

# What is Deep Learning?

[Jason Brownlee](#)

Last Updated on August 14, 2020

**Deep Learning** is a subfield of machine learning concerned with algorithms inspired by the structure and function of the brain called **artificial neural networks**.

If you are just starting out in the field of deep learning or you had some experience with neural networks some time ago, you may be confused. I know I was confused initially and so were many of my colleagues and friends who learned and used **neural networks in the 1990s and early 2000s**.

The leaders and experts in the field have ideas of what deep learning is and these specific and nuanced perspectives shed a lot of light on what deep learning is all about.

In this post, you will discover exactly what deep learning is by hearing from a range of experts and leaders in the field.

**Kick-start your project** with my new book [Deep Learning With Python](#), including *step-by-step tutorials* and the *Python source code* files for all examples.

Let's dive in.



What is Deep Learning?

Photo by [Kiran Foster](#), some rights reserved.

## Deep Learning is Large Neural Networks

[Andrew Ng](#) from Coursera and Chief Scientist at Baidu Research formally founded [Google Brain](#) that eventually resulted in the productization of deep learning technologies across a large number of Google services.

He has spoken and written a lot about what deep learning is and is a good place to start.

In early talks on deep learning, [Andrew](#) described deep learning in the context of traditional artificial neural networks. In the 2013 talk titled "[Deep Learning, Self-Taught Learning and Unsupervised Feature Learning](#)" he described the idea of deep learning as:

*Using brain simulations, hope to:*

- *Make learning algorithms much better and easier to use.*
  - *Make revolutionary advances in machine learning and AI.*
- I believe this is our best shot at progress towards real AI*

Later his comments became more nuanced.

The core of deep learning according to Andrew is that we now have fast enough computers and enough data to actually train large neural networks. When discussing why now is the time that deep learning is taking off at ExtractConf 2015 in a talk titled "[What data scientists should know about deep learning](#)", he commented:

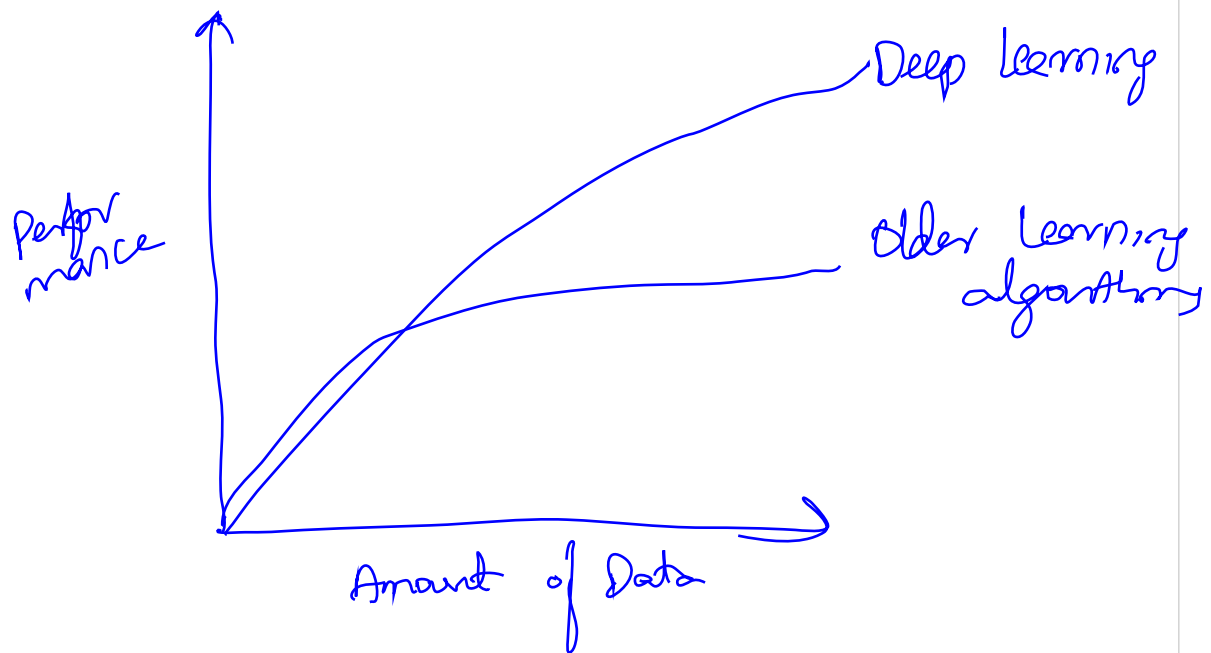
*very large neural networks we can now have and ... huge amounts of data that we have access to*

He also commented on the important point that it is all about scale. That as we construct larger neural networks and train them with more and more data, their performance continues to increase. This is generally different to other machine learning techniques that reach a plateau in performance.

