**Deep Learning Simply**

**Deep Origins**

The story of deep learning, as viewed by the general public, is a recent phenomenon. In reality it is actually a long and old story rooted in humankind’s fascination with understanding reasoning: where it occurred, the processes­­-­ associated with it, their interconnections, if and how it could be recreated.

Though it is now common knowledge that

1. Thinking occurs in the brain
2. Our brains have billions of interconnected neurons or brain cells
3. These brain cells or neurons work collectively as a connected “intelligence” network to enable us to think, to feel, and, to act

this common knowledge is actually the result of millennia of “not so common” research and investigation; research that has been punctuated by short, intense bursts of progress and long, debilitating periods of regress.

Historical evidence shows that the brain has been a constant object of study from antiquity (B.C.E). Scientists from ancient times, across several civilizations, studied the brain to comprehend its functions and used that knowledge for medical purposes (reference: images of Indian historical figures who studied the brain including Charaka, Sushruta, trepanation in Harappa: <https://bit.ly/38tt9L5>; Babylonian, Egyptian efforts).

In the early Common Era (C.E.), scientists and researchers who shared this fascination with the workings of the brain continued their efforts, and by the first century, a general physical description of the brain was available to the scientific community.  Basic structures such as the soft and hard layers encasing the brain were identified, and the brain was divided into functional regions, known as ventricles.

Building upon this work, over the next few centuries, physicians concluded that mental activity occurred in the brain rather than the heart, thus concurring with what some pre-common era scientists had already suggested.  What seems like a small step today was actually a huge step forward for that time. Some scientists also concluded that the brain was the seat of the soul -- one of three "souls" found in the body, each associated with a principal organ.  Other scientists were of the opinion that the brain was a cold, moist organ formed of sperm!

By A drawing of a person

Description automatically generatedthe Middle Ages, the anatomy of the brain was broadly believed to be consolidated around three principal divisions.  Each division localized the site of a different mental activity.  Imagination was believed to be located in the anterior ventricle and memory in the posterior ventricle, with reason in between.  Yet, there was no consensus on where sensory processing and storing of the inputs and the processing of the inputs received from the five senses (what was then called simply as “common sense”) resided. At that time, input from the five senses together was known simply as “common sense.”  There was no consensus on where the storage and processing of input from common sense took place. Eleventh century scientists proposed that this “common sense” was housed in a "faculty of fantasy," one that received "all the forms which are imprinted on the five senses." Memory then preserved what common sense received.

But by the 14th century, opinions had changed. It was now believed that common sense lay in the *middle* of the brain.  Historical records also suggest that many fundamental questions regarding the functions of the brain (including those related to common sense) remained open to debate. Indeed, this is true to this day, but at a far more nuanced level of detail. Such differences of opinion only underscore how little was known, then, of the brain's anatomy, let alone its physiology or function and the same holds true today.

In the Renaissance period, physicians began to dissect the brain with much greater frequency, more so at the end of the fifteenth century. This, mid-sixteenth century anatomy illustration demonstrates such a dissection. Of particular note is the famous Leonardo da Vinci who dissected and drew the brain.  A picture containing text, map

Description automatically generatedDa Vinci's images were considerably more anatomical in nature than a lot of his peers’. He systematically examined the relationship between the brain, olfactory nerves, and optical nerves through experiments with wax injections, which helped him model the ventricles of the brain.

Leonardo sketched the brain from many different perspectives, looking closely at the ventricles and the origins of the nerves in the medulla.  Records suggest that the more he looked, the less sure he became about the function of each section! It is now known that one of his quests was to find the location of “common sense”.  The other was to locate the seat of the soul!

SixA close up of a sign

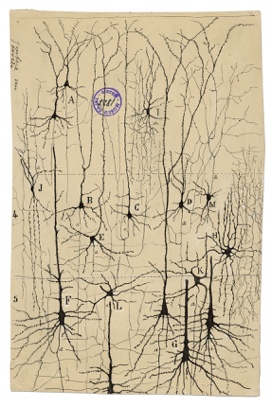
Description automatically generatedteenth and early seventeenth-century anatomists contributed a great deal to the physical description of the brain. Terms such as cerebrum, cerebellum and medulla were popularized. But they made few significant advances in their *understanding* of its functioning.  It was not until the 17th century that views on the anatomy of the brain changed significantly. A close up of text on a white background

Description automatically generatedScientists began to advocate for more careful explorations of the cortex and the ventricles (two images of the brain, from the late sixteenth and mid-seventeenth centuries). The brain, finally, had a modern physiology, grounded in research and verifiable experiments. The basic concept and principles of neurology were established, and scientists believed that the soul no longer had a home in the brain.

Then, in the late 19th century, Spanish physician Santiago Cajal (reference: figure of Santiago) formally identified neurons by staining and studying brain tissue in great detail. Cajal’s two brilliant insights — that every neuron in the brain is separate and that neurons communicate across synapses — came to be known as the neuron doctrine. Because that gap between neurons is too small to see through a light microscope, Camillo Golgi and other rigorous scientists of Cajal’s day at first dismissed the neuron doctrine as a fantasy. It would take another half-century until a new instrument, the electron microscope, could finally confirm what Cajal had seen in his mind’s eye — and carefully sketched out in thousands of stunning pen-and-ink diagrams.

He published his observations in 1894. Subsequently, in the early parts of the 20th century, researchers began the process of understanding exactly how these neuronal cells *functioned*.

A person wearing a suit and tie

Description automatically generated *Cajal’s graceful drawings of neurons show them as separate, individual cells. He was the first to realize that the nervous system is not a network of continuous fibers, as was widely believed at the time.*

Then a left shift event occurred! Computers arrived and gained momentum, triggering ideas of transitioning research and experimentation from the physical world to the world of computers. Brain behavior research was considered an ideal candidate for computerized research, given the myriad constraints involved in acquiring specimens, performing physical experiments, and deducing conclusions from results and verifying them.

Thus, in the mid 20th century, spurred by curiosity, inspired by the progresses made in the understanding of the brain and sensing the opportunity presented by their access to computing resources, Computer Scientists (a newly minted term) began experimenting with computerized *simulations* of the cognitive process of the brain. A subset of them chose to simulate cognitive processes by creating “artificial neurons” and then connecting them in “artificial neural networks,” thus aiming to loosely mimic the physical organization and functioning of brain cells as seen in experiments involving small, localized sections of the brain.

Warren McCulloch and Walter Pitts created a computer model using a combination of mathematics and algorithms to mimic the thought process. In their model, small layers of artificial neurons passed input information to other neuronal layers until the final layer output values. Then in his paper “The Perceptron: A Perceiving and Recognizing Automaton”, Rosenblatt showed the new avatar of the McCulloch-Pitts neuron – the ‘Perceptron,’ which had true learning capabilities and could perform binary [classification](https://machinelearningknowledge.ai/glossary/binary-classification/) on its own. The Perceptron inspired a revolution in the research of shallow neural network for years to come. The Forward Propagation Artificial Neural Network had arrived!

An old photo of a person

Description automatically generated 

Frank Rosenblatt

A screenshot of a cell phone

Description automatically generated

Henry J Kelley

Following this, in 1960, Henry J. Kelley improved this model by developing the basics of Continuous Back Propagation. This approach introduced the notion of feedback control into the “learning” process. Although this key concept of back propagation existed in the early 1960s, it became of practical use only around 1985.

