

Transformer

XXXXX

xxx,

Northwestern Polytechnical University

June 2, 2025



(+86)123-4567-890

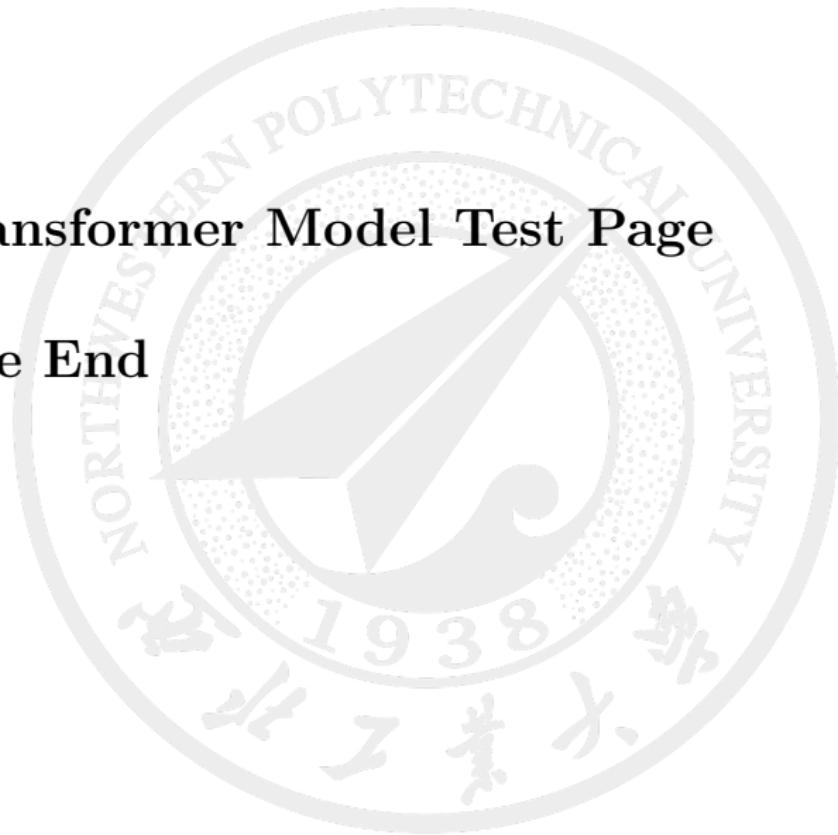


xxx@gmail.com



github.com/xxx

- 1 Transformer Model Test Page
- 2 The End





Introduction to Transformer

What is Transformer?

The Transformer is a deep learning model introduced in 2017^a, which relies entirely on attention mechanisms to draw global dependencies between input and output. It has become the foundation of most modern NLP models.

^aSee Vaswani et al., 2017 for details.

Why is it Important?

Transformers enable efficient parallel training and have revolutionized NLP, vision, and multimodal tasks.

Fun Fact

The phrase "Attention is All You Need" comes from the original Transformer paper.



Mathematical Formulation

Scaled Dot-Product Attention

The core of Transformer is the attention mechanism, defined as:

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

where Q (queries), K (keys), and V (values) are projections of the input, and d_k is the dimension of the keys.



Python Example: Simple Transformer Block

Code Example

```
import torch
import torch.nn as nn

class SimpleSelfAttention(nn.Module):
    def __init__(self, d_model):
        super().__init__()
        self.attn = nn.MultiheadAttention(d_model,
                                         num_heads=4)
    def forward(self, x):
        attn_output, _ = self.attn(x, x, x)
        return attn_output
# Example usage:
x = torch.rand(10, 32, 64) # (seq_len, batch, d_model)
model = SimpleSelfAttention(64)
y = model(x)
```



