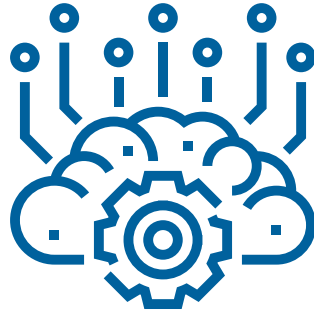


A LAYMEN'S GUIDE TO

NEURAL NETWORKS

CONTENTS

*	4	WHAT ARE NEURAL NETWORKS?
*	5	NEURAL NETWORKS: HOW THEY WORK
*	6	A NEW MODEL
*	8	EARLY NEURAL NETWORKS
*	9	THE ARTIFICIAL NEURON
*	11	DIMENSIONALITY REDUCTION
*	12	DIFFERENT TOPOLOGIES
*	13	HOW THEY'RE BEING USED
*	14	HOW THEY MIGHT BE USED IN THE FUTURE



When you're talking about machine learning and artificial intelligence these days, you're likely to find yourself talking about neural networks. Over the past few years, as scientists ponder big advances in artificial intelligence, neural networks have played a significant role.

But what are these technologies, and how do they work?

Understanding neural networks better will get you further toward comprehending how computers are coming to life all around us and starting to make evermore complicated decisions in all sorts of scenarios. In many ways, neural networks are some of the fundamental building blocks that are going to offer us smart homes, smart services and smarter computing in general.

WHAT ARE

NEURAL NETWORKS?

**AN ARTIFICIAL
NEURAL NETWORK
IS A TECHNOLOGY
THAT FUNCTIONS
BASED ON THE
WORKINGS OF THE
HUMAN BRAIN.**

In particular, the artificial neural network acts to simulate in some ways the activity and build of biological neurons in the brain.

Neural networks are built in various different ways, in calculated models that are used to pursue machine learning projects where computers can be trained to “think” in their own ways.

Based on what we know about the human brain, and based on what we can do with state-of-the-art technologies, we can make progress toward figuring out how to make a computer “act like a brain.” But engineers are not replicating human cognitive behavior. The human brain is a black box that we still don’t understand.

In some senses, neural networks are a strange hybrid of “making it” and “faking it.” They do perform much like the layers of neurons in the brain, but they still work based on enormous amounts of training data, so that in the end, they are only really semi-intelligent, at least when compared to our own human brains.

NEURAL NETWORKS:

HOW THEY WORK

To understand how neural networks work, it's important to understand how the neurons work in the human brain.

Biologically, a neuron – composed of a nucleus, dendrites and an axon – takes an electrical impulse and uses it to send signals through the brain. That's how we get all of our sensations and stimuli to which we create central nervous system responses.

Biological models show the unique build of this type of cell, but often don't really map out the activity paths that guide neurons to send signals on through various levels.

Like a neuron, a neural network has these multiple levels. Particularly, the neural network typically has an input layer, hidden layers and an output layer. Signals make their way through these layers to trigger machine learning outcomes.

The fundamental way that artificial neural networks work is by using a series of weighted inputs. This is based on the biological function of neurons in the brain that take in a variety of impulses and filter them through those different levels. They do this so they can interpret the signals they are receiving and turn them out at their destination as understandable ideas and concepts.

Think of the brain – and the neural network – as a “thought factory”: inputs in, outputs out. But by mapping what goes on in those in-between areas, the scientists behind the advancement of neural networks can get a lot closer to “mapping out” the human brain – although the general consensus is that we have a long way to go.

In a neural network, this discovery and modeling takes the form of computational data structures that are getting composed of the input layer, the hidden layers and the output layer. The key to these layers of neurons is a series of weighted inputs that combine to give the network layer its “food” and determine what it will pass on to the next layer.

Scientists often talk about feedforward neural networks, in which information moves in one direction only – from the input layer through hidden layers to the output layer – as a major model. They also calculate all of the weighted inputs in order to contemplate how the system takes in information. This model can help someone who is just approaching neural networks to understand how they work – it's a chain reaction of the passage of data through the network layers.

A NEW

MODEL

In addition to understanding how artificial neural networks mimic human brain activity, it's also very helpful to consider what's new about these technologies.

One of the best ways to do that is to consider what technologies were groundbreaking within the last two decades, before artificial neural networks and similar technologies came along.

You could describe that era of new internet, the era of Moore's law and the increasing power of hardware, as the age of deterministic programming.

