BRAIN, ALGORITHMS AND BEYOND: RECENT ADVANCES AND CHALLENGES IN AI

Prof. Dr. Rizwan Ahmed Khan
IBA, Karachi

INTELLIGENCE &THE NEURON AND THE NEURAL-NET

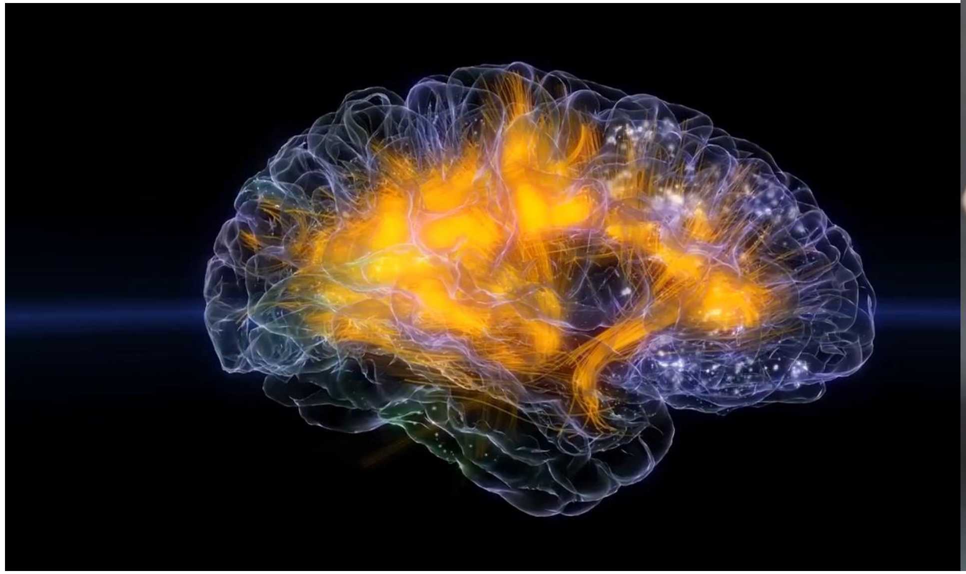
Every decision begins with millions of neurons firing in patterns of logic and inhibition

- Basic building block of Neural network is a neuron (~100 Billion Neurons) and ~100 trillion Synapses

- Electrical signal spike travels via axon.

- The Axon of neuron is connected to the dendrites of many other neuron.

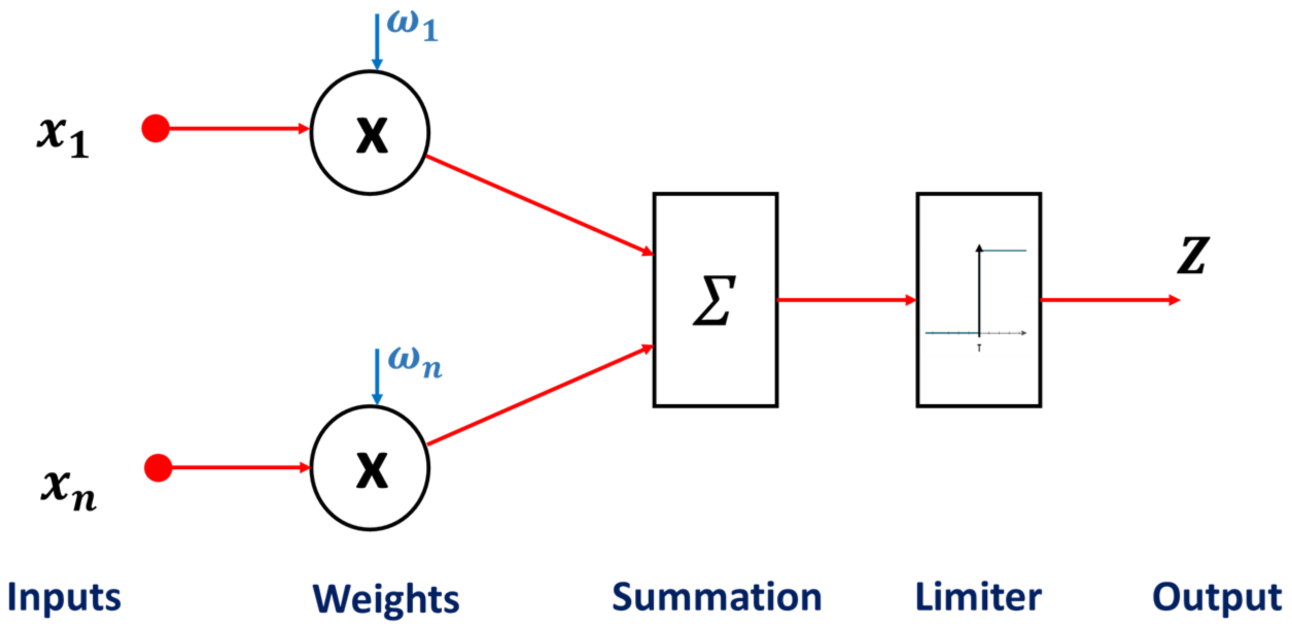
- When axon is stimulated (more than threshold) it dumps vesicles into inner synaptic space.

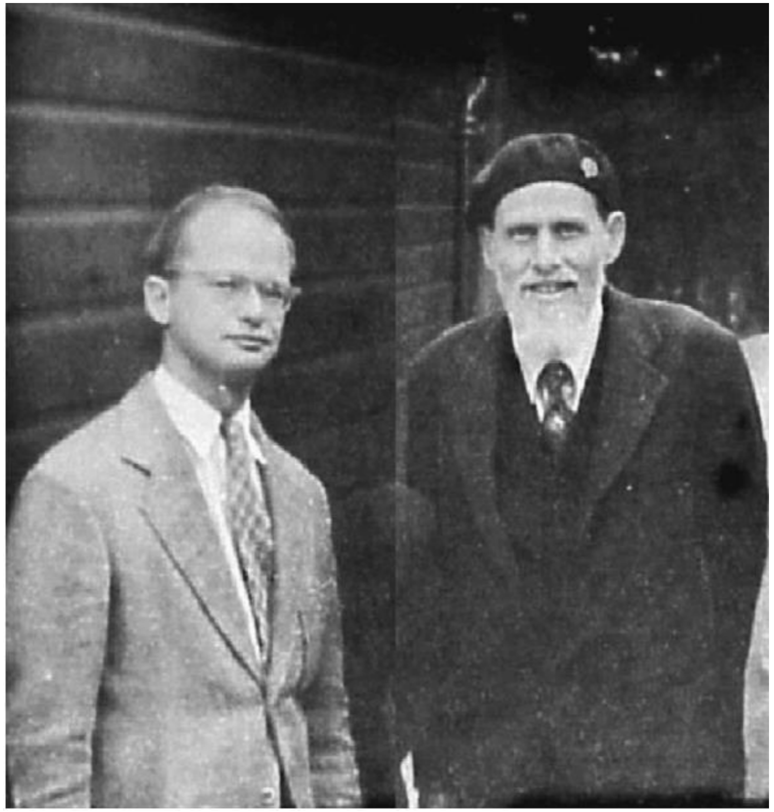


Neuroscape, University of California San Francisco, https://neuroscape.ucsf.edu/

THE ARTIFICIAL NEURON - MCCULLOCH & PITTS – 1943

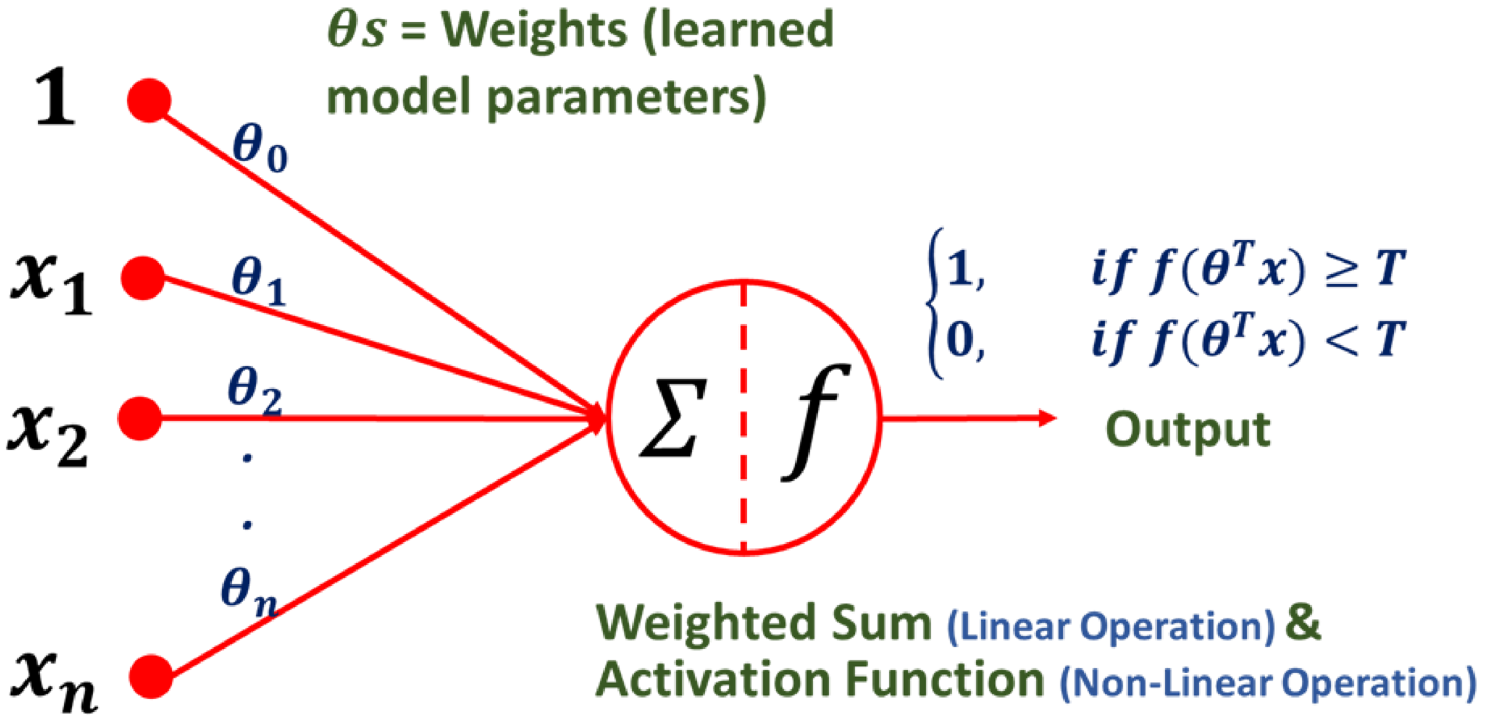
McCulloch, a **neurophysiologist**, and Pitts, a **neuroscientist**, came together in 1943 to mathematical model how the brain might work, marking the **birth of computational neuroscience and the earliest spark of artificial intelligence.**





