

Artificial Intelligence

CSC 116

DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE



1956
FIRST AI CONFERENCE

1956 Dartmouth Conference:
The Founding Fathers of AI



John McCarthy



Marvin Minsky



Claude Shannon



Ray Solomonoff



Alan Newell



Herbert Simon



Arthur Samuel



Oliver Selfridge



Nathaniel Rochester



Trenchard More

Goal of this conference

The conference's goal was to define and develop the new field of AI by exploring whether machines could be made to think like humans.

- **Established the term:** Defined the term "artificial intelligence" to describe the project.
- **Set the agenda:** The conference set the research direction for the new field, which would lead to subfields like machine learning and natural language processing.

Artificial Intelligence



Has Intelligence



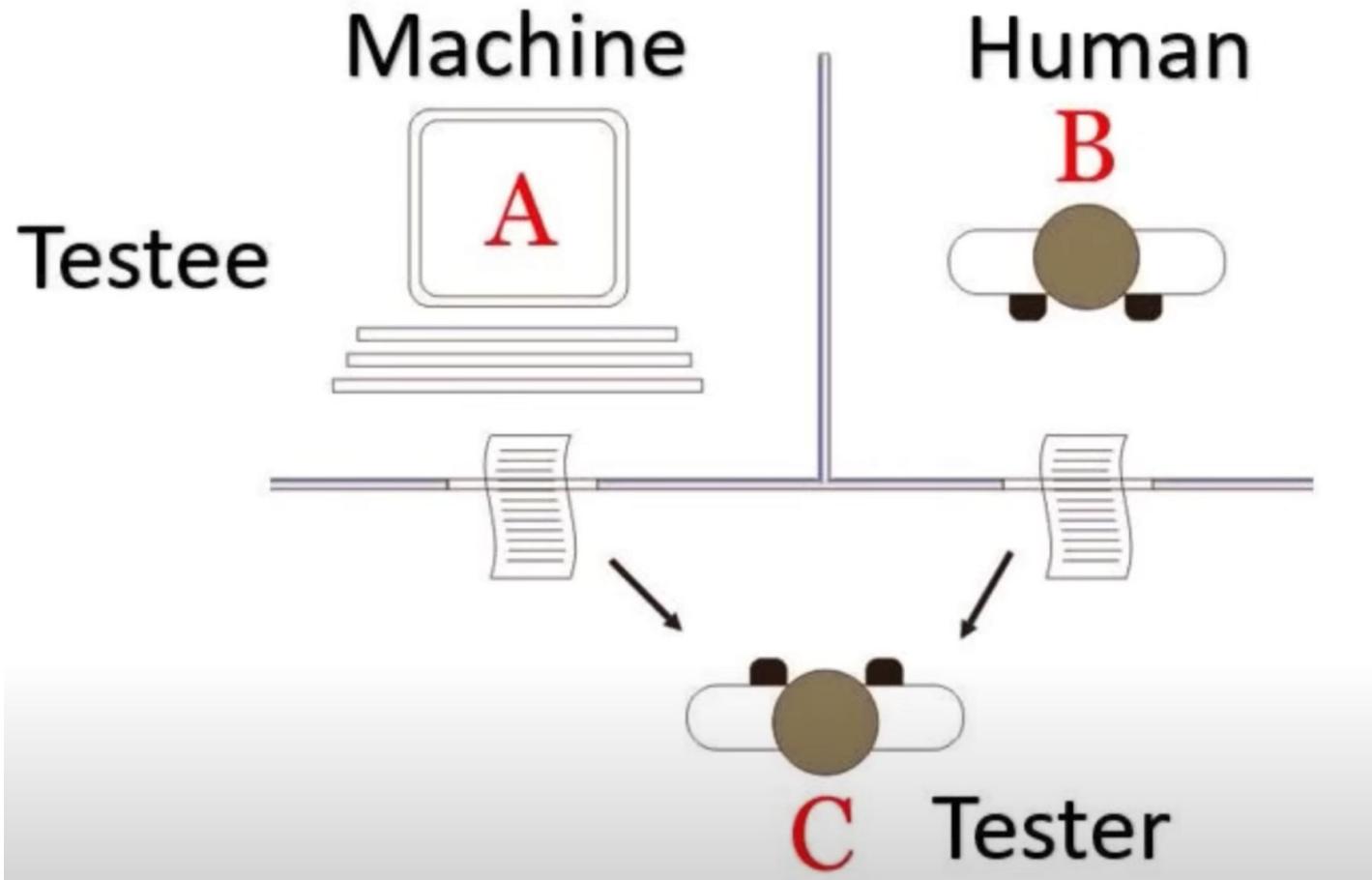
No Intelligence

paramecium

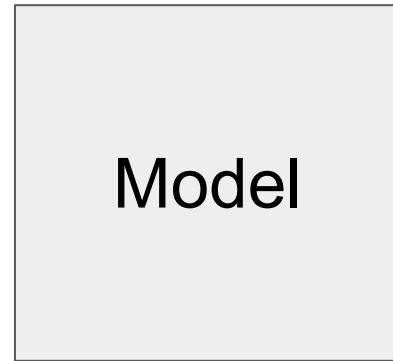


[https://www.reddit.com/r/microscopy/com
ments/1jxulzg/anyone_have_any_idea_w
hat_is_happening_lol/](https://www.reddit.com/r/microscopy/comments/1jxulzg/anyone_have_any_idea_what_is_happening_lol/)

What is AI?



Input



Output



1, Type of your interest: **Yes**

2, From Which Country?: **Japan**

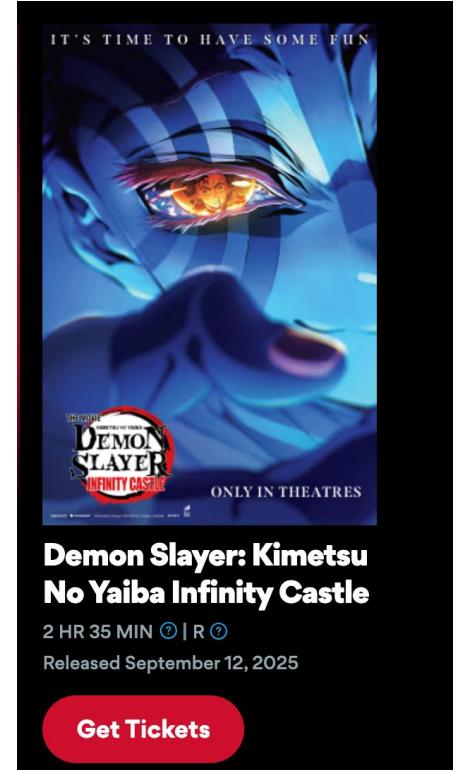
3, How much for a ticket? : **15 dollars**

