■■■■■■■■■■■

■■: Attention Is All You Need ■■: Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, ■ukasz Kaiser, Illia Polosukhin (Google Brain/Research & University of Toronto) ■■■■■■: 2024■

■■: ■■■■■■■■■

■■■■■■ChatGPT■GPT-4■Claude■■AI■■■■■■■■■■■■■"Transformer"■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■Google■■■2017■■■■■

■■■■■■■■■■■■■■■■■■"Attention Is All You Need"■■■

■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■AI■"■■■"——■■■■■■■■■■■■■■■■■GPT■BERT■Claude■■■■■■■■■■■Transforn

---

■■■■: ■■■■■■■■■

■■■■■■■■■■■■■■■■■■2017■■■■■■■■■■■■■■■■

■■■■■■■■■■■

2017■■■■■■■■■■■■■■■■■■■■■■■■■■■"■■■■■■"■RNN■■■■■■■■■■■■■——■■■■■■■■LSTM■■■■■■■■■GRU■■

■■RNN■■■■■■■——■■■■■■■■"■"■■■■■■■■■■■

■■■■■■■■■■■■■■■■■■■■■

■■■■■■■■■■

■■RNN■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■……■■■■■■■■100■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■

■■■■■■■■■■■■■■■■■■GPU■■■■■■■■■■■■■■■■■■■■■■■■RNN■■■"■■"■■■■■■■■GPU■■■■■■■■■■■■■■■■■■■■■■■■■■■■

■■■■■■■■■■■

RNN■■■■■■■■■"■■"■■■■■■■■■■■■■■■"■■"■■■■■■■■■■■LSTM■GRU■■■■■■■■■■■■■■■■■■■■■■

■■■■■■■■■■"■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■"■■■■■■■"■"■■■■■■■■■■■■■"■"■■■■■■

■■■■■■■■

■Transformer■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■"■■■■■"■

[illegible]

[illegible]

[illegible]2017[illegible]RNN[illegible][illegible][illegible]RNN[illegible]

[illegible]Google[illegible][illegible]RNN[illegible]

[illegible]: [illegible]

Transformer[illegible]

Transformer[illegible][illegible]

[illegible]RNN[illegible]100[illegible]1[illegible]100[illegible]99[illegible]99[illegible]

[illegible]Transformer[illegible]1[illegible]100[illegible]"[illegible]"[illegible]

Transformer[illegible]

[illegible]Transformer[illegible]"[illegible]-[illegible]"[illegible]Encoder-Decoder[illegible]

1. [illegible][illegible]"[illegible]"[illegible]

2. [illegible][illegible]

Transformer[illegible]6[illegible]

1. ■■■■■■■■■Multi-Head Self-Attention■

■■Transformer■■■■■■■■■■■■■■■■■■■■■■■■■■

■■■■■■■■■■"The cat sat on the mat"■■■■■■■■■■■■■■■■■■■■"■■■■■■■■■■■■■■"

- ■■"cat"■■■■■■■"sat"■"The"■■■■■■■■"■■■■"■

- ■■"mat"■■■■■■■"on"■"the"■■■■

■■■■■■■■■■■■■■■■■■■■Query■■■■■■Key■■■■■Value■■■■■

- Query■■■■"■■■■■■"

- Key■■■■"■■■■■"

- Value■■■"■■■■■■■■"

■■■■■■■Query■■■■■■Key■■■"■■■■■"■■■■■■■■■■■■■■Value■■■■■■■■■■■■■■■

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

■■"■■■■d_k"■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■softmax■■■■■■■■■■■■■■1■■■■■0■■■■■■■■■■■■■■■■■■■

"■■"■■8■"■"■■■

2. ■■■■■■■Feed-Forward Network■

■■"■■"■

3. [illegible]

[illegible]"[illegible]"[illegible]"[illegible]"[illegible]

[illegible]

RNN[illegible]——[illegible]Transformer[illegible]"cat sat"[illegible]"sat cat"[illegible]

[illegible]"[illegible]"[illegible]

$$PE_{(pos, 2i)} = \sin(pos / 10000^{2i/d_{model}})$$

$$PE_{(pos, 2i+1)} = \cos(pos / 10000^{2i/d_{model}})$$

[illegible]k[illegible]PE_{pos+k}[illegible]PE_{pos}[illegible]