Source: Semantic Scholar

FROM MCCULLOCH-PITTS TO THE MODERN ARTIFICIAL NEURON

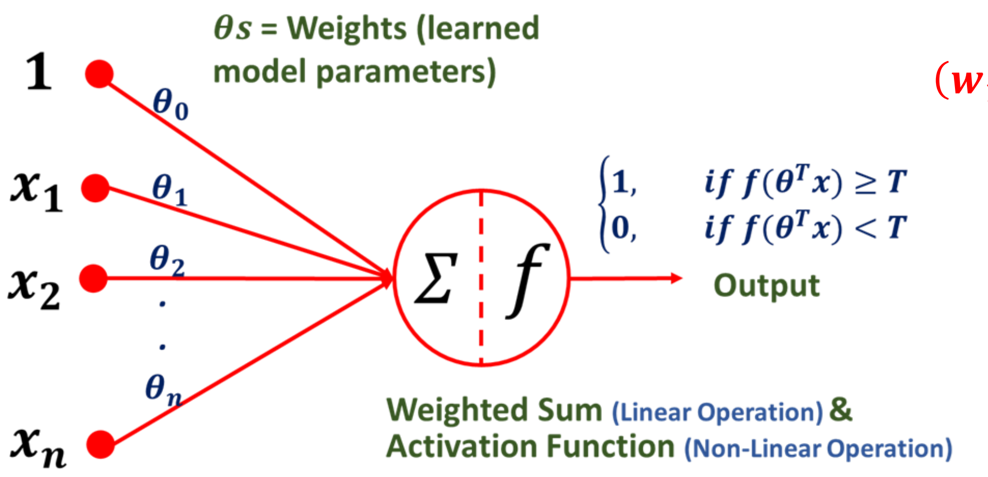


Inputs

Neuron

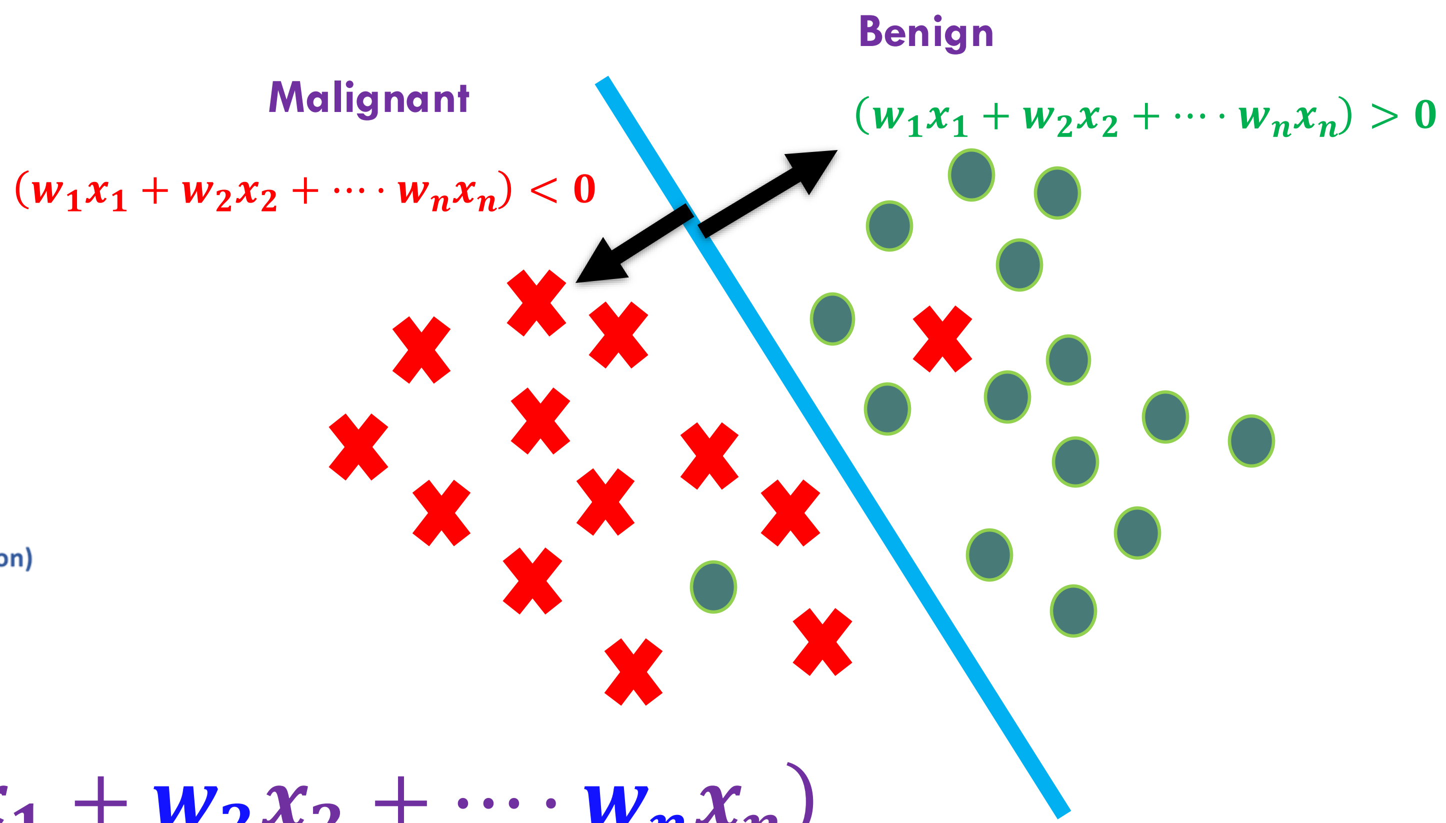
All these "weights updates" constitutes "**learning**".

FROM MCCULLOCH-PITTS TO THE MODERN ARTIFICIAL NEURON



Inputs

Neuron

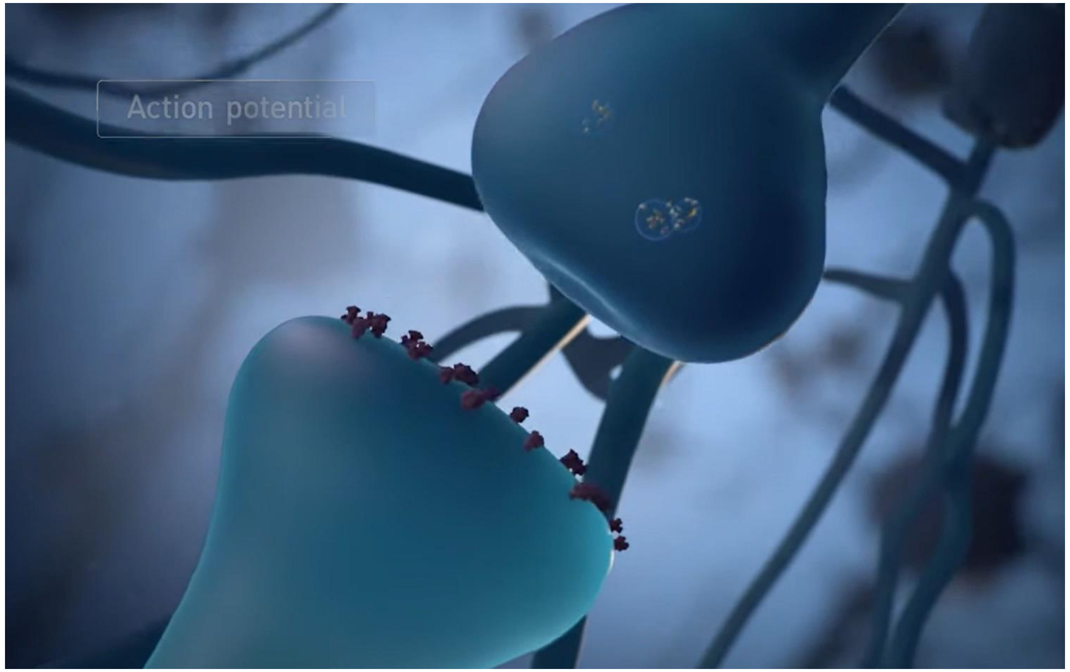


$$f(w_1x_1 + w_2x_2 + \cdots w_nx_n)$$

Intelligence (artificial) lies in discovering the right weights

How to find weights?

HOW DO THESE NEURONS LEARN?



- When neurons keep firing together, their connection grows stronger. **Repeated activity forges a lasting bond**, the brain's way of learning through experience - *Donald Hebb*

HEBBIAN LEARNING – 1949

- The next big question emerged, **how do these neurons learn?**
- **Donald Hebb** (neuropsychologist -1949) proposed a simple but powerful idea: when one neuron consistently activates another, their connection grows stronger (synaptic plasticity).
- The core idea is often summarized as: **"Neurons that fire together, wire together."**

$$\Delta w_{ij} = \eta \, x_i \, x_j$$

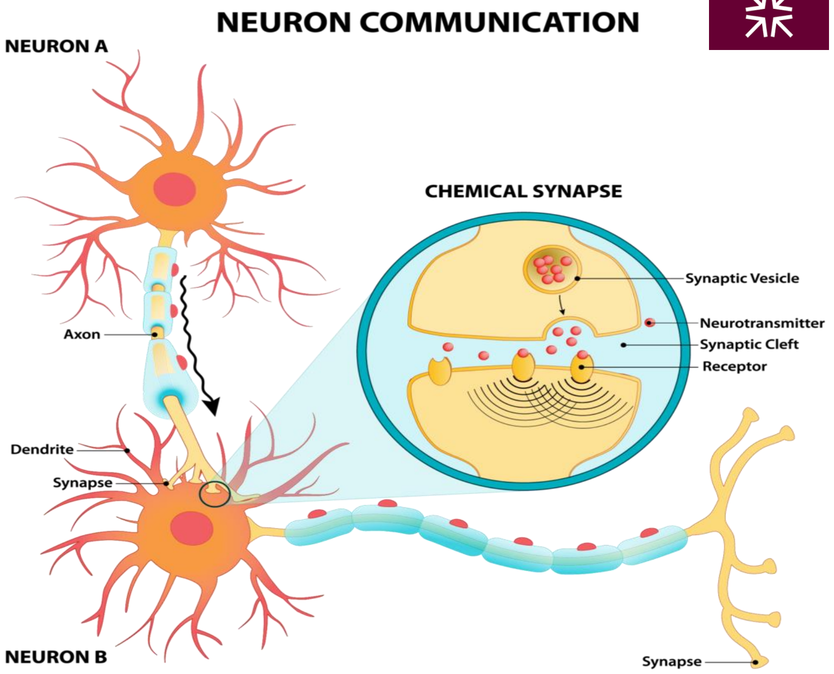
Where:

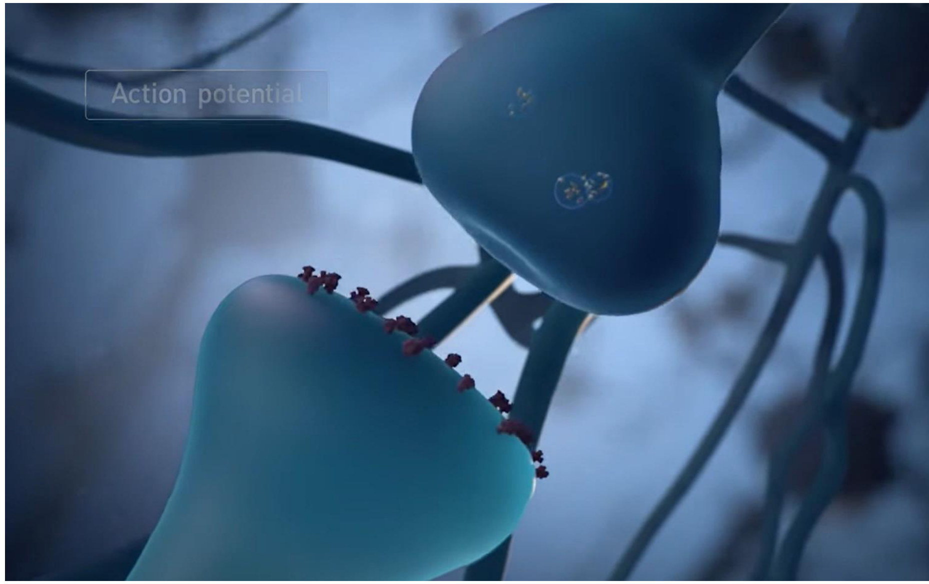
Δw_{ij} → change in synaptic weight from neuron j to neuron i

η → learning rate (small constant controlling speed of adaptation)

x_j → **pre-synaptic activity** (input neuron firing)

x_i → **post-synaptic activity** (output neuron firing)





DELTA RULE (50S – 60S)

Widrow and Hoff (electrical engineer) took "Hebbian learning", **neuroscience study to computers.** Don't need pre-synaptic, post-synaptic calculations. Need more concrete rule.

$$\Delta w_i = \eta \, (d - y) \, x_i$$

Where:

Δw_i → change in weight for input i

η → learning rate

d → desired output

y → actual output

x_i → input from feature/neuron i (pre-synaptic activity)

WHAT AI ORIGINALLY AIMED FOR

1950s–1980s | The Founding Vision

Symbolic AI / GOFAI

Logic, rules, and explicit reasoning: **mimic human thinking step by step**

DT, Expert System, Bayesian Net., Rule based etc.

Turing's Vision (1950)

A machine that converses indistinguishably from a human: **the Imitation Game**

LLMs, Transformers, RNN/LSTM

Minsky & McCarthy

General problem solving with common sense: **machines that truly understand**

Knowledge Graphs, Neuro-symbolic AI, Cognitive Arch.

The Grand Dream: AGI

Artificial General Intelligence : **learn and reason across any domain like a human**

domain agnostic learning …. RBM ?

THE PERCEPTRON – 1958

The first model of **neural network** that could learn to classify inputs.

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 beings, 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.

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.