4, Coupon? : **No**

5, Weather: good or not? **Good**

6, HW Completed? **No**

Predicting
whether
you will go
watch it



IT'S TIME TO HAVE SOME FUN

THE MOVIE ADAPTED FROM YOUR FAVORITE ANIME

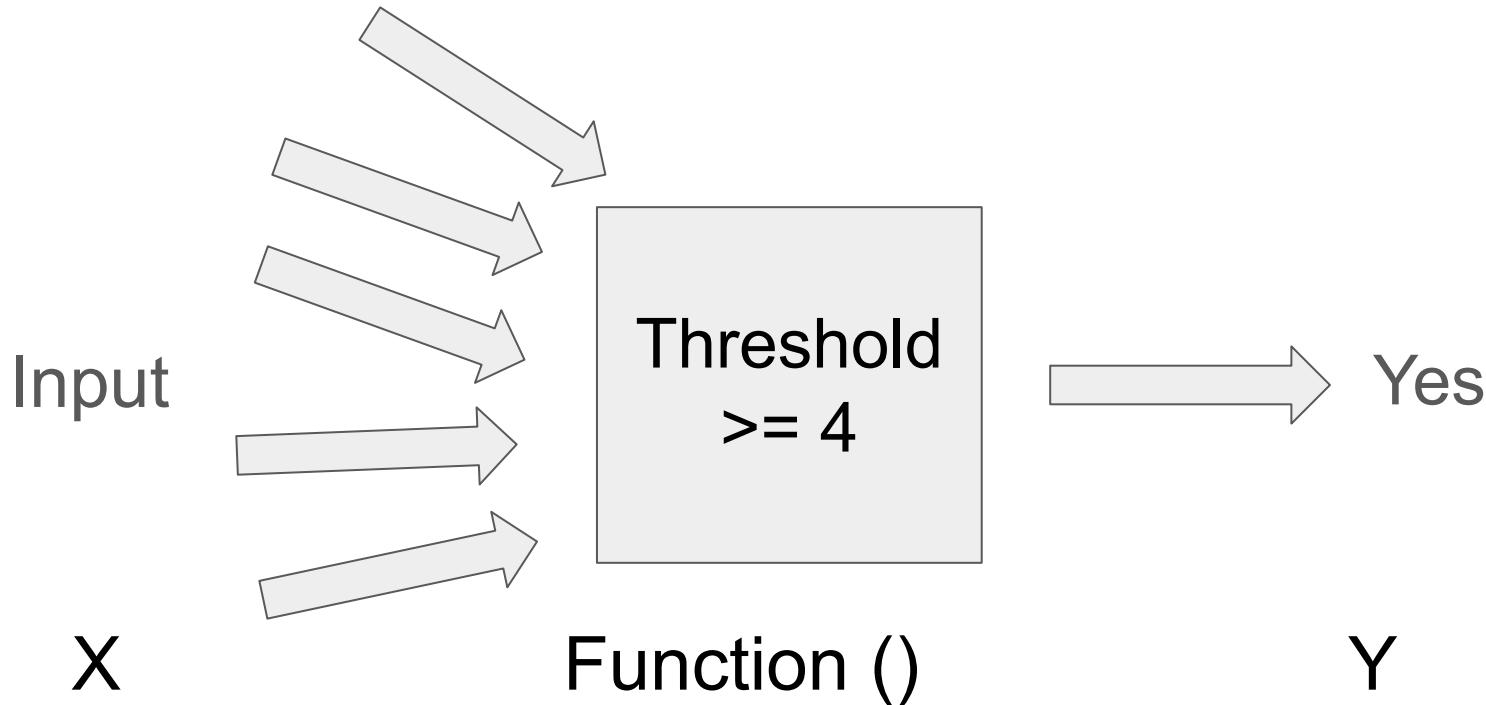
DEMON SLAYER
INFINITY CASTLE

ONLY IN THEATRES

2 HR 35 MIN | R

Released September 12, 2025

Get Tickets



Symbolish & Empiricism

A: You interested in this type of movie

B: There is no free version online

A and B => True

A or B => False

Weights

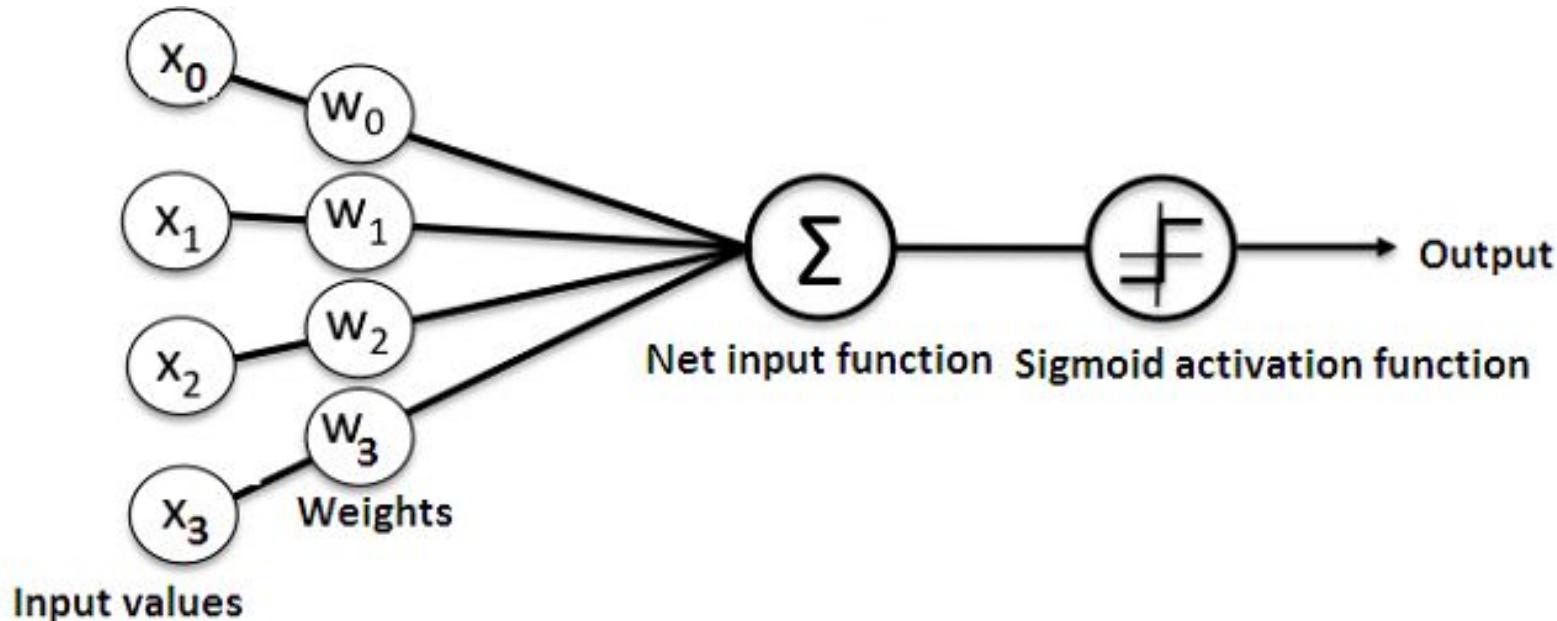
- 1, Type of your interest: Yes **50%**
- 2, From Which County?: Japan **3%**
- 3, How much for a ticket? : 15 dollars **5%**
- 4, Coupon? : No **0.5%**
- 5, Weather: good or not? Good **0.5%**
- 6, HW Completed? No **1%**
- 7, any free version online? Yes = 0, No = 1, **40%**

