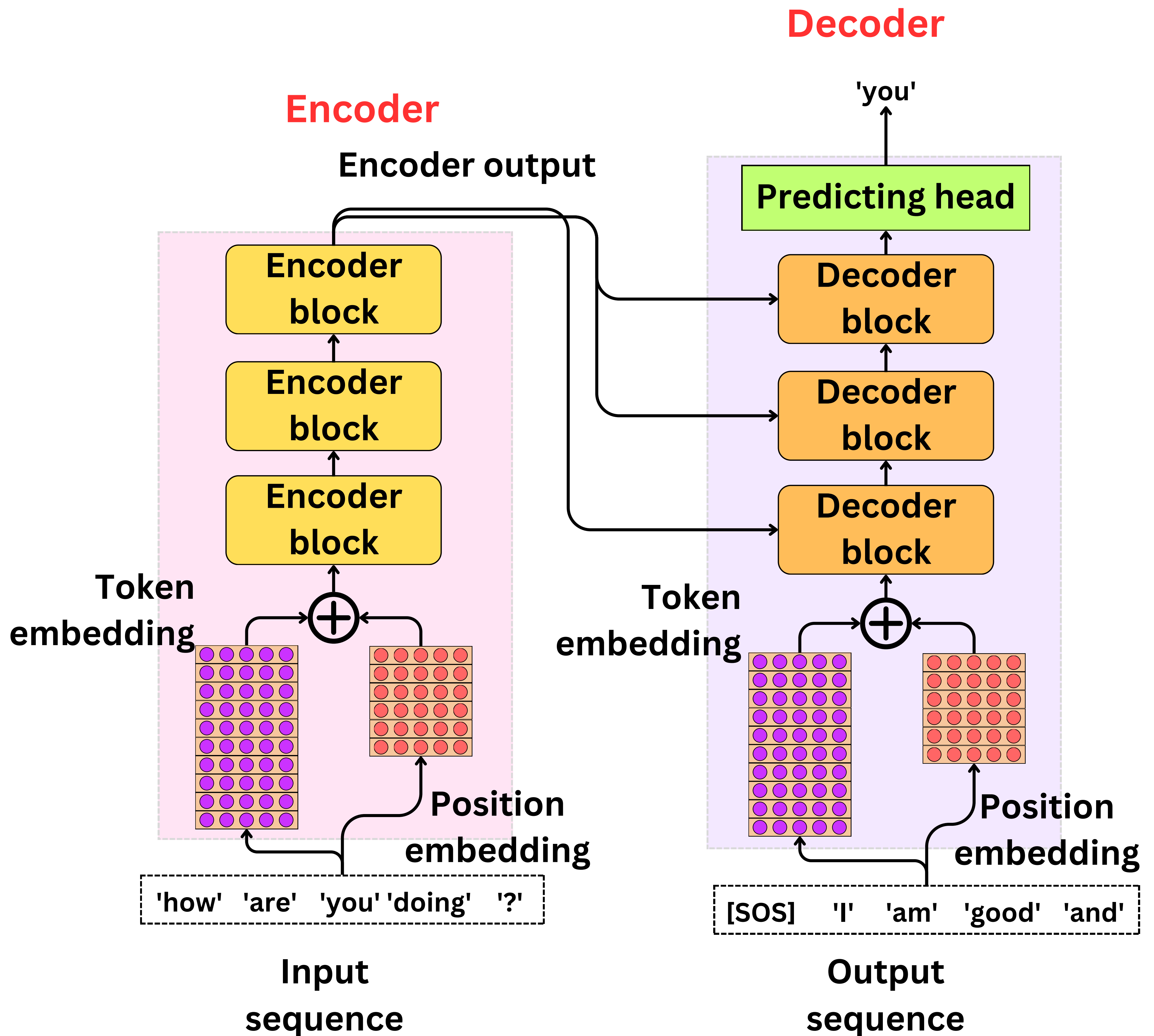


The Transformer Architecture

TheAiEdge.io



