Transformers

Navidut Tauhid

UCC Consultant and Cloud Architect

https://www.linkedin.com/in/naveedtauheed/

Overview

Transformers:

- Processing sequential data
- Architecture
- Key components
- Use cases

Transformers

Deep learning architecture

Generative AI models

Natural language processing (NLP)

Transformers

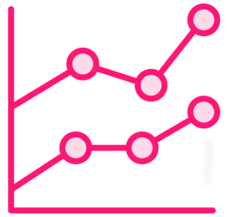
Context and understanding

Sequential data analysis

Ordered and correlated

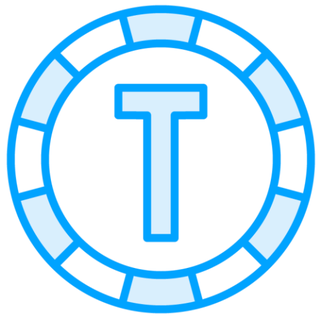
Daily conversation

Sequential words



Word relation

Transformers



Text Tokenization

Analyze Relationships

Transformers

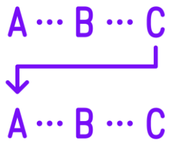
Pre-training and fine-tuning

Crucial stages

BERT and GPT

Natural language processing (NLP)

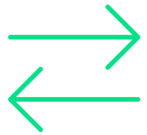
Transformers



Pre-training

Language structures

Large dataset



BERT: predicts both ways

GPT: predicts next word

Transformers

Fine-tune labeled data

Leverage understanding

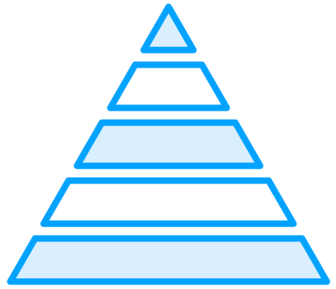
Specific tasks

Sentiment analysis

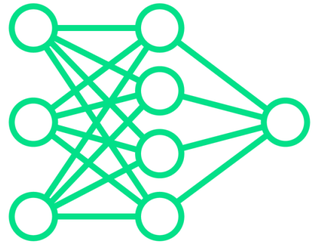
Save time and use less data

Text summarization

Transformer Architecture



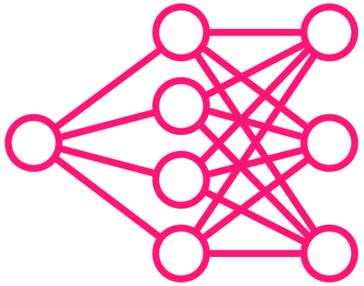
Based on
Encoder-decoder



Multiple layers
Neural network

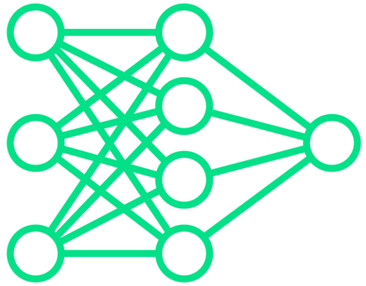
Transformer Architecture

Layer focuses on aspects



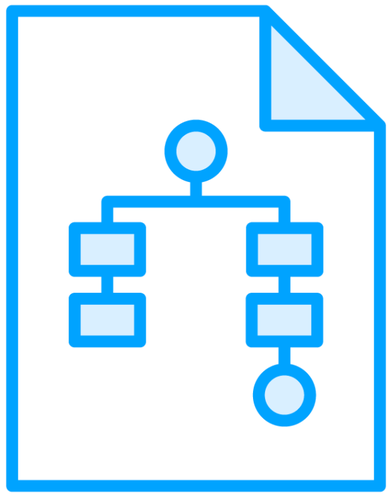
Encoder processes input

Self-attention captures features



Decoder generates output

Transformer Architecture



Self-attention:

- Understands the input sequence
- Example sentence: "The cat sat on the mat"
- Learns relationships: cat and sat

Transformer Architecture

Multi-head attention:

Parallel self-attention

Example: BERT-12, GPT3-96

Robust relationships

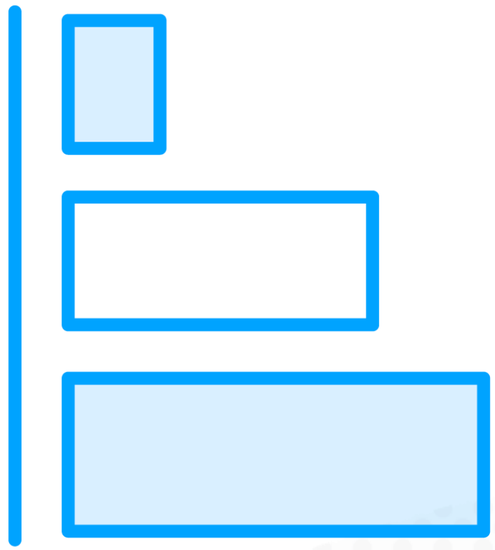
Transformer architecture:

Positional encoding

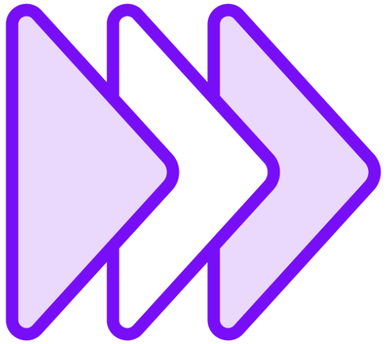
Understand word order

Faster processing

Relative positional information



Transformer Architecture



Feed-forward neural networks

- Context-aware representations
- Used in translation
- Question answering
- Multiple layers with non-linear functions

Demo

Review of successful transformer use cases:

- Natural language processing (NLP)
- Computer vision
- Audio and multi-modal processing

Summary

Transformers

Components:

- Self-attention
- Multi-head attention

Purpose:

- Understand the context
- Generate meaningful outputs

Fields:

- Text generation
- Language translation
- Chatbots
- Natural language processing (NLP)