Yusen: $1 * 50\% + 3\% * 1 + 5\% + 0 + 0.5\% + 1 + 40\% = 99.5\%$

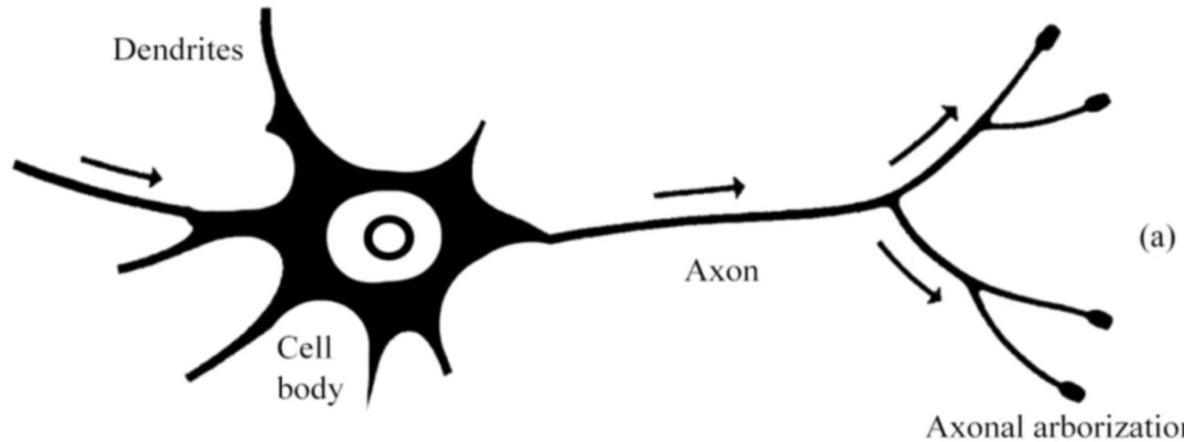
You: $0 + 3 + 5 + 0.5 + 0.5 + 1 + 40 = 50\%$

Threshold >= 95%

Perceptron



$$w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5 - b$$



NEW NAVY DEVICE LEARNS BY DOING

Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser

WASHINGTON, July 7 (UPI)—The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.

The embryo—the Weather Bureau's \$2,000,000 "704" computer—learned to differentiate between right and left after fifty attempts in the Navy's demonstration for newsmen.

The service said it would use this principle to build the first of its Perceptron thinking machines that will be able to read and write. It is expected to be finished in about a year at a cost of \$100,000.

Dr. Frank Rosenblatt, designer of the Perceptron, conducted the demonstration. He said the machine would be the first device to think as the human brain. As do human be-

ings, Perceptron will make mistakes at first, but will grow wiser as it gains experience, he said.

Dr. Rosenblatt, a research psychologist at the Cornell Aeronautical Laboratory, Buffalo, said Perceptrons might be fired to the planets as mechanical space explorers.

Without Human Controls

The Navy said the perceptron would be the first non-living mechanism "capable of receiving, recognizing and identifying its surroundings without any human training or control."

The "brain" is designed to remember images and information it has perceived itself. Ordinary computers remember only what is fed into them on punch cards or magnetic tape.

Later Perceptrons will be able to recognize people and call out their names and instantly translate speech in one language to speech or writing in another language, it was predicted.

Mr. Rosenblatt said in principle it would be possible to build brains that could reproduce themselves on an assembly line and which would be conscious of their existence.

1958 New York Times...

In today's demonstration, the "704" was fed two cards, one with squares marked on the left side and the other with squares on the right side.

Learns by Doing

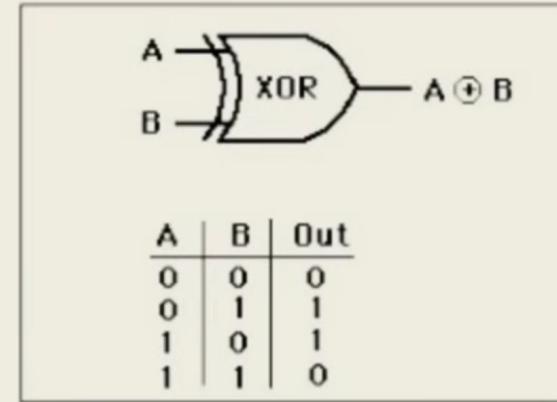
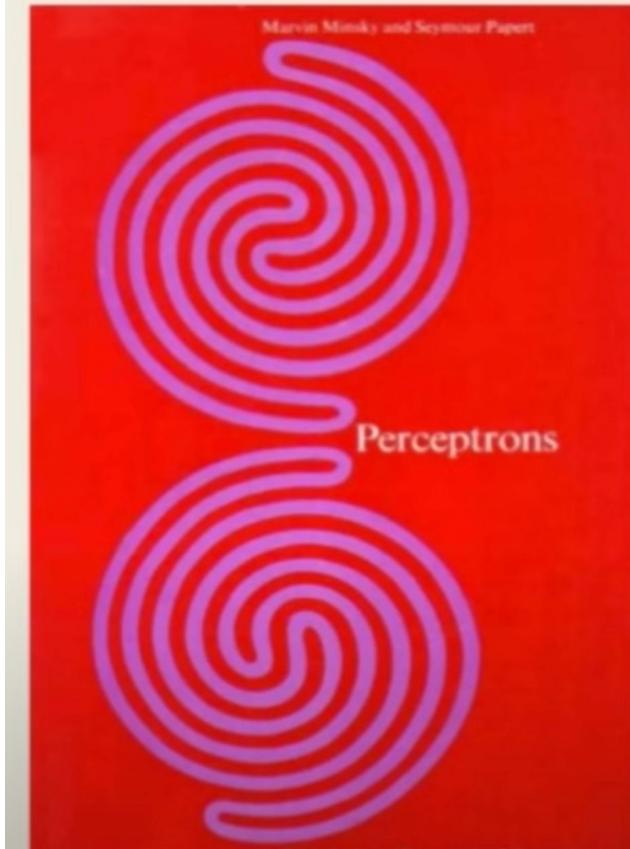
In the first fifty trials, the machine made no distinction between them. It then started registering a "Q" for the left squares and "O" for the right squares.

Dr. Rosenblatt said he could explain why the machine learned only in highly technical terms. But he said the computer had undergone a "self-induced change in the wiring diagram."

The first Perceptron will have about 1,000 electronic "association cells" receiving electrical impulses from an eye-like scanning device with 400 photo-cells. The human brain has 10,000,000,000 responsive cells, including 100,000,000 connections with the eyes.

1969

1969: Perceptrons can't do XOR!



<http://hyperphysics.phy-astr.gsu.edu/hbase/electronic/ietron/xor.gif>



XOR

x	y	$x \oplus y$
0	0	0
0	1	1
1	0	1
1	1	0

AND

If you are an
American and you
are **under 20** are
allowed to be
selected.