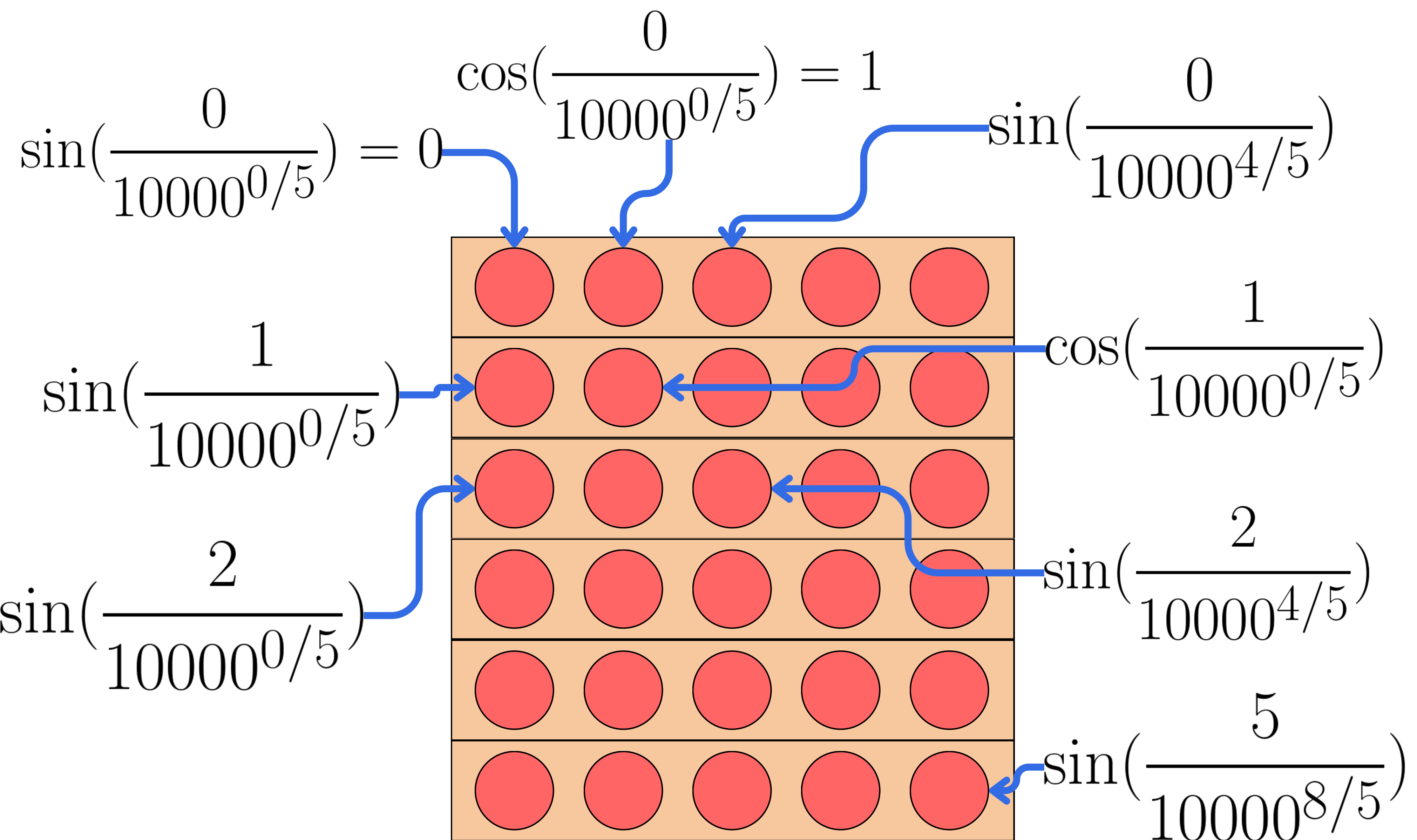
The Overall Architecture

The Transformer Architecture

TheAiEdge.io

