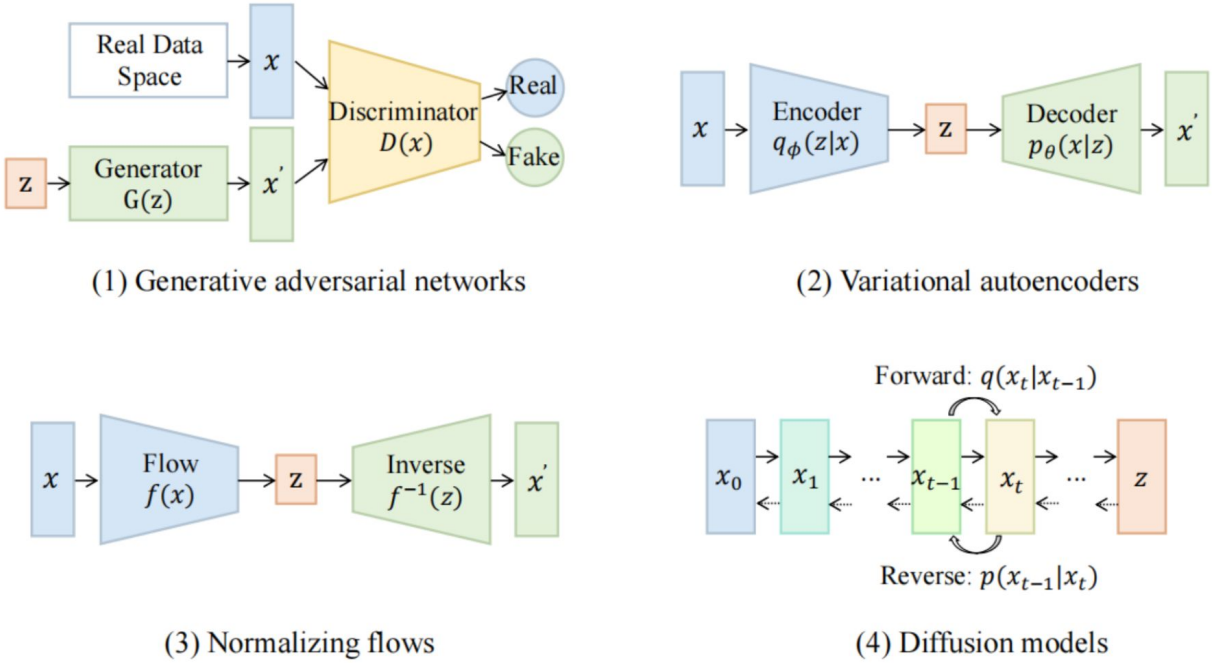


Fig. 7. Categories of vision generative models.