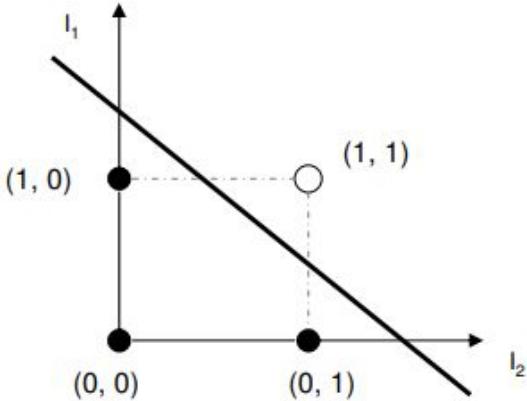
0	0	0
1	0	0
0	1	0
1	1	1

OR

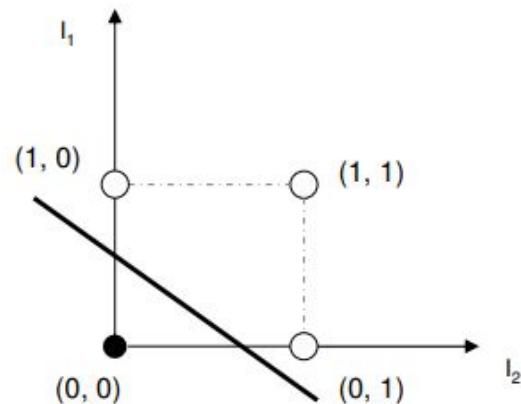
If you are an
American or you are
under 20 are allowed
to be selected.

0	0	0
1	0	1
0	1	1
1	1	1

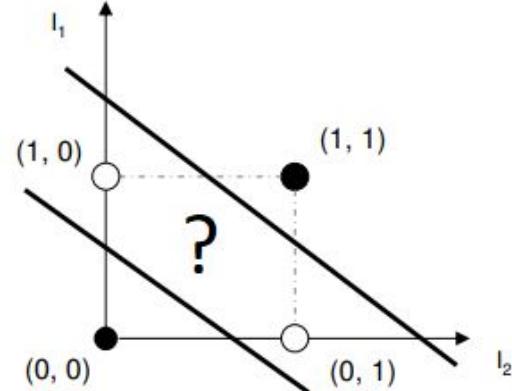
AND		
I_1	I_2	out
0	0	0
0	1	0
1	0	0
1	1	1



OR		
I_1	I_2	out
0	0	0
0	1	1
1	0	1
1	1	1



XOR		
I_1	I_2	out
0	0	0
0	1	1
1	0	1
1	1	0





Overview

Marvin Lee Minsky was an American cognitive and computer scientist concerned largely with research in artificial intelligence. He co-founded the Massachusetts Institute of Technology's AI laboratory and wrote extensively about AI and philosophy. [Wikipedia](#)

Born: August 9, 1927, [New York, NY](#)

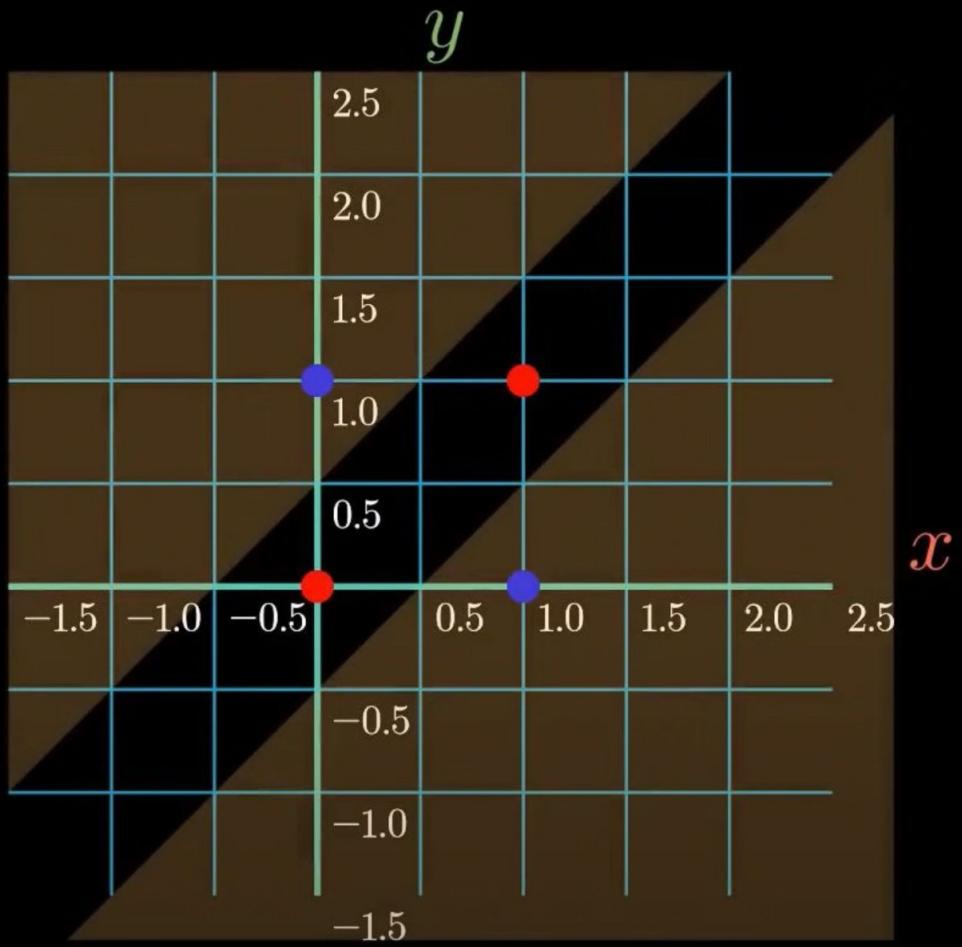
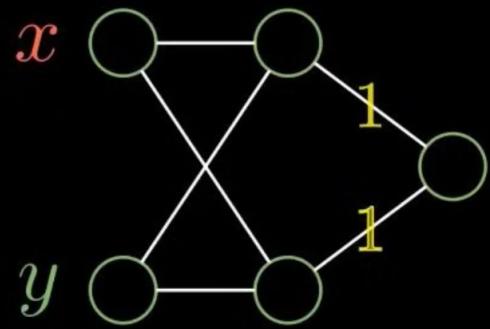
Died: January 24, 2016 (age 88 years), [Boston, MA](#)

Education: Princeton University (1951–1954) · [See more](#)

Awards: Turing Award, Japan Prize · [See more](#)

