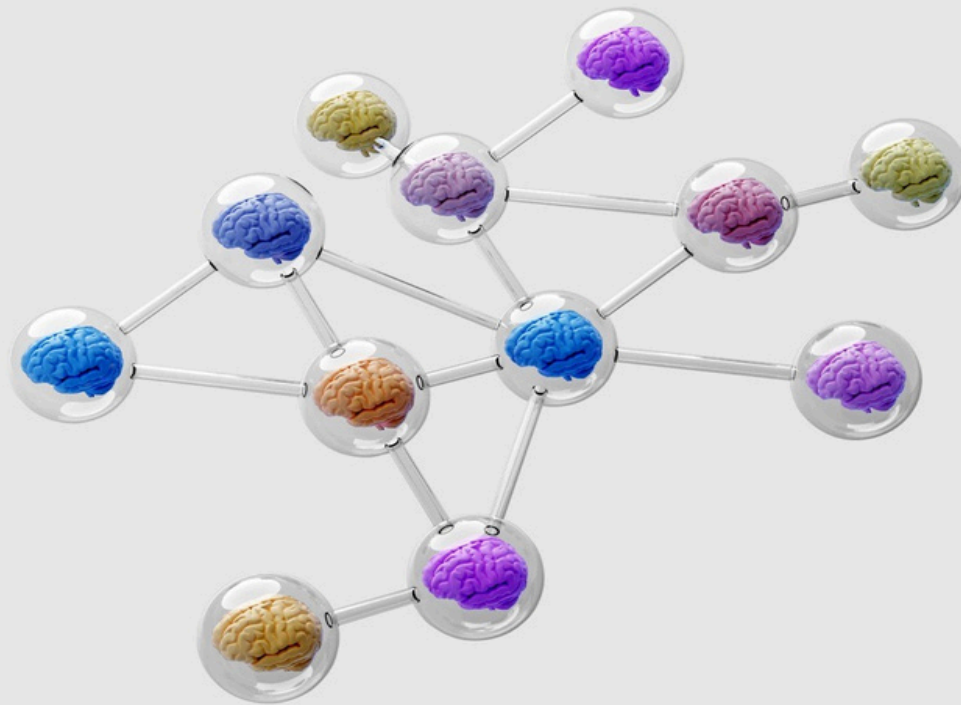


What is Deep Learning ?

Par M. Thirapounnho – GOAT LEARNING® Co-founding Partner



Deep learning is a type of machine learning that is more capable, autonomous, and accurate than traditional machine learning.

Deep learning has been around for a while, but most of us never used a deep learning–based tool until the release of OpenAI’s ChatGPT, in late 2022. (And even as we marveled at ChatGPT’s outputs, most of us didn’t know it was using deep learning to generate them.) Like its predecessors DALL-E, Google’s Imagen and PaLM, Stable Diffusion, and others, ChatGPT relies on large deep learning models trained on massive data sets to generate content based on prompts. But unlike its predecessors, ChatGPT works via an open-access API, which means the general public can experience the power of deep learning for the first time.

- The world of artificial intelligence and machine learning (for which deep learning is the next evolutionary step) is undergoing a **generational transformation**, from an idea studied by scientists to a tool used by all kinds of people for all kinds of tasks. McKinsey analysis has shown that between 2015 and 2021, the cost to train an image classification system (which runs on deep learning models) fell by 64 percent. Training times improved by 94 percent in the same period. We’ve also found that generative AI (gen AI) could add the equivalent of up to \$4.4 trillion annually to the global economy. These profound changes are all powered by deep learning.

But what actually is deep learning? And how does it make all this possible? Read on to find out.

What is machine learning?

Before we move to deep learning, let’s get the basics down. Machine learning is a form of artificial intelligence that can adapt to a wide range of inputs, including large data sets and human instruction. These algorithms can detect patterns and learn how to make predictions and recommendations by processing data and experiences, rather than by receiving explicit programming instruction. The algorithms also adapt in response to new data and experiences to improve over time. The volume and complexity of the data that is now being generated, too vast for humans to reckon with, has increased the need for machine learning—and has enhanced its potential. In the years since its widespread deployment, **machine learning has had impact in a number of industries**, including medical-imaging analysis and high-resolution weather forecasting.

How is deep learning different from machine learning?

Deep learning is a **more advanced version of machine learning that is particularly adept at processing a wider range of data resources** (text, as well as unstructured data including images), requires even less human intervention, and can often produce more accurate results than traditional machine learning. Deep learning uses neural networks—based on the ways neurons interact in the human brain—to ingest and process data through multiple neuron layers that recognize increasingly complex features of the data. For example, an early neuron layer might recognize something as being in a specific shape; building on this knowledge, a later layer might be able to identify the shape as a stop sign. Similar to machine learning, deep learning uses iteration to self-correct and to improve its prediction capabilities. Once it “learns” what an object looks like, it can recognize the object in a new image.



27-33 Rue du Colonel Pierre Avia
75015 Paris – FRANCE
www.goat-learning.com

What's the relationship between deep learning and gen AI?

ChatGPT made AI visible—and accessible—to the general public for the first time. ChatGPT, and other language models like it, were trained on deep learning tools called transformer networks to generate content in response to prompts. Transformer networks allow gen AI tools to weigh different parts of the input sequence differently when making predictions. Transformer networks, comprising encoder and decoder layers, enable gen AI models to learn relationships and dependencies between words in a more flexible way compared with traditional machine and deep learning models. That's because transformer networks are trained on huge swaths of the internet (for example, all traffic footage ever recorded and uploaded) instead of a specific subset of data (certain images of a stop sign, for instance). Foundation models, as further discussed below, trained on transformer network architecture—like OpenAI's ChatGPT or Google's BERT—are able to transfer what they've learned from a specific task to a more generalized set of tasks, including generating content. At this point, you could ask a model to create a video of a car going through a stop sign.

- Foundation models can create content, but they don't know the difference between right and wrong, or even what is and isn't socially acceptable. When ChatGPT was first created, it required a great deal of human input to learn. OpenAI employed a large number of human workers all over the world to help hone the technology, cleaning and labeling data sets and reviewing and labeling toxic content, then flagging it for removal. This human input is a large part of what has made ChatGPT so revolutionary.

What kinds of neural networks are used in deep learning?

There are three types of artificial neural networks used in deep learning :

I : Feed-forward neural network.

In this simple neural network, first proposed in 1958, information moves in only one direction: forward from the model's input layer to its output layer, without ever traveling backward to be reanalyzed by the model. That means you can feed, or input, data into the model, then "train" the model to predict something about different data sets. As just one example, feed-forward neural networks are used in banking, among other industries, to detect fraudulent financial transactions. Here's how it works: first, you train a model to predict whether a transaction is fraudulent based on a data set you've used to manually label transactions as fraudulent or not. Then you can use the model to predict whether new, incoming transactions are fraudulent so you can flag them for closer study or block them outright.

II : Convolutional neural network (CNN).

CNNs are a type of **feed-forward neural network** whose connectivity connection is inspired by the organization of the brain's visual cortex, the part of the brain that processes images. As such, CNNs are well suited to perceptual tasks, like being able to identify bird or plant species based on photographs. Business use cases include diagnosing diseases from medical scans or detecting a company logo in social media to manage a brand's reputation or to identify potential joint marketing opportunities.



27-33 Rue du Colonel Pierre Avia
75015 Paris – FRANCE
www.goat-learning.com

Here's how they work:

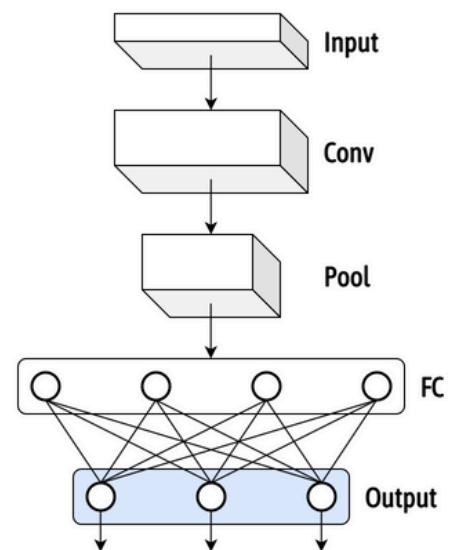
- First**, the CNN receives an image—for example, of the letter “A”—that it processes as a collection of pixels.
- In the hidden layers, the CNN identifies, unique features—for example, the individual lines that make up the letter “A.”
- The CNN can then classify a different image as the letter “A” if it finds that the new image has the same unique features previously identified as making up the letter.
- **Recurrent neural network (RNN).** RNNs are artificial neural networks whose connections include loops, meaning the model both moves data forward and loops it backward to run again through previous layers. RNNs are helpful for predicting a sentiment or an ending of a sequence, like a large sample of text, speech, or images. They can do this because each individual input is fed into the model by itself as well as in combination with the preceding input.

Continuing with the banking example, RNNs can **help detect fraudulent financial transactions** just as feed-forward neural networks can, but in a more complex way. Whereas feed-forward neural networks can help predict whether one individual transaction is likely to be fraudulent, recurrent neural networks can “learn” from the financial behavior of an individual—such as a sequence of transactions like a credit card history—and measure each transaction against the person’s record as a whole. It can do this in addition to using the general learnings of the feed-forward neural network model.

What is a foundation model?

Foundation models are deep learning models trained on transformer network architecture: vast quantities of unstructured, unlabeled data. Foundation models can be used for a wide range of tasks, either out of the box or adapted to specific tasks through fine-tuning. Fine-tuning involves a relatively short period of training on a labeled data set, which is typically much smaller than the data set on which the model was initially trained. This additional training allows the model to learn and adapt to the nuances, terminology, and specific patterns found in the smaller data set. Examples of foundation models include DALL-E 2, GPT-4, and Stable Diffusion.

Convolution Neural Network



What is a large language model?

Large language models are a class of foundation models that can process massive amounts of unstructured text. These models can learn the relationships between words or portions of words, also known as tokens. This enables large language models to generate natural language text, or perform tasks like summarization or knowledge extraction. Google’s Gemini runs on a large language model called LaMDA.

Bibliography

Goodfellow, I., Bengio, Y., & Courville, A. – Deep Learning (MIT Press)

<https://www.deeplearningbook.org>

Stanford University – CS231n: Convolutional Neural Networks for Visual Recognition

<http://cs231n.stanford.edu>

– Understanding and Improving Transformer Models

<https://ai.googleblog.com/2020/08/understanding-and-improving.html>

OpenAI – Introducing GPT-4

<https://openai.com/research/gpt-4>

MIT Technology Review – What is a Foundation Model?

<https://www.technologyreview.com/2022/08/22/1057349/what-is-a-foundation-model-ai/>

NVIDIA – Deep Learning Overview

<https://www.nvidia.com/en-us/deep-learning-ai/>

Harvard University – "AI and Deep Learning: Transforming the Future" (Harvard Gazette)

<https://news.harvard.edu/gazette/story/2023/03/ai-and-deep-learning-transforming-the-future/>

Nature – “A survey of deep learning for scientific discovery” (2022)

<https://www.nature.com/articles/s42256-021-00407-0>

DeepMind – Research publications and breakthroughs

<https://www.deepmind.com/research>