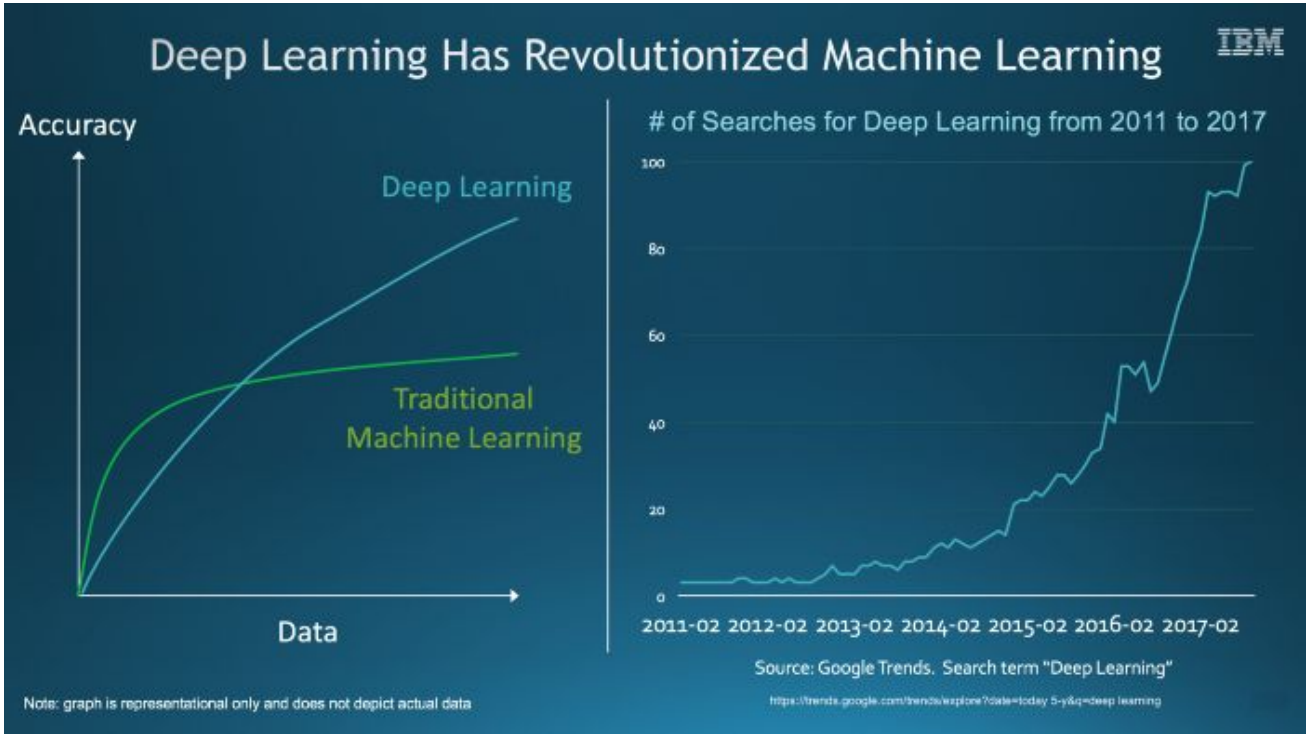
https://arxiv.org/abs/2303.04226

GAN - Generative Adversarial Network

https://www.nytimes.com/2022/09/16/learning/are-ai-generated-pictures-art.html

https://www.smithsonianmag.com/smart-news/how-a-real-photo-of-a-flamingo-won-an-artificial-intelligence-photography-competition-180984558.

Aprendizagem Profunda (Deep Learning)



https://medium.com/@sanchittanwar75/introduction-to-machine-learning-and-deep-learning-bd25b792e488

IA Revolution

Each productivity revolution kills old jobs, but unleashes new ones

Through each of the past 4 industrial revolutions, **the rapid advent of technology disrupts not only jobs but whole industries**. Nonetheless each time **jobs (and industries) previously not thought of were created**, and **humanity benefited from the leap of productivity**.

	Industrial revolution (1760-1840)	The 2nd industrial revolution (1870 - 1914)	Digital revolution (1969 - 2000)	Industry 4.0 (2008 - 2020)	The 5th revolution? (present)
					
Powered by	**Steam**	**Electrification**	**Automation**	**Interconnectivity**	**Artificial Intelligence**
What happened	Subsistence farmers lost out to factories; cottage textile artisans to mills; sail boats to steamships	Rapid scientific discovery, mass production, assembly line, telephone, and finally airplanes; emergence of corporates	Automation, digitisation, electronic devices etc. Manual, repetitive jobs lost out to IT jobs; emergence of MNCs	Digitalisation, IoT, AR and VR, cyber security, cloud, mobility, blockchain and machine learning; further automation of jobs	Everything that is described in this report (and more)

Equipped with better education and more information than our predecessors, we are more likely to **take charge of our own fate in the current wave**.