You might have noticed by now that all of the work in Artificial Neural Networks that we have discussed thus far deal with what we would today call “shallow” networks, i.e. Artificial Neural Networks that just have a few connected layers between the input and the output as well as being comprised of a relatively small number of connected neurons as a whole. But Deep Learning is defined as the process of processing signals and encoding knowledge using Deep Neural Networks, which are large, deeply stacked, multi-layered networks of artificial neurons comprised of millions or more networked artificial neurons.

So when did this shift from simple shallow networks to complex deep networks occur?

The earliest efforts in developing deep learning algorithms actually date back to 1965, when Alexey Grigoryevich Ivakhnenko and Valentin Grigorʹevich Lapa used models with polynomial activation functions, which were subsequently analyzed statistically. Then, just as things seemed to be gaining momentum, in 1970, Marvin Minsky and Seymour Papert published the book “Perceptrons” in which they show that Rosenblatt’s Perceptron could not solve complicated functions like XOR. For such function Perceptrons should be placed in multiple hidden layers which compromises the Perceptron Learning Algorithm. This was a setback for “bottom up” Artificial Intelligence (AI).

Around the same time, there was mounting sentiment that millions had been spent on AI research with the hope that it would provide a strategic technological advantage in the Cold War, but little had come out of it. There was strong criticism of AI research programs from the US Congress. To add to this, in 1973, leading mathematician Professor Sir James Lighthill gave a damning health report on the state of AI in the UK.

Lighthill’s view was that machines would only ever be capable of an "experienced amateur" level of chess. Common sense reasoning and supposedly simple tasks like face recognition would always be beyond their capability. Funding for the industry was slashed, ushering in what became known subsequently as the AI winter. It was a period of intense setback in the research and development of Artificial Neural Networks and Artificial Intelligence in general. Severe cutbacks in funding plagued deep learning and AI research as a whole.

However, despite the lean times, passionate researchers carried on the work without funding. Kunihiko Fukushima developed an artificial neural network, called Neocognitron, in 1979, which used a multi-layered and hierarchical design. This design enabled computers to learn to recognize visual patterns. The backpropagation method was enhanced to feed errors in the output to influence and control the training of the models. This approach became widely popular when Seppo Linnainmaa wrote his master’s thesis, including FORTRAN code illustrating the backpropagation technique. The backpropagation concept was applied to neural networks and shown to work but made little impact. AI work then was primarily focused on top-down, rule based expert systems. Bottom up AI was still languishing.

The first signs of revival were displayed when in 1985 Hinton and Williams demonstrated backpropagation in a neural network which could provide interesting distribution representations. Yann LeCun followed this up by providing the first practical demonstration of backpropagation at Bell Labs in 1989 by combining convolutional neural networks with backpropagation to read handwritten digits. The combination of convolutional neural networks with a backpropagation system was then used to read the numbers of handwritten checks, spurring business interest.

Interestingly, though this was a period of renewed interest, the 1985-90s actually are considered by the scientific AI community as the *second lull* in artificial intelligence. Hardware and software issues plagued the progress of research in neural networks and deep learning. Deep Learning algorithms, while producing good results in lab conditions, struggled to scale well to industrial proportions, were quite unstable and seemed unable to generate consistent results.

Despite these adverse conditions, passionate individuals continued to move the needle. Vladimir Vapnik and Dana Cortes developed the support vector machine, which was a data driven system for mapping and recognizing similar data in 1995. Long short-term memory (LSTM) was developed in 1997 by Juergen Schmidhuber and Sepp Hochreiter developed Recurrent Neural Networks (RNN).

Then the next significant deep learning advancement happened due to progress help from unexpected quarters!

Around 1999 computers began to take advantage of graphics processing units (GPUs), which were introduced to accelerate the massive amounts of mathematical operations needed for fast image processing and display. Faster processing meant increased computational speeds of 1000 times over a 10-year span. Deep learning researchers were quick to note that this was exactly what deep neural networks needed and pounced upon GPUs! This quickly led to neural networks competing with support vector machines. Neural networks began to offer better results using the same data, though still a tad slower when compared to support vector machines.

Then in 2001, a research report compiled by the META Group (now called Gartner) was published outlining the challenges and opportunities of massive, three-dimensional, data growth (Volume, Velocity and Variety). This report marked the onslaught of the Big Data and Data Driven Science phenomenon. It described the increasing volume and speed of data as increasing the range of data sources and types.

On another front, the cloud phenomenon occurred giving many organizations and individuals democratized access to large compute resource. Deep Learning, which was all data driven and needed massive compute power, suddenly was back in black in a democratized fashion. Fei-Fei Li, an AI professor at Stanford (who launched ImageNet in 2009) began assembling a free, open database of more than 14 million labeled images to the community at large. These images were the inputs needed to train deep neural nets. The speed of GPUs had increased significantly by 2011, making it possible to train deep convolutional neural networks *without the need for layer by layer pre-training*. Deep learning now held advantages in efficiency, efficacy and speed.

Then in 2012, Google Brain released the results of an unusual, free-spirited project called the Cat Experiment which explored the difficulties of unsupervised learning. Deep learning deployed supervised learning, which means the convolutional neural net is trained using labeled data like the images from ImageNet. This experiment used a neural net which was spread over 1,000s of computers where ten million unlabeled images were taken randomly from YouTube, as inputs to the training software.

This event is considered the tipping point for deep learning. It made students, researchers and corporations around the world sit up and take notice, and it triggered a wave of scientific efforts and breakthroughs backed by corporate investment. Consequently, today, deep learning powers several marquee apps like Amazon’s Alexa, Tesla’s Autopilot, Google’s Translation engine and many more.

The true appeal of deep learning is that it has improved the accuracy of a great number of computational tasks from 95 percent to 99 percent or better— that tricky few percent that can make an automated service feel as though it works by magic. Although the concrete, interactive code examples throughout this book will dispel this apparent wizardry, deep learning has indeed imbued machines with superhuman capability on complex tasks as diverse as face recognition, text summarization, and elaborate board games. Given all these prominent advances, it is unsurprising that “deep learning” has become synonymous with “artificial intelligence” in the popular press, the workplace, and the home.

