

# 1 Attention

Motivation:

## 1.1 Tokens: Inputs for Transformer/Attention Layers

Definition:

### 1.1.1 Positional Encoding

Definition:

## 1.2 Attention

Process:

- 1.

Notes:

### 1.2.1 Residuals, Norms, and FFN

Notes:

- 

### 1.2.2 Self-Attention vs. Cross-Attention

Notes:

### 1.2.3 Multi-Head Attention

## 1.3 Transformers

Notes:

### 1.3.1 Transformer Block

Notes:

### 1.3.2 Transformers are GNNs

**Summary:** Transformers are a special case of GNN

## 1.4 Examples

### 1.4.1 Tokens

Example:

## 2 LLMs

### 2.1 Transformers & LLMs

#### Summary:

- 2.1.1 Inputs: Tokenizing Text & Embedding Layers
- 2.1.2 Outputs: Auto-Regressive Decoding of Tokens
- 2.1.3 Sizes of Text Datasets for LLMs
- 2.1.4 Text to Text Tasks
- 2.1.5 Transformers and Masking: Encoders and Decoders
- 2.1.6 Masking Language Modelling (Self-Supervised)

## 2.2 Scaling LLMs

Motivation:

### 2.2.1 Techniques

**Summary:** Table format

### 2.2.2 High-Level Impacts

**Summary:**

### 3 Transformers