

# Positional Encoding

## What Is Positional Encoding?

**Positional Encoding** is a technique used in Transformer models to add information about **word order (position in a sentence)** to token embeddings.

In simple terms:

*Positional encoding tells the model where each word appears in a sentence.*

## Why Positional Encoding Is Needed

Transformers rely on **self-attention**, which processes all tokens in a sentence simultaneously. As a result, Transformers do **not naturally understand word order**.

### Problem Without Positional Encoding

The following two sentences would appear identical to the model:

I love AI  
AI love I

Although the words are the same, the meanings are different due to word order.

**Positional encoding solves this problem by injecting position information.**

## How Positional Encoding Works (Intuition)

The process works as follows:

- Each token has a word embedding
- A positional vector is created for each position (0, 1, 2, ...)

- The positional vector is added to the word embedding

**Final embedding:**

$$\text{Final Embedding} = \text{Word Embedding} + \text{Positional Encoding}$$

This allows the model to know:

- What the word is
- Where the word is

## Simple Example

Sentence:

The bank approved the loan

Word positions:

The(0), bank(1), approved(2), the(3), loan(4)

Even though the word “the” appears twice:

“the” at position 0  $\neq$  “the” at position 3

This difference occurs because their positional encodings are different.

## Mathematical Formulation (Sinusoidal Positional Encoding)

The original Transformer model uses **sinusoidal positional encoding**.

For a position pos and dimension  $i$ :

$$PE(\text{pos}, 2i) = \sin\left(\frac{\text{pos}}{10000^{2i/d}}\right)$$

$$PE(\text{pos}, 2i + 1) = \cos\left(\frac{\text{pos}}{10000^{2i/d}}\right)$$

## Meaning of Symbols

- $\text{pos} \rightarrow$  position of the word in the sentence
- $d \rightarrow$  embedding dimension

These sine and cosine functions create unique patterns for each position.

**You do not need to memorize the formula—only understand the idea.**

## Why Sine and Cosine Are Used

Sine and cosine functions are chosen because they:

- Assign a unique encoding to each position
- Allow the model to infer relative positions
- Work for any sentence length
- Require no learning (fixed encodings)

## Learned vs Sinusoidal Positional Encoding

Type	Description
Sinusoidal	Fixed, mathematical (original Transformer)
Learned	Trainable position embeddings (used in BERT, GPT)

Modern Transformer-based models often use **learned positional embeddings**, but the core concept remains the same.