

Large language models are artificial neural networks (algorithms) that have gone from a recent development to widespread use within a few years. They have been instrumental in the development of ChatGPT, the next evolutionary step in artificial intelligence. Generative AI was combined with large language models to produce a smarter version of artificial intelligence.

Large language models (LLMs) are based on artificial neural networks, and recent improvements in deep learning have supported their development.

A large language model also uses semantic technology (semantics, the semantic web, and natural language processes). The history of large language models starts with the concept of semantics, developed by the French philologist, Michel Bréal, in 1883. Bréal studied the ways languages are organized, how they change as time passes, and how words connect within a language.

Currently, semantics is used for languages developed for humans, such as Dutch or Hindi, and artificial programming languages, such as Python and Java.

Natural language processing, however, is focused on translating human communications into a language understood by computers, and back again. It uses systems that can provide an understanding of human instructions, allowing computers to understand written text, recognize speech, and translate between computer and human languages.

### How Natural Language Processing Was Almost Lost Before It Started

From 1906 to 1912, Ferdinand de Saussure taught Indo-European linguistics, general linguistics, and Sanskrit at the University of Geneva. During this time he developed the foundation for a highly functional model of languages as systems.

Then, in 1913, he died, before organizing and publishing his work.

Fortunately, Albert Sechehaye and Charles Bally, two instructors who were also Saussure's colleagues, recognized the potential of his concepts and decided they were important enough to save. The two instructors collected his notes for his future manuscript, and then made the effort to gather the notes of Saussure's students. Based on these, they wrote Saussure's book, titled *Cours de Linguistique Générale* (translated to *Language as a Science*, which eventually evolved into natural language processing (NLP), which was published in 1916.

*Language as a Science* laid the foundation of the [structuralist approach](#), and later, natural language processes.

### The Need for Language Translation Jump-Starts Natural Language Processing

After the end of World War II (1945), the field of [natural language processing](#) received a great deal of attention. Peace talks and the desire for international trade prompted recognition of the importance of understanding one another and promoted the hopes of creating a machine that could translate languages, automatically.

Not too surprisingly, the goal of building a language translation machine wasn't as easy as first assumed. However, while human languages are filled with chaos and broken rules, the language of mathematics is not. Language translation machines could be adapted quite successfully to mathematics, with its unchangeable rules.

### Machine Learning and the Game of Checkers

Arthur Samuel of IBM developed a computer program for [playing checkers](#) in the early 1950s. He completed a number of algorithms that allowed his checker-playing program to improve and described it as "[machine learning](#)" in 1959.

### The Mark 1 Perceptron Uses Neural Networks

In 1958, Frank Rosenblatt, of the Cornell Aeronautical Laboratory, merged Hebb's algorithmic model of neural networks with Samuel's work on machine learning, creating the first artificial neural network, called the Mark 1 Perceptron. Although language translation was still a goal, computers were being built primarily for mathematical purposes (much less chaotic than languages). These huge computers, built with vacuum tubes, and used as calculators, were not manufactured, but built individually, as were their software programs.

The Perceptron was also unique because it used software designed for the IBM 704, and established that similar computers could share standardized software programs.

Unfortunately, the [Mark 1 Perceptron](#) could not recognize many kinds of basic visual patterns (such as faces), resulting in broken expectations and cuts to neural network research and machine learning.

### ELIZA Uses Natural Language Programming

In 1966, an MIT computer scientist, Joseph Weizenbaum, developed ELIZA, which is described as the first program using NLP. It could identify keywords from the input it received, and respond with a pre-programmed answer.

Weizenbaum was attempting to prove his assumption that the communications between humans and machines were fundamentally superficial, but things didn't work out as planned. To simplify the experiment and minimize disputes, Weizenbaum developed a program using "[active listening](#)," which did not require a database storing real-world information, but would reflect back a person's statements to carry the conversation forward.

He was surprised (perhaps even horrified) that people, including Weizenbaum's own secretary, described the computer program as having human-like feelings. Weizenbaum [wrote](#): "My secretary, who had watched me work on the program for many months and therefore surely knew it to be merely a computer program, started conversing with it. After only a few interactions with it, she asked me to leave the room." He [later added](#), "I had not realized ... that extremely short exposures to a relatively simple computer program could induce powerful delusional thinking in quite normal people."