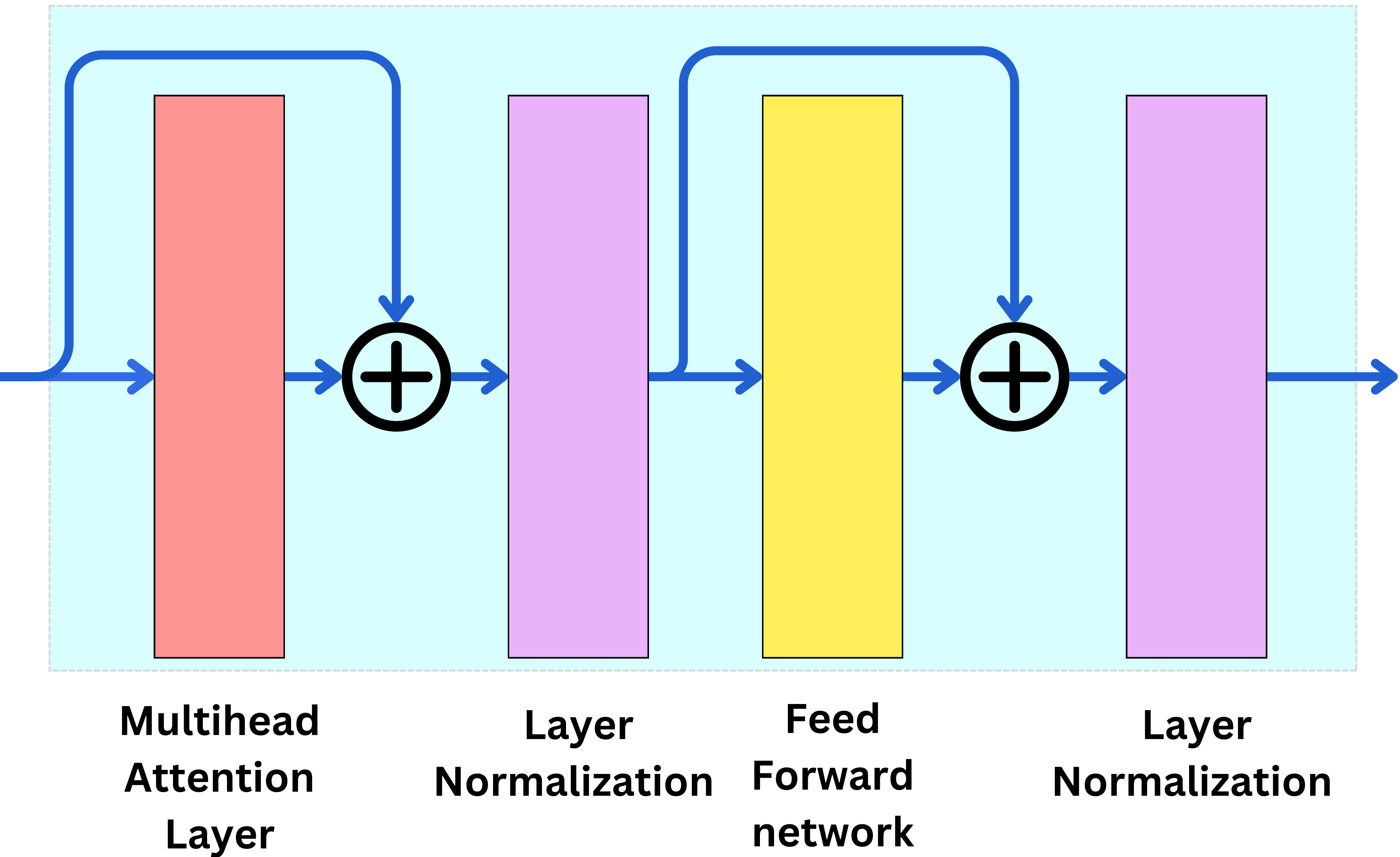
$$PE_{(pos, 2i)} = \sin(pos/10000^{2i/d_{\text{model}}}) \quad \text{even } i$$