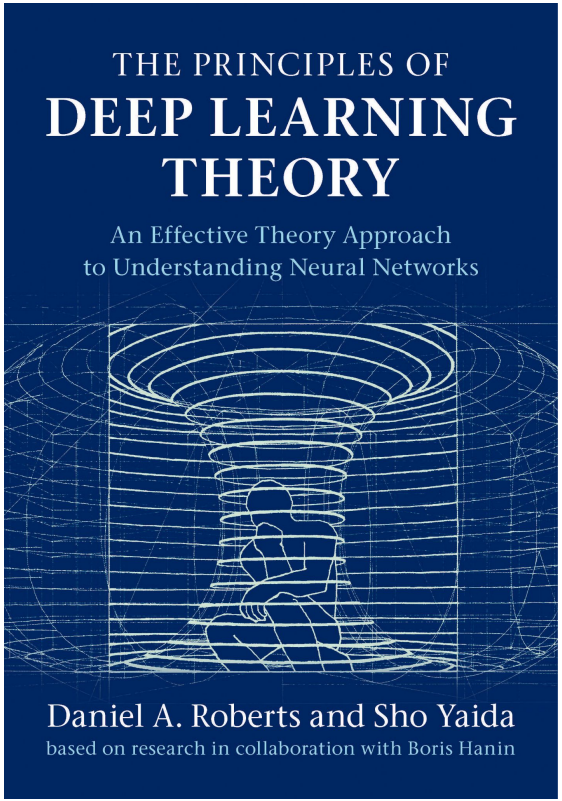
© Momentum Works

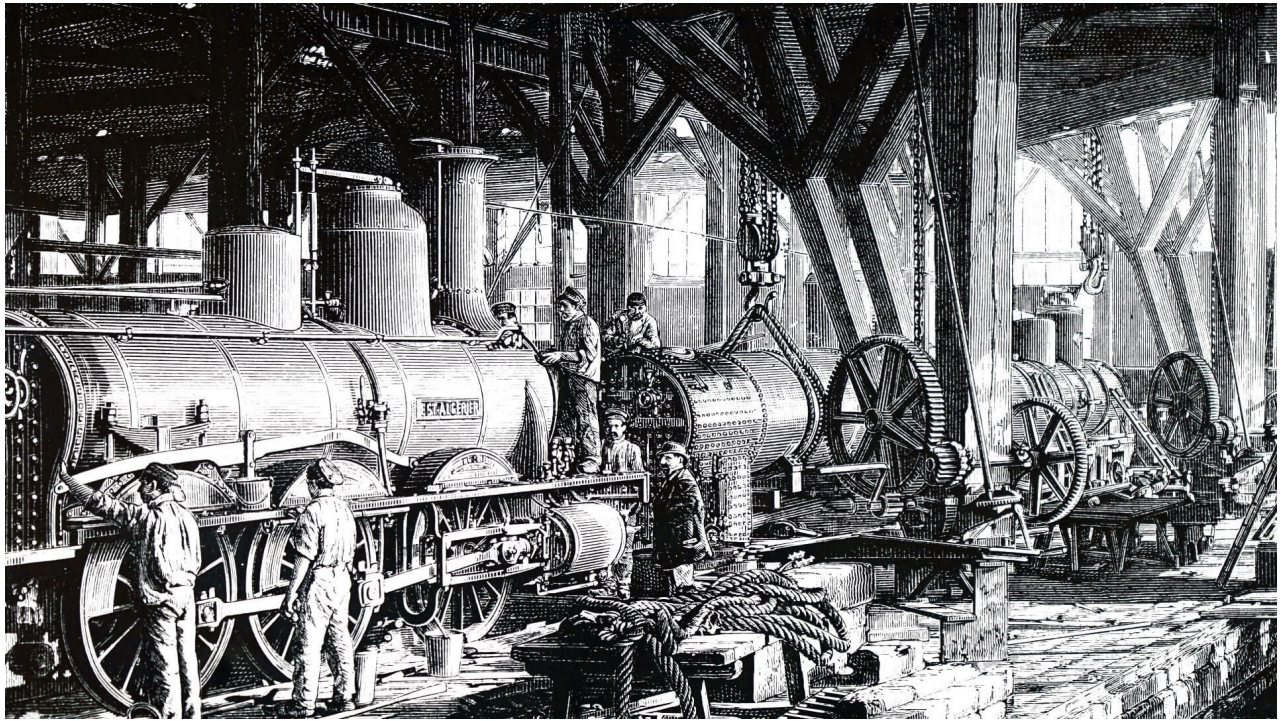
https://thelowdown.momentum.asia/the-emergence-of-large-language-models-llms/

Deep Learning Theory !?

