

Deep Learning – Neural Networks

Author: Mária Hartyányi, iTStudy Hungary Training and Research Centre

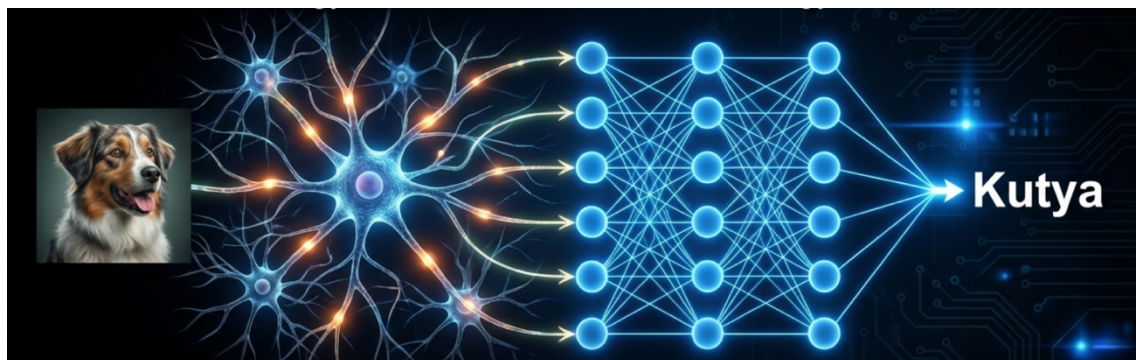
Reviewer: Zsolt Lengyel, iTStudy Hungary Training and Research Centre

Learning objective

Responsible use of artificial intelligence requires an understanding of the mechanisms that govern how it works. The aim of this material is to provide a clear picture of the operating principles of advanced artificial intelligence systems without requiring deep mathematical or programming knowledge. By the end of the material, teachers will be able to interpret the basic concepts needed for conscious use (such as artificial neural network, deep learning, training data, and parameter), distinguish between models in terms of resource requirements and performance, and explain their limitations.

Introduction

The development of the most effective machine learning methods was inspired by the study of processes taking place in the human brain. Neuroscientists observed that during learning, the connections between neurons change: some connections strengthen while others weaken.



The multilayer artificial neural network – a model of the natural neural network

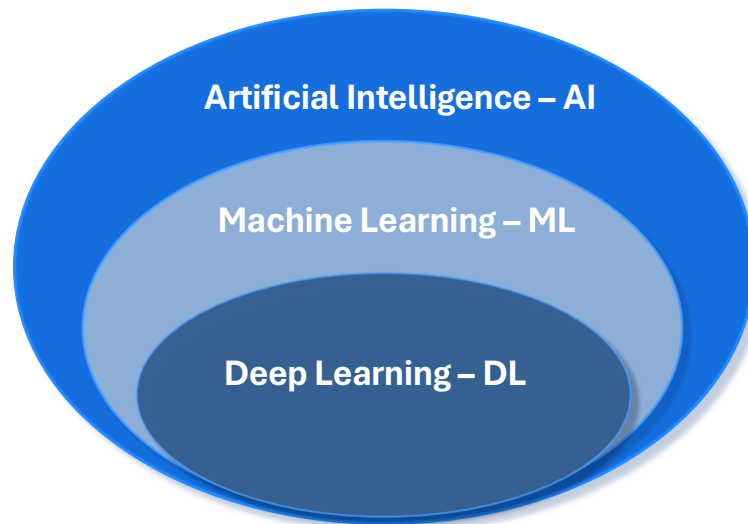
This insight led to the development of artificial neural networks (ANN). Although ANN mathematical models describe the processes taking place in the brain in a highly simplified way, they brought a breakthrough in the development of artificial intelligence. Among other things, these algorithms are what make it possible for us to communicate with artificial intelligence applications in natural human language.

From machine learning to deep learning

Artificial intelligence is a broad concept that encompasses any computer-based solution that mimics some capability associated with human intelligence.

One subfield of this is machine learning, in which the system does not operate solely on predefined rules, but recognizes patterns from data.