2. See bit.ly/ aiindex18 for a review of machine performance relative to humans.

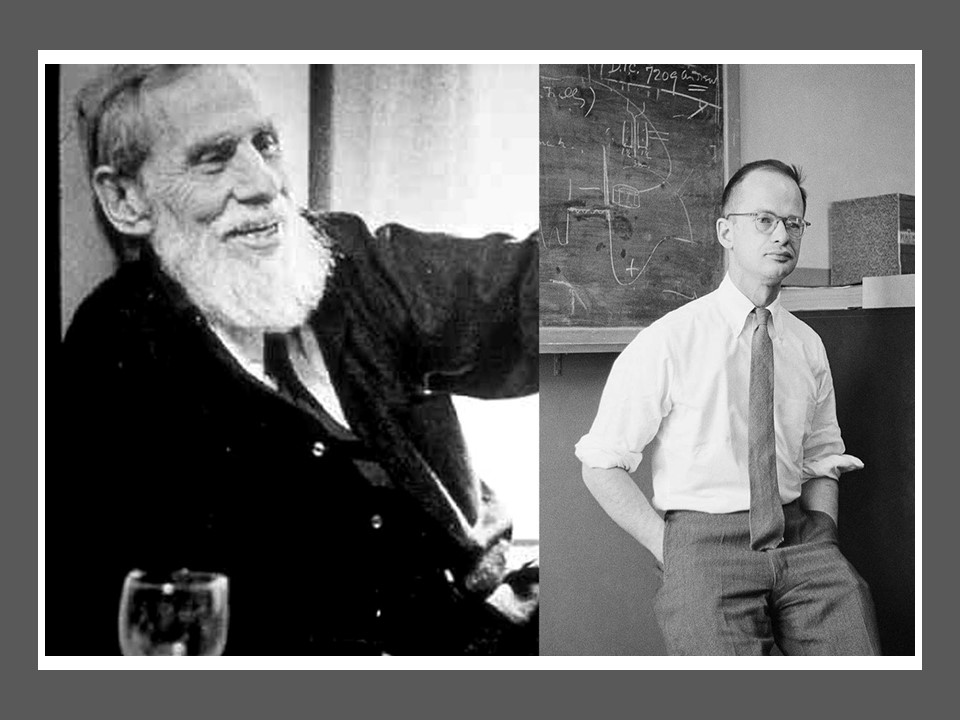
In summary, these are exciting times, because, as you’ll discover over the course of this book, perhaps only once in a lifetime has a single concept has cause disruption so widely in such a short period of time. We are delighted that you too have developed an interest in deep learning and we can’t wait to share our enthusiasm for this unprecedented, transformative technique with you.

A Summary Deep Learning Timeline

**1943**

1943

[**McCulloch Pitts Neuron – Beginning**](https://machinelearningknowledge.ai/timeline/mcculloch-pitts-neuron-the-beginning/)

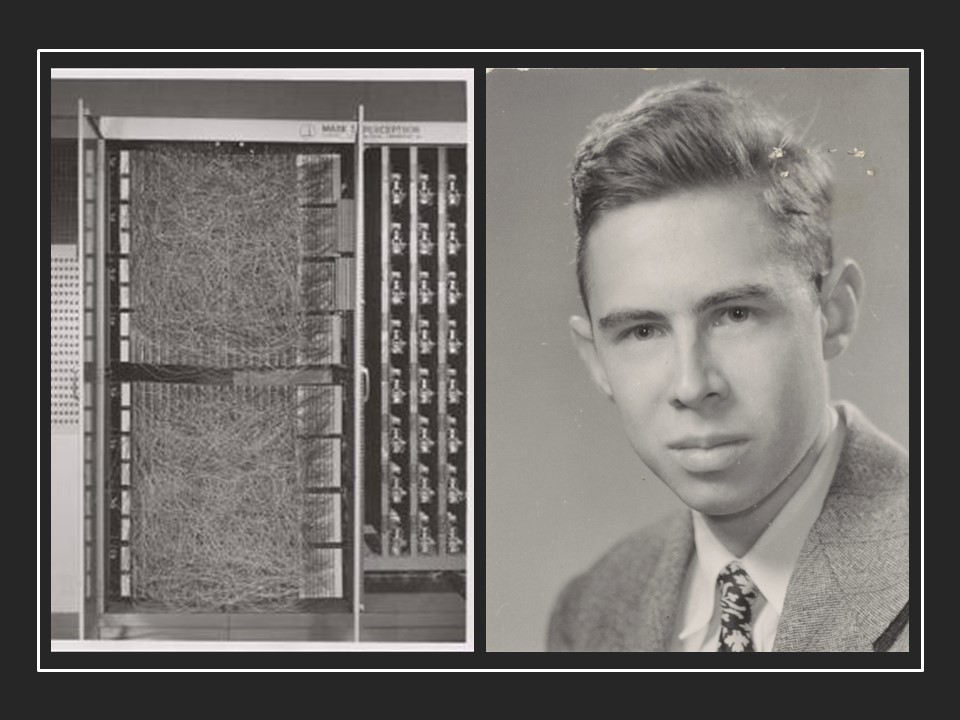


Walter Pitts and Warren McCulloch showed in their paper, *“A Logical Calculus of the Ideas Immanent in Nervous Activity,”* the mathematical model of biological [neuron](https://machinelearningknowledge.ai/glossary/artificial-neuron/). This McCulloch-Pitts Neuron had very limited capability and had no learning mechanism. Yet it laid the foundation for artificial neural network & deep learning.

**1957**

1957

[**Frank Rosenblatt Creates Perceptron**](https://machinelearningknowledge.ai/timeline/perceptron-neuron-that-can-learn/)

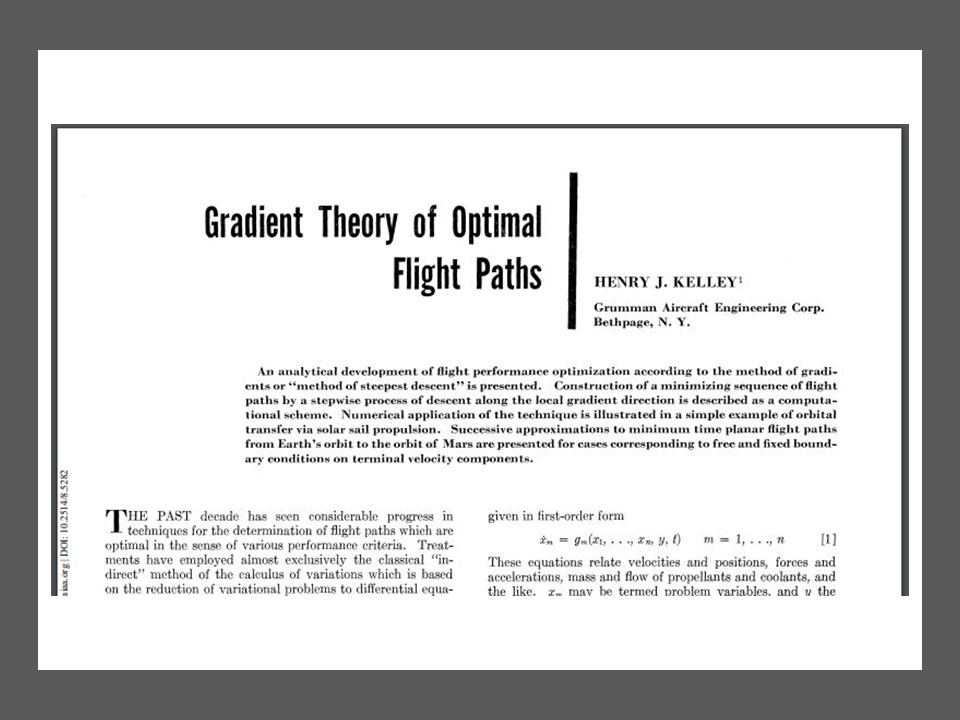


In his paper “The Perceptron: A Perceiving and Recognizing Automaton”, Rosenblatt showed the new avatar of the McCulloch-Pitts neuron – ‘Perceptron’ that had true learning capabilities and could perform [binary classification](https://machinelearningknowledge.ai/glossary/binary-classification/) on its own. This inspired the revolution in research of shallow neural network that continued until the first AI winter.