1958 New York Times…

Frank Rosenblatt

THE PERCEPTRON – 1958

Net Input:

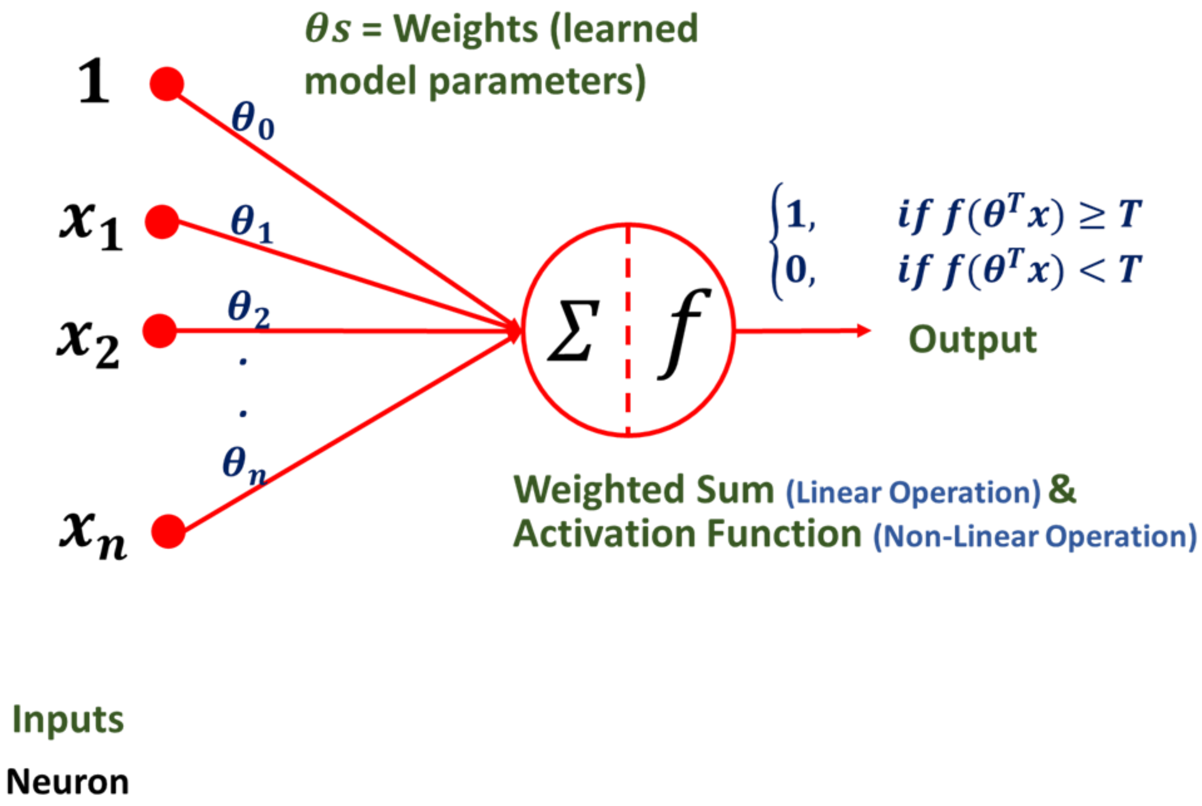
$$u = \sum_i w_i x_i + b$$

Activation (Step Function):

$$y = f(u) = \begin{cases} 1, & u \geq 0 \\ 0, & u < 0 \end{cases}$$

Perceptron Learning Rule:

$$w \quad \leftarrow \quad w + \eta(d - y)x$$

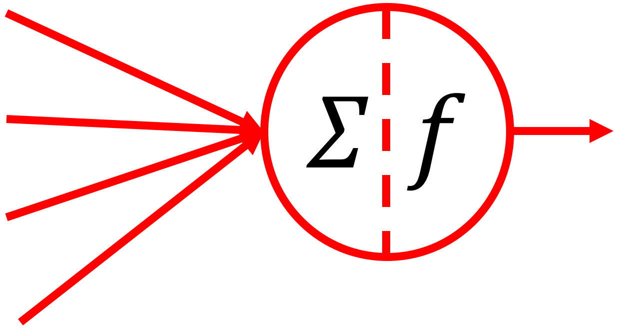


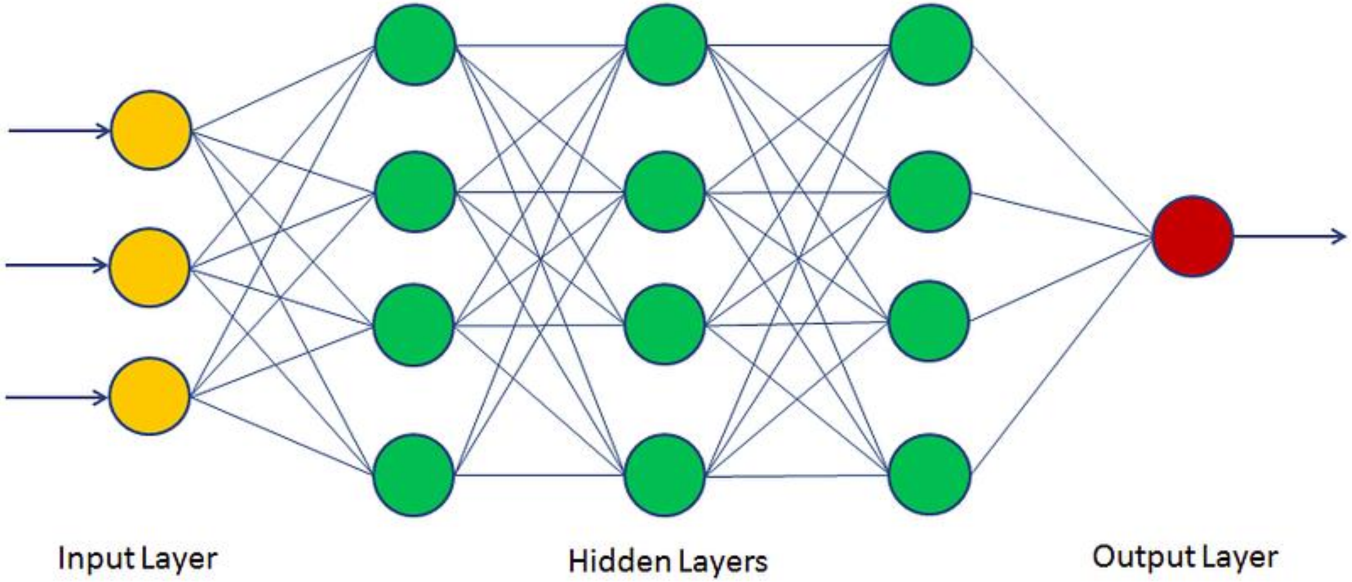
GRADIENT DESCENT – 1959

- The method of **Gradient Descent** is much older (19th century, used by Cauchy in 1847 for optimization).
- 1959, **Arthur Samuel** used **gradient-descent–like methods** in his checkers-playing program.
- His program is considered one of the first self-learning AI systems (**temporal-difference learning**, which much later influenced reinforcement learning).
- It adjusted weights based on **minimizing error** between predicted and actual outcomes of moves.

Arthur Samuel

MULTI-LAYER PERCEPTRON (MLP) (60s – 70s)





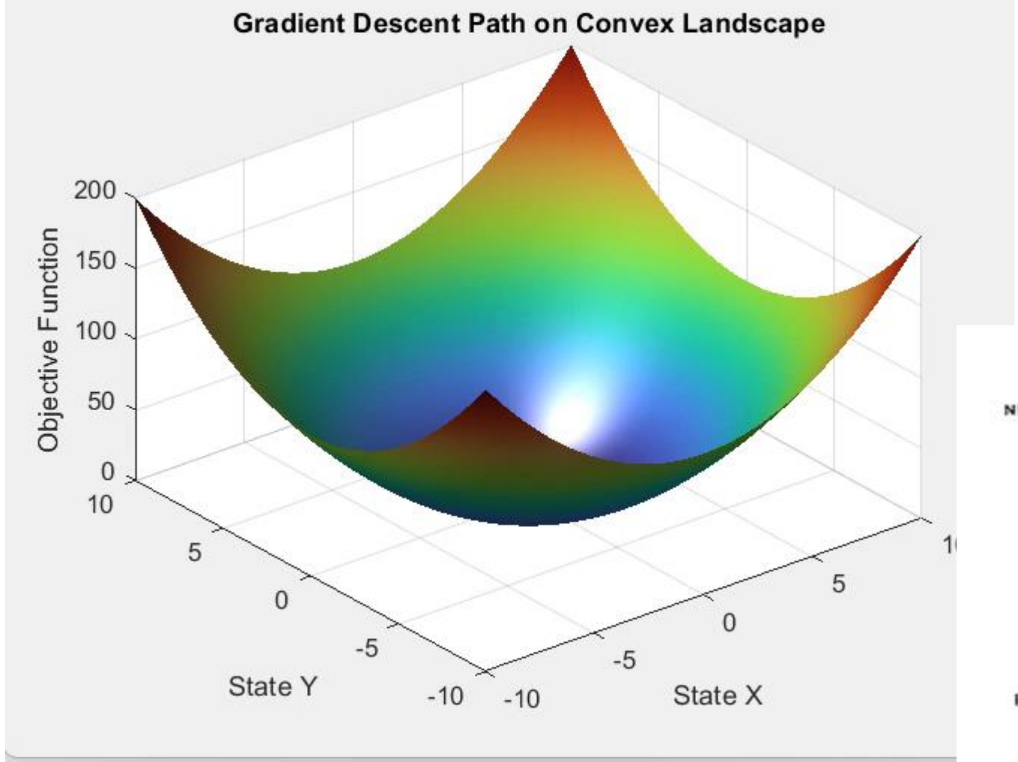
- Perceptron's idea extended to include **multiple layers**, hoping to solve more complex, non-linear problems.
- Researchers knew multiple layers would help, but
  - they lacked the training methods.
  - we don't know what features to use either.

} **Obstacles**

BACKPROPAGATION – 1974

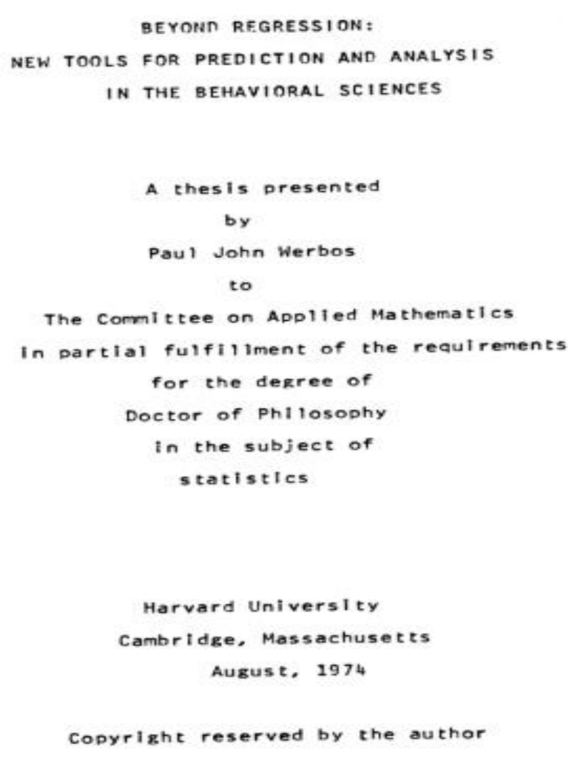
- **Backpropagation** (“*back propagation of errors*” the method of <u>propagating error backwards</u> through the network) as a training algorithm for multilayer perceptrons.

- **Cauchy (1847)** gave the idea of gradient descent for optimization, while Werbos (1974) gave backpropagation to compute gradients efficiently in multilayer networks.



$$w_i \leftarrow w_i - \eta \frac{\partial E}{\partial w_i}$$

Paul Werbos



BEYOND REGRESSION:
NEW TOOLS FOR PREDICTION AND ANALYSIS
IN THE BEHAVIORAL SCIENCES

A thesis presented
by
Paul John Werbos
to
The Committee on Applied Mathematics
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
statistics

Harvard University
Cambridge, Massachusetts
August, 1974

Copyright reserved by the author

BACKPROPAGATION – 1974

- Critics argued (and still do) that backprop doesn't resemble how real brains learn, so its inspiration from neuroscience was weak.