Transformer Architecture Diagram

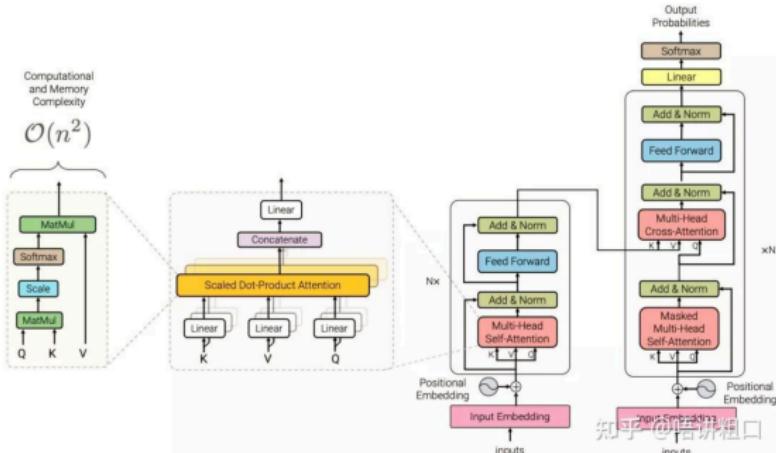


Figure: The architecture of the original Transformer model.



Comparison Table

Transformer vs. RNN vs. CNN

Model	Parallel	Long-range Dependency	Application
RNN	No	Weak	Sequence Modeling
CNN	Partial	Limited	Vision, Sequence
Transformer	Yes	Strong	NLP, Vision, Multimodal

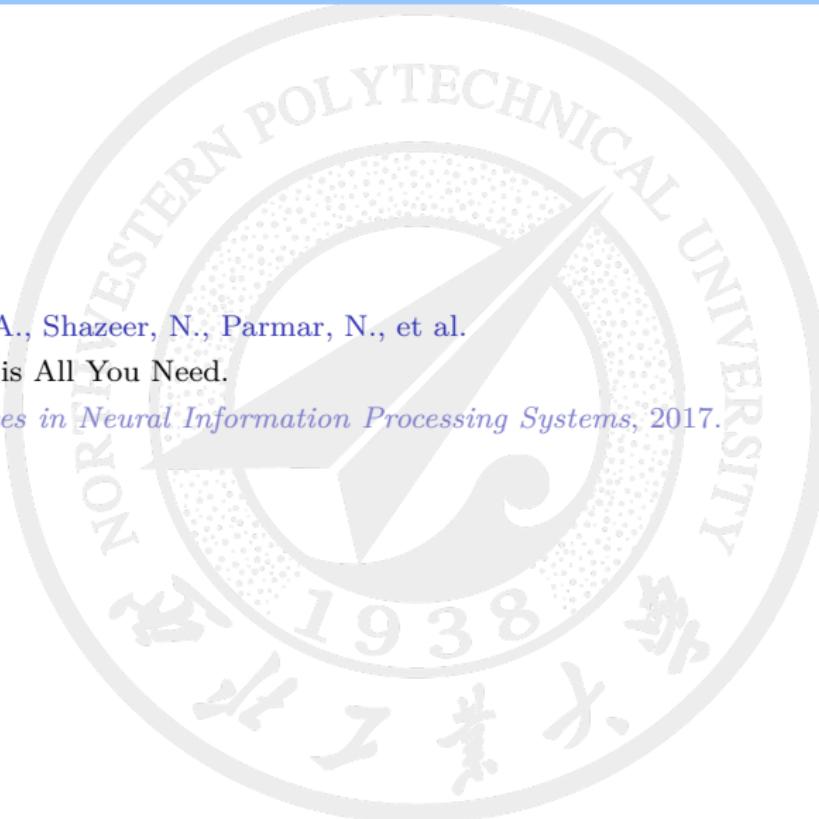
References I



Vaswani, A., Shazeer, N., Parmar, N., et al.

Attention is All You Need.

In *Advances in Neural Information Processing Systems*, 2017.





Thank You!

Any Questions?

✉ xxxx@gmail.com
⌚ github.com/xxxx
📞 (+86)123-456-789