**1960**

1960

[**The First Backpropagation Model**](https://machinelearningknowledge.ai/timeline/the-first-backpropagation-model/)

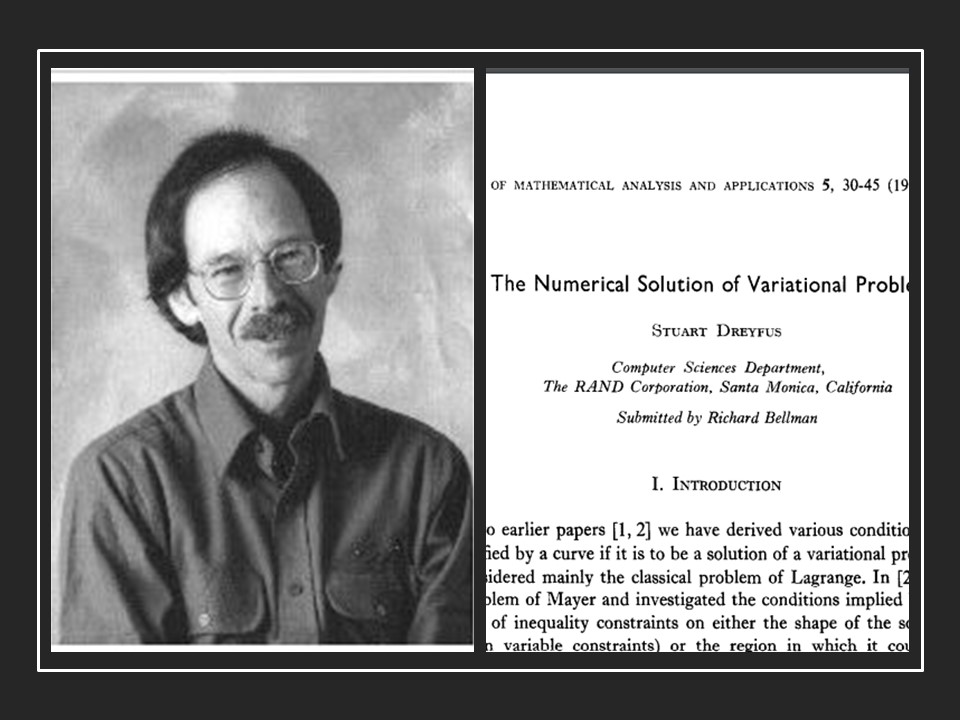


Henry J. Kelley, in his paper “Gradient Theory of Optimal Flight Paths,” showed the first ever version of the continuous backpropagation model. His model is in the context of Control Theory, yet it laid the foundation for further refinement in the model and would be used in ANN in future years.

**1962**

1962

[**Backpropagation With Chain Rule**](https://machinelearningknowledge.ai/timeline/backpropagation-with-chain-rule/)

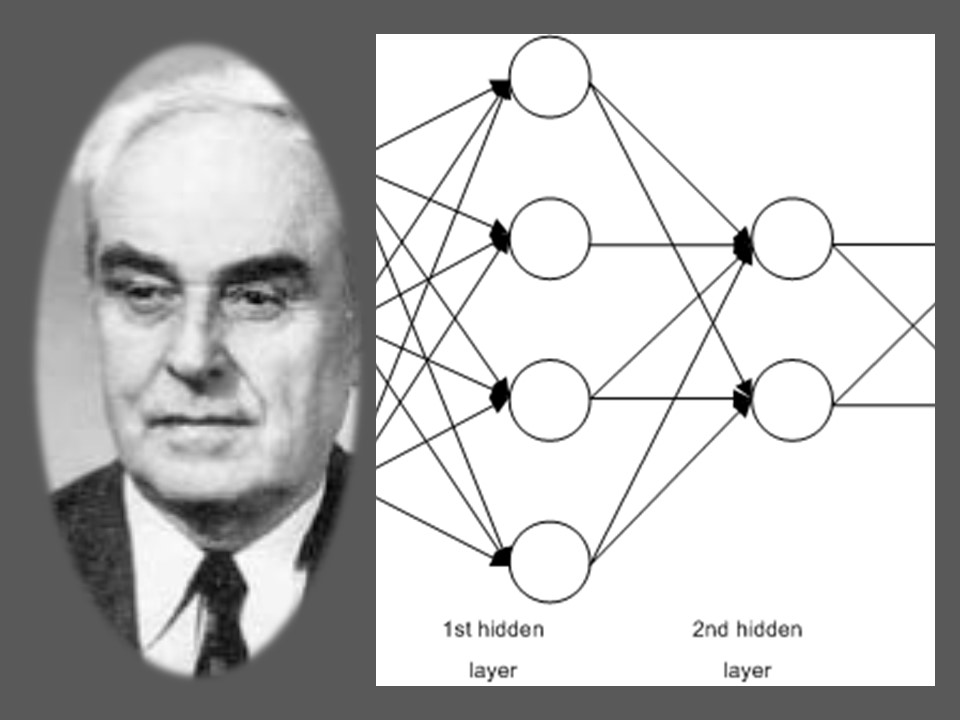


Stuart Dreyfus, in his paper *“The numerical solution of variational problems,”* showed a backpropagation model that used the simple derivative chain rule, instead of the dynamic programming which earlier backpropagation models were using. This was yet another small step that strengthened the future of deep learning.

**1965**

1965

[**Birth Of Multilayer Neural Network**](https://machinelearningknowledge.ai/timeline/birth-of-multilayer-neural-network/)