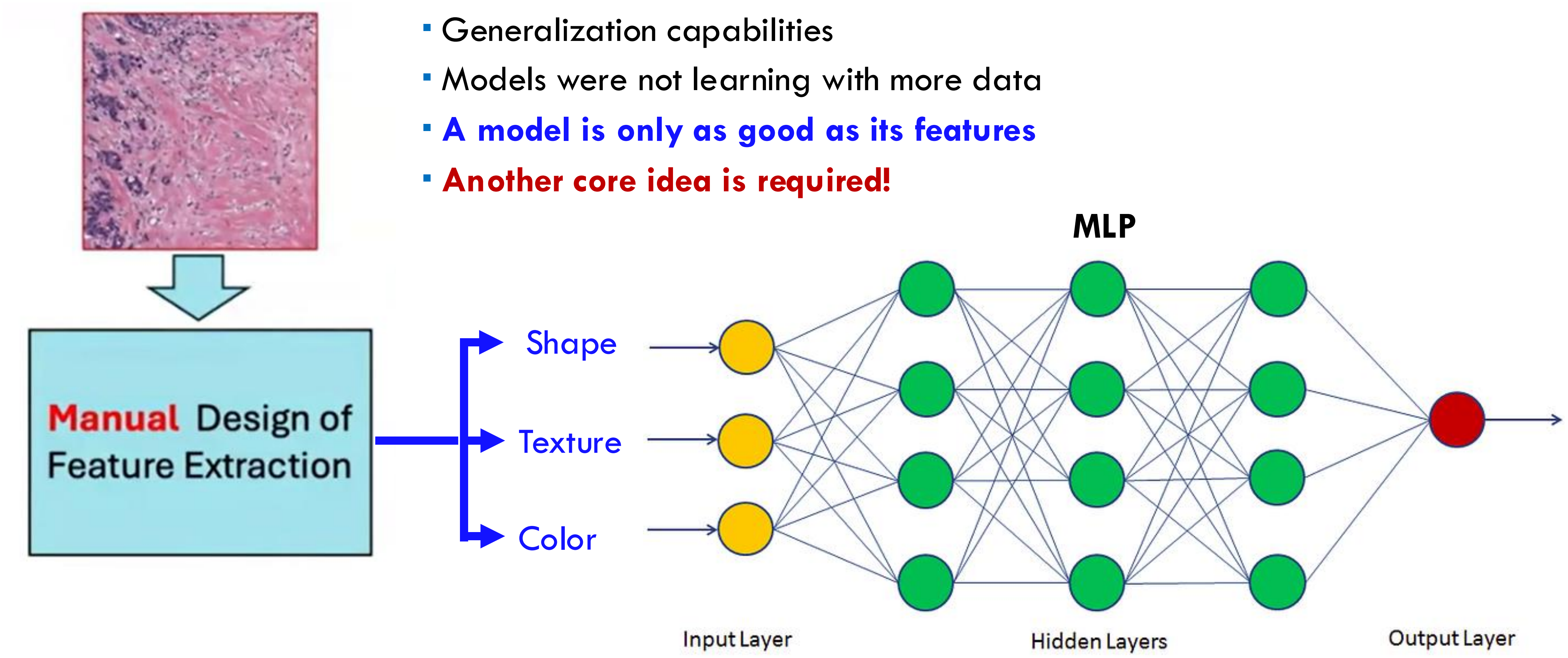
- Paul Werbos published his work in "*economics/systems theory*", not in AI or cognitive science and also didn't have the practical neural-network training demonstrations (like Rumelhart, *Hinton & Williams did in 1986*), so the AI community didn't immediately recognize its value. Other issues : AI winter, lack of research in NeuralNet, compute power.

Backpropagation is the backbone of deep learning, without it, there would be no deep neural networks powering today's AI

HISTORIC OBSTACLE: MANUAL FEATURES

Humans had to manually design the features fed into models.

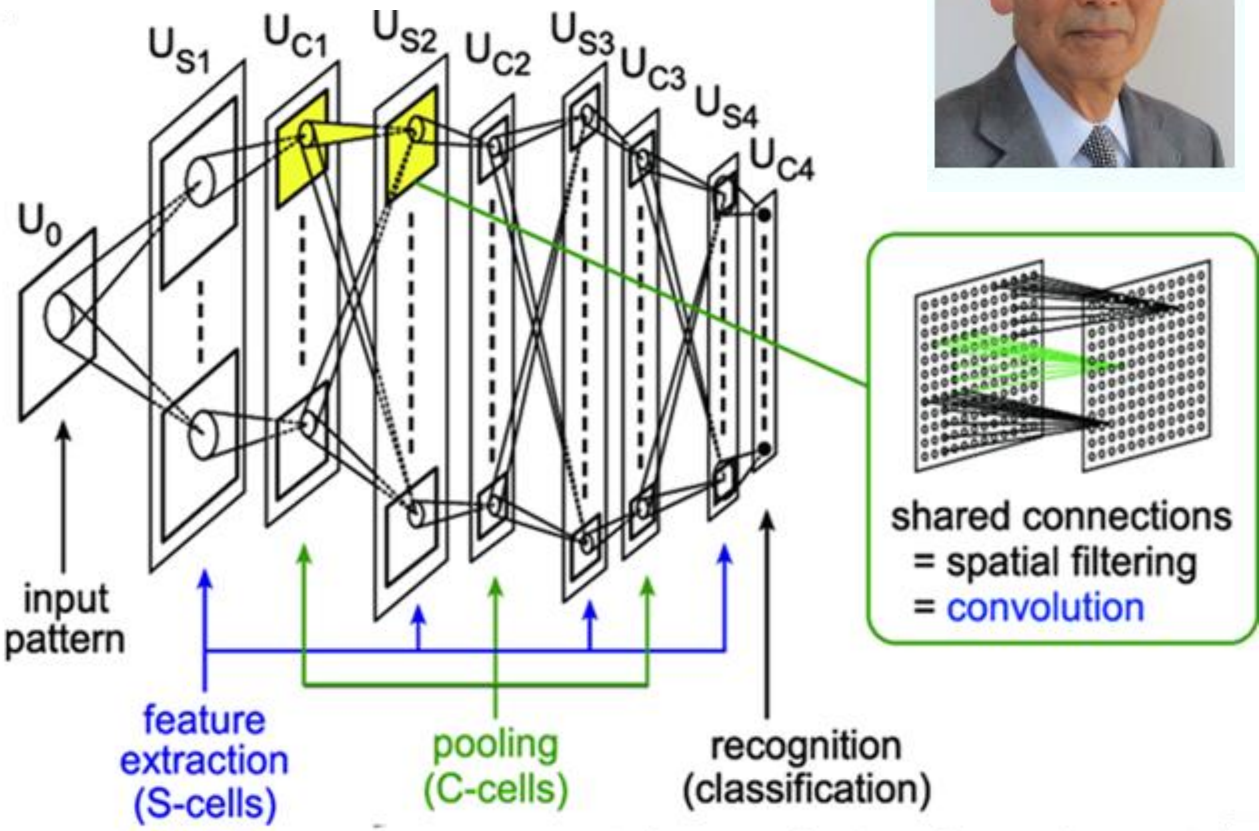
- Domain expertise required
- Generalization capabilities
- Models were not learning with more data
- **A model is only as good as its features**
- **Another core idea is required!**



NEOCOGNITRON – 1980

Fukushima (1980): **Automated feature extraction (core idea)** as done in visual cortex, inspired by work of Hubel and Wiesel – 1959 (Nobel prize 1981(Physiology); hierarchy of **complex and simple cells**).

1. **Simple Cells (S-cells)**
   - Detect **local patterns** (like edges, lines, corners).
2. **Complex Cells (C-cells)**
   - Pool responses from multiple neighboring simple cells.
   - Provide **invariance to small shifts and distortions** (translation tolerance).

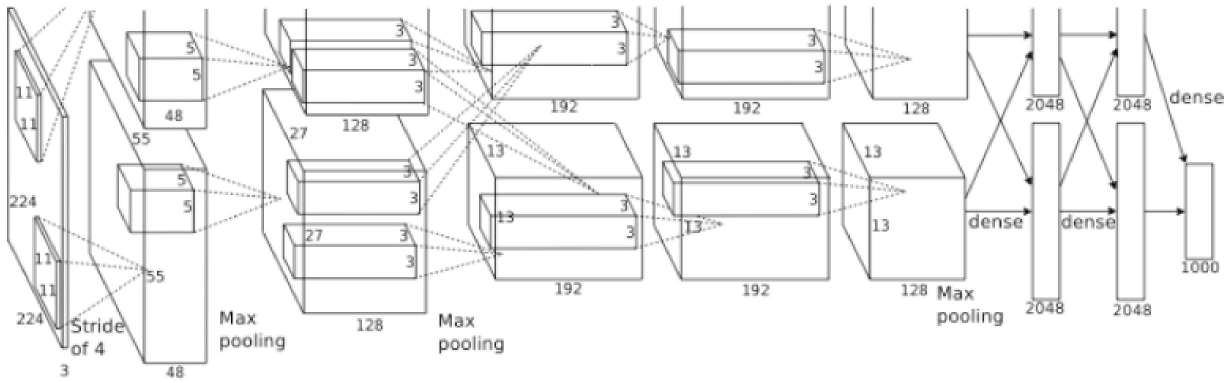


https://www.cs.princeton.edu/courses/archive/spr08/cos598B/Readings/Fukushima1980.pdf

WHY NEOCOGNITRON FAILED?

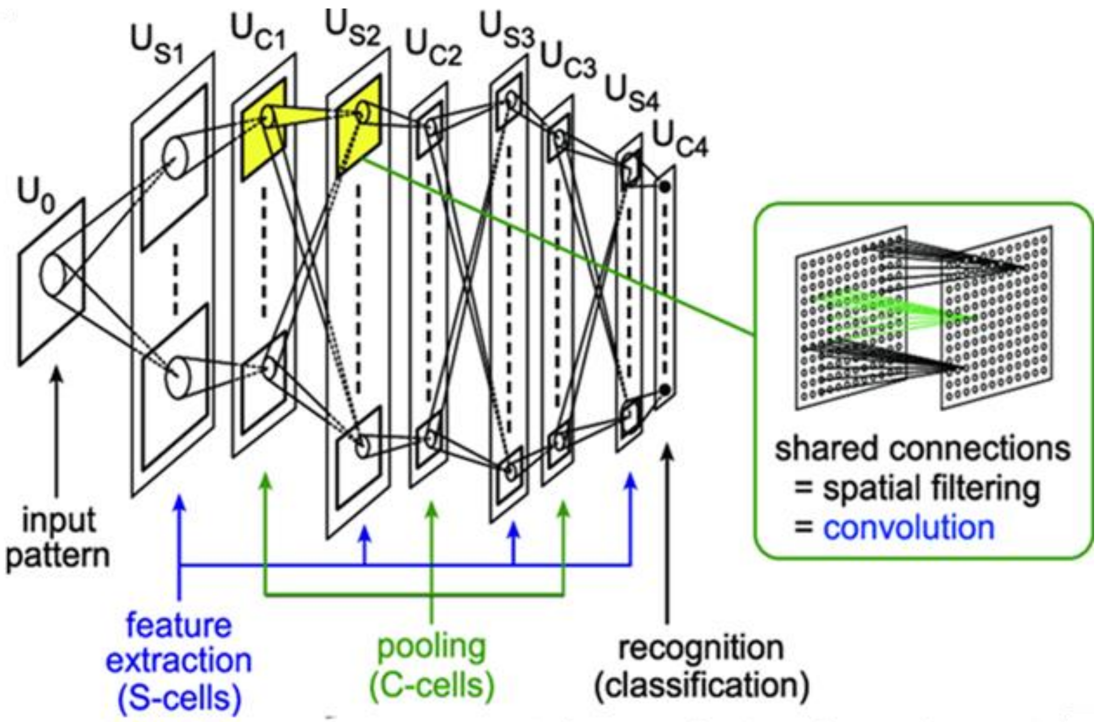
1. Lack of effective learning algorithm
   - Fukushima used **unsupervised learning (Hebbian-like rules)** for feature extraction, and **hand-tuning** for higher layers.
2. Computation complexity
   - 1980s hardware was far too slow to train large hierarchical networks.
3. Limited training data
4. Manual configuration
   - The Neocognitron didn't use backpropagation (Werbos's idea was still obscure)

IMPACT OF NEOCOGNITRON



Looks a lot like AlexNet more than 32 years later!

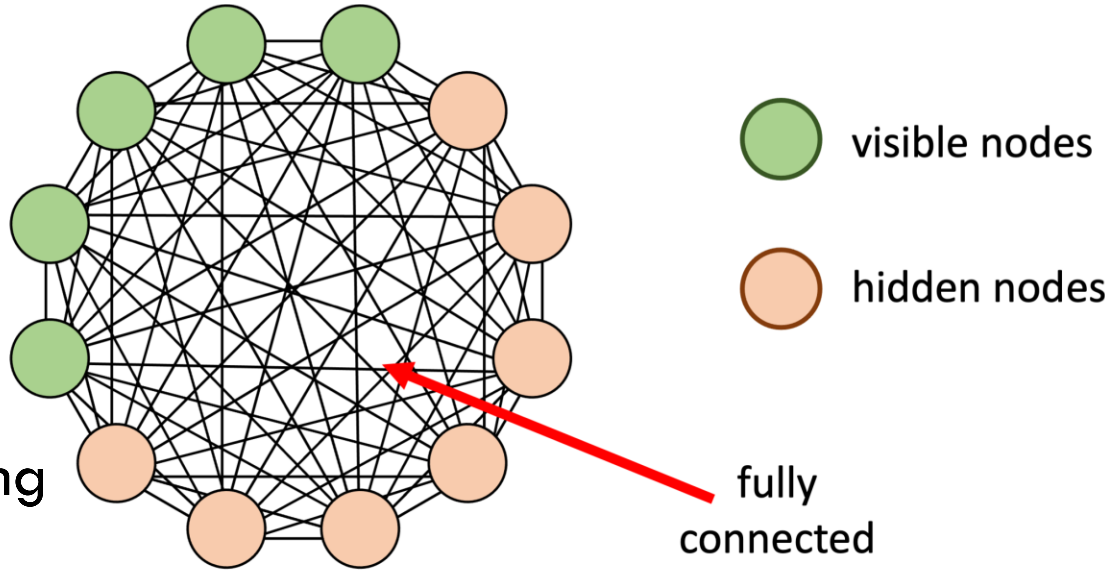
Krizhevsky, Alex; Sutskever, Ilya; Hinton, Geoffrey. "**ImageNet classification with deep convolutional neural networks**". Communications of the ACM: 84–90. doi:10.1145/3065386. ISSN 0001-0782



https://www.cs.princeton.edu/courses/archive/spr08/cos598B/Readings/Fukushima1980.pdf

BOLTZMANN MACHINES – 1985

- Introduced probabilistic, energy-based learning, leading *to unsupervised pretraining methods* that helped deep learning take off in the 2000s.
- It's a stochastic **RNN** that can learn complex distributions.
- Boltzmann Machines are theoretically intriguing due to their ***Hebbian-style learning.***
- Boltzmann Machine learns the **data's probability distribution** by reducing the gap between what it sees in the data and what it generates itself through energy minimization and sampling.



