

# :Transformer AI

■: Attention Is All You Need ■: Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, Illia Polosukhin (Google■) ■■■■■: 2026■1■14■

ChatGPT GPT-4, Attention Is All You Need

2017, Google introduced "Transformer" (RNN)

■ 2017, "Seq2Seq" (Seq2Seq) RNN, ■

RNN: [REDACTED]

```
    return LSTM(hidden_size)(x, h0)
```

"it" "animal" "street", RNN

RNN Transformer: RNN, "Attention Is All You Need"

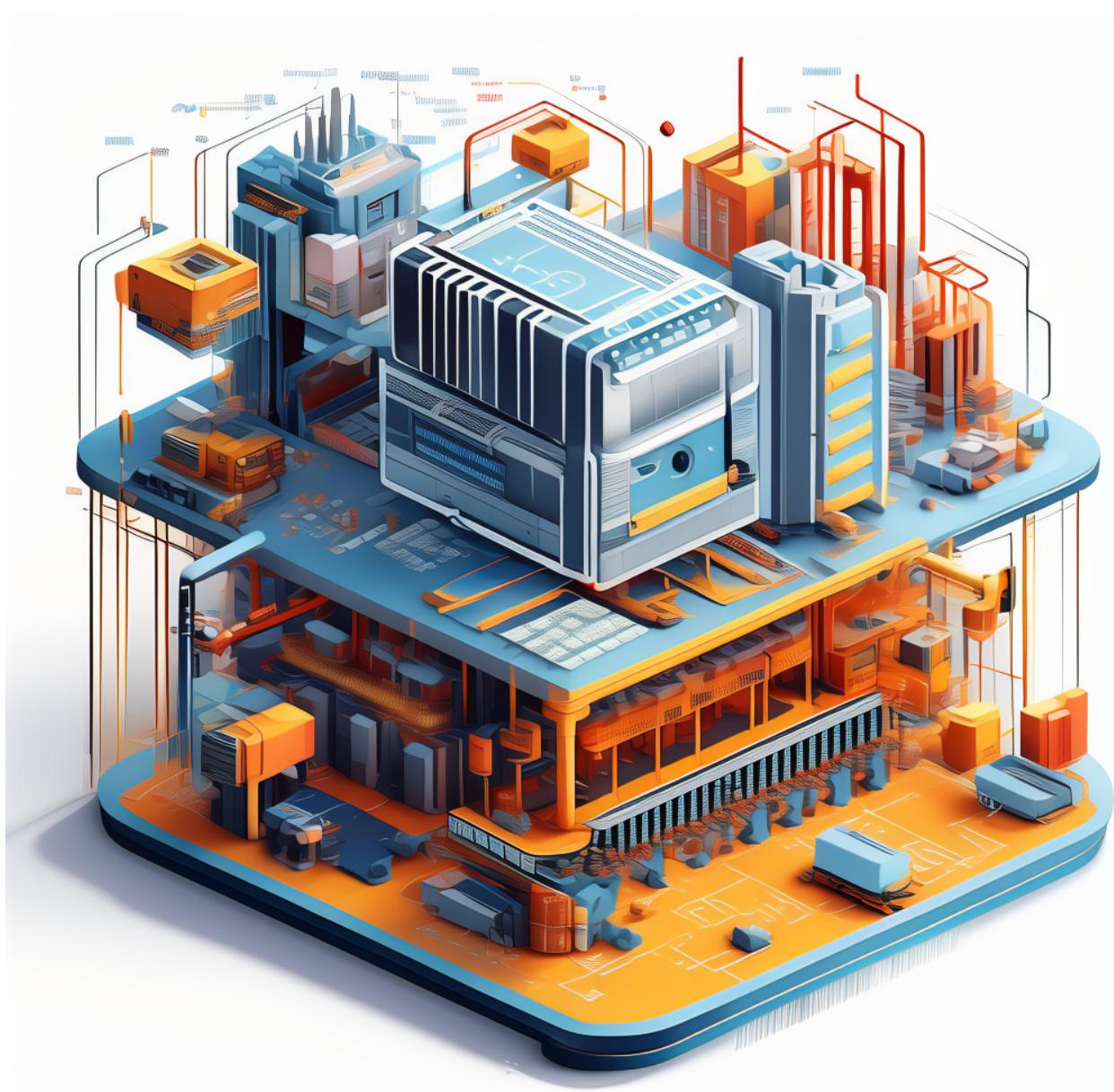
Transformer: [REDACTED]