*for most flavors of the old generations of learning algorithms ... performance will plateau. ... deep learning ... is the first class of algorithms ... that is scalable. ... performance just keeps getting better as you feed them more data*

He provides a nice cartoon of this in his slides:

## Why Deep Learning?



How Do Data Science techniques scale with amount of data?

Why Deep Learning?

Slide by [Andrew Ng](#), all rights reserved.

Finally, he is clear to point out that the benefits from deep learning that we are seeing in practice come from supervised learning. From the 2015 ExtractConf talk, he commented:

*almost all the value today of deep learning is through supervised learning or learning from labeled data*

Earlier at a talk to Stanford University titled "[Deep Learning](#)" in 2014 he made a similar comment:

*one reason that deep learning has taken off like crazy is because it is fantastic at supervised learning*

Andrew often mentions that we should and will see more benefits coming from the unsupervised side of the tracks as the field matures to deal with the abundance of unlabeled data available.

[Jeff Dean](#) is a Wizard and Google Senior Fellow in the Systems and Infrastructure Group at Google and has been involved and perhaps partially responsible for the scaling and adoption of deep learning within Google. Jeff was involved in the **Google Brain project** and the development of **large-scale deep learning software DistBelief** and later **TensorFlow**.

In a 2016 talk titled "[Deep Learning for Building Intelligent Computer Systems](#)" he made a comment in the similar vein, that **deep learning** is **really all about large neural networks**.

*When you hear the term **deep learning**, just think of a **large deep neural net**. **Deep** refers to the number of layers typically and so this kind of the popular term that's been adopted in the press. I think of them as **deep neural networks** generally.*

He has given this talk a few times, and in a [modified set of slides for the same talk](#), he **highlights the scalability of neural networks** indicating that **results get better with more data and larger models**, that **in turn require more computation to train**.

Results Get Better With More Data, Larger Models, More Compute

Important Property of Neural Networks  
Results get better with  
more data +  
bigger models +  
more computation  
(Better algorithms, new insights, and  
improved techniques always help, too!)

## Deep Learning is Hierarchical Feature Learning

In addition to scalability, another often cited benefit of deep learning models is their ability to perform automatic feature extraction from raw data, also called feature learning.

[Yoshua Bengio](#) is another leader in deep learning although began with a strong interest in the automatic feature learning that large neural networks are capable of achieving.

He describes deep learning in terms of the algorithms ability to discover and learn good representations using feature learning. In his 2012 paper titled "[Deep Learning of Representations for Unsupervised and Transfer Learning](#)" he commented:

*Deep learning algorithms seek to exploit the unknown structure in the input distribution in order to discover good representations, often at multiple levels, with higher-level learned features defined in terms of lower-level features*

An elaborated perspective of deep learning along these lines is provided in his 2009 technical report titled "[Learning deep architectures for AI](#)" where he emphasizes the importance the hierarchy in feature learning.

*Deep learning methods aim at learning feature hierarchies with features from higher levels of the hierarchy formed by the composition of lower level features. Automatically learning features at multiple levels of abstraction allow a system to learn complex functions mapping the input to the output directly from*

*data, without depending completely on human-crafted features.*

In the soon to be published book titled "[Deep Learning](#)" co-authored with Ian Goodfellow and Aaron Courville, they define deep learning in terms of the depth of the architecture of the models.

*The hierarchy of concepts allows the computer to learn complicated concepts by building them out of simpler ones. If we draw a graph showing how these concepts are built on top of each other, the graph is deep, with many layers. For this reason, we call this approach to AI deep learning.*

This is an important book and will likely become the definitive resource for the field for some time. The book goes on to describe multilayer perceptrons as an algorithm used in the field of deep learning, giving the idea that deep learning has subsumed artificial neural networks.