Centro de Informática UFPE

Universidade Federal de Pernambuco





https://ai.meta.com/blog/advancing-ai-theory-with-a-first-principles-understanding-of-deep-neural-networks/

https://deeplearningtheory.com/

cin.ufpe.br

Noble Prize

The Nobel Prize in Physics 2024

Ill. Niklas Elmehed © Nobel Prize Outreach

John J. Hopfield

Prize share: 1/2

Ill. Niklas Elmehed © Nobel Prize Outreach

Geoffrey Hinton

Prize share: 1/2

The Nobel Prize in Physics 2024 was awarded jointly to John J. Hopfield and Geoffrey E. Hinton "for foundational discoveries and inventions that enable machine learning with artificial neural networks"

The Nobel Prize in Chemistry 2024

Ill. Niklas Elmehed © Nobel Prize Outreach

David Baker

Prize share: 1/2

Ill. Niklas Elmehed © Nobel Prize Outreach

Demis Hassabis

Prize share: 1/4

Ill. Niklas Elmehed © Nobel Prize Outreach

John Jumper

Prize share: 1/4

The Nobel Prize in Chemistry 2024 was divided, one half awarded to David Baker "for computational protein design", the other half jointly to Demis Hassabis and John Jumper "for protein structure prediction"

Future.....??

OpenAI

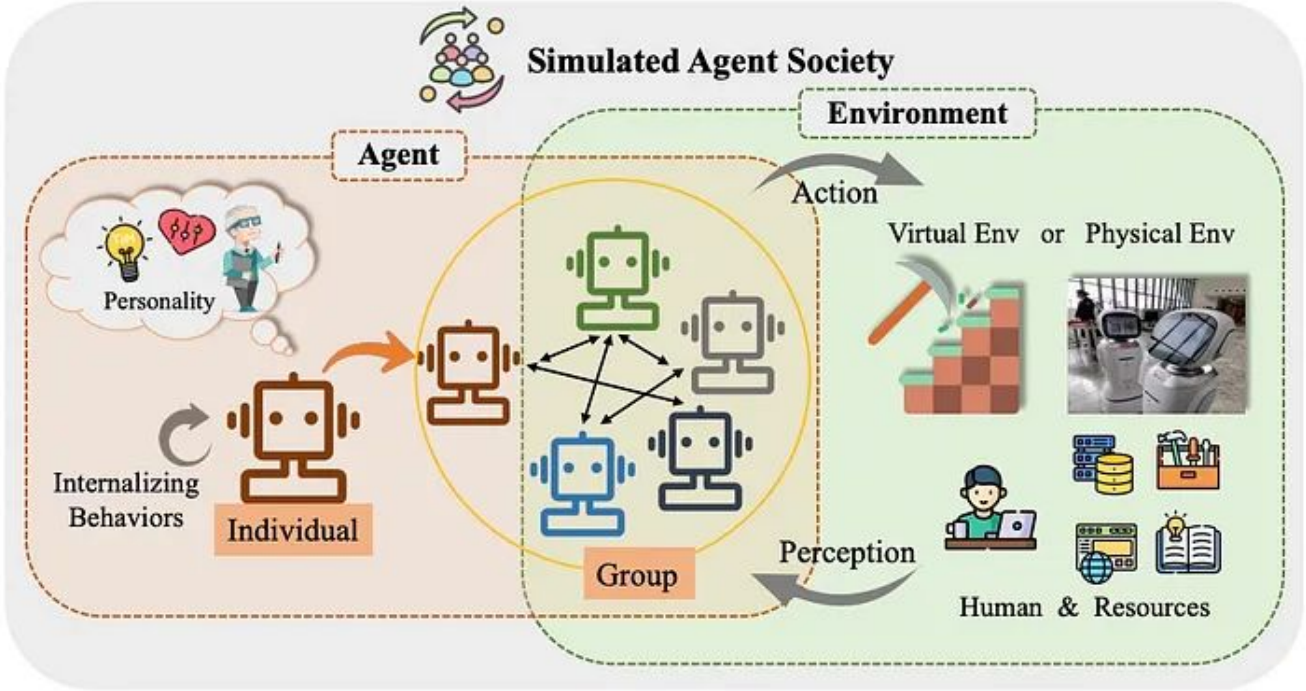
- Research
- Safety
- For Business
- For Developers
- ChatGPT
- Sora
- Stories
- Company
- News

Announcing The Stargate Project

https://openai.com/index/announcing-the-stargate-project/

AI Agents

Future.....??



https://medium.com/@henryhengluo/intro-of-ai-agent-ai-agent-projects-summary-52f4a364ab86

Robotics

Future.....??

https://www.1x.tech/discover/announcement-1x-unveils-neo-beta-a-humanoid-robot-for-the-home

Future.....??