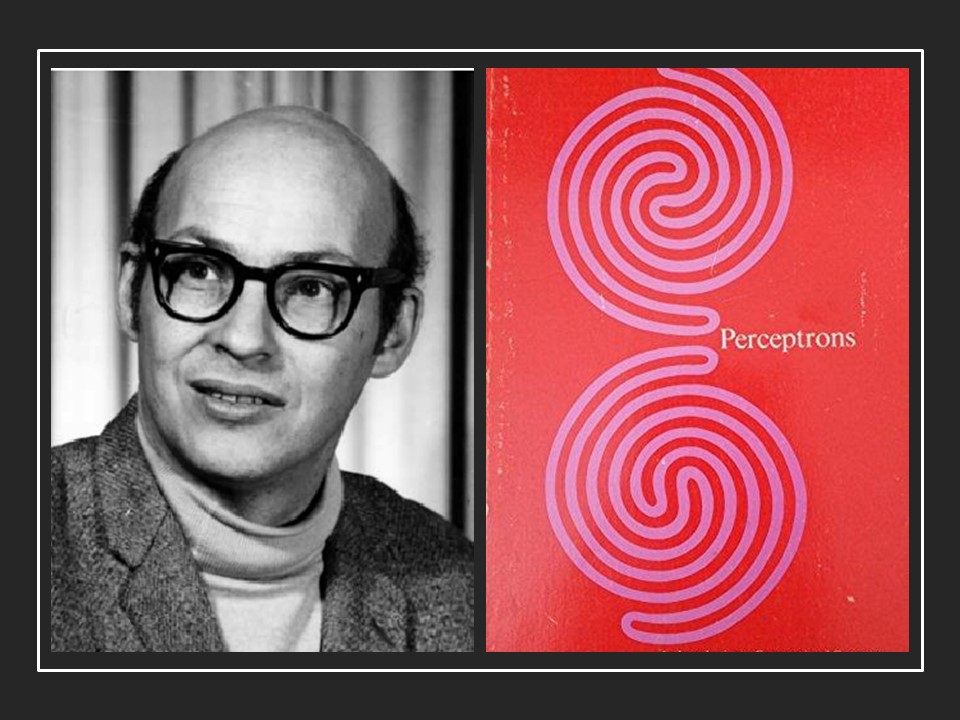


Alexey Grigoryevich Ivakhnenko and Valentin Grigorʹevich Lapa created a hierarchical representation of neural network that used a polynomial [activation function](https://machinelearningknowledge.ai/glossary/activation-function/) and are trained using the Group Method of Data Handling (GMDH). It is considered the first ever multi-layer perceptron ,and Ivakhnenko is considered the father of deep learning.

**1969**

1969

[**The Fall Of Perceptron**](https://machinelearningknowledge.ai/timeline/the-fall-of-perceptron/)



Marvin Minsky and Seymour Papert published the book “Perceptrons” in which they show that Rosenblatt’s Perceptron cannot solve complicated functions like XOR. For such functions Perceptrons should be placed in multiple hidden layers which compromises Perceptron’s learning algorithm. This setback triggered the winter of neural network research.

**1970**

1970

[**Backpropagation Is Computer Coded**](https://machinelearningknowledge.ai/timeline/backpropagation-gets-coded-in-computer/)

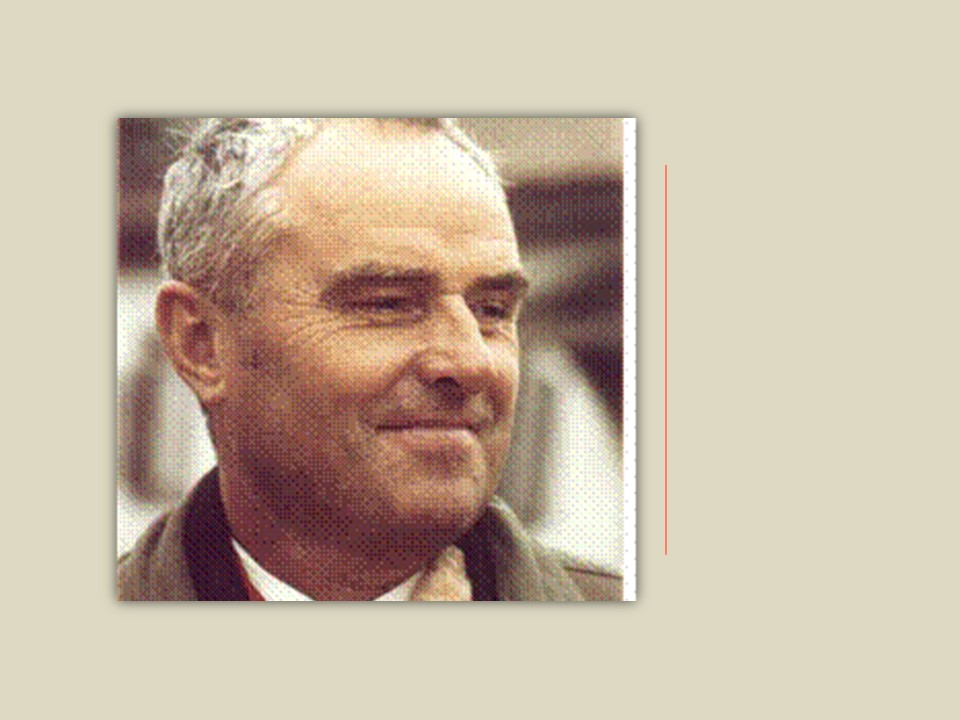


Seppo Linnainmaa published general method for automatic differentiation for backpropagation and also implemented backpropagation in computer code. The research in backpropagation has progressed greatly, yet it would not be implemented in neural networks until the next decade.

**1971**

1971

[**Neural Network Goes Deep**](https://machinelearningknowledge.ai/timeline/neural-network-goes-further-deep/)



Alexey Grigoryevich Ivakhnenko continued his research in Neural Network. He created an 8-layer Deep Neural Network using the Group Method of Data Handling (GMDH).

**1980**

1980

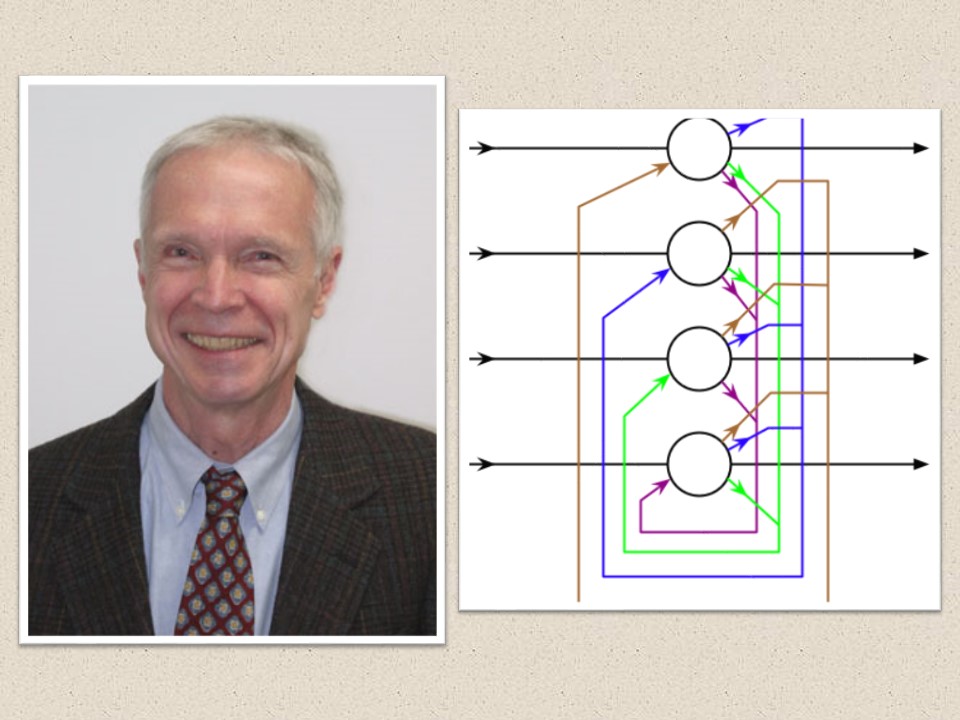
[**Neocognitron – First CNN Architecture**](https://machinelearningknowledge.ai/timeline/neocognitron-the-first-cnn-architecture/)

Kunihiko Fukushima created Neocognitron, the first convolutional neural network architecture. which could recognize visual patterns such as handwritten characters.