*The quintessential example of a deep learning model is the feedforward deep network or multilayer perceptron (MLP).*

[Peter Norvig](#) is the Director of Research at Google and famous for his textbook on AI titled "[Artificial Intelligence: A Modern Approach](#)".

In a 2016 talk he gave titled "[Deep Learning and Understandability versus Software Engineering and Verification](#)" he defined deep learning in a very similar way to Yoshua, focusing on the power of abstraction permitted by using a deeper network structure.

*a kind of learning where the representation you form have several levels of abstraction, rather than a direct input to output*

**Why Call it "Deep Learning"?**  
**Why Not Just "Artificial Neural Networks"?**

[Geoffrey Hinton](#) is a pioneer in the field of artificial neural networks and co-published the first paper on [the backpropagation algorithm](#) for training multilayer perceptron networks.

He may have started the introduction of the phrasing "deep" to describe the development of large artificial neural networks.

He co-authored a paper in 2006 titled "[A Fast Learning Algorithm for Deep Belief Nets](#)" in which they describe an approach to training "deep" (as in a many layered network) of restricted Boltzmann machines.

*Using complementary priors, we derive a fast, greedy algorithm that can learn deep, directed belief networks one layer at a time, provided the top two layers form an undirected associative memory.*

This paper and the related paper Geoff co-authored titled "[Deep Boltzmann Machines](#)" on an undirected deep network were well received by the community (now cited many hundreds of times) because they were successful examples of greedy layer-wise training of networks, allowing many more layers in feedforward networks.

In a co-authored article in Science titled "[Reducing the Dimensionality of Data with Neural Networks](#)" they stuck with the same description of "deep" to describe their approach to developing networks with many more layers than was previously typical.

*We describe an effective way of initializing the weights that allows deep autoencoder networks to learn low-dimensional codes that work much better than principal components analysis as a tool to reduce the dimensionality of data.*



In the same article, they make an interesting comment that meshes with Andrew Ng's comment about the recent increase in compute power and access to large datasets that has unleashed the untapped capability of neural networks when used at larger scale.

*It has been obvious since the 1980s that backpropagation through deep autoencoders would be very effective for nonlinear dimensionality reduction, provided that computers were fast enough, data sets were big enough, and the initial weights were close enough to a good solution. All three conditions are now satisfied.*

In a talk to the Royal Society in 2016 titled "[Deep Learning](#)", Geoff commented that Deep Belief Networks were the start of deep learning in 2006 and that the first successful application of this new wave of deep learning was to speech recognition in 2009 titled "[Acoustic Modeling using Deep Belief Networks](#)", achieving state of the art results.

It was the results that made the speech recognition and the neural network communities take notice, the use "deep" as a differentiator on previous neural network techniques that probably resulted in the name change.

The descriptions of deep learning in the Royal Society talk are very backpropagation centric as you would expect. Interesting, he gives 4 reasons why backpropagation (read "deep learning") did not take off last time around in the 1990s. The first two points match comments by Andrew Ng above about datasets being too small and computers being too slow.

What Was Actually Wrong With Backpropagation in 1986?

- we all drew wrong conclusions about why it failed.
- Real reasons were
  - 1) ~~our~~ labelled datasets were 1000s of times small
  - 2) computers were million times slow
  - 3) initialised weights in stupid way
  - 4) we used wrong type of non-linearity

What Was Actually Wrong With Backpropagation in 1986?

Slide by [Geoff Hinton](#), all rights reserved.

## Deep Learning as Scalable Learning Across Domains

Deep learning excels on problem domains where the inputs (and even output) are analog. Meaning, they are not a few quantities in a tabular format but instead are **images of pixel data, documents of text data or files of audio data**.

[Yann LeCun](#) is the director of Facebook Research and is the father of the **network architecture that excels at object recognition in image data called the Convolutional Neural Network (CNN)**. This technique is seeing great success because **like multilayer perceptron feedforward neural networks, the technique scales with data and model size and can be trained with backpropagation**.

