Title: Self-attention or Self-attention mechanism explained in the paper by the title “Attention is all you need”.

This paper presents a new type of Artificial Neural Network, known as the Transformer model. It can be used with sequences of the data, for instance the sentences in language.Before, most of the AI models for language tasks rely on large structures that are referred to as recurrent neural networks (RNNs) and convolutional neural networks (CNNs). The Transformer is different from these because the Transformer doesn’t use them. Still, it strictly relies on a factor known as “attention.”

The main parts of the Transformer are:

Encoder: This part takes in the input (like a sentence in English) and processes it.

Decoder: This part of the transformer produces the output (like a translation into German).

Attention layers: These let the model concentrate on distinct input components while generating each output component.

Multi-head attention: This lets the model look at the input in several different ways at once.

Positional encoding: Since the model doesn't process words in order, this helps it understand the order of words in a sentence.

A significant plus of the Transformer model is that it takes much less time to train than other forms of models. This is so because its operation can attend to all the segments of a sentence simultaneously and not have to go through each word in that sentence. The researchers applied the proposed model to MAP tasks – meaning the translation of text from one language to another.

They tried the moving of the vowel and the change of the direction from left to right The change of gender from masculine to feminine. As far as the translation from English to French is concerned, in both cases the Transformer performed the best, outstripping all models preceding it. It was also much faster to train but the accuracy and quality of the articles was questionable.

To check if the Transformer could be used for other purposes as well, they also tested it for English constituency parsing. This is a task in which the model attempts at identifying the syntactic pattern of sentences. It also performed well in this, although it was not clearly trained to do so.