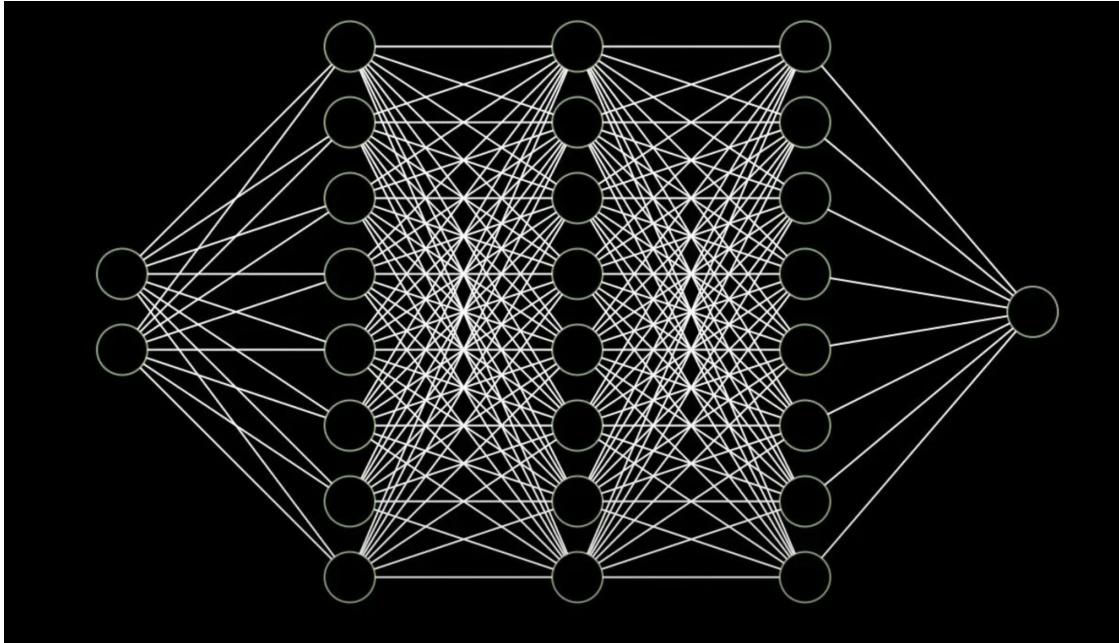
Parents: Fannie Resier, Henry Minsky

Spouse: Gloria Rudisch Minsky (m. 1952–2016)

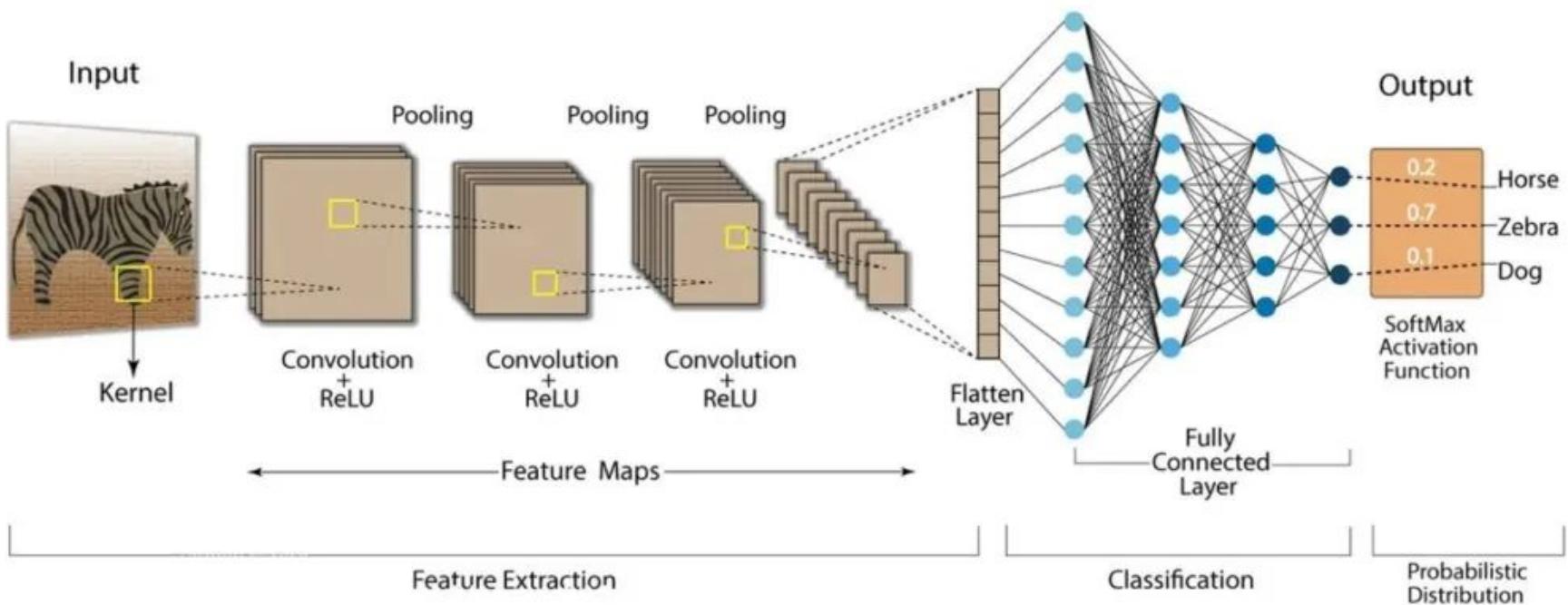


(MLP)