• -

Transformer "Encoder-Decoder":

**(Encoder):**  ,  , 

[REDACTED], RNN [REDACTED] Transformer [REDACTED]



*Transformer* 

### 1: (Self-Attention)

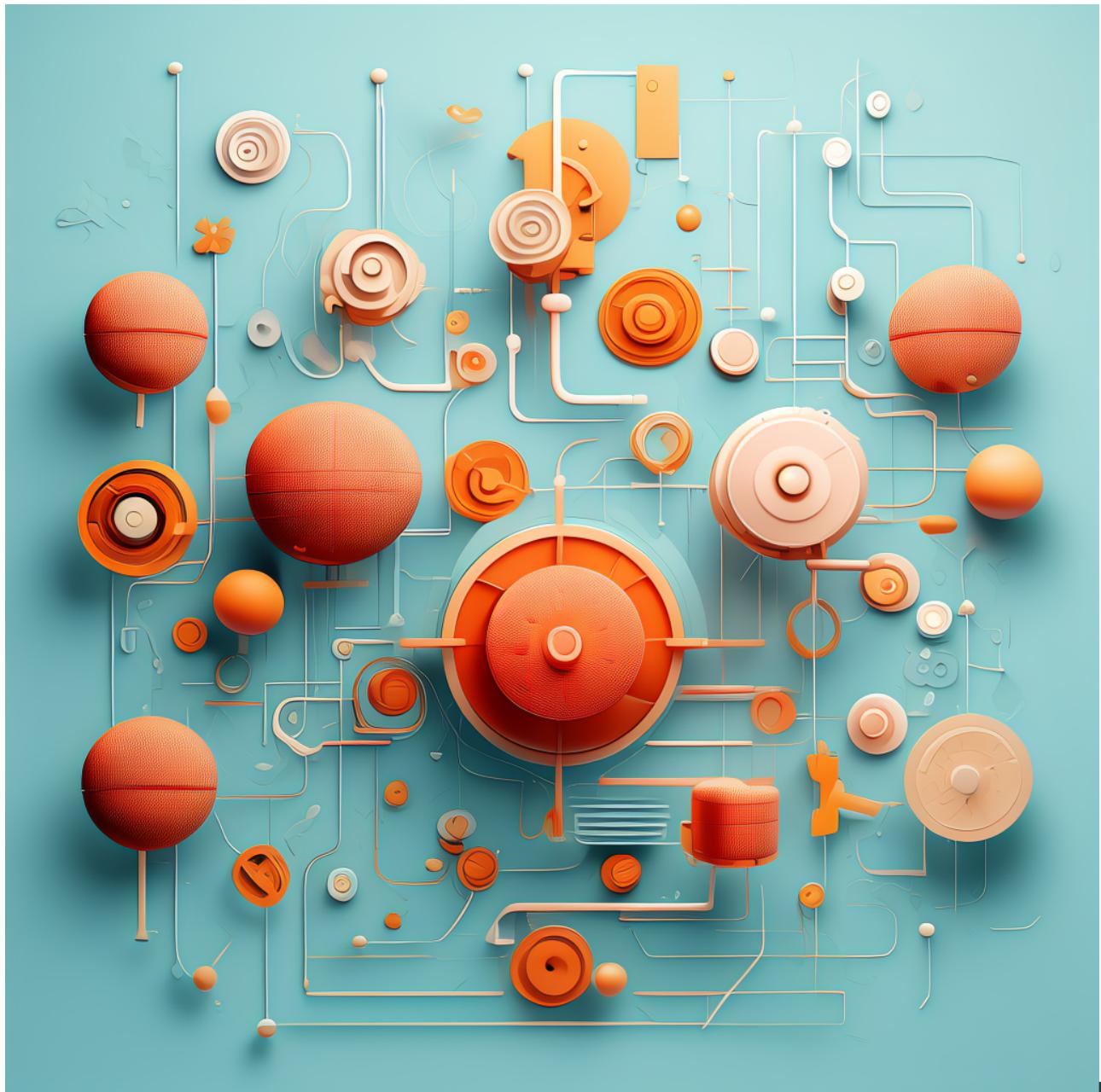
—“I’m not going to let you do that again.”