* David H. Ackley, Geoffrey E. Hinton, Terrence J. Sejnowski ; **A Learning Algorithm for Boltzmann Machines**, Cognitive Science 1985

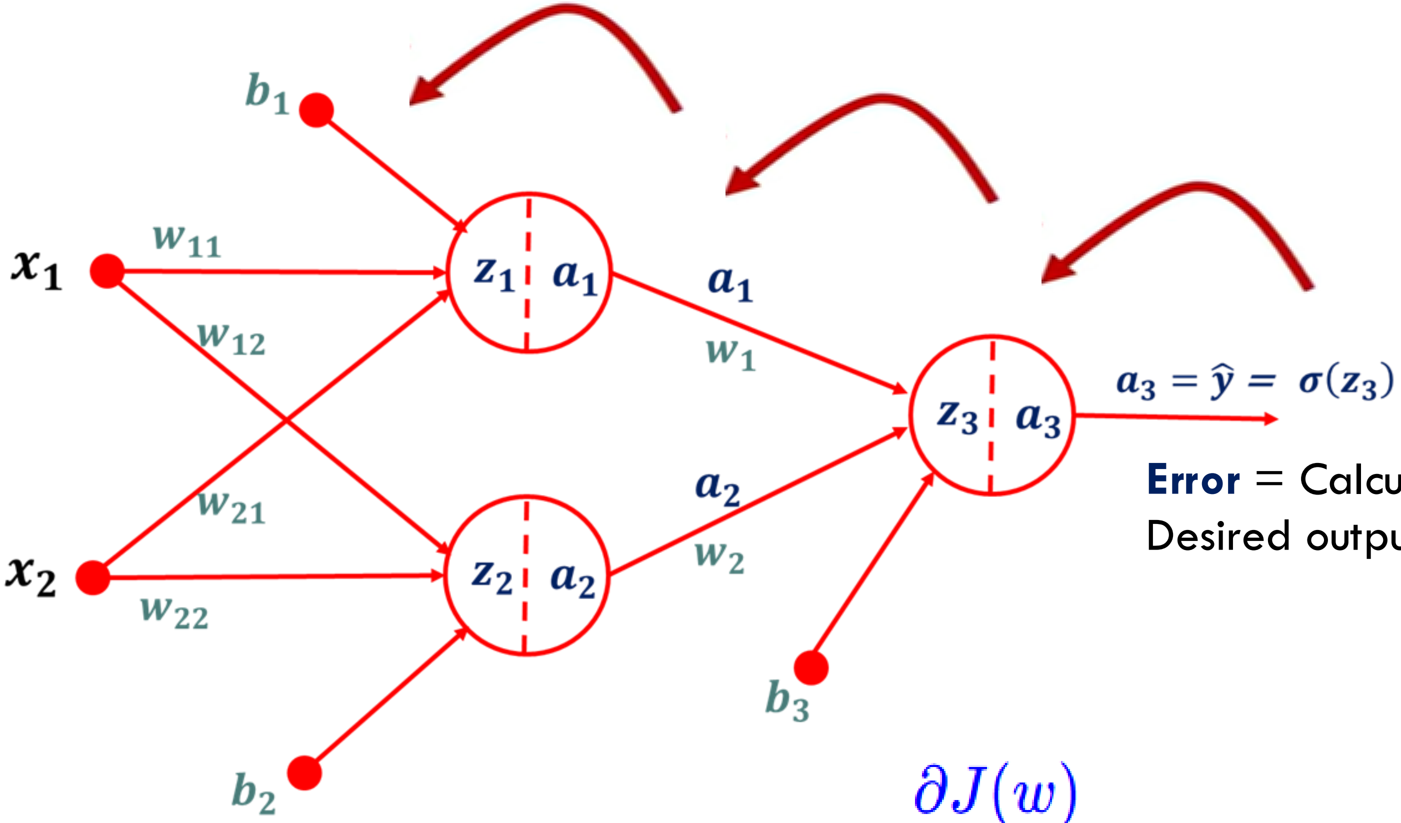
BACKPROPAGATION RE-DISCOVERED – 1986

- **Core Idea:** Backpropagation trains multilayer perceptrons efficiently by sending error signals backward using the chain rule.
- **Breakthrough:** Hidden layers enable networks to learn internal representations, not just input–output mappings.
- **Impact:** Solved the XOR problem, addressing a major criticism from Minsky & Papert (1969).
- **Representation Learning:** Networks can automatically discover features instead of relying on handcrafted ones.

Historical Role: Ended the "NeuralNet dark age", paving the way for modern deep learning.

Rumelhart, D., Hinton, G. & Williams, R. **Learning representations by back-propagating errors.** Nature 323, 533–536 (1986). https://doi.org/10.1038/323533a0

(ERROR) BACKPROPAGATION RE-DISCOVERED – 1986



$$J = -\left(y \log \hat{y} + (1 - y) \log(1 - \hat{y})\right)$$

Error = Calculated – Desired output

The **credit assignment problem** was solved through **backpropagation with gradient descent.** The **errors to be propagated backward** through multiple layers, enabling efficient training of deep networks.

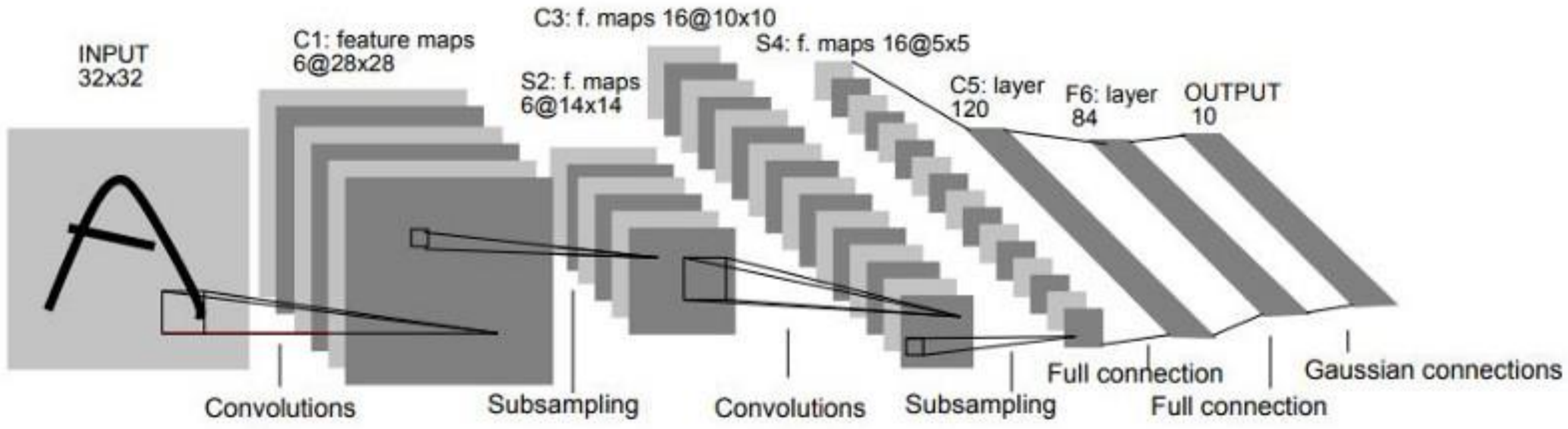
$$w_i := w_i - \alpha \frac{\partial J(w)}{\partial w_i}$$

Credit assignment / weight update

$$\frac{\partial J}{\partial w_{ij}} = \frac{\partial J}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_3} \cdot \frac{\partial z_3}{\partial a_j} \cdot \frac{\partial a_j}{\partial z_j} \cdot \frac{\partial z_j}{\partial w_{ij}}$$

CONVOLUTIONAL NEURAL NETWORK (CNN) – 1989

Convolutional Neural Network (CNN) – Yann LeCun et. al (1989*) introduced network that uses convolutional layers, revolutionizing image and spatial data processing (predecessor of Neocognitron).

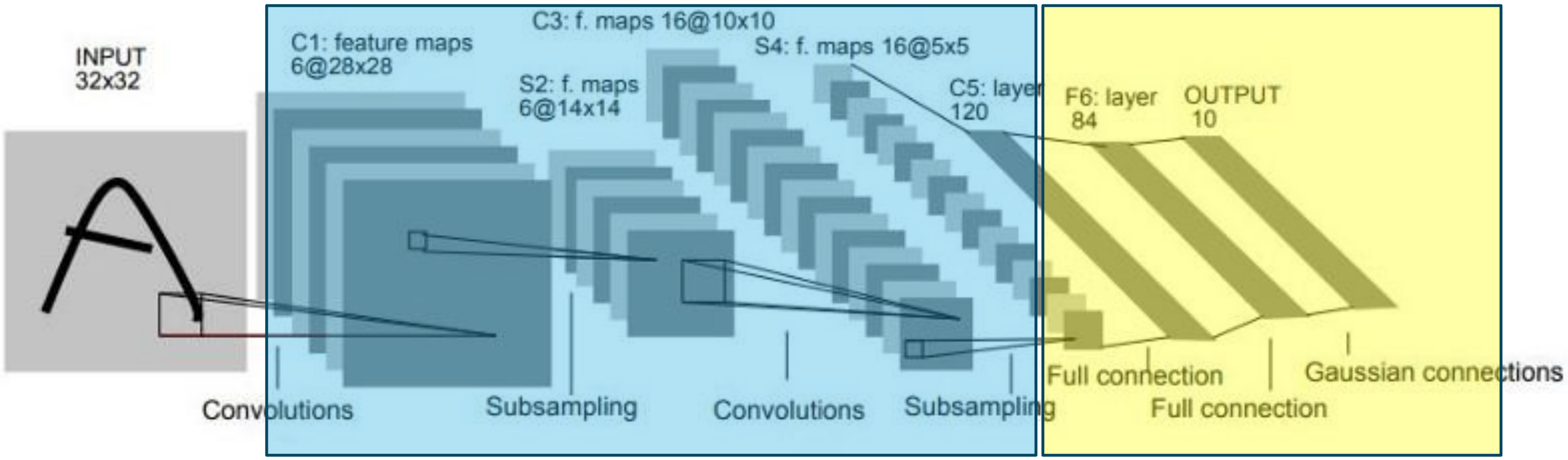


*Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel; Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Comput* 1989; 1 (4): 541–551. doi: https://doi.org/10.1162/neco.1989.1.4.541

CONVOLUTIONAL NEURAL NETWORK (CNN) – 1989

Core Ideas of LeNet

1. Learning to filter (filter weights were learned)
2. Shared Weights (Convolution)
3. Divide & Conquer



Divide [Feature Extraction] **Conquer [MLP]**

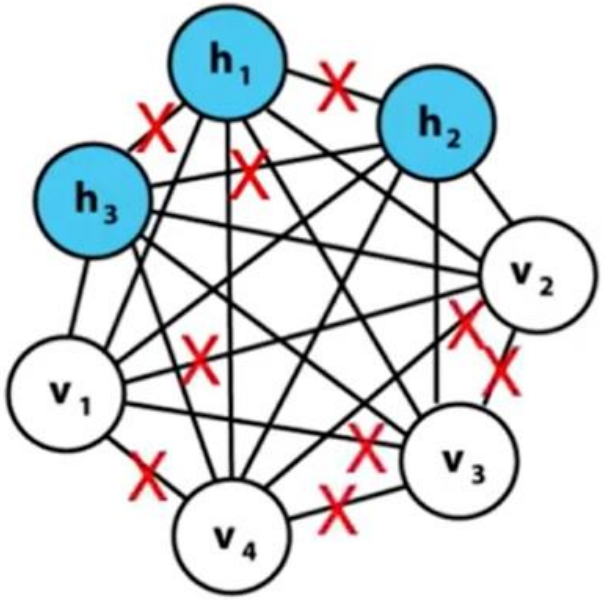
EARLY CONVOLUTIONAL NEURAL NETWORKS (CNN)

[redacted]

Version	Year	Key Enhancements
LeNet-1	1989	Early CNN; used for digit recognition on low-resolution images (16×16); featured simple architecture with convolution + subsampling.
LeNet-2	1990	Improved feature extraction and training strategy; used modified learning algorithms.
LeNet-3	1993	Incorporated better weight sharing and enhanced structure; not widely published but used internally.
LeNet-4	1993	Used discrete cosine transform (DCT) for input preprocessing; had more hidden units.
LeNet-5	1998	Most well-known version; used for digit recognition on MNIST dataset; introduced structured convolution + pooling layers, ReLU-like activations, and fully connected output layers.