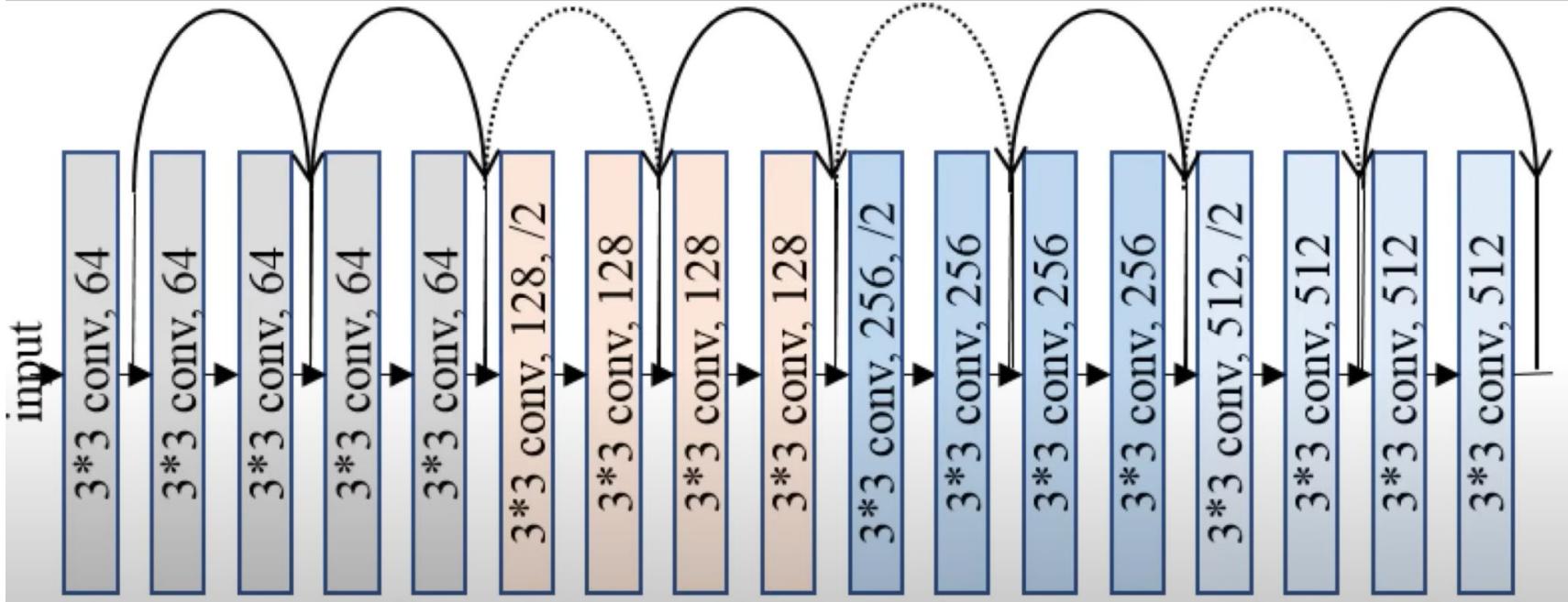
Multi-Layer Perceptrons



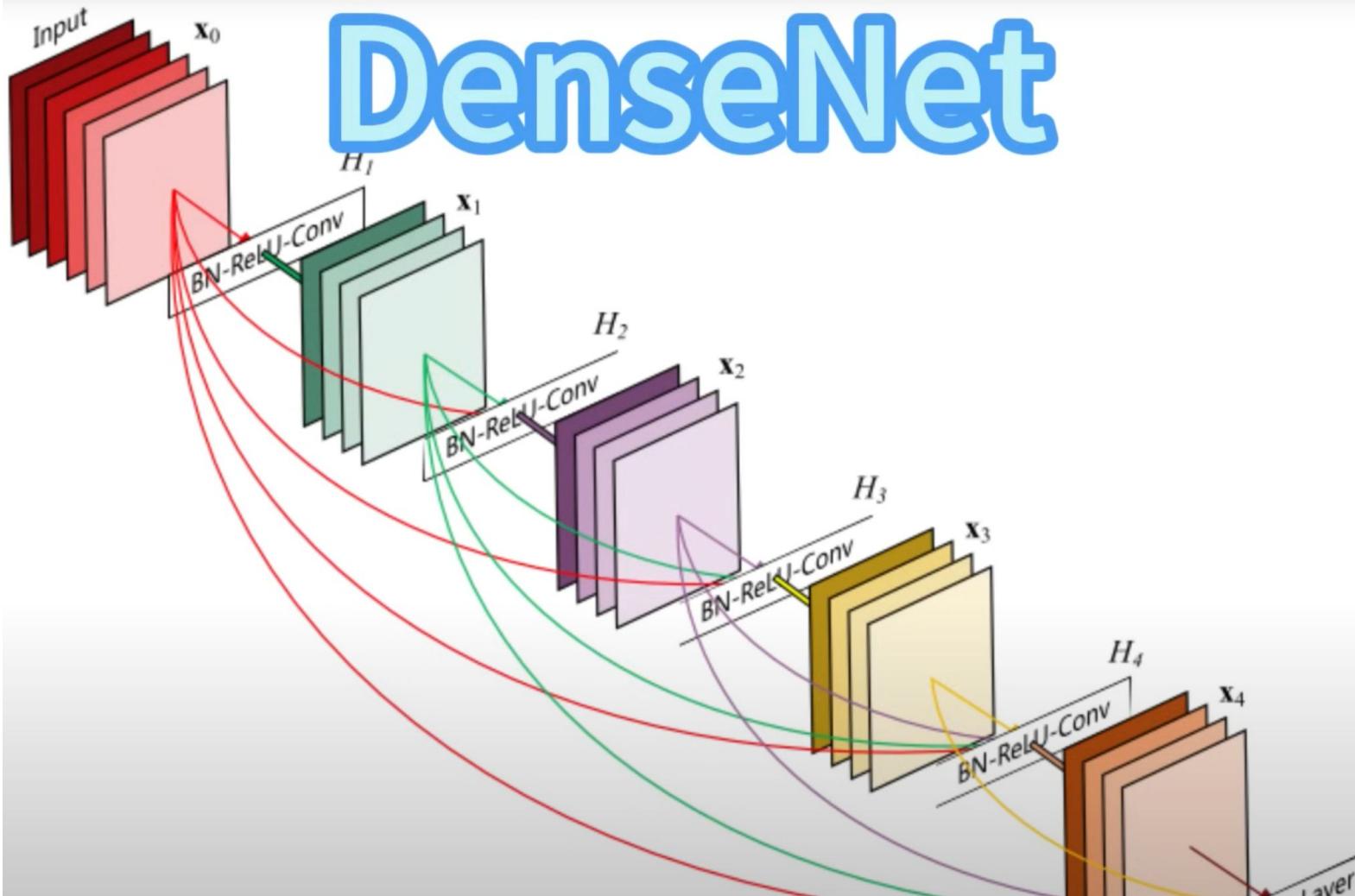
Convolution Neural Network (CNN)



ResNet



DenseNet



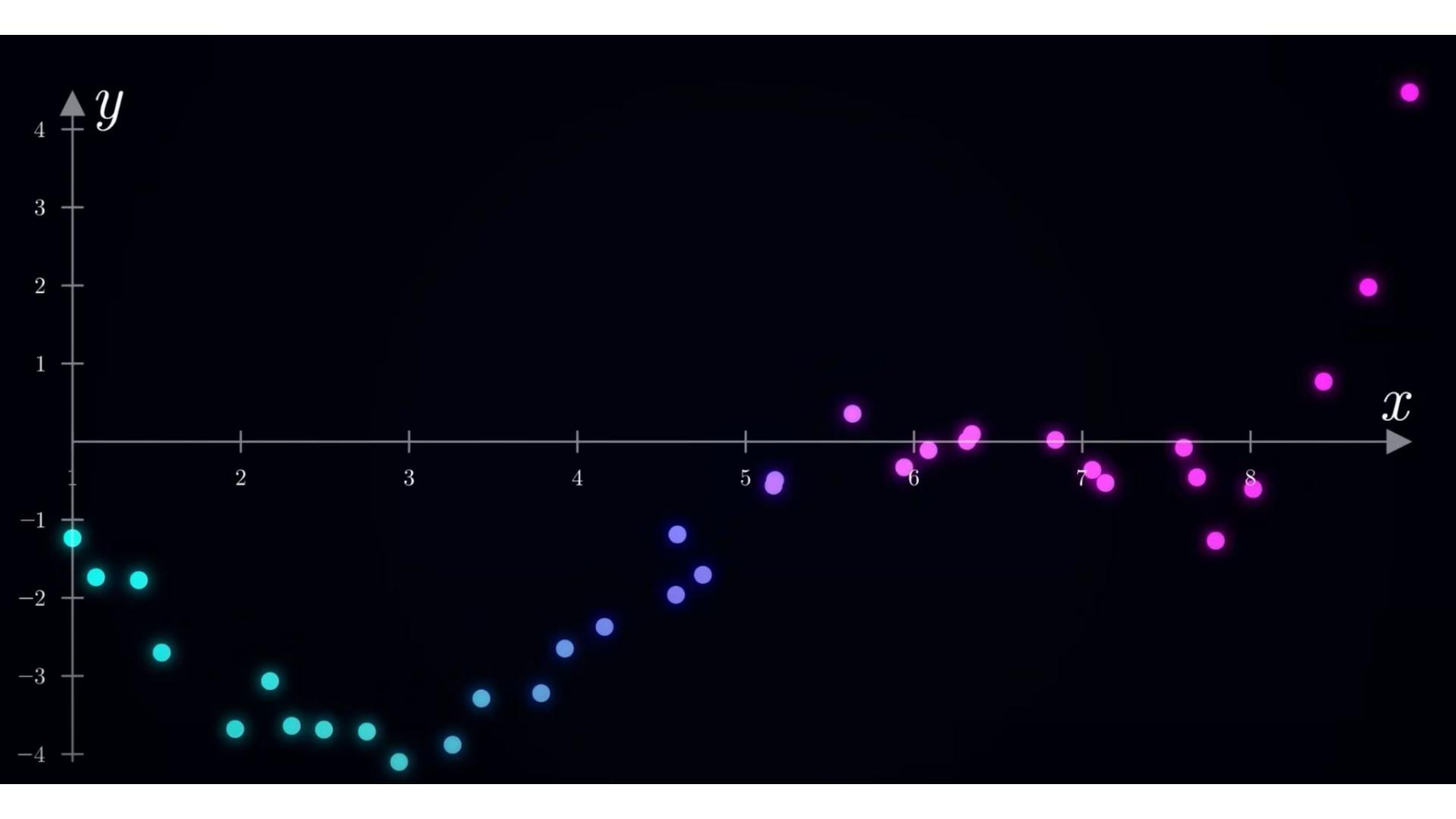
All the model architectures are only designed to efficiently find the weights using **Gradient Descent**