**1982**

1982

[**Hopfield Network – Early RNN**](https://machinelearningknowledge.ai/timeline/early-recurrent-neural-network/)

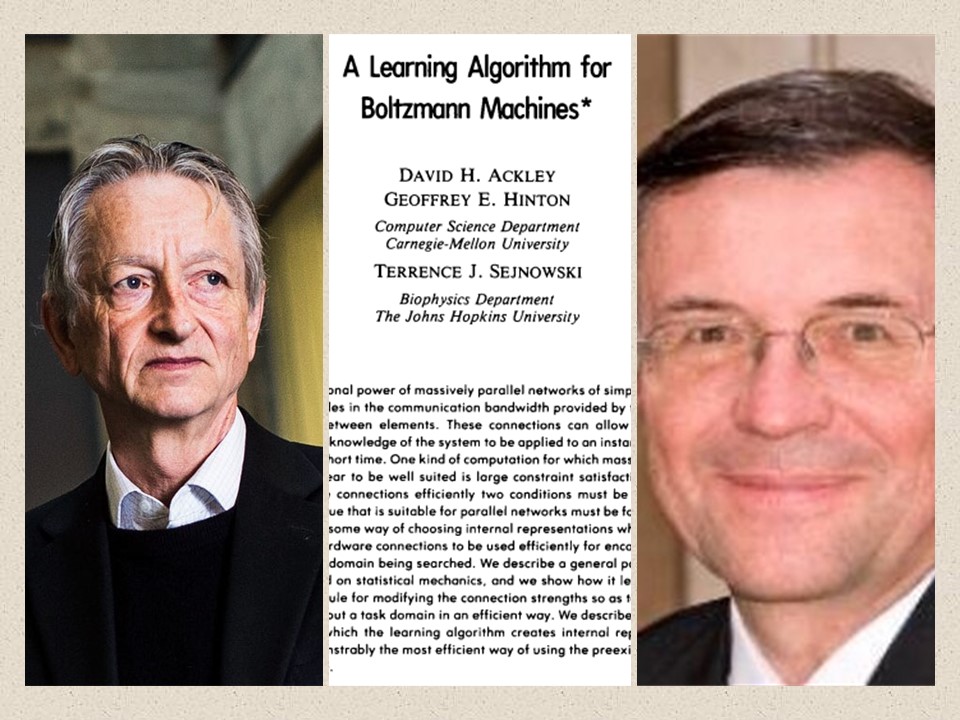


John Hopfield created the Hopfield Network, a recurrent neural network. It serves as a content-addressable memory system, and would be instrumental for further RNN models of modern deep learning era.

**1985**

1985

[**Boltzmann Machine**](https://machinelearningknowledge.ai/timeline/boltzmann-machine/)



Geoffrey Hinton and Terrence Sejnowski created the Boltzmann Machine, which is a stochastic recurrent neural network. This neural network has only an input layer and a hidden layer but no output layer.

**1986**

1986

[**NetTalk – ANN Learns Speech**](https://machinelearningknowledge.ai/timeline/nettalk-neural-network-learns-speech/)

Terry Sejnowski created NeTalk, a neural network which learns to pronounce written English text by being shown text as input and matching phonetic transcriptions for comparison.

1986

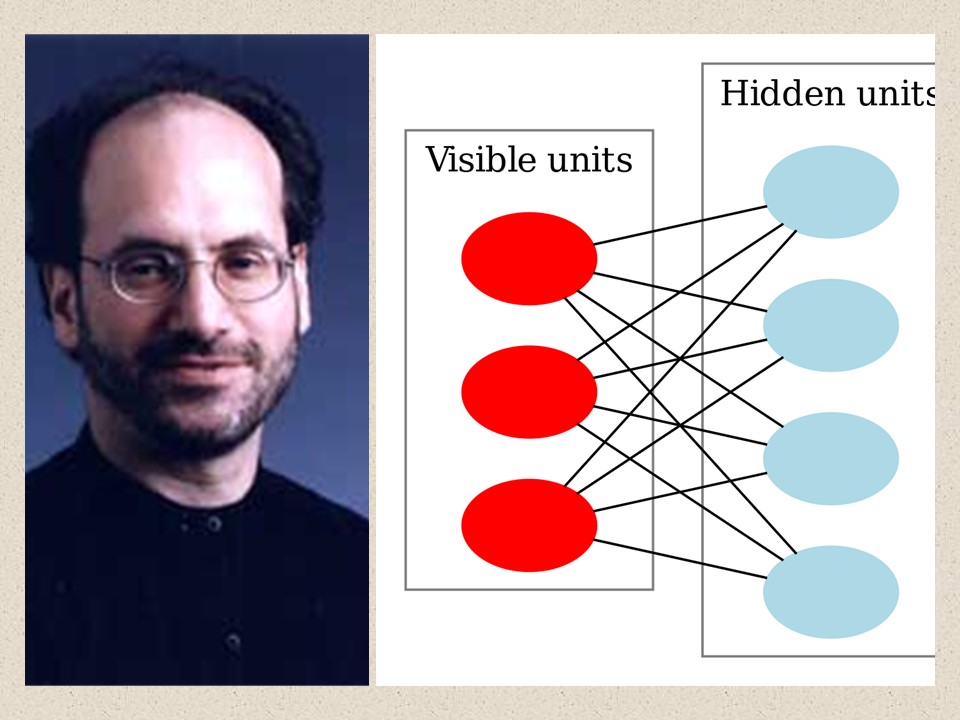
[**Backpropagation In Neural Network**](https://machinelearningknowledge.ai/timeline/finally-backpropagation-meets-neural-network/)



Geoffrey Hinton, David Rumelhart and Ronald Williams, in their paper “Learning Representations by back-propagating errors,” showed a new and better learning procedure for neural network by using backpropagation. It opened the gates for training complex deep neural network easily, which was the main obstruction in the earlier days of neural network research in this area.

1986

[**Restricted Boltzmann Machine**](https://machinelearningknowledge.ai/timeline/restricted-boltzmann-machine/)



Paul Smolensky came up with a variation of the Boltzmann Machine, one where there is not intra layer connection in input and hidden layer. It is known as Restricted Boltzmann Machine (RBM). It would become popular in years to come especially for building recommender systems.