[View all posts](#)

10 of 10

$$\text{Attention}(Q, K, V) = \text{softmax}(Q \cdot K^T / \sqrt{dk}) \cdot V$$

██████████: ██████████! ██████████" "█████████████████, ██████████ RNN ██████████



**Query Key Value**

## 2: (Multi-Head Attention)

—

Transformer 8" "(Attention Heads),": - 1 ( ) - 2 ( ) - 3

88

### 3: (Positional Encoding)

```

PE(pos, 2i) = sin(pos / 10000^(2i/d))
PE(pos, 2i+1) = cos(pos / 10000^(2i/d))

```

pos ( ), i : ,

10

[REDACTED], Transformer [REDACTED]:

■■■■■ (Feed-Forward Network):

**(Residual Connection & Layer Normalization):**

███████████(Masking)███████████,███████████i███████████,███████████i-1███████████,"███████████"

66,6500

10 of 10

Transformer:

1:

(big) WMT

:Transformer 41.8 BLEU .

["base", Transformer(["base"])]

12:

Transformer (big) 8 P100 GPU 3.5 Transformer

: - Transformer (base) :  $3.3 \times 10^{18}$  FLOPs - ConvS2S :  $7.7 \times 10^{19}$  FLOPs( )

20

13:

|                |                          |        |                |
|----------------|--------------------------|--------|----------------|
|                |                          |        |                |
| Self-Attention | $O(n^2 \cdot d)$         | $O(1)$ | $O(1)$         |
| RNN            | $O(n \cdot d^2)$         | $O(n)$ | $O(n)$         |
| CNN            | $O(k \cdot n \cdot d^2)$ | $O(1)$ | $O(\log_k(n))$ |

10 of 10

**Q(1):** RNN vs. n

**O(1):**  RNN  $\rightarrow$  n  $\rightarrow$  n

Transformer

## 4: Transformer

Transformer: A sequence-to-sequence model.

- Transformer: "its" → "it's"
- Transformer: "I am" → "I'm"
- Transformer: "I am" → "I am"

Transformer, Encoder-Decoder

## 5: Transformer

Transformer: A sequence-to-sequence model (English Constituency Parsing)

Transformer: 4 Transformer, 92.7 F1, Transformer, Transformer

---

### 1. Transformer

Transformer: BLEU:

#### 1.1. Encoder-Decoder

Transformer, (Encoder-Decoder), "RNN"

Transformer: Encoder-Decoder, RNN

#### 1.2. Encoder-Decoder

Transformer, Transformer, RNN

Transformer, Transformer, NLP: - **BERT**(2018): Transformer, 11 NLP -  
**GPT**(2018-2023): GPT-1, GPT-4, T5, BART, XLNet

, Transformer, AI

#### 1.3. AI

Transformer, AI, AI:

- Vision Transformer (ViT), Transformer, CNN
- Speech Transformer, HMM, RNN
- Decision Transformer
- AlphaFold 2, Transformer, 50

AI, " ",

#### 1.4. AI

Transformer, AI:

- Google 2018 Transformer, Transformer
- Google BERT(Transformer)
- ChatGPT, Claude AI, Transformer
- GitHub Copilot, Transformer

, AI, Transformer

---

### 2. Transformer

## 1:

$O(n^2 \cdot d)$ ,  $n$  (number of nodes),  $d$  (number of dimensions).

GPT BERT 512 1024 token (Sparse Transformer Longformer Linformer),

2:

Transformer [██████████] 450 [██████████], 8 GPU [██████████], [██████████]

GPT-3 token, AI " " " "

3:

Transformer

4

[REDACTED], [REDACTED]:

