

Attention Is All You Need

The Transformer model introduces a novel architecture that relies entirely on attention mechanisms, dispensing with

Key Points:

1. Self-attention allows weighing token importance.
2. Multi-head attention captures multiple relationships.
3. Positional encoding keeps track of word order.
4. Efficient training compared to RNNs.
5. Foundation for GPT, BERT, and many modern LLMs.