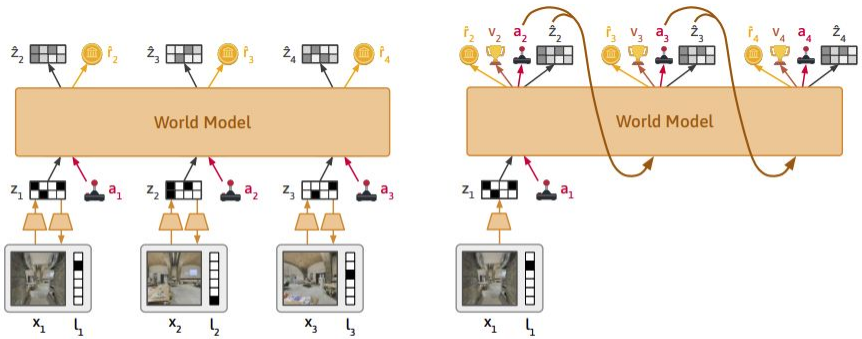
Autonomous Vehicles

https://www.wired.com/story/pentagon-inches-toward-letting-ai-control-weapons/

https://www.tesla.com/autopilot

Future.....??

World Model



(a) World Model Learning (b) Learning to Act by Latent Imagination

Figure 3. During world model learning, the model compresses observations of image frames and text to a latent representation. The model is trained to predict the next representation and reconstruct observations from the representation. During policy learning, imagined rollouts are sampled from the world model and the policy is trained to maximize imagined rewards.

https://arxiv.org/pdf/2308.01399

DIFFUSION MODELS ARE REAL-TIME GAME ENGINES

Dani Valevski* Google Research
Yaniv Leviathan* Google Research
Moab Arar* Tel Aviv University†
Shlomi Fruchter* Google DeepMind

ABSTRACT

We present *GameNGen*, the first game engine powered entirely by a neural model that enables real-time interaction with a complex environment over long trajectories at high quality. GameNGen can interactively simulate the classic game DOOM at over 20 frames per second on a single TPU. Next frame prediction achieves a PSNR of 29.4, comparable to lossy JPEG compression. Human raters are only slightly better than random chance at distinguishing short clips of the game from clips of the simulation. GameNGen is trained in two phases: (1) an RL-agent learns to play the game and the training sessions are recorded, and (2) a diffusion model is trained to produce the next frame, conditioned on the sequence of past frames and actions. Conditioning augmentations enable stable auto-regressive generation over long trajectories.

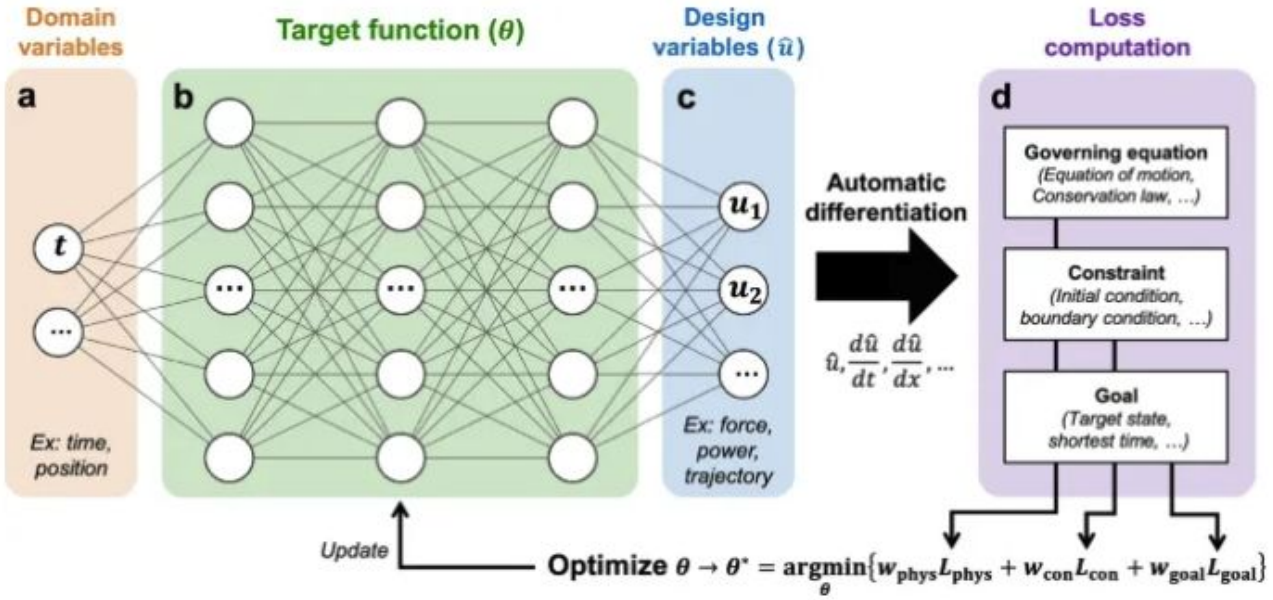
Figure 1: A human player is playing DOOM on **GameNGen** at 20 FPS. See `https://gamengen.github.io` for demo videos.

