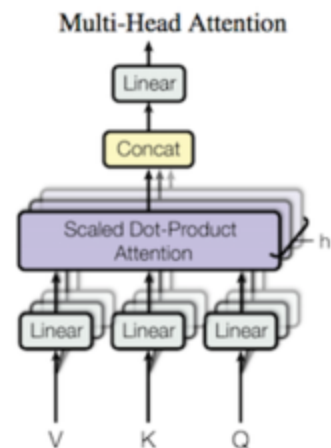
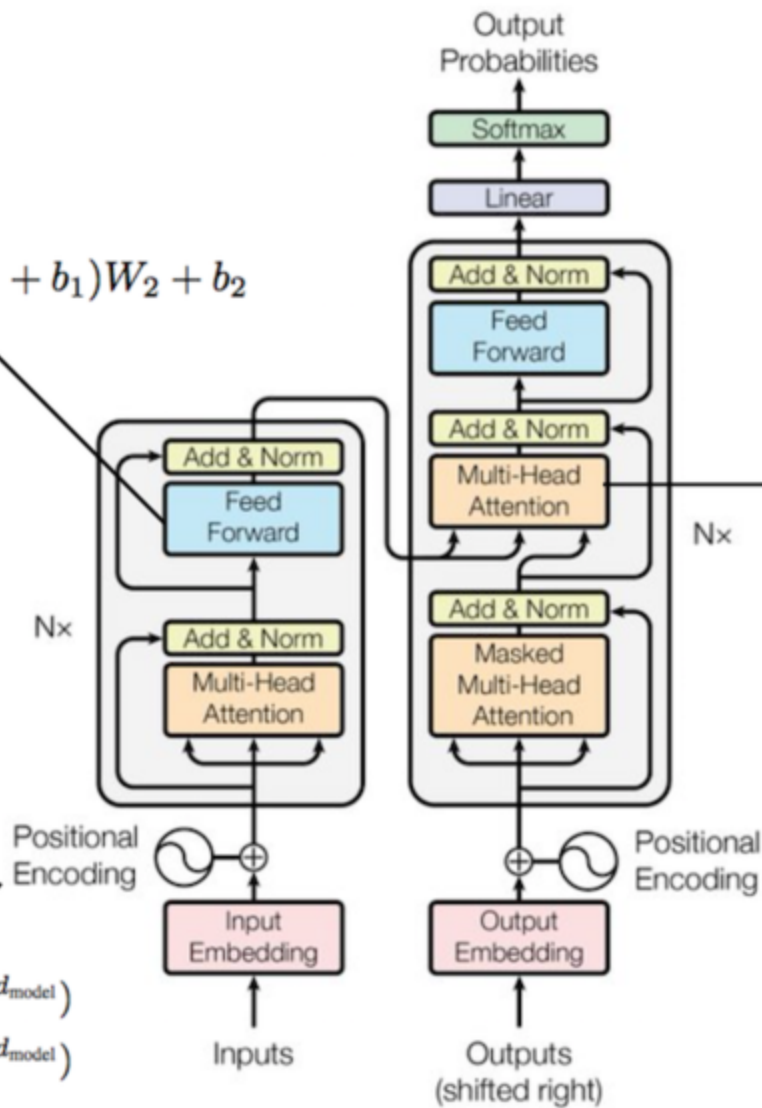


$$\text{FFN}(x) = \max(0, xW_1 + b_1)W_2 + b_2$$

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

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



Scaled Dot-Product Attention

