UDC: 004.8

# GENERATIVE PRE-TRAINED TRANSFORMER 3 (GPT-3)

*Makey P.I.*

*Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus*

Lazarenko *A. M. – senior lecturer of the department of foreign languages*

**Annotation.** In our time, the Internet has become an integral part of our life. Thanks to new technologies, our experience of using the global network has become even better. One of these technologies is GPT-3

**Keywords.** Artificial intelligence, gadget, invention, machine learning.

***Introduction***. Today, the Internet has become a part of everyday life. Millions of people around the world use the world Wide Web for correspondence, calls, meetings, as well as for ordering any goods. The features of one of these technologies: GPT-3, which makes the use of the internet more convenient are described in this paper.

GPT-3 (Generative Pre-trained Transformer 3) is a language model that was created by OpenAI, artificial intelligence research laboratory in San Francisco. It requires a small amount of input text to generate large volumes of relevant and sophisticated machine-generated text [1].

***The main part.*** GPT-3's deep learning neural network is a model with over 175 billion machine learning parameters. To put things into scale, the largest trained language model before GPT-3 was Microsoft's Turing NLG model, which had 10 billion parameters. In early 2021, GPT-3 is the largest neural network ever produced. As a result, GPT-3 is better than any prior model for producing text that is convincing enough to seem like a human could have written it [2].

In terms of where it fits within the general categories of AI applications, GPT-3 is a language prediction model. This means that it is an algorithmic structure designed to take one piece of language and transform it into what it predicts is the most useful following piece of language for the user [3].

It can do this thanks to the training analysis it has carried out on the vast body of text used to "pre-train" it. Unlike other algorithms that, in their raw state, have not been trained, OpenAI has already expanded the huge amount of compute resources necessary for GPT-3 to understand how languages work and are structured. The compute time necessary to achieve this is said to have cost OpenAI $4.6 million [3].

To learn how to create language constructions such as sentences, it uses semantic analytics — it studies not only words and their meanings, but also gathers understanding of how the use of words differs depending on other words also used in the text.

It's also a form of machine learning termed unsupervised learning because the training data does not include any information on what is a "right" or "wrong" response, as in the case with supervised learning. All of the information it needs to calculate the probability that its output will be what the user needs is gathered from the training texts themselves [3].

This is done by studying the usage of words and sentences, then taking them apart and attempting to rebuild them itself [3].

For example, during training, algorithms may encounter the phrase "there is a red door in the house". Then it is given a phrase again, but with a missing word – for example "there is a red X in the house.

He then scans the entire text in its training data – hundreds of billions of words arranged into a meaningful language – and determines which word should be used to recreate the original phrase.

To begin with, it will probably make mistakes potentially millions of times. But eventually it will find the right word. By checking its original input data, it will know that it has the correct output data, and "weight" is assigned to the algorithm process that provided the correct answer. This means that it gradually "learns" which methods are most likely to lead to the correct answer in the future.

The scale of this dynamic "weighing" process is what makes GPT-3 the largest artificial neural

network ever created. It was noted that in some way there is nothing new in what it does, since transformer models of language prediction have been around for many years. However, the number of weights that the algorithm dynamically stores in its memory and uses to process each request is 175 billion – ten times more than its closest competitor, manufactured by Nvidia.

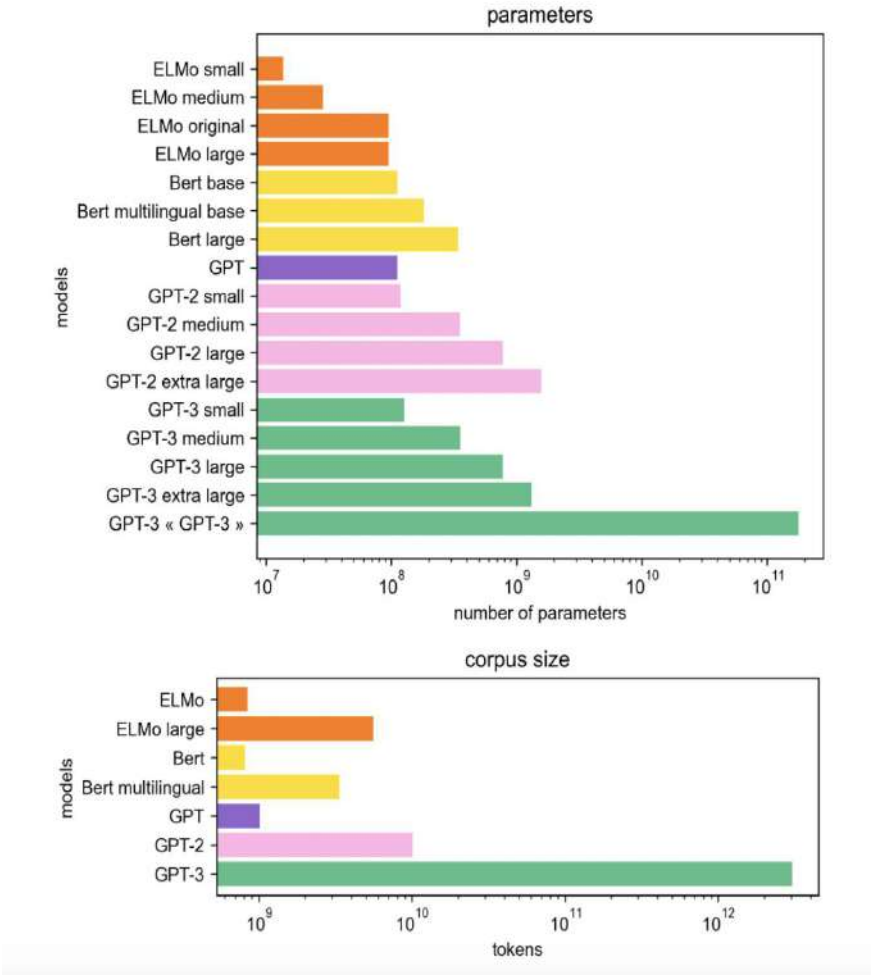


Figure 1 – Comparison of the number of parameters among the most popular models

***Conclusion.*** GPT-3 has attracted a lot of attention since last summer because it is by far the largest and possibly the most powerful language model created at the time of writing this article. However, GPT-3 still suffers from a number of limitations (GPT-3 has limited input and output sizes, has leaks of any form of memory, cannot 'Understand' Semantics, etc.) that make it far from an ideal language model or AGI example. If you want to use GPT-3 for research or commercial purposes, you can apply to use the Open AI API, which is currently in closed beta. Otherwise, you can always work directly with GPT2, which is publicly available and open source thanks to HuggingFace's transformers library. All in all, it's a fair conclusion that GPT-3 produces results that are leaps and bounds ahead of what we have seen previously.

## *References*

*1. What is GPT-3 and why is it so powerful?: [Electronic resource] - https://towardsdatascience.com/what-is-gpt-3-and-why-is-it-so-powerful-21ea1ba59811*

*2. GPT-3: [Electronic resource] - https://www.techtarget.com/searchenterpriseai/definition/GPT-3*

*3. What Is GPT-3 And Why Is It Revolutionizing Artificial Intelligence?: [Electronic resource] - https://www.forbes.com/sites/bernardmarr/2020/10/05/what-is-gpt-3-and-why-is-it-revolutionizing-artificial-intelligence/?sh=586d8e5a481a*