https://arxiv.org/pdf/2408.14837

cin.ufpe.br

Future.....??

Fundamental Concepts of Physics-Informed Neural Networks (PINNs)



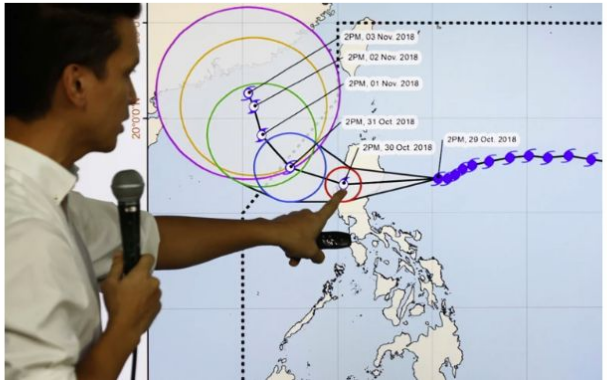
https://medium.com/kinomoto-mag/exploring-the-capabilities-of-physics-informed-neural-networks-pinns-integrating-data-science-a2e11fc31e19

Future.....??

EDITORIAL | 27 September 2023

AI will transform science — now researchers must tame it

A new *Nature* series will explore the many ways in which artificial intelligence is changing science — for better and for worse.



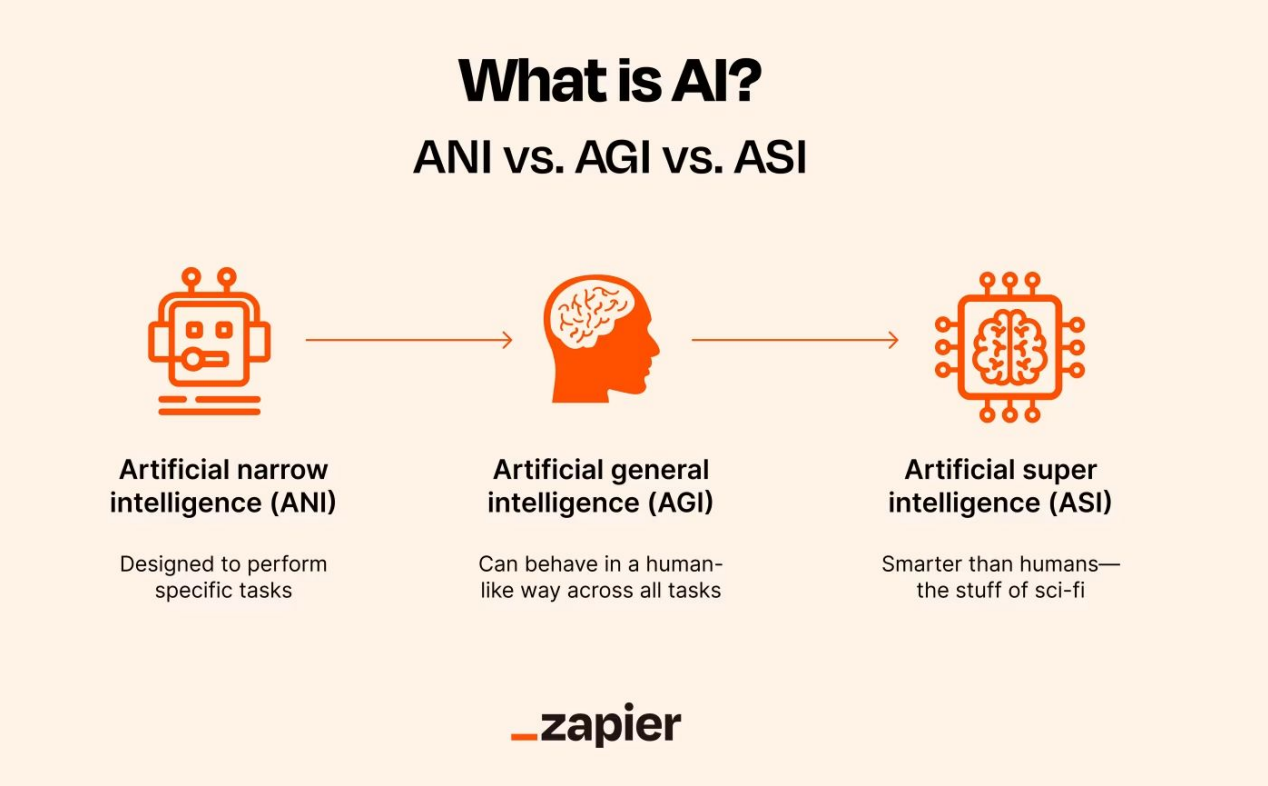
You have full access to this article via **Federal University of Pernambuco**
Download PDF