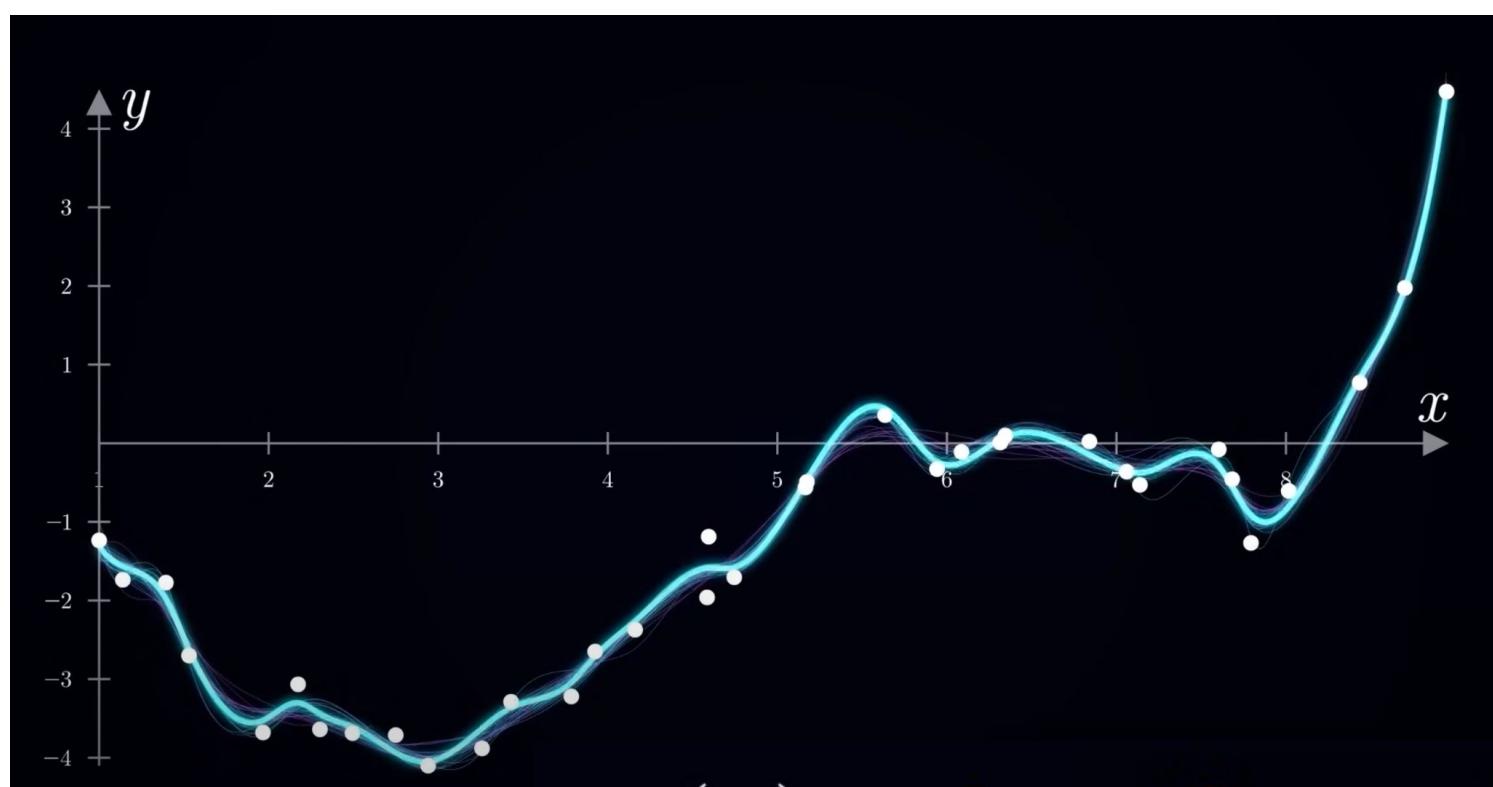
$$w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5 - b$$

How to train the Model??



Well-Labeled Datasets





$$y(x) = k_0 + k_1 x + k_2 x^2 + \\ k_3 x^3 + k_4 x^4 + k_5 x^5$$

1. Randomly generate all the weights
2. Calculate the Loss (errors)
3. Graduate Descent
4. To Minimum the Loss



k0=7.9394
k1=3.7451
k2=-4.8679
k3=1.6327
k4=-0.2165
k5=0.0098
Loss=26.5

Curve Fitter 6000

k0

k1

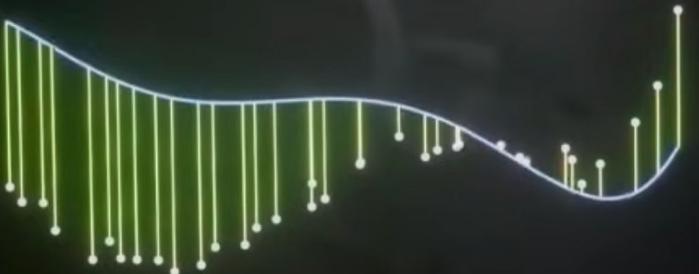
k2

k3

k4

k5



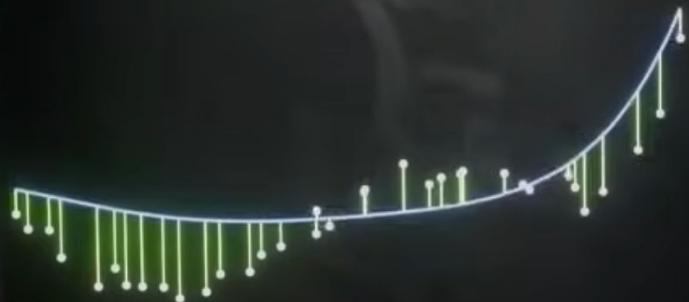


k0=4.1777
k1=2.4798
k2=-3.1837
k3=1.0569
k4=-0.1402
k5=0.0064
Loss=11.5

Curve Fitter 6000

k0 k1 k2 k3 k4 k5



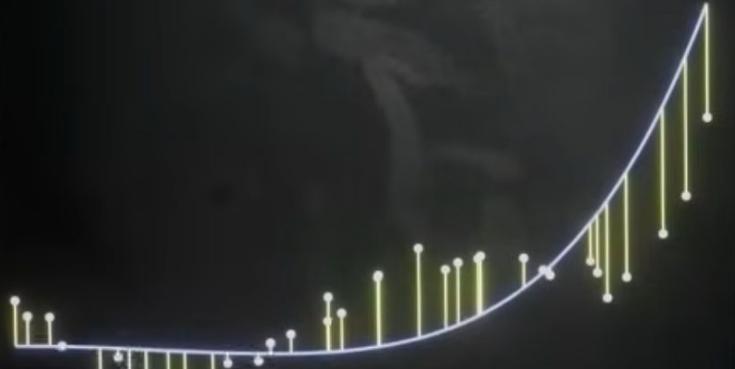


k0=-0.5134
k1=0.9618
k2=-1.1445
k3=0.3662
k4=-0.0486
k5=0.0024
Loss=9.7

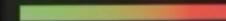
Curve Fitter 6000

k0 k1 k2 k3 k4 k5





$k_0 = -2.7823$
 $k_1 = 0.2276$
 $k_2 = -0.1582$
 $k_3 = 0.0320$
 $k_4 = -0.0043$
 $k_5 = 0.0004$
Loss = 18.2