People who learned programming skills in the 1980s learned very basic linear programming concepts. Those programming concepts continued to evolve year by year until deterministic programming could form pretty impressive technologies like early chatbots and decision support software. However, we didn't think of these technologies as being "artificial intelligence." We saw them mainly as tools to calculate and quantify data.

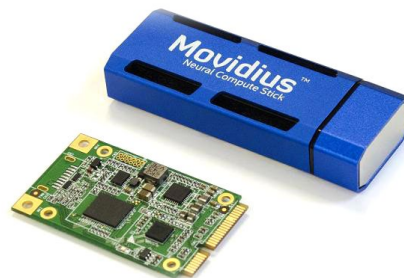
More recently, the age of big data came along. Big data gave us amazing insights and helped computers to evolve in many ways, but the model was still deterministic – data in, data out.

What's new about artificial neural networks is that instead of deterministic inputs, the machine uses probabilistic input systems (i.e. the weighted inputs that go into each artificial neuron) that are much more sophisticated. One of the best ways to illus-

trate this is by looking at new chips that companies like Intel are now manufacturing.

Traditional chips work purely on a binary model. They are made to process deterministic programming input.

The new chips are essentially "made for machine learning" – they're made to directly process the types of probabilistic weighted inputs that make up neural networks and that give those artificial neurons their power.



An [article in Engadget](#) talks about the new "chip wars" where companies like Qualcomm and Apple are building these new types of multi-core processors. The strategy starts with mixing high-powered cores with other energy-saving cores. Then there's the strategy of porting special kinds of data – usually data associated with image processing or video work – into GPUs instead of CPUs.

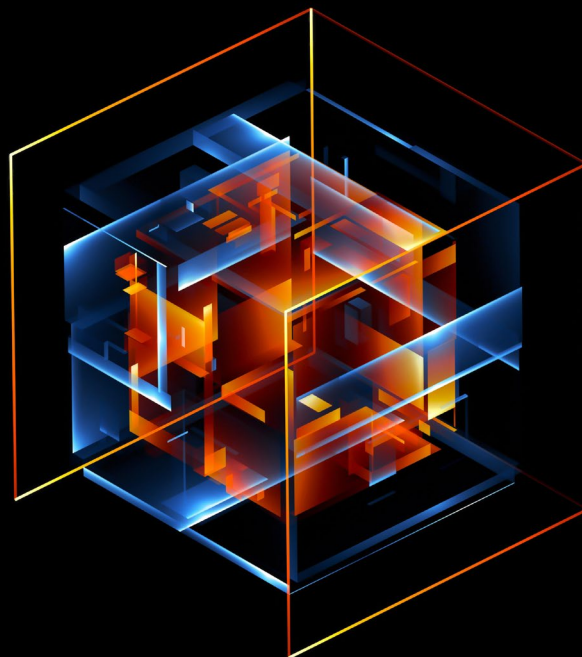
The emergence of new types of microprocessors is based on an idea that originated earlier in the development of computer graphics – in the early days of VGA and pixelation, the graphics processing unit or GPU emerged as a way to handle specific types of computing related to rendering and working with graphic images.

One of the best ways to understand the GPU is to contrast it to the CPU or central processing unit. The first computers had CPUs in order to process memory and input in an application. However, when computers started to be sophisticated enough to run complex computer graphics, engineers designed the GPU specifically for this type of high activity computing.

GPUs are different than CPUs in that the GPUs tend to have many smaller processing cores that can work simultaneously on tasks. The CPU, on the other hand, has a few high-powered cores that work sequentially. Think of it as the difference between a few capable processors that can triage and schedule application threads, or a chip with many small low-powered processing cores that can all handle small tasks at once to make the overall work quick and effective.

The article also talks about other kinds of specialization – for instance, a “neural engine” in an Apple GPU – and how so much of the work of the new processor groups consists of specific delegation, allowing devices to truly multi-task – to send tasks to the chips and cores through which they are best served.

**WHAT'S NEW ABOUT
ARTIFICIAL NEURAL
NETWORKS IS
THAT INSTEAD OF
DETERMINISTIC INPUTS,
THE MACHINE USES
PROBABILISTIC INPUT
SYSTEMS THAT ARE MUCH
MORE SOPHISTICATED.**



EARLY

NEURAL NETWORKS?

NEURAL NETWORKS DIDN'T MIRACULOUSLY EMERGE AS THE MODERN JUGGERNAUTS OF COGNITIVE ABILITY THAT THEY ARE NOW. INSTEAD, THEY WERE BUILT FROM VARIOUS INCREMENTAL TECHNOLOGY ADVANCES OVER THE YEARS.