Related Articles

Science and the new age of AI: a Nature special
AI and science: what 1,600 researchers think
AI tools as science policy advisers? The potential and the pitfalls
How to stop AI deepfakes from sinking society — and science
ChatGPT is a black box: how AI research can break it open

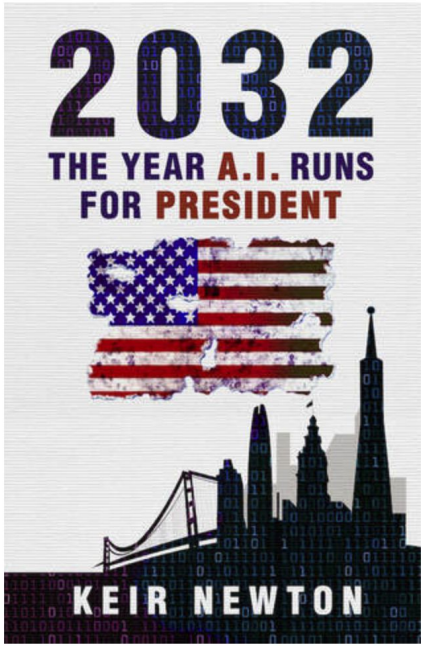
https://www.nature.com/articles/d41586-023-02988-6

Future.....??



https://zapier.com/blog/artificial-general-intelligence/

Future.....??



ONE NATION UNDER A.I.? In our near-future, America remains deeply divided and broken. Suddenly, a tech billionaire reveals an innovation that could save the nation: "Algo", an A.I. candidate for president. Algo is heralded as an efficient, rational, and seemingly incorruptible leader, underpinned by the brainpower of a million minds. To many, it is like a dream come true. But... who is really pulling its strings?

The year is 2032, and Silicon Valley launches its most audacious project yet: Algo, the world's most powerful AI that will run for president. Algo is the brainchild of the eccentric billionaire Jamin Lake. Jamin vows the A.I. has been programmed to follow only a single benevolent mantra: to deliver "the most good for the most people". In a nation mired in seemingly endless crisis and division, it feels like a real chance for a better future.

Isaac Raff returns to a visibly fraying San Francisco to meet Algo. He doesn't trust a word his old friend Jamin says. As the disillusioned pioneer behind the groundbreaking tech upon which Algo is based, Isaac knows how powerful – and how dangerous – putting an AI built in Jamin's image in any position of power could be. And despite his initial curiosity, he soon begins to discover a much darker side to his old friend's utopian promises...

https://www.2032book.com/