

What are Transformers in AI?

Blaine Swieder

18 July 2025

We get our very first definition of a transformer from the 2017 paper, *Attention Is All You Need*. Below is a definition:

Definition 0.1. A **Transformer** is a type of AI model that was introduced in 2017 that has changed how we work with language, images, and audio by using a smart method known as attention.

Now, what are the components of a transformer? These are as follows:

- **Encoder:** Understands the input (e.g., a sentence).
- **Decoder:** Generates an output (e.g., a translated sentence)
- **GPT:** This type of transformer only utilizes decoders (for text generation).
- **BERT:** This type of transformer only utilizes encoders (for understanding tasks).

But, why use these so called transformers? Well these are the advantages of the transformer:

- Handle long-range dependencies.
- Parallel Processing which means faster training.
- It is state-of-the-art in NLP, vision, speech, etc.
- It is the foundation for models such as GPT, BERT, T5, and even image generators like DALL-E.

Now, what are some examples?

- **BERT:** Bidirectional understanding
- **GPT:** Text Generation
- **T5:** Text-to-text tasks
- **ViT:** Vision Transformer (for image classification)
- **DALL-E:** Image generation from text.