■■■■: ■■■■■■■

■■■■■■■■■■■■■

■WMT 2014■■■■■■■■Transformer■■■28.4■BLEU■■■■■■■■■■■■■■■■■■■■■■■■■■■■2■■■■■■■■■■2■■■■■■■■■■

■■■■■■■■■Transformer■■■41.8■BLEU■■■■■■■■■■■■■■

■■■■BLEU■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■

■■■■■■■■■■■■

■■■■■■■■■■■■■■■■■Transformer■■■■■8■P100 GPU■■■■■■3.5■■■■■■■■■■

■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■Transformer■■■■■■■■■■■■■■■■■■■■■■■■■■■■

■■■■■■■■■Transformer■■■■■RNN■■■■■■■■■■■■■■■■GPU■■■■■■■■■■■■■■■■■■

■■■■■■■■■■■■

■■■

- ■■■■■■■■■■■■■■■■■■"making...more difficult"■■"making"■"difficult"■■■■■■■■■■■■■■■■■■■■■■■■■■
- ■■■■■■■■■■■■■■■■■■■"its"■■■■■■■■■■■■■■■■"Law"■
- ■■■■■■■■■■■■■■■■

■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■

■■■■■■■■■■

■■■■Transformer■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■4■■■■■■■■Transformer■■■■■■■■■■■■■■■■■■■■■■

■■■■Transformer■■■■■"■■"■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■

■■■■■■■■■■■■■

■■■

- ■■■■■■■■■■■■■■■■■■■■■■■■■■8■■■■■■■■■■■■■■

- ■■■■■■■■■■■■■■■■■■■■■■dropout■■

- ■■■■■■■■■■■■■■■■■■■■■■■■

■■■■: ■■■■■■■

■■■■

■■■■■■■■■■■■"■■■■■"■■■■■■■■■■■■■■■■■■■■RNN■■■Transformer■■■■■■■■■■■■■■■■■■■■■

■■■■■■■■■■"Attention Is All You Need"■■■■■■■■■■■■■■■■"■■■■■■■"■■■■■"■■■■■■■■■■■■"■■■■■■■■■■■■■■■■■■■

■■■■■■

Transformer■■■■■■■■■"■■■"■■■■■■■■■■■■■■■GPU■■■

■■■■■■■■

RNN■■■■■■■■■O(n)■■n■■■■■■■■■■■■■O(log n)■■■Transformer■■■O(1)■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■

■■■■

■■■"■■"■■■■■Transformer■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■"■"■■■■■■■■■■■■■■■■■■■■■■■■■

■■■■■

■■■■■■■

■■■■■■■■■■■$O(n^2 \cdot d)$■■■n■■■■■■d■■

■■■■■■■■■■■■■■■■■■■■Longformer■Big Bird■■■■■■■■■■■■

■■■■

■■

■■■■■■

■■■

■■■■■■■

■■

■■■■

■■■■■■■■■■■■■■■**■■■■■■■■■■■■■■■■■■■■**

Transformer■■■"■■■"■

■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■

■■

■■■■■■■■■■■■■■■■2017■■■■■■■■■■2018■■BERT■■2019■■GPT-2■■2020■■GPT-3■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■

■■■■2017■■■■■■■■■■■■■■"■■■■■■■■■■■■■■"■■■■■■■■■■■■■■■AI■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■AI■■■

■■

■■■■■■■Transformer■■■Transformer■■■RNN■■■■■■■■■■■

■■■ ■ ■■■■: ■■■■■■■■/■■■■ ■■ ■■■■: ■15■■ ■■ ■■■■: ■■■■■■■■■■■■■■■■■■■■■

■■■ ■ ■■■■: `papers/downloaded_paper.pdf` ■■ ■■■■: 116.8■ ■■ ■■■■: ■■■API ■■■ ■ ■■■■: claude-opus-4-5-20251101 (via Claude Agent SDK) ■ ■■■■: 2026■01■17■ 11:01:09

■■■■ *GitHub Actions + Claude Agent SDK* + ■■■■ ■■■■

■■■■ GitHub Actions + Claude Agent SDK + ■■■■ ■■■■