Deep learning is a narrower subfield of machine learning. Every deep learning system is both a machine learning and an artificial intelligence solution, but not every machine learning method uses a neural network.



Crucial English abbreviations: AI, ML, DL

The first ANN – the Perceptron

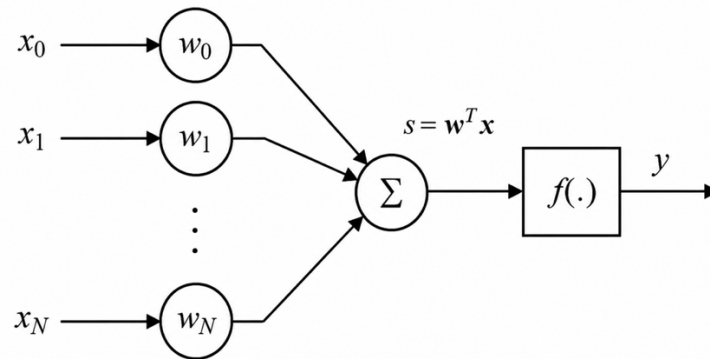
One of the early milestones in the history of neural networks was the perceptron, a learning-capable analog machine presented in 1958 in the United States by Frank Rosenblatt, a professor at Cornell University.



Rosenblatt's Mark I Perceptron, an analog machine operating with a 20x20 photocell array
(Source: Computer History Museum)

The image shows part of Rosenblatt's electromechanical machine, the Mark I Perceptron. During learning, the strength of the connections was stored and adjusted by electromechanical devices. Under controlled conditions, the system was able to distinguish simple shapes and standard letters.

The perceptron worked with a *supervised learning algorithm*. During training, the operator provided the correct answer for each input (which category the presented sample belonged to). The perceptron compared its own output with the correct answer and, in case of an error, adjusted the weights of the connections.



The mathematical model of the perceptron learning algorithm

The figure shows the mathematical model of the first, single-layer artificial neural network, i.e. the perceptron. This simple diagram illustrates that behind a machine solution capable of recognizing shapes there is no magic, only mathematical calculations. The input data are called artificial neurons. Each input (x_i) has an associated weight (w_i), and the result of the calculation is the weighted sum (S). From this value, the activation function (f) generates an output, which in our case is 0 or 1, depending on whether the machine guesses correctly or not. The w_i multipliers and other functions are called the parameters of the model.

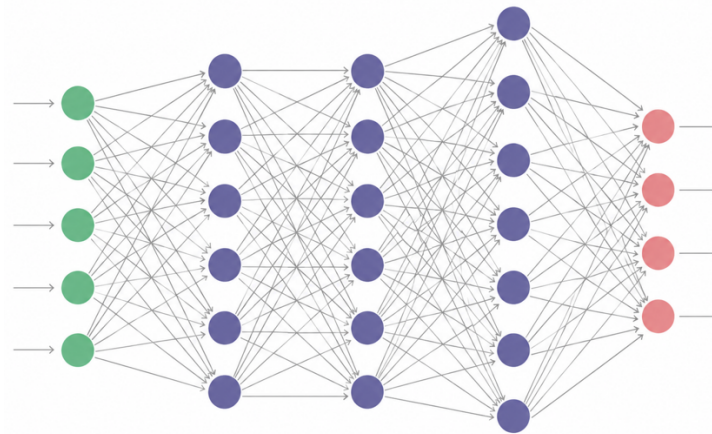
The single-layer neural network did not live up to the excessive expectations; from the 1970s onward, funding and interest in research declined, and computer performance was also insufficient for more complex tasks. This period is known as the **AI Winter**.

Multilayer neural networks and their parameters

With the emergence of greater computing capacity in the 1980s, the development of neural networks gained new momentum. The breakthrough came from connecting multiple neurons in sequence and performing processing not in one but in several layers.

The learning algorithms of artificial neural networks that operate with more than one layer are collectively known as deep learning (DL).