**1989**

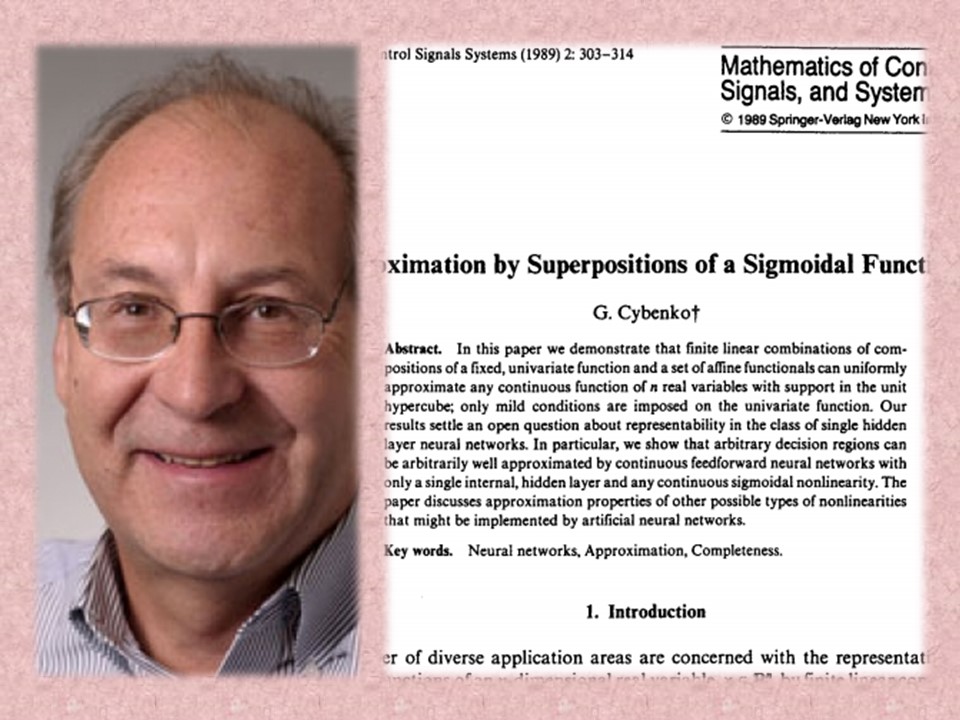
1989

[**CNN Using Backpropagation**](https://machinelearningknowledge.ai/timeline/cnn-using-backpropagation/)

Yann LeCun used backpropagation to train convolutional neural network to recognize handwritten digits. This was a breakthrough moment as it laid the foundation of modern computer vision using deep learning.

1989

[**Universal Approximators Theorem**](https://machinelearningknowledge.ai/timeline/neural-networks-are-universal-approximators/)



George Cybenko published the earliest version of the Universal Approximation Theorem in his paper “*Approximation by superpositions of a sigmoidal function.”* He proved that feed forward neural network with a single hidden layer containing a finite number of neurons could approximate any continuous function. It further added credibility to Deep Learning.

**1991**

1991

[**Vanishing Gradient Problem Appears**](https://machinelearningknowledge.ai/timeline/vanishing-gradient-problem-appears/)

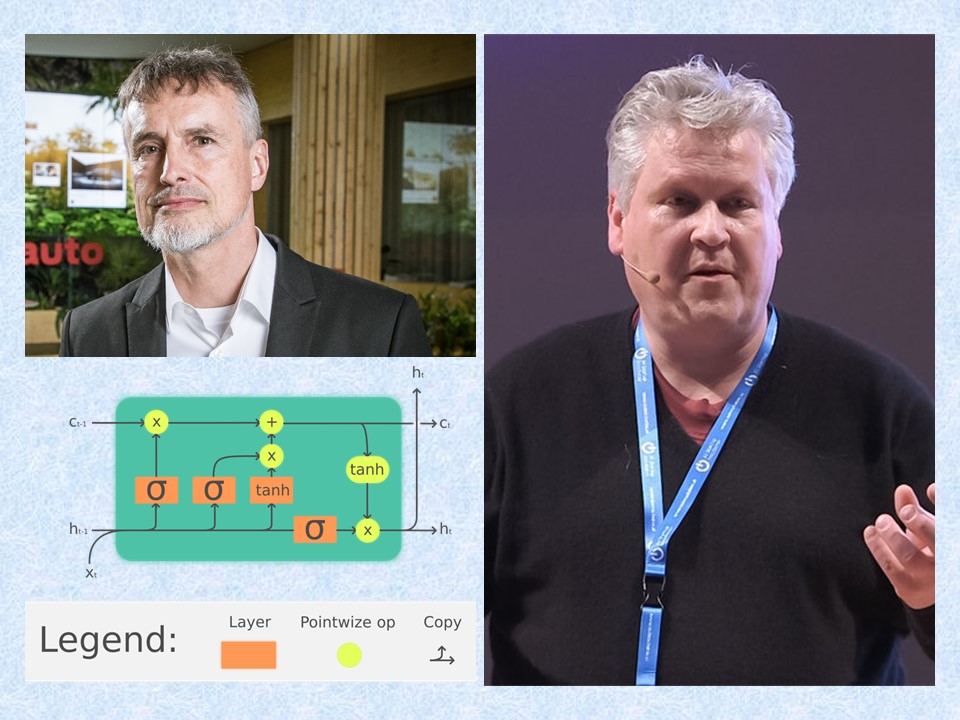


Sepp Hochreiter identified the problem of the vanishing gradient, which lengthens the learning of time of a deep neural network, making deep neural networks almost impractical. This problem continued to annoy the deep learning community for many years to come.

**1997**

1997

[**The Milestone Of LSTM**](https://machinelearningknowledge.ai/timeline/the-milestone-of-lstm/)

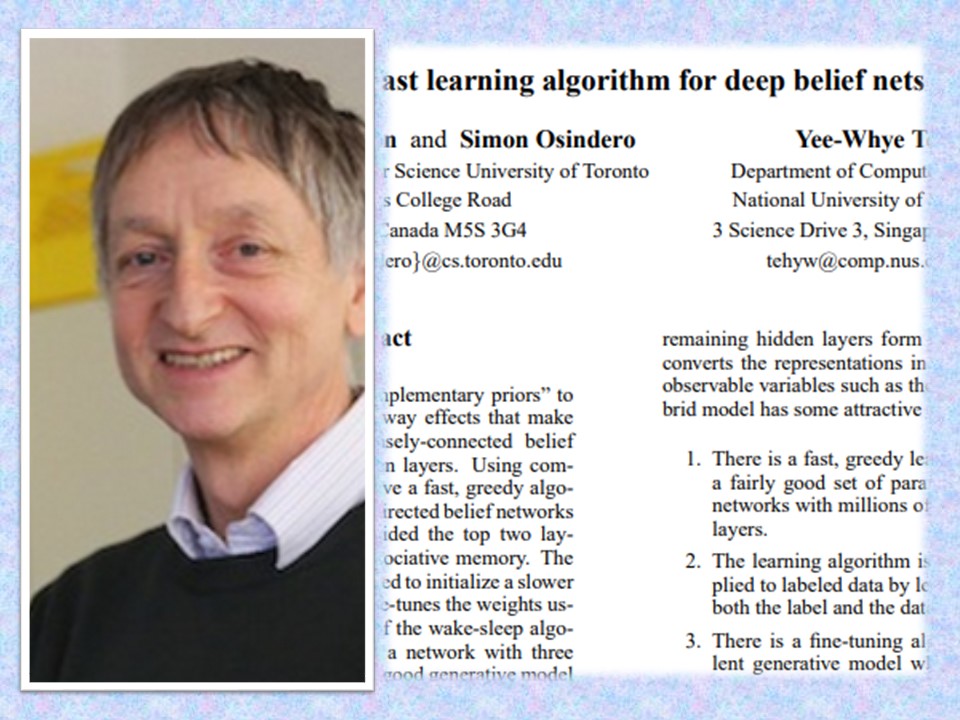


Sepp Hochreiter and Jürgen Schmidhuber published a milestone paper on “Long Short-Term Memory” (LSTM). It is a type of recurrent neural network architecture revolutionized deep learning.