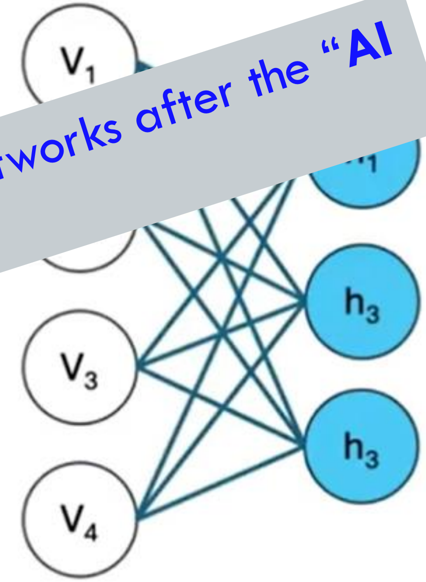
LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE.

RESTRICTED BOLTZMANN MACHINE (RBM) – 2006



If we remove internal connections b/w hidden and visible nodes

Boltzmann machine, … structure like brain, learns probability distribution

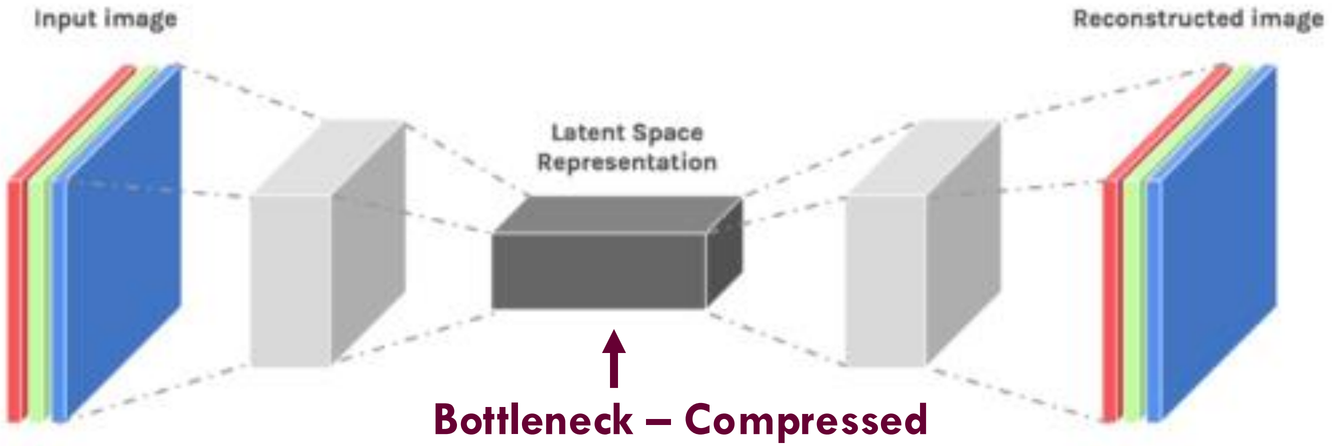


It reignited interest in neural networks after the "**AI winter.**"

RBM, Looks like MLP but its not, built to learn probability distribution

- Solved the training barrier for **deep networks** (unsupervised pretraining) and vanishing gradient problem (Idea: don't initialize randomly and train end-to-end. Instead, pre-train each layer greedily, one at a time, using RBMs).
- Introduced **generative modeling** into practical neural nets.
- Established the principle of learned feature hierarchies → cornerstone of modern deep learning.

AUTOENCODER – 2000s



Bottleneck – Compressed Representation

Autoencoders, Minimum Description Length and Helmholtz Free Energy

Geoffrey E. Hinton
Department of Computer Science
University of Toronto
6 King's College Road
Toronto M5S 1A4, Canada

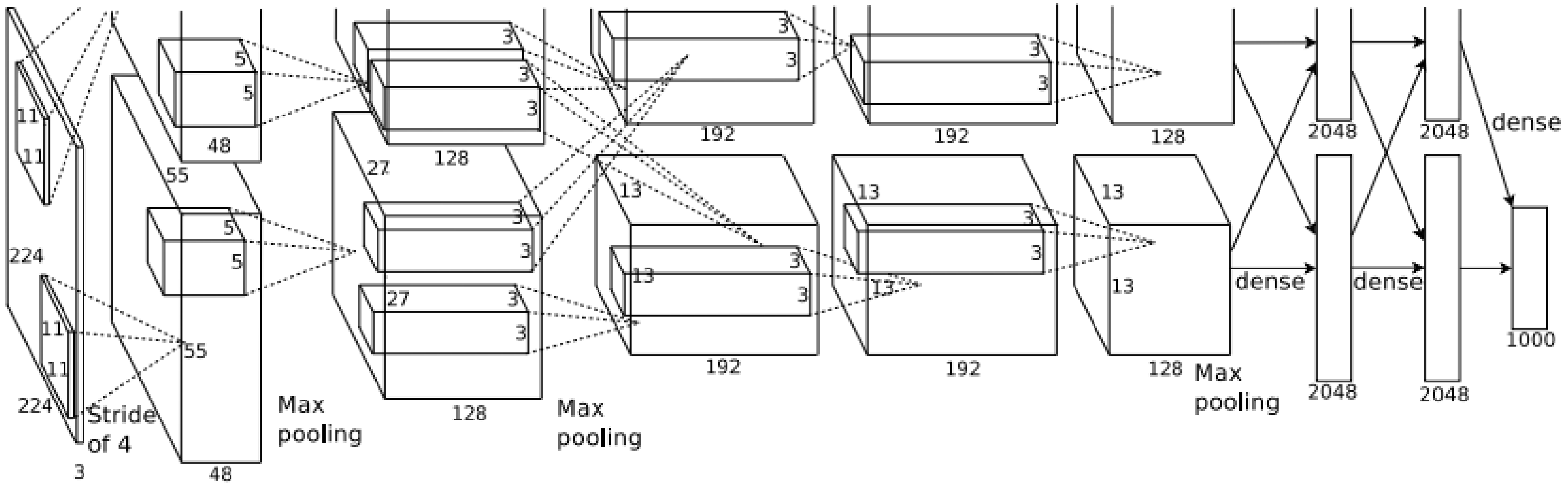
Richard S. Zemel
Computational Neuroscience Laboratory
The Salk Institute
10010 North Torrey Pines Road
La Jolla, CA 92037

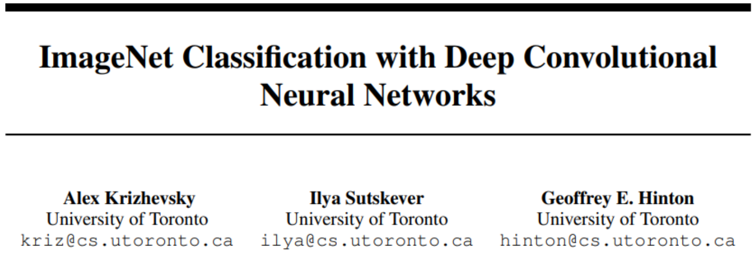
Geoffrey E. Hinton et al. - NeuralPS'93

<u>In the mid-2000s:</u>

- **RBMs** were used to pre-train deep networks layer by layer (Geoff Hinton, 2006).
- Later, it was shown that **stacked autoencoders** can do the same pretraining, but with easier optimization (no sampling needed).
- This transition from **RBM pretraining** → **autoencoder pretraining** was one of the **steppingstones to modern deep learning**

SUCCESS OF CNNs – ALEXNET - 2012

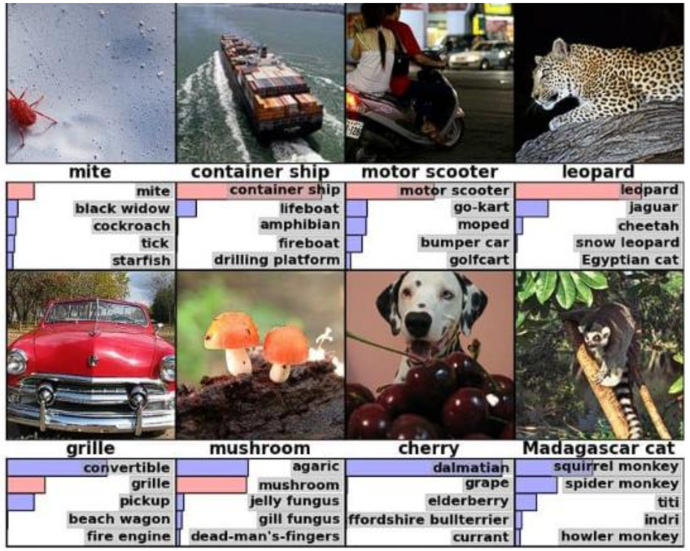


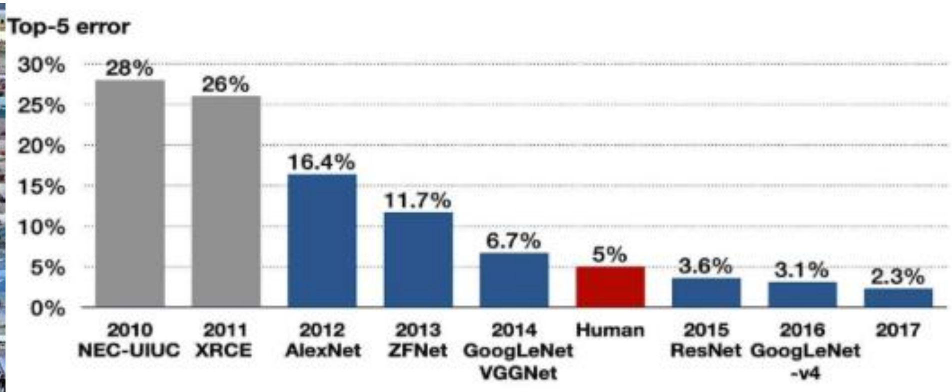

ImageNet Classification with Deep Convolutional Neural Networks

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





ageNet winners by year –

SUCCESS OF CNNs – 2012 -

Model	Year	Authors / Group	Key Contributions	Impact
LeNet-5	1998	Yann LeCun et al.	First CNN for digit recognition; convolution + pooling; tanh/sigmoid activations.	Proved CNNs effective on vision (MNIST, checks).
AlexNet	2012	Alex Krizhevsky et al.	8-layer CNN; ReLU activation; dropout; GPU training; data augmentation.	Won ImageNet 2012, triggered modern deep learning era.
VGGNet	2014	Oxford VGG group (Simonyan & Zisserman)	Very deep (16–19 layers); uniform **3×3 conv filters**; simplicity in design.	Showed depth improves performance; widely used backbone.
GoogLeNet (Inception)	2014	Google (Szegedy et al.)	Inception modules (multi-scale conv); reduced parameters with 1×1 conv.	Efficient deep networks; won ImageNet 2014.
ResNet	2015	Microsoft (He et al.)	Introduced **residual connections (skip connections)** enabling 152+ layers.	Won ImageNet 2015; solved vanishing gradient problem.
DenseNet	2017	Huang et al.	Dense connectivity (each layer connected to all later layers).	Improved parameter efficiency and gradient flow.
EfficientNet	2019	Google (Tan & Le)	Compound scaling of depth, width, and resolution.	Achieved SOTA accuracy with fewer parameters.