For example, Marvin Minsky, who lived from 1927 to 2016, was one of the early pioneers of these types of intelligent systems. In an instructive YouTube video from his later years, Minsky explains the concept of early neural networks in a way that's remarkably similar to the logic gates of yesterday's motherboards – in the sense that he describes the neuron inputs as handling logical functions. He also elaborates on some of the background through which neural networks came into being – for example, the work of algorithm scientists Claude Shannon and remarkable mathematician and AI pioneer Alan Turing in the mid-20th century.

In general, as Minsky points out, neural networks are in some ways just an extension of the logic handling practices built into earlier technology – but instead of using circuit logic gates, the neurons can handle more sophisticated inputs. They can tie things together and spit out much more elaborate results, and that makes computers look like they're thinking in a more human way.

When you look at the question of what neural networks can do, one short answer is – what can't they do?

In addition to recommendation engines and the sorting and selecting of good data from a big background field, neural networks are learning to really imitate human intelligence to a high degree. What if your smart home or Alexa device could talk to you, knowing a lot about who you are and your preferences, instead of just responding to questions about the weather? What if internet-connected devices could follow you wherever you go, and market your products and services based on a deep knowledge of your personality and background?

These are just some of the future uses that are going to be absolutely remarkable to us when we finally start implementing them in consumer markets. However, if you've read through all of the above and understand how neural networks came about, you'll be more familiar and more comfortable with that robotic intelligence that's suddenly in your devices, and even in your appliances, homes and cars.

THE ARTIFICIAL

NEURON

Looking at an artificial neuron and its structure may help with understanding how neural networks are designed. After all, neural networks are collections of these artificial neurons, with their own computations and digital structure.

The artificial neuron has various weighted inputs, along with a transfer function or activation function, that allow it to “fire” down the line. The output of the artificial neuron acts as the axon of the biological neuron.

Engineers use various different types of activation functions to help with determining the output of an artificial neuron and a neural network in general.

Nonlinear activation functions can help apply trajectories to data outputs. A sigmoid function helps with determining outputs between zero and one.

A function called ReLu is often used in constitutional neural networks and may be modified to accommodate negative values.

Essentially, the artificial neuron propagates its output to the next level, and data goes along, being fed into

these weighted inputs in order to help the machine come up with an informed result.

In a simple feedforward neural network, which is one of the simplest forms of artificial neural network, data only passes one direction – from input to output – however, engineers have come up with a philosophy of backpropagation, which helps to fine-tune the neural network’s artificial learning processes.

Backpropagation is a very important part of machine learning often used with supervised machine learning – it’s short for “backward propagation of errors.”

Backpropagation takes a loss function and calculates a gradient to adjust those weighted inputs of the neuron – the loss function means the output of the network is compared to a desired output in training. Error values are then propagated to adjust the weights. Essentially, the program “looks back” at what could be adjusted to make the model fit the data better, and by adjusting the weights, it makes sure that the model amplifies the right signals.



DIMENSIONALITY REDUCTION

**BACKPROPAGATION
IS NOT THE ONLY
METHOD THAT
ENGINEERS
USE TO REFINE
OR OPTIMIZE
MACHINE LEARNING
PROJECTS.
ANOTHER ONE
IS CALLED
DIMENSIONALITY
REDUCTION.**

To understand how this works, imagine a detailed map with many different location points, for example, the border of a U.S. state, which is usually not smooth – (forget about Colorado and Wyoming) – but composed of intricate borderlines.

Now imagine that the map is scaled back to include fewer locational points. That border which looked very intricate now becomes smooth and simplified – you don't see a lot of the curves and ridges of rivers or coastlines or mountains or anything else that made up the original lines. You get a much smoother and simpler image – and it's easier for the machine to connect the dots.

Engineers can work with dimensionality reduction and various programming tools to control the complexity of a machine learning project. This can help with many of the problems that cause machine learning outputs to vary quite a bit from a desired output.



DIFFERENT TOPOLOGIES

In addition to simple feed-forward neural networks, there are other types of networks that are often used in today's artificial intelligence engineering world.

Here are a few of the most common types of setups.

1. Recurrent Neural Network

A recurrent neural network includes a specific function that helps preserve memory through the layers of the network in order to generate results.

Since every artificial neuron remembers some of the information it is passed or given, many of the results of a recurrent neural network will deliver straight-forward connections based on the original inputs – experts provide an instance where the machine learning program learns to predict a word based on the previous word in text.