■:■Transformer■,■DALL-E■Stable Diffusion■Sora■

1

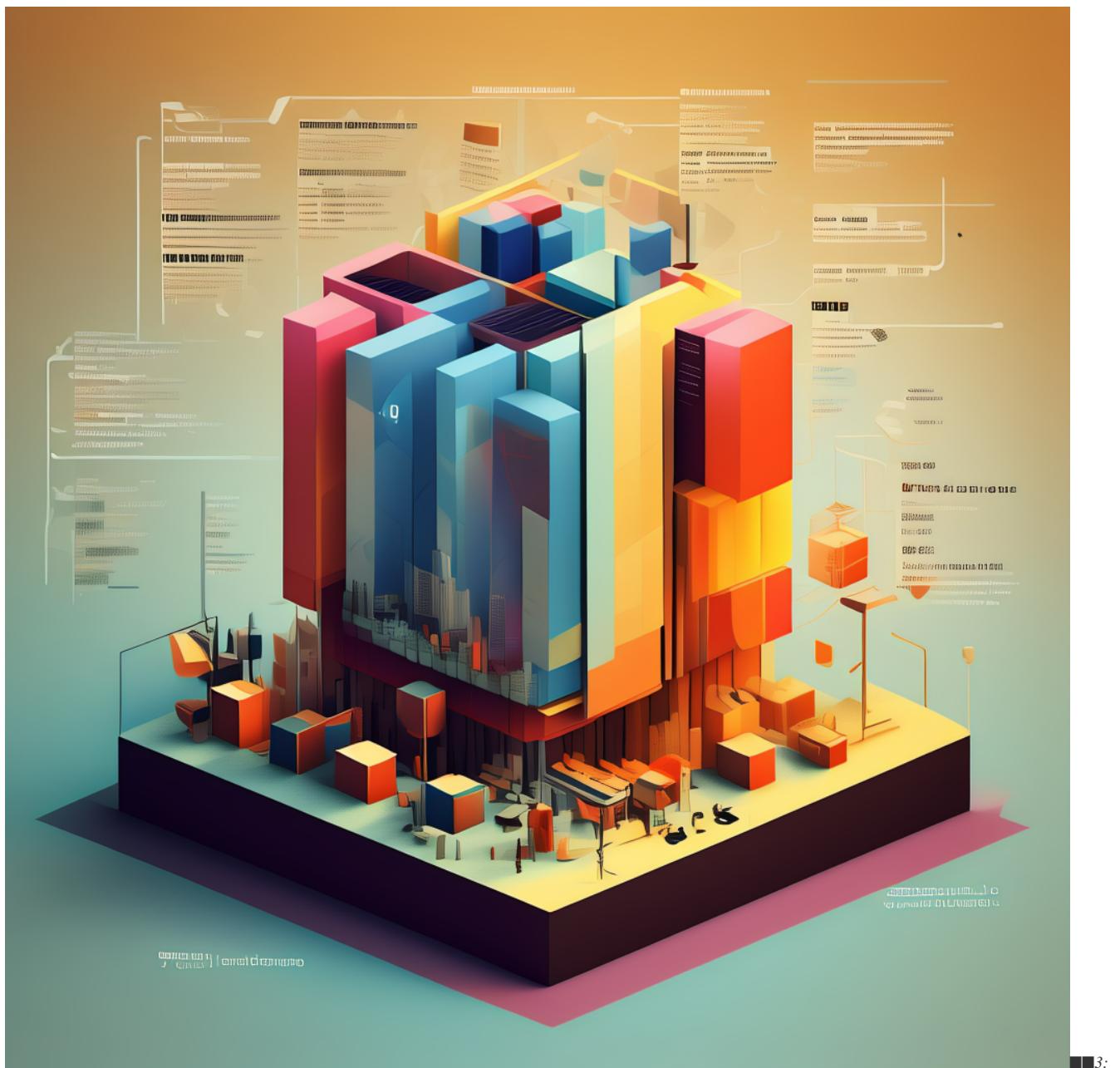
Transformer

Transformer

11

Transformer

AI "Transformer" BERT GPT, AI AI , 2017



Transformer  AI 

Transformer - BERT (2018) GPT-3 (2020) - The Annotated Transformer (<https://github.com/tensorflow/tensor2tensor>)

██████████: - Transformer: █████ / █████ - Attention Mechanism: ████████ - Self-Attention: ████████ - Multi-Head Attention: ████████ - Encoder-Decoder: █████-█████ - Positional Encoding: █████ - BLEU: ██████████ (██████████)

■ ■ ■ ■ ■ Claude Code (lunwen skill) ■ ■ ■ ■ ■

■ ■ ■ ■ ■: 2026 | ■ ■ ■ Transformer ■ ■ (Vaswani et al., 2017)