The original version of [ELIZA](#) recently became open-source and is available [here](#).

## Machine Learning Carries on as a Separate Industry

The years between [1974 and 1980](#) are referred to as "the first AI winter." AI researchers had to deal with two very basic limitations: small amounts of data storage and painfully slow processing speeds. Most universities had abandoned neural network research and a schism developed between AI and machine learning. Before this schism, ML was used primarily to train artificial intelligence.

However, the machine learning industry, which included several researchers and technicians, reorganized itself into a separate field. After the schism, the ML industry shifted its focus to probability theory and statistics, while continuing to work with neural networks. ML was used to answer phones and perform a variety of automated tasks.

## Small Language Models

Early development of the first (small) [language models](#) was started in the 1980s by IBM, and they were/are designed to predict the next word in a sentence. Part of their design includes a "dictionary," which determines how often certain words occur within the text the model was trained on. After each word, the algorithm recalculates statistically what the following word should be. (This limited statistical model does not support the creativity offered by ChatGPT.)

## NLP Is combined with Machine Learning and Research Funding Returns

By the late 1980s, computational power had increased significantly. Additionally, machine algorithms had improved and a revolution in natural language processing came about. It was the result of both the steady increase of computational power and the shift to machine learning algorithms. (Prior to the 1980s, most NLP systems used complicated, "handwritten" rules.)

During the 1990s, the use of statistical models for NLP analyses increased dramatically because of their speed and the tremendous flow of text moving through the internet.

## The World Wide Web Provides a Massive Source of Data

Tim Berners-Lee thought of the World Wide Web (WWW) in 1989, and it was made available to the public in 1991. The World Wide Web makes it possible for large library models to access massive amounts of data for research.

The creation of the World Wide Web made the internet searchable and provided large language models with access to massive amounts of information. The [World Wide Web](#) offers a platform to create, store, locate, and share information on a variety of topics. During the mid-1990s, the WWW initiated new levels of use on the internet, promoting interest in online shopping and what was called “surfing” the internet.

## GPUs and Large Language Models

Large language models require complex training, which includes the use of huge amounts of data containing billions of words and phrases. Training large language models can be described as training the individual pieces of a massive jigsaw puzzle, with each puzzle piece representing a portion of the LLM’s understanding. GPUs ([graphics processing units](#)) provide a solution to these problems.

A GPU is an electronic circuit that was originally designed to accelerate the processing of computer graphics and images. GPUs can process several pieces of data simultaneously, which makes them extremely useful for machine learning, gaming applications, video editing, and 3D graphics.

As the GPU’s memory capacity and speed increased, they played a significant role in developing sophisticated language models.

## Deep Learning and Large Language Models

In the 1990s, the arrival of deep learning supported even more advanced language models. A large language model is a very large deep learning model that is pre-trained on massive amounts of data. [Deep learning](#) is a form of machine learning, which is also a neural network, but with additional layers.

In 2011, deep learning began becoming popular. By 2018, deep learning algorithms were being used in every industry, from photography to online detail. Some of the ways deep learning applications were used include Apple’s Siri, automated drug design, and NLP for sentiment analysis.

## The Generative Adversarial Neural Network

In 2014, Ian Goodfellow introduced the [Generative Adversarial Neural Network](#) (a concept prompted by a conversation with friends while at a bar). The design uses two neural networks, which play against one another in a game. The game’s goal is for one of the networks to imitate a photo, tricking the opposing network into believing the imitation is real. The opposing network is looking for flaws – evidence the photo is not real. The game continues to be played until the photo is so close to perfect it tricks its opponent.

## Large Language Models Support Smarter AI

Near the end of 2022, OpenAI released ChatGPT and changed the world of AI, dramatically. They offered a powerful new chatbot capable of communicating in normal, human-like English, and able to complete a wide range of tasks, including developing new software and writing speeches.

**Generative AI**, with the support of large language models, produced a new level of intelligent behavior in chatbots. OpenAI's "smarter chatbots" quickly became a powerful tool useful for research, good writing, and generating realistic images or videos.

A design for large language models used for the new chatbots, called OpenChatKit, was **open-sourced** on March 10, 2023, by Together Computer