2. Convolutional Neural Networks

Convolutional neural networks are extremely popular for image processing and computer vision. In these types of neural networks, each of the layers deals with certain portions of an image or graphic. For example, after an input layer, the next layer of a convolutional neural network may deal with specific contours and ridges – another will deal with individual features of a unified whole – essentially the neural network will classify an image.

In popular science, you can see the results of these types of neuron that works in programs that can identify cats or dogs or other animals, or people or objects from a field of digital vision. You can also see the limitations, for example in this hilarious result showing [Chihuahuas and muffins](#).

3. Self-Organizing Neural Networks

Self-organizing neural networks put together their own parts through iterations, and with their agile builds, they're popular for recognizing patterns in data, including in some medical applications.

Another way to describe self-organizing neural networks is that they are made to adapt to the inputs that they're given – for instance, a paper on these types of networks talked about a “grow-when-required” model that helps the network to learn various inputs like body motion patterns. In general, self-organizing neural networks are based on a principle that relates task management to the structure of an input space – think of this as a way to make neural networks more agile in adapting to the roles that they are given. Another example is self-organizing feature maps that help to sort all of the types of feature selection and extraction that are used to fine-tune machine learning models.

HOW THEY'RE

BEING USED

Neural networks are being used in many different industries.

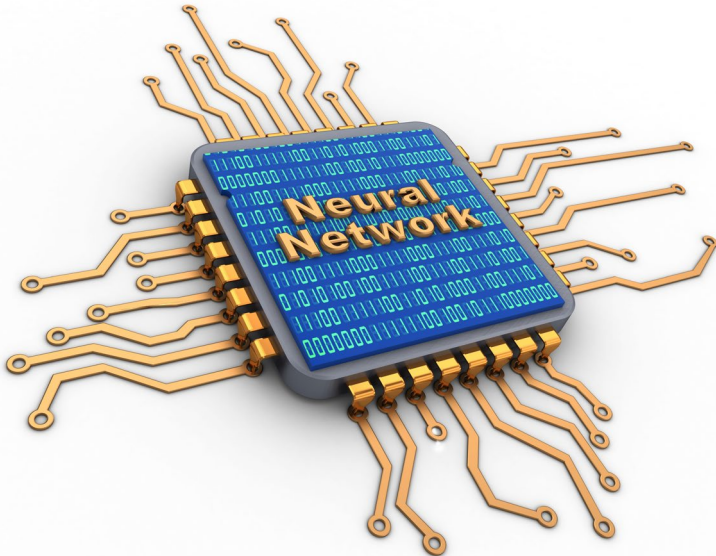
They're especially useful in the health care industry, where doctors are using machine learning projects with neural networks to diagnose disease, generate optimal medical cures or treatments, and learn more about patients and communities of patients.

Experts describe the value of neural networks in medicine this way: Whenever a doctor is acting like a machine – whether it's evaluating a digital image, projecting a diagnosis from enormous volumes of data, or listening for abnormalities in an audio stream, the doctor is performing tasks that machine learning projects can accomplish with excellent results.

By contrast, neural networks are not as good at mimicking the more behavioral parts of human thought – our social and emotional outputs and other reactions that have to do with extremely abstract inputs.

Neural networks are being used in retail to help companies understand what customers want. Some recommendation engines for music services at e-commerce stores are based on this type of technology.

Neural networks are important in transportation for fleet management. They are used in shipping for cold chain logistics. They are used in many kinds of private or public offices to help with workflows and delegation. Neural networks are really at the cutting edge of popular enterprise technology and consumer-facing technologies that will help us learn more about what we can do with computers.



HOW THEY MIGHT

BE USED IN THE FUTURE

Along with all the exciting applications that we use neural networks for today, there are even more amazing possibilities coming down the pike.

One way of thinking about this is to read about the singularity proposed by prize-winning theorist Ray Kurzweil, who is [now employed at Google](#) and working on the vanguard of IT in the future. The singularity idea posits that computers will actually learn to think like humans, and eventually outpace us in cognitive ability.

A less extreme way to think about this is that neural networks will start to be used for all sorts of services

and projects – they'll become better able to pass the Turing test, that is, trick humans into thinking that computers are interacting with them as humans. Physical robots will be able to read body language and talk to you as if they were another human being.

All of this is going to revolutionize call-center work, cashiering and everything else that involves human interactions. It's essentially going to remake our world – so stay tuned and look for more on what the average person can do to learn more about these exciting technologies.