GENERATIVE ADVERSARIAL NETWORKS (GANs) – 2014 -

- **We've only seen discriminative models so far**
  - Given an image **X**, predict a label **Y**
  - Estimates **P(Y|X)**
- **Discriminative models have several key limitations**
  - Can't model **P(X)**, i.e. the probability of seeing a certain image
  - Thus, can't sample from **P(X)**, i.e. **can't generate new images**
- **Generative models (in general) cope with all of above**
  - Can model **P(X)**
  - Can generate new images

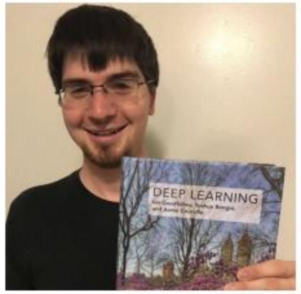
Generative Adversarial Nets

Ian J. Goodfellow, Jean Pouget-Abadie*, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair†, Aaron Courville, Yoshua Bengio‡
Département d'informatique et de recherche opérationnelle
Université de Montréal
Montréal, QC H3C 3J7

Abstract

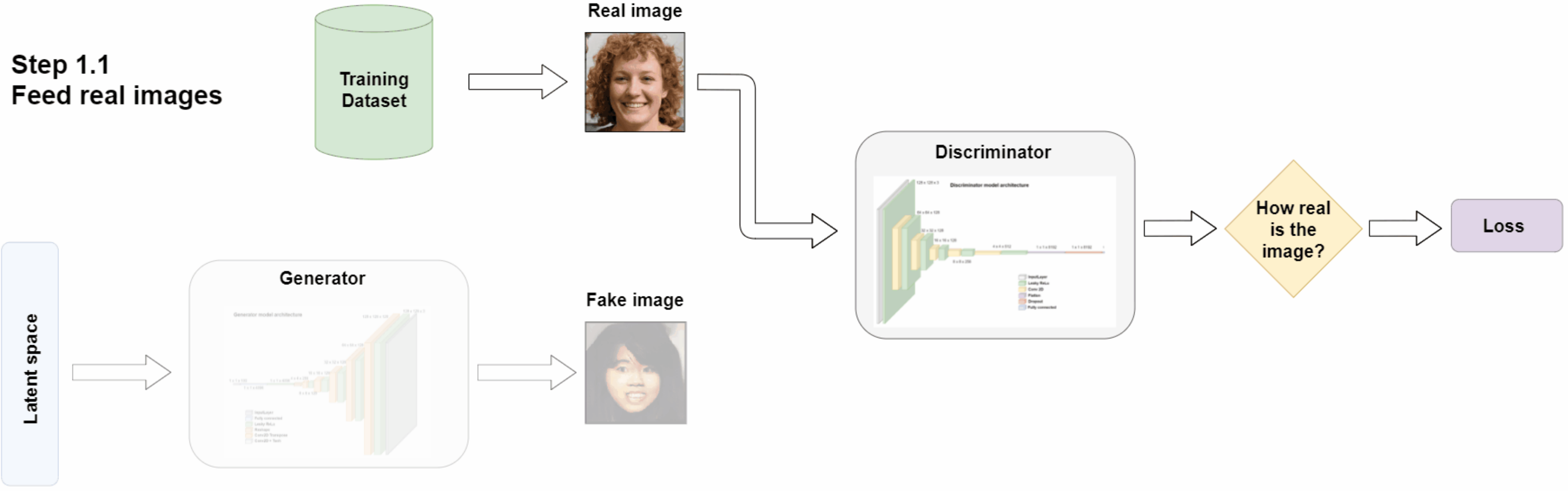
We propose a new framework for estimating generative models via an adversarial process, in which we simultaneously train two models: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G. The training procedure for G is to maximize the probability of D making a mistake. This framework corresponds to a minimax two-player game. In the space of arbitrary functions G and D, a unique solution exists, with G recovering the training data distribution and D equal to $\frac{1}{2}$ everywhere. In the case where G and D are defined by multilayer perceptrons, the entire system can be trained with backpropagation. There is no need for any Markov chains or unrolled approximate inference networks during either training or generation of samples. Experiments demonstrate the potential of the framework through qualitative and quantitative evaluation of the generated samples.

Goodfellow et al. 2014 (arXiv)



Ian Goodfellow

GENERATIVE ADVERSARIAL NETWORKS (GANs) – 2014 -



1. Forward pass (through L layers):
   For each layer $l = 1, 2, \ldots, L$:

$$h_l = \sigma(W_l h_{l-1} + b_l), \quad h_0 = x$$

   Final output:

$$\hat{y} = h_L$$

2. Loss computation:

$$\mathcal{L} = L(\hat{y}, y)$$

3. Backward & parameter update:
   For each layer $l = 1, 2, \ldots, L$:

$$W_l \leftarrow W_l - \alpha \nabla_{W_l} \mathcal{L}, \quad b_l \leftarrow b_l - \alpha \nabla_{b_l} \mathcal{L}$$

Real or Fake?

Karras et al. 2019; A Style-Based Generator Architecture for Generative Adversarial Networks

TRANSFORMERS – 2017

- (**Final core idea**) A novel architecture that relies on **self-attention mechanism**, revolutionizing NLP and forming basis for BERT and GPTs.
- Core idea coming from **corporation**, not from academia, new direction.
- Most of the CNNs, LLM, Foundation models are coming from industry, any implications?

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

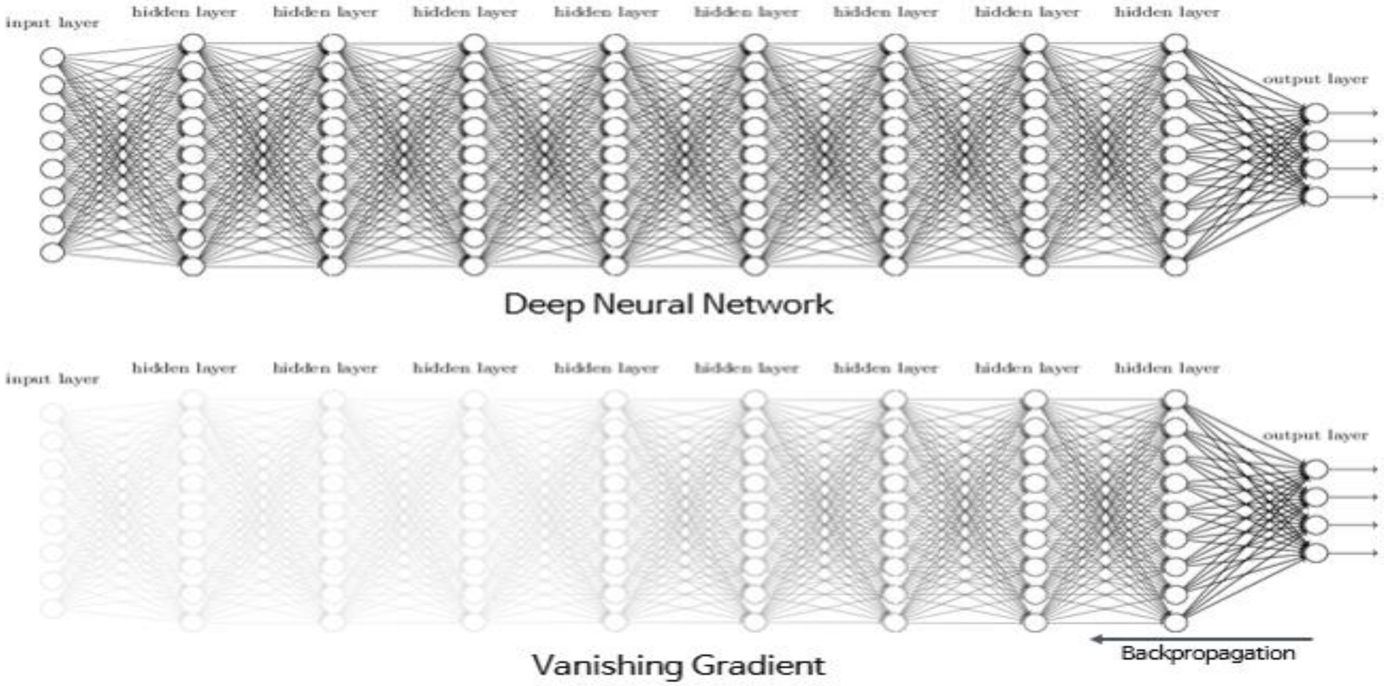
The dominant sequence transduction models are based on complex recurrent or convolutional neural networks in an encoder-decoder configuration. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, 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 41.0 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.

Vaswani et al. 2017 (arXiv)

TRANSFORMERS – 2017

1. **Long-range Dependencies → Captures global context in sequences better than traditional architectures.**

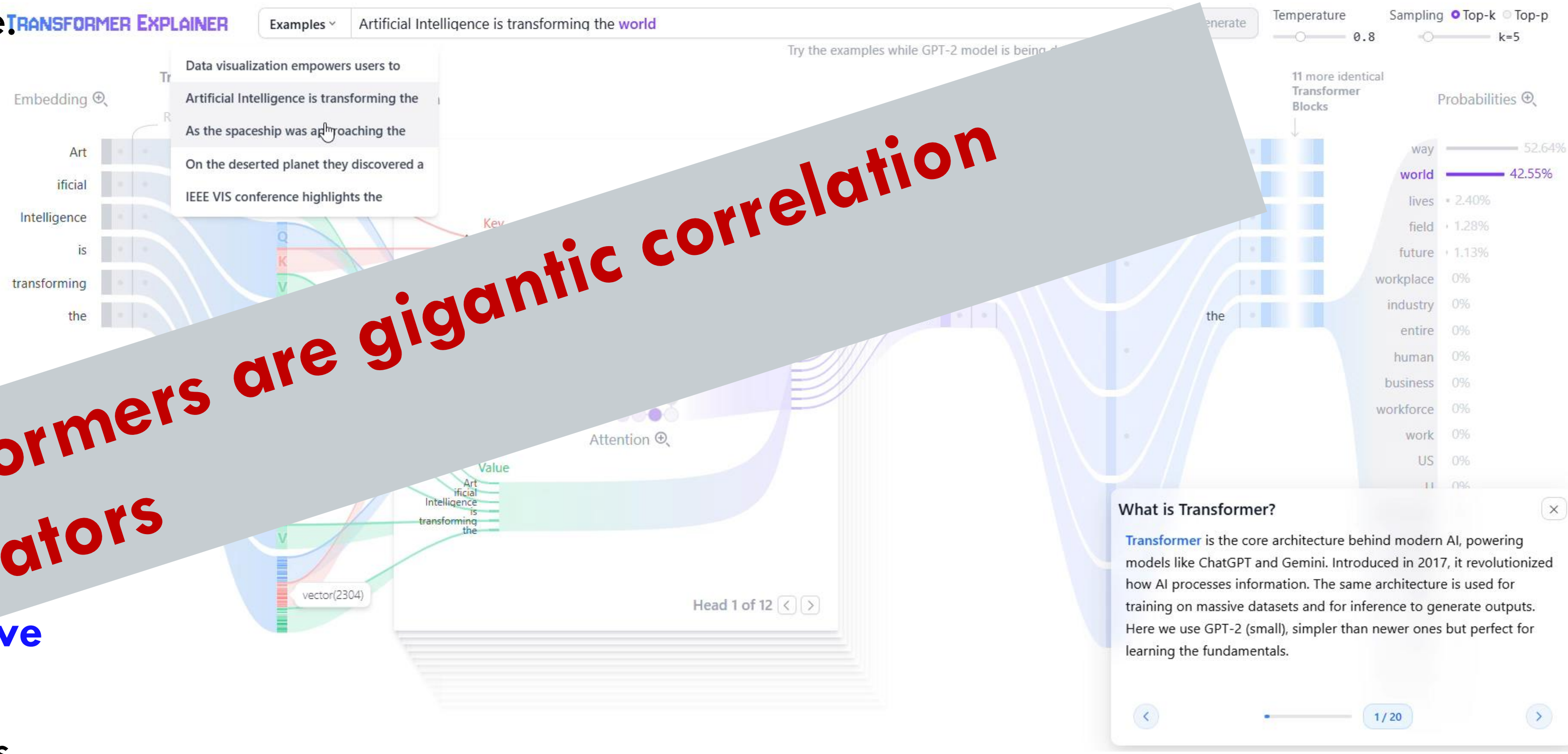
- **RNNs** and **LSTMs** cannot retain information over long sequences due to vanishing gradients problem (decay of information through time). Thus, difficult to capture dependencies between distant words in a sentence.
- It means deep layers remains untouched, no training.
- **Problems** : Hallucinations , lack of generalization etc.



TRANSFORMERS – 2017

2. Attention Mechanism ↔ Selective Attention

- Brain amplifie relevant signal, suppresses noise.
- Transformer attention focusses on important tokens similarly.



Transformers are gigantic correlation calculators

3. Q–K–V Mechanism ↔ Working Memory

- Mirrors **content-addressable memory**
- Query retrieves relevant stored infor (key–value pairs).

4. Multi-Head Attention ↔ Multip gnitive Streams

- Brain uses multiple simultaneous attention filters.
- Multi-head attention = parallel feature extraction channels.

TRANSFORMERS

📊 Statistical pattern matchers at massive scale

🔮 Trained on next-token prediction: **not explicit reasoning**

🏆 Extraordinary at language, perception & generation

⚠️ Not truly reasoning — interpolating learned patterns

KEY FACTS

2017
Year introduced

175B
GPT-3 parameters

~100T
Training tokens (GPT-4 est.)

Self-Attn
Core mechanism

THE DEBATE: ARE WE ON THE RIGHT PATH?

