

## Presentation NLP Project.

### Introduction:

Google's Transformer is a deep learning architecture that revolutionized the field of natural language processing and other sequence to sequence tasks. It was first introduced in the paper titled "Attention is all you need" by Vaswani et al. in 2017. The transformer architecture has since become a foundation for many subsequent advancements in NLP and related fields.

### The Transformer Architecture:

At its core, the transformer architecture is a dynamic and versatile framework that addresses the limitations of conventional sequence-based models such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs). Conceived by Vaswani et al. in 2017, transformers have since become a cornerstone of cutting-edge NLP models due to their remarkable parallelization capabilities, scalability, and proficiency in capturing long-range dependencies within sequences.

1. **Encoder:** Through self-attention mechanisms, the encoder processes input sequences and transforms them into a collection of high-dimensional feature representations. This unique attention mechanism empowers each word to weigh the significance of all other words within the sequence, thereby capturing intricate contextual relationships across varying distances.
2. **Decoder:** The decoder, building upon the encoder's output, generates predictions, often following an autoregressive approach. The decoder also employs self-attention to focus on relevant portions of the input sequence. This ensures that the generated text maintains coherence and contextuality by considering the previous tokens.

#### **Self-Attention and Positional Encoding:**

By allowing each word to flexibly allocate attention to other words, the model gains the ability to assign varying degrees of importance to different elements within the sequence. This self-attention mechanism underpins the model's contextual understanding, transcending conventional limitations associated with word order.

To counter the absence of inherent notions of word order, transformers incorporate positional encodings into input embeddings.

#### **Parameters Utilized in Training:**

In the process of training the GPT-2 language model on the Legal Ontario website dataset, several key parameters were meticulously selected to optimize the model's learning process and performance. These include:

GPT2 model has been used for training our corpus [4], although the architecture is not explicitly out, these are primarily based on the [3] transformer architecture of Google. [7] The gpt-2 is based on 1.5 billion parameters and is trained on a dataset of 9 million web pages. The main purpose of it is to

predict the next word given all of the previous words. As it is trained on a diverse dataset it is better for giving general responses. The gpt-2 model is trained with absolute positional embeddings and should be padded therefore.[8] The model has four versions gpt2, gpt2-Medium, gpt2-Large, and gpt2-XL. As the names suggest these are the sizes which models were trained and therefore the larger the better, but these consume a lot of resources and should be used only as per requirement. The gpt2 model also advises to use the model with care as sometimes the language can be disturbing or offensive to some and can propagate historical and current stereotypes.

Params used in the training:

model_name	gpt2
per_device_train_batch_size	4
num_train_epochs	1.0
save_steps	2000
block_size	256

By tuning and configuring these parameters, the training process achieves a delicate balance between computational efficiency and model accuracy, ultimately leading to a GPT-2 model that excels in generating contextually coherent and relevant text.

Conclusion:

As we navigate through the intricacies of training a GPT-2 language model using the Legal Ontario website dataset, we gain a deeper appreciation for how the interplay between the transformer's architectural elegance and the judicious selection of parameters shapes the model's performance and language generation capabilities. The convergence of these factors underscores the transformer architecture's transformative impact on NLP and its ability to unlock new dimensions in language understanding and expression.