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



The Position Embedding

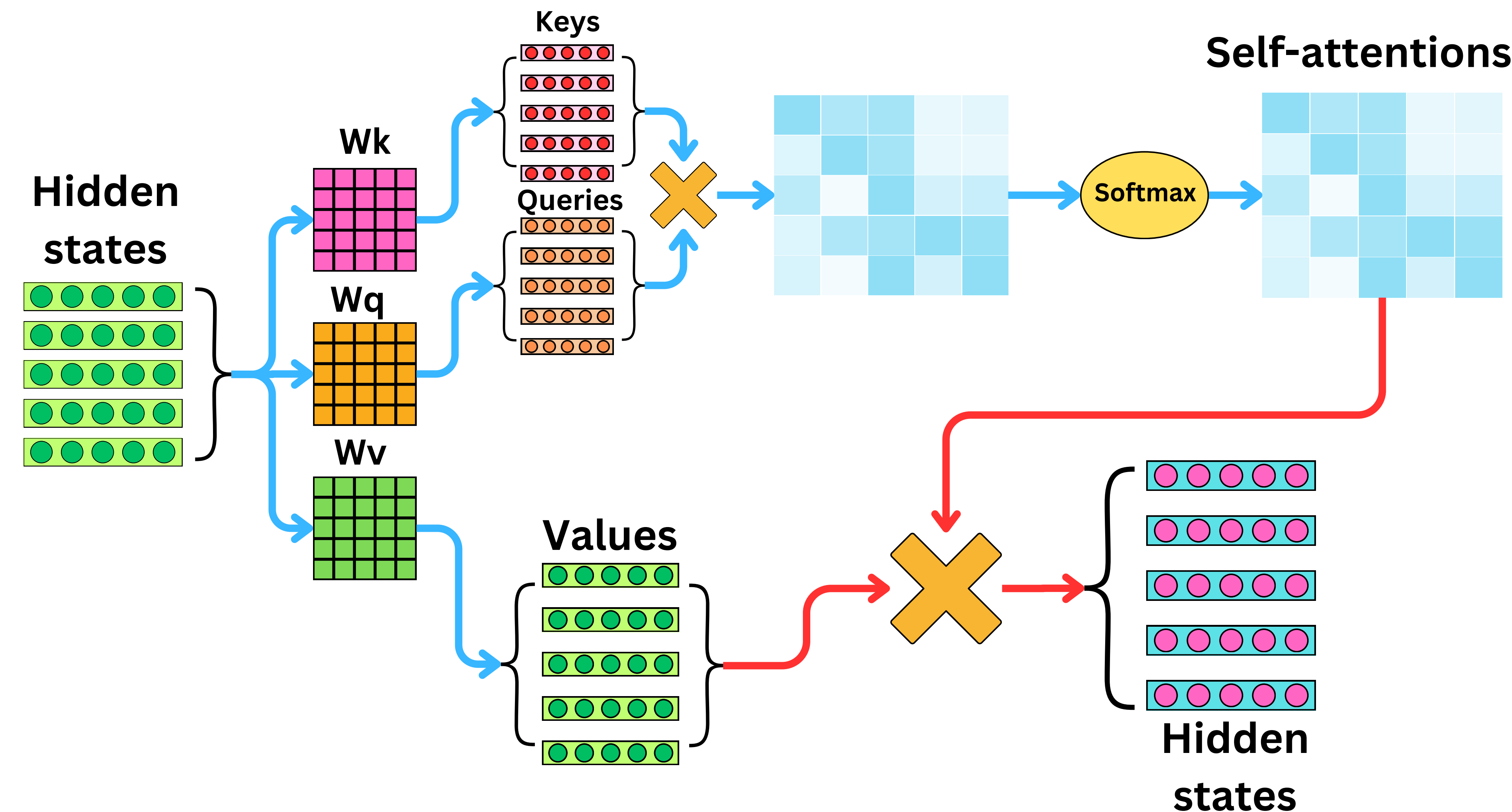
Encoder block



The Encoder Block