**2006**

2006

[**Deep Belief Network**](https://machinelearningknowledge.ai/timeline/deep-belief-network/)



Geoffrey Hinton, Simon Osindero and Yee-Whye Teh published the paper “*A fast learning algorithm for deep belief nets,*” in which they stacked multiple RBMs together in layers and called them Deep Belief Networks. The training process is much more efficient for large amount of data.

**2008**

2008

[**GPU Revolution Begins**](https://machinelearningknowledge.ai/timeline/the-deep-learning-gpu-revolution-starts/)



Andrew NG’s group in Stanford began advocating for the use of GPUs for training Deep Neural Networks to speed up training. The use of GPUs allows for efficient training on large volumes of data.

**2011**

2011

[**Combat For Vanishing Gradient**](https://machinelearningknowledge.ai/timeline/the-combat-for-vanishing-gradient-problem/)

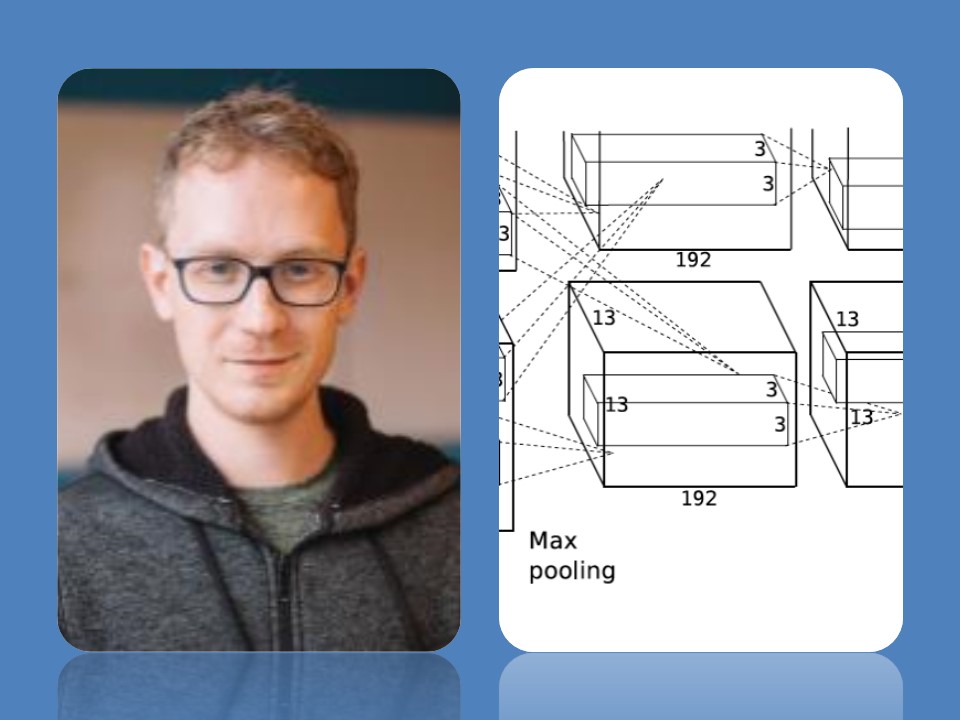


Yoshua Bengio, Antoine Bordes, and Xavier Glorot, in their paper “Deep Sparse Rectifier Neural Networks,” showed that the Rectified Linear Unit (ReLU) activation function can avoid the vanishing gradient problem. This meant that the deep learning community had a tool other than GPUs to avoid the issue of long and impractical training times.

**2012**

2012

[**AlexNet Starts Deep Learning Boom**](https://machinelearningknowledge.ai/timeline/alexnet-starts-the-deep-learning-boom/)

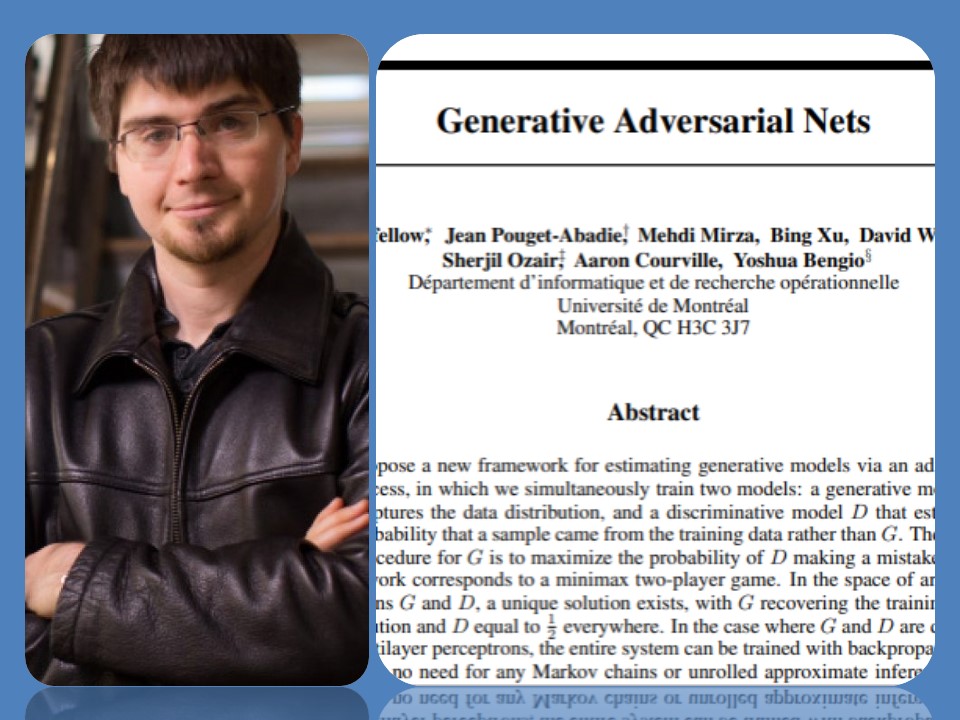


AlexNet, a GPU implemented CNN model designed by Alex Krizhevsky, won Imagenet’s image [classification](https://machinelearningknowledge.ai/glossary/classification/) contest with an accuracy of 84%. It was a huge jump over the 75% accuracy that earlier models had achieved. This win triggered a new, global deep learning boom.

**2014**

2014

[**The Birth Of GANs**](https://machinelearningknowledge.ai/timeline/generative-adversarial-neural-network-is-created/)



The Generative Adversarial Neural Network (GAN) is created by Ian Goodfellow. GANs open a whole new doors of application of deep learning in fashion, art, and science due to its ability to synthesize seemingly real data.

**2016**

2016

[**AlphaGo Beats Human**](https://machinelearningknowledge.ai/timeline/alphago-beats-human/) **Champion**



Deepmind’s deep [reinforcement learning](https://machinelearningknowledge.ai/glossary/reinforcement-learning/) model beats human champion Lee Sedol in the complex game of Go. Go is more complex than chess, and this feat captured human imagination and took the promise of deep learning to a whole new level.

**2019**

2019

[**Godfathers Win Turing Award**](https://machinelearningknowledge.ai/timeline/godfathers-of-deep-learning-win-turing-award/)



Yoshua Bengio, Geoffrey Hinton, and Yann LeCun won the 2018 Turing Award for their immense contribution to advancements in the area of deep learning and artificial intelligence. This was a defining moment for those who had worked relentlessly on neural networks when the entire machine learning community had moved away from it in the 1970s.