

Name: Vinutha.B.S

USN:1BG23CS188

# TRANSFORMERS

Transformers are **deep learning models** specifically designed to handle sequential and structured data, such as natural language, time-series, images, and even video frames.

They are built around a key innovation called the **attention mechanism**, particularly **self-attention**, which allows the model to selectively focus on the most relevant parts of the input sequence.

Transformers were introduced in **2017 by Vaswani et al.** in their influential paper, *“Attention is All You Need.”*

This model completely changed the landscape of **Natural Language Processing (NLP)** and later expanded into fields like **computer vision, audio processing, and reinforcement learning.**

Transformers are **deep learning models** that rely on a mechanism called **self-attention** to process and understand sequences of data (like text, images, or audio).

Before transformers, models like **Recurrent Neural Networks (RNNs)** and **Long Short-Term Memory (LSTM)** were commonly used for sequential tasks.

However, these models had problems:

- **Slow training:** They processed one word (or step) at a time.
- **Limited memory:** They struggled to remember long-distance relationships in sequences.

We need transformers because:

- They process the entire sequence **in parallel**, making training much faster.
- They can **capture long-range dependencies** more effectively.
- They are more scalable and flexible for complex tasks.

## What Does a Transformer Do?

A transformer:

1. **Takes sequential input** (like a sentence or a series of image patches).
2. **Applies self-attention** to learn which parts of the input are important to each other.
3. **Processes the input in layers** to build a deep understanding of the data.
4. **Outputs predictions**
  - Translate languages
  - Classify images
  - Generate text

## Advantages

- **Parallelization:** Unlike RNNs, transformers can process all tokens simultaneously, enabling faster training.
- **Long-Range Dependency Handling:** They capture relationships between distant words better than traditional models.
- **Scalability:** Suitable for large datasets and complex tasks.

## Popular Transformer Models

- BERT (Bidirectional Encoder Representations from Transformers)
- GPT (Generative Pre-trained Transformer)
- T5 (Text-to-Text Transfer Transformer)
- ViT (Vision Transformer)

## **Conclusion**

Transformers have set a new standard in AI research and applications. Their ability to learn complex patterns and dependencies in data has made them the foundation of state-of-the-art models across various domains.