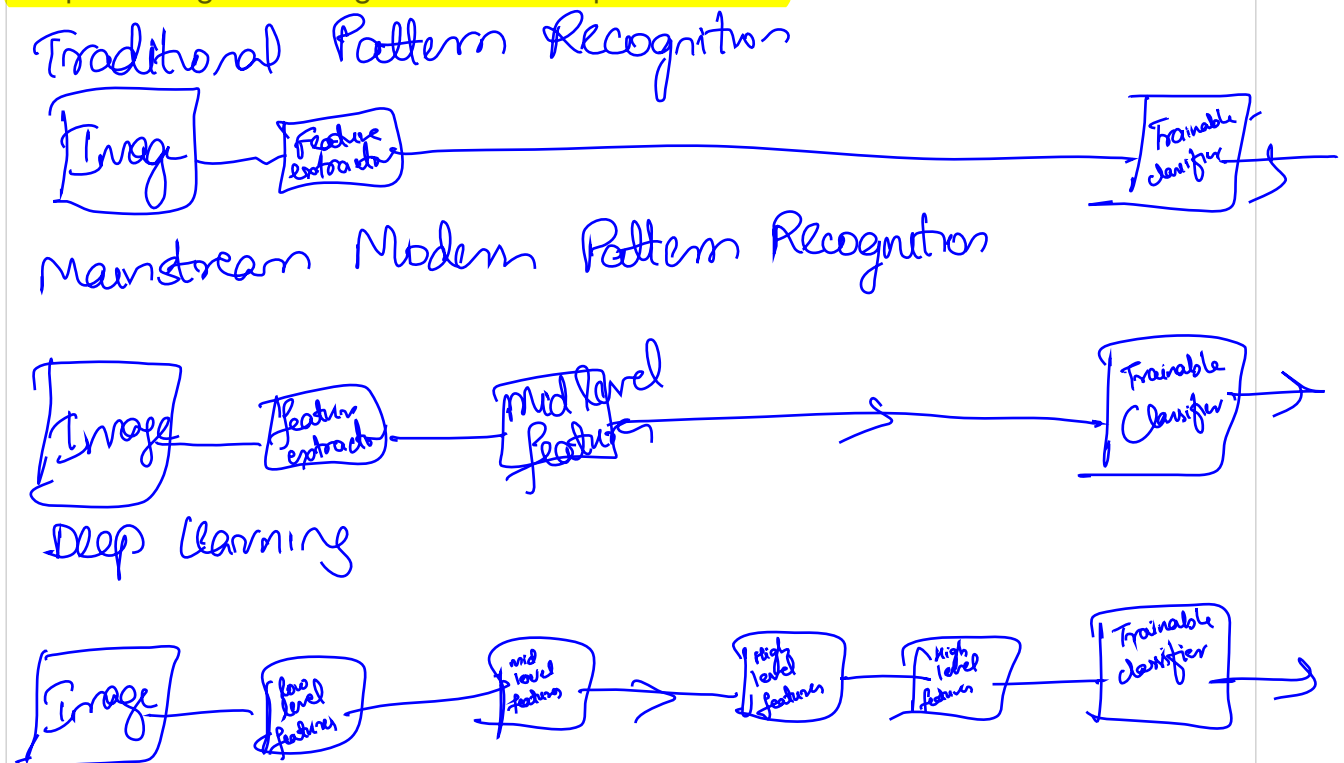
This **biases his definition of deep learning as the development of very large CNNs, which have had great success on object recognition in photographs**.

In a 2016 talk at Lawrence Livermore National Laboratory titled "[Accelerating Understanding: Deep Learning, Intelligent](#)

Applications, and GPUs" he described deep learning generally as learning hierarchical representations and defines it as a scalable approach to building object recognition systems:

deep learning [is] ... a pipeline of modules all of which are trainable. ... deep because [has] multiple stages in the process of recognizing an object and all of those stages are part of the training"

Deep Learning = Learning Hierarchical Representations



Deep Learning = Learning Hierarchical Representations

Slide by Yann LeCun, all rights reserved.

Jurgen Schmidhuber is the father of another popular algorithm that like MLPs and CNNs also scales with model size and dataset size and can be trained with backpropagation, but is instead tailored to learning sequence data, called the Long Short-Term Memory Network (LSTM), a type of recurrent neural network.

We do see some confusion in the phrasing of the field as "deep learning". In his 2014 paper titled "Deep Learning in Neural Networks:

[An Overview](#)" he does comment on the problematic naming of the field and the differentiation of deep from shallow learning. He also interestingly describes depth in terms of the complexity of the problem rather than the model used to solve the problem.