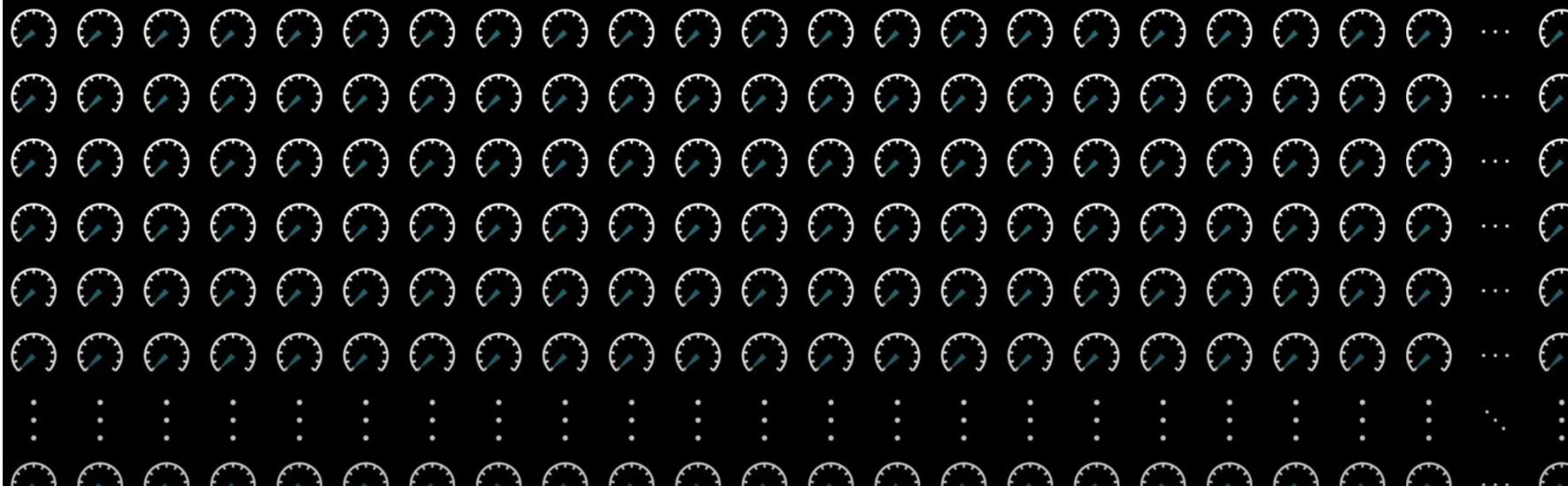
Curve Fitter 6000

$k_0 \quad k_1 \quad k_2 \quad k_3 \quad k_4 \quad k_5$





Total parameters: 175,181,291,520



Model Types

1, Regression Model

2, Classification

Natural Language Processing (NLP)

It is a branch of artificial intelligence that enables computers to understand, interpret, and generate human language.

Transformer

arXiv:1706.03762v7 [cs.CL] 2 Aug 2023

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

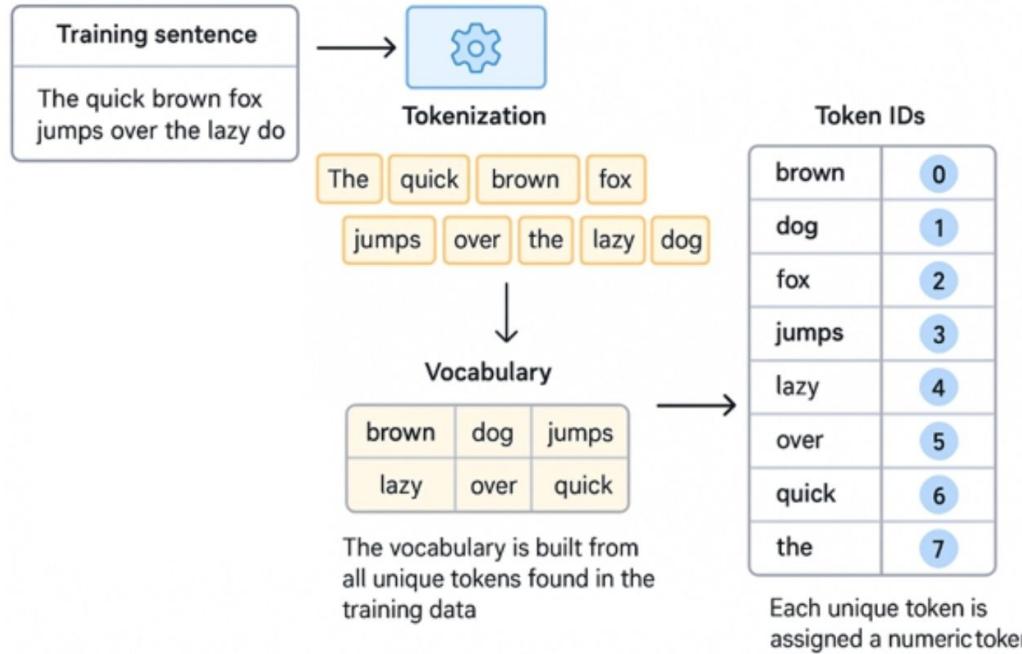
Łukasz Kaiser*
Google Brain
lukaszkaiser@google.com

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 best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

An LLM is trained to predict the next word



<https://tiktokenizer.vercel.app/>

Self-Attention

*“The animal didn’t cross the street because **it** was too tired.”*

Token	Weight
The	0.01
animal	0.62
didn't	0.02
cross	0.03
the	0.01
street	0.02
because	0.03
it	0.05
was	0.07
too	0.06
tired	0.07

Generative Pre-train Transformer (GPT)

Predicting the next word:

Hi, How are _you! _



You need 30mins to predict the next word if you use your local PC

Dataset	Sampling prop.	Epochs	Disk size
CommonCrawl	67.0%	1.10	3.3 TB
C4	15.0%	1.06	783 GB
Github	4.5%	0.64	328 GB
Wikipedia	4.5%	2.45	83 GB
Books	4.5%	2.23	85 GB
ArXiv	2.5%	1.06	92 GB
StackExchange	2.0%	1.03	78 GB

Table 1: **Pre-training data.** Data mixtures used for pre-training, for each subset we list the sampling proportion, number of epochs performed on the subset when training on 1.4T tokens, and disk size. The pre-training runs on 1T tokens have the same sampling proportion.