Transformer Encoder (Self-Attention + FFN)

- What it is?

A building block that processes sequences using attention instead of recurrence.

- Goal:

To model long-range dependencies efficiently and in parallel.

- Math:

Self-attention:

$$Z = \text{Attention}(x)$$

feed-forwardnetwork:

$$FFN(z) = \max(0, zW_1 + b_1)W_2 + b_2$$

- Conclusion:

Transformer encoders replace RNNs with faster, more expressive architectures.