*At which problem depth does Shallow Learning end, and Deep Learning begin? Discussions with DL experts have not yet yielded a conclusive response to this question. [...], let me just define for the purposes of this overview: problems of depth > 10 require Very Deep Learning.*

[Demis Hassabis](#) is the founder of [DeepMind](#), later acquired by Google. DeepMind made the breakthrough of combining deep learning techniques with reinforcement learning to handle complex learning problems like game playing, famously demonstrated in playing Atari games and the game Go with Alpha Go.

In keeping with the naming, they called their new technique a Deep Q-Network, combining Deep Learning with Q-Learning. They also name the broader field of study "Deep Reinforcement Learning".

In their 2015 nature paper titled "[Human-level control through deep reinforcement learning](#)" they comment on the important role of deep neural networks in their breakthrough and highlight the need for hierarchical abstraction.

*To achieve this, we developed a novel agent, a deep Q-network (DQN), which is able to combine reinforcement learning with a class of artificial neural network known as deep neural networks. Notably, recent advances in deep neural networks, in which several layers of nodes are used to build up progressively more abstract representations of the data, have made it possible for artificial neural networks to learn concepts such as object categories directly from raw sensory data.*

Finally, in what may be considered a defining paper in the field, Yann LeCun, Yoshua Bengio and Geoffrey Hinton published a paper in Nature titled simply "[Deep Learning](#)". In it, they open with a clean definition of deep learning highlighting the multi-layered approach.

*Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction.*

Later the multi-layered approach is described in terms of **representation learning and abstraction.**

*Deep-learning methods are representation-learning methods with multiple levels of representation, obtained by composing simple but non-linear modules that each transform the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level. [...] The key aspect of deep learning is that these layers of features are not designed by human engineers: they are learned from data using a general-purpose learning procedure.*

This is a nice and generic a description, and could easily describe most artificial neural network algorithms. It is also a good note to end on.

## Summary

In this post you discovered that deep learning is just very big neural networks on a lot more data, requiring bigger computers.

Although early approaches published by Hinton and collaborators focus on greedy layerwise training and unsupervised methods like autoencoders, modern state-of-the-art deep learning is focused on training deep (many layered) neural network models using the

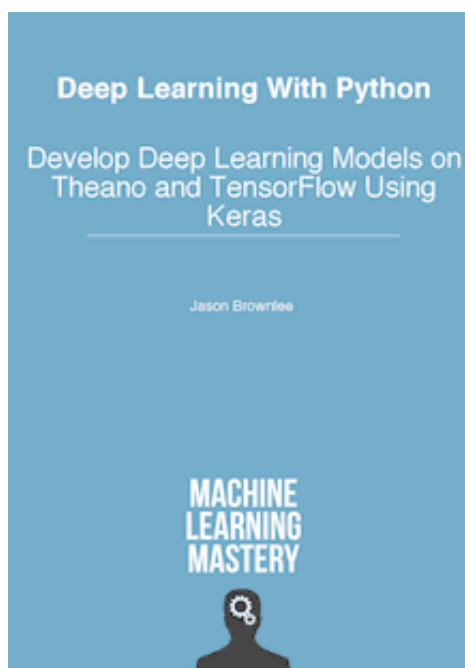
backpropagation algorithm. The most popular techniques are:

- Multilayer Perceptron Networks.
- Convolutional Neural Networks.
- Long Short-Term Memory Recurrent Neural Networks.

I hope this has cleared up what deep learning is and how leading definitions fit together under the one umbrella.

**If you have any questions about deep learning** or about this post, ask your questions in the comments below and I will do my best to answer them.

## Develop Deep Learning Projects with Python!



### What If You Could Develop A Network in Minutes

...with just a few lines of Python

Discover how in my new Ebook:

[Deep Learning With Python](#)

It covers **end-to-end projects** on topics like:

*Multilayer Perceptrons, Convolutional Nets and Recurrent Neural Nets, and more...*

### Finally Bring Deep Learning To Your Own Projects

Skip the Academics. Just Results.

[See What's Inside](#)

## **About Jason Brownlee**

Jason Brownlee, PhD is a machine learning specialist who teaches developers how to get results with modern machine learning methods via hands-on tutorials.