Input → LayerNorm → Attention → ⊕ → LayerNorm → MLP → Output

LayerNorm
Non-Linear / Scalar Product
$$k = 1/\sqrt{\sum(x-\mu)^2}$$
$$k \circ (x - \mu)$$
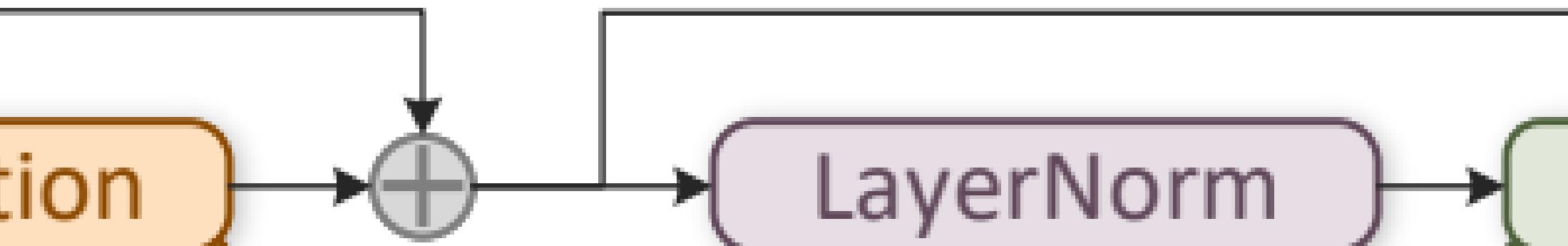
Attention
Matmul Product: $Q \cdot K^{\mathrm{T}}$
Non-Linear: Softmax

Non-Linear: e^x
Scalar Product: $\frac{1}{\sum e^{x_i}} \circ e^x$

MLP
Non-Linear: GELU