OPTIMISTS

Hinton, Altman

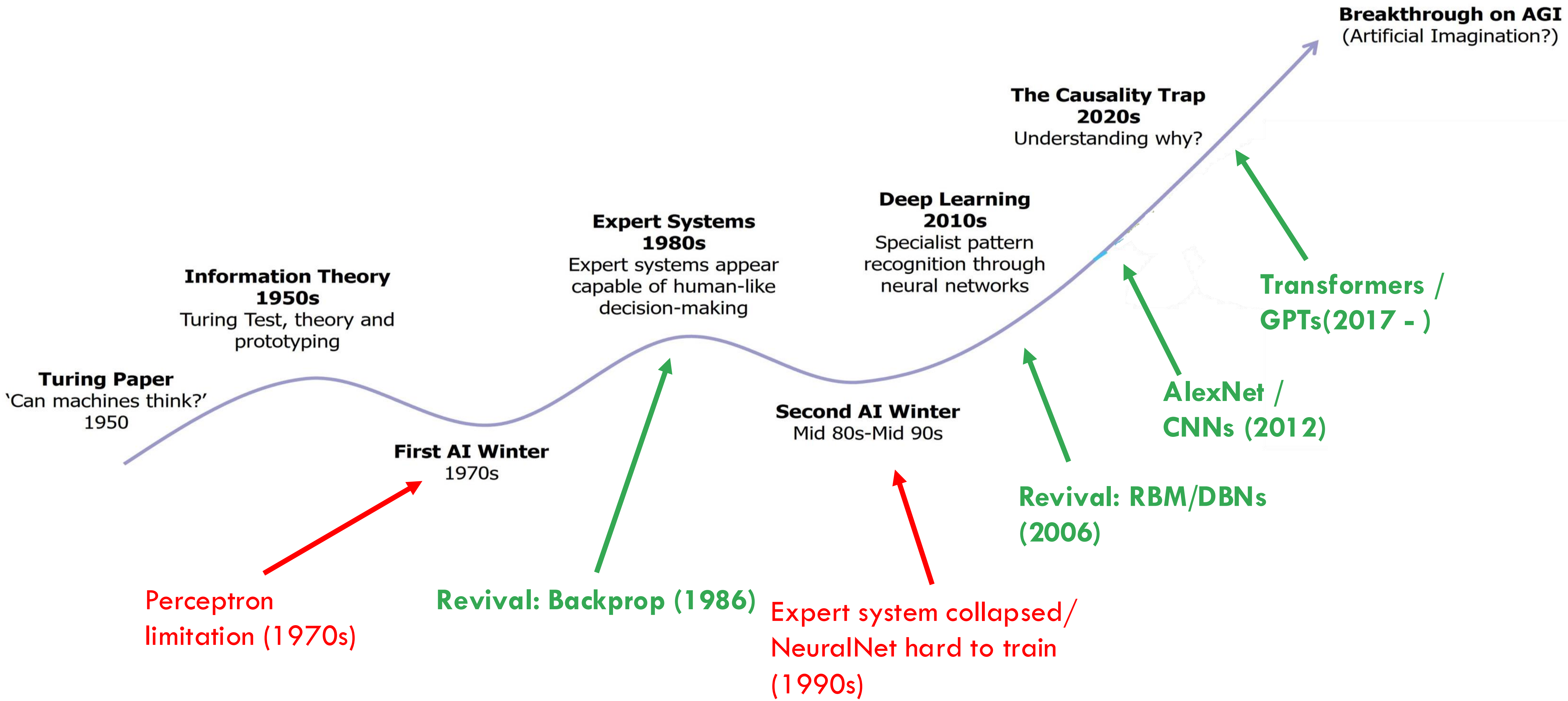
- → Scaling + data + architecture will reach **AGI**
- → Emergent capabilities suggest something deeper
- → **May not need explicit reasoning to be intelligent**
- → **We may be closer to AGI than most think**

SKEPTICS

Gary Marcus, Chomsky, early Minsky school

- → **Sophisticated autocomplete** not true understanding
- → Lacks grounding, causality, real world models
- → **Brittle failures** reveal pattern matching not reasoning
- → We may be on the wrong — or incomplete path

AI WINTER - SUMMER TIMELINE



CHALLENGES

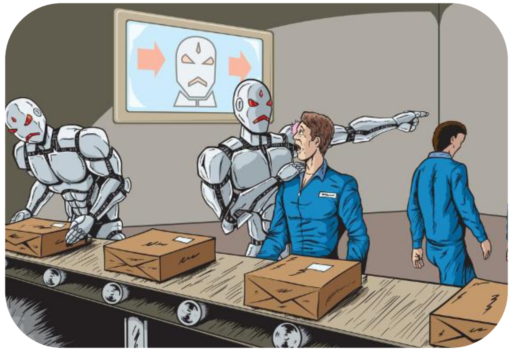
Ethical Bias & Fairness

Overhyped Expectations / Scaling limits

Data Privacy & Security Risks

Regulation & Governance Gaps

Energy Consumption & Environmental Impact



Job Displacement & Skill Gaps

Explainability / Transparency ("Black Box" issue)

Misinformation / Deepfakes / Trust Erosion

Thank you

Dr. Rizwan Ahmed Khan
Professor, SMCS,
Department of Computer Science, **IBA Karachi.**
Adjunct Faculty, **Universite Jean Monnet, France.**

RizwanKhan@iba.edu.pk
https://sites.google.com/site/drkhanrizwan17/home

HUBEL AND WIESEL – 1959 (NOBEL PRIZE (1981) IN PHYSIOLOGY)

1959 marks the year when **David Hubel and Torsten Wiesel** (neurophysiologists) began their landmark experiments on the **visual cortex of cats** at Harvard Medical School.

This study revealed:

1. Neurons in the **primary visual cortex (V1)** respond selectively to **edges and oriented lines.**
2. Visual processing occurs **hierarchically,** from simple to complex features.

Receptive fields of single neurons in the cat's striate cortex. J Physiol. 1959 148(3):574–591.

doi: 10.1113/jphysiol.1959.sp006308; https://pmc.ncbi.nlm.nih.gov/articles/PMC1363130/

NEOCOGNITRON - 1980

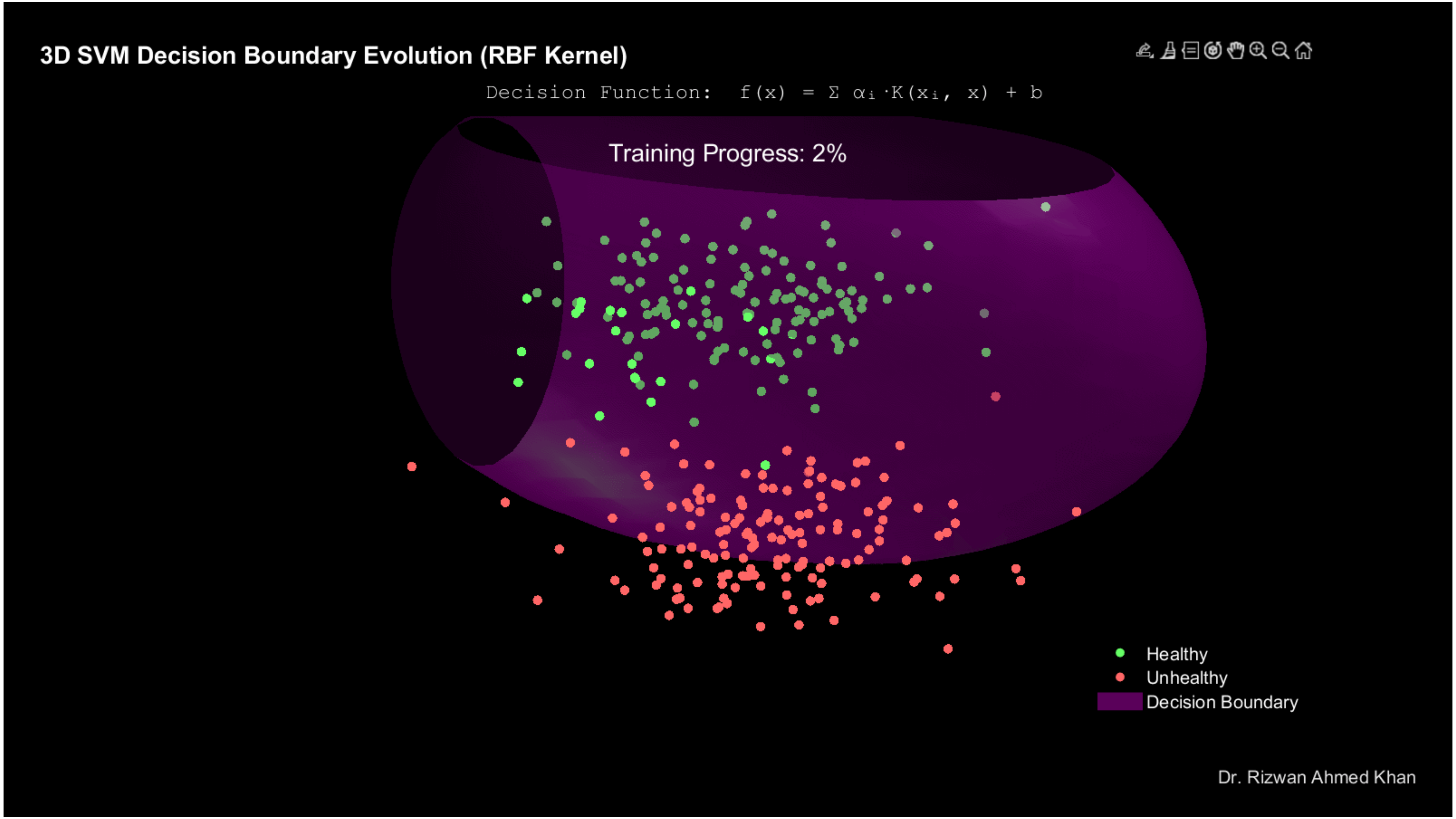
https://www.cs.princeton.edu/courses/archive/spr08/cos598B/Readings/Fukushima1980.pdf

NEOCOGNITRON - 1980

https://www.cs.princeton.edu/courses/archive/spr08/cos598B/Readings/Fukushima1980.pdf

NUMBER CRUNCHING

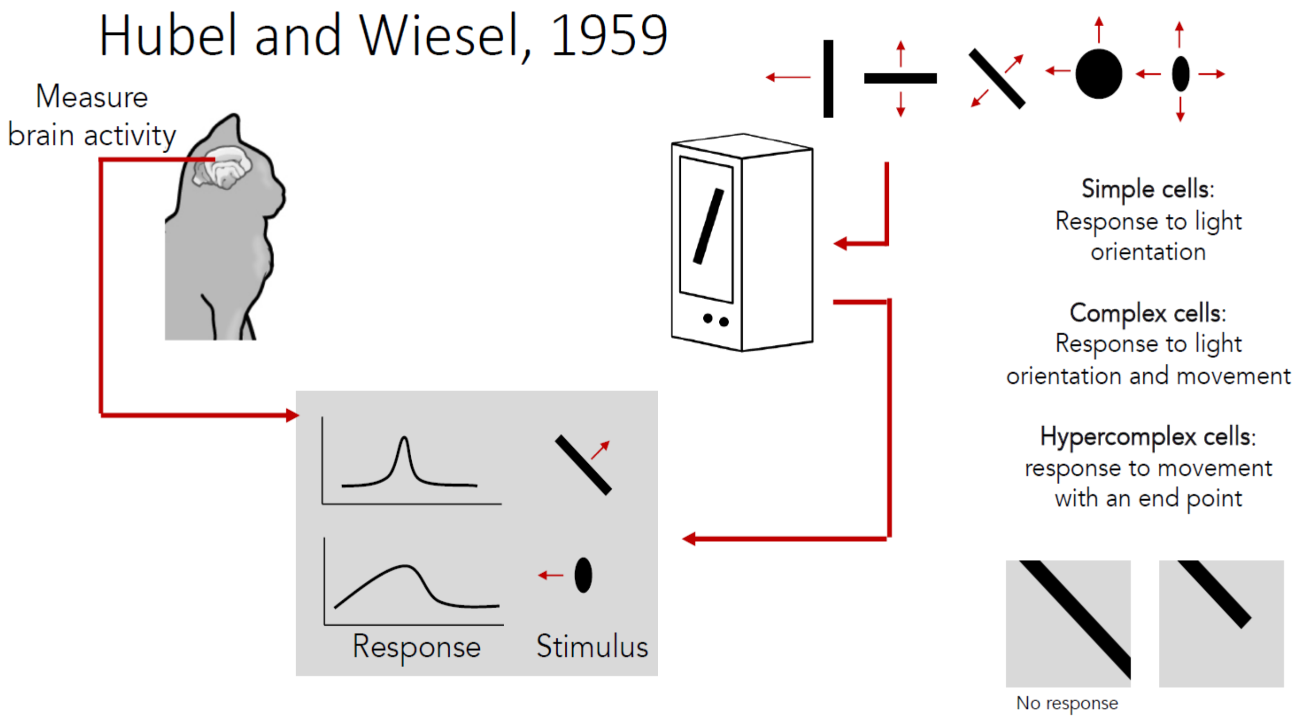
$$H = \{h \mid h : X \rightarrow Y\}$$



HANDCRAFTED FEATURES -1980s

In the 1980s, most of machine learning and computer vision relied on **handcrafted features** (edges, corners, textures), and the main research challenge was designing the "right" features for each task.

Around this time, **Fukushima's Neocognitron (1980)** introduced the idea of a hierarchical, multi-layered neural network that could **automatically learn features** from data, an early inspiration for today's convolutional neural networks (CNNs).



Receptive fields of single neurons in the cat's striate cortex. J Physiol. 1959 148(3):574–591.

doi: 10.1113/jphysiol.1959.sp006308; https://pmc.ncbi.nlm.nih.gov/articles/PMC1363130/