There is still no magic behind: the number of model parameters increases, the algorithms become more complex, but training still consists of mathematical operations, carried out on increasingly powerful machines. **AI does not think, it calculates — it makes decisions and gives answers on a statistical basis.**



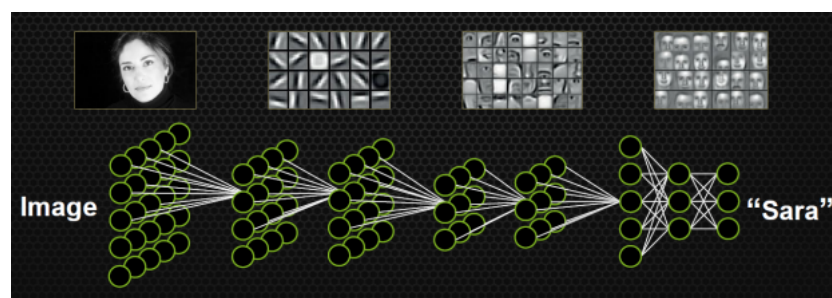
Input layer

Hidden layers

Output layer

Multilayer artificial neural network

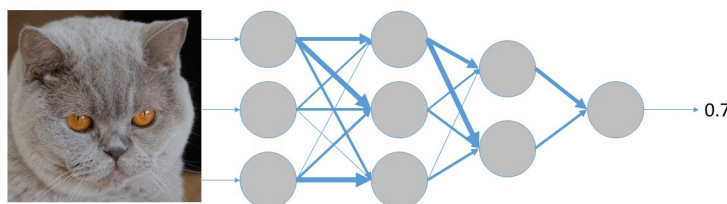
Between the input and output layers are so-called hidden layers; the first of these recognizes simple features, while the subsequent layers build increasingly complex patterns from them.



How a neural network works (Source: <https://developer.nvidia.com/>)

An image recognizer, for example, may first distinguish edges, then shapes, and finally objects.

Effective training of multilayer networks became widely usable thanks to backpropagation algorithm. This method works backward from the output error to determine how the connection weights of each layer should be adjusted.



Is this a cat? A model of a neural network operating with backpropagation

The figure shows a very simple model capable of determining whether a dog or a cat is in the picture. The magnitude of the parameters (weight values) is indicated by the thickness of the blue lines. The number shown in the red frame indicates the probability with which the model states that the picture

contains a cat. At least 1 million parameters are needed for such a simple model to function.

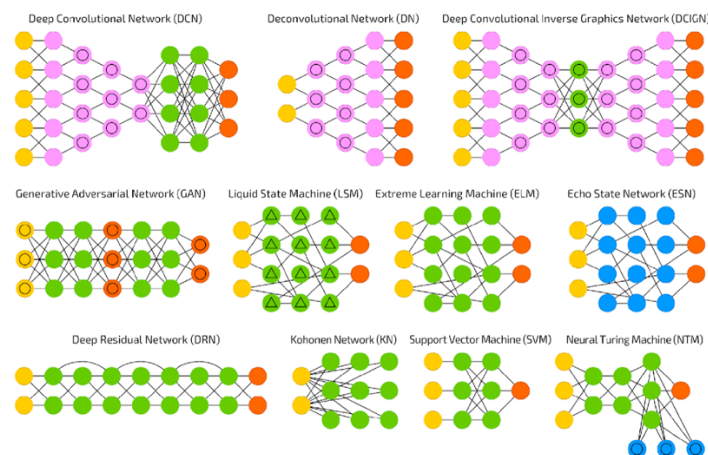
The size of deep learning models is usually characterized by two figures, the number of:

- parameters and
- hidden layers.

However, these two figures alone are not sufficient to describe the model's performance.

The width of the layers is also important, as are the number of inputs, the type of architecture, the size and quality of the training data, the activation functions, and the training algorithms.

The number of parameters in today's text-generating models ranges from a few million (small, task-specific models) to 1–2 trillion. Developers typically do not disclose the parameter counts of the largest, cutting-edge models (the latest versions of GPT, Claude, and Gemini) for business reasons.



A few examples of neural networks (Source: <https://www.deep-learning-site.com/>)

Deep learning models

In practice, neural networks can have very diverse structures depending on the problem to be solved. Some models work with sequences and time-based data, while there are specialized networks for image processing, compression, and pattern recognition. Following the emergence of the perceptron, numerous different neural network architectures developed, and today's modern artificial intelligence systems are built on further-developed versions of these.

Most AI applications offer the option of choosing between several models, and different models are optimized for different capabilities, speeds, and task types. One model may answer the same question faster, while another may perform better on complex logical tasks, programming, analyzing long documents, or detailed planning. It's as if we were choosing between different “engines” within the same application.

How do they differ?

Accuracy and reasoning: a more advanced reasoning model can solve multi-step problems more reliably.

Speed: for simple questions, a faster model may be more practical.

Programming and mathematics: some models are particularly strong at coding, debugging, or calculations.

Reading comprehension: when processing long documents, contracts, or teaching materials, models can differ in how well they follow the underlying context.

Creativity: the quality and style of the responses can also differ in writing, brainstorming, or creating educational materials.

Tool use: certain functions, such as web search, file analysis, or the use of other tools, may also depend on the current configuration.

The resource requirements of deep learning

From the simple artificial neuron handling a few inputs, we have now arrived at systems containing many billions of connections that process text, images, and sound. The current heightened expectations are based on the successes achieved with the deep learning algorithm.

Training deep learning models, however, involves enormous costs. It requires high-performance graphics processors and long runtimes, often lasting several months, with significant energy consumption. Using a finished model is generally much cheaper than training it; the largest cost of development is usually associated with data preparation and the training process.

Prompt to the ChatGPT GPT-5.6 Sol model

How long did it take to train the GPT-5.5 model?

OpenAI – the developer of ChatGPT – has not disclosed the exact duration of GPT-5.5's training, so it is not possible to state reliably how many days or months it took. From the published documents, it is only known that:

- part of the training data extended up to February 2026;
- web-based data collection took place roughly between 2018 and December 2025;
- the model was trained on datasets containing more than 10 trillion text tokens, more than 1 billion images, and more than one million hours of audio and video material.

Takeaways for teachers

1. It matters which model we work with, and when!

Most systems offer us the option to choose, and we can also ask at any time which model is currently being used!

If, for example, we ask “*What is the capital of France?*”, it hardly matters which model we use — we will get the correct answer.

If, however, we ask: “*Analyze this eighty-page application, identify the contradictions, and then suggest improvements based on the EU evaluation*”

criteria" it can matter a great deal whether we are working with an advanced or a weaker model.

On the other hand, if we use a powerful model for simple tasks, we will pay more, we will use up our available credits faster!

2. The models' uncertainty stems from their underlying algorithms, and for this reason it cannot be eliminated.

Although the algorithms are constantly improving, because the uncertainty arises from their fundamental operating principle, it can only be reduced, never fully eliminated.

In object recognition applications, the system often also displays percentage-based confidence values.



AI trained for object recognition, which frames a shape within the area observed by the camera and indicates next to it the probability that the object might be a person, a boat, etc.
(Source: <https://arteries.hu/>)

Even the response generated by a text generator is only true with a certain probability, but the probability value is not indicated.

Asking ChatGPT

How can we trust AI's answers if it is often wrong?

Answer

You shouldn't trust it blindly, but it can still be a very useful tool. It's worth thinking of it as an extremely knowledgeable, fast, and helpful assistant who usually answers well but can sometimes be confidently wrong.

The level of trust also depends on the type of question involved.

- **High reliability:** mathematics (if there is no calculation error), programming, grammar, general scientific knowledge, explaining concepts.

- **Medium reliability:** historical details, lesser-known facts, professional summaries.
- **Lower reliability** *without external verification*: breaking news, current legislation, stock market data, rare statistics, specific quotations, exact references.

The difference between models from a few years ago and today's models can be as great as the difference between a 2005 mobile phone and a modern smartphone.

When someone says, "I've already tried AI and it didn't work well," their opinion is often based on earlier experiences. And if this experience does relate to the latest AI systems, it is likely that a weaker model or a free version is being used, which significantly affects the result. Subscription-based services typically deliver results that are orders of magnitude better than free ones.

Copyright © 2026 VITA Consortium

This work may be reused and adapted under the terms of the Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).



Funded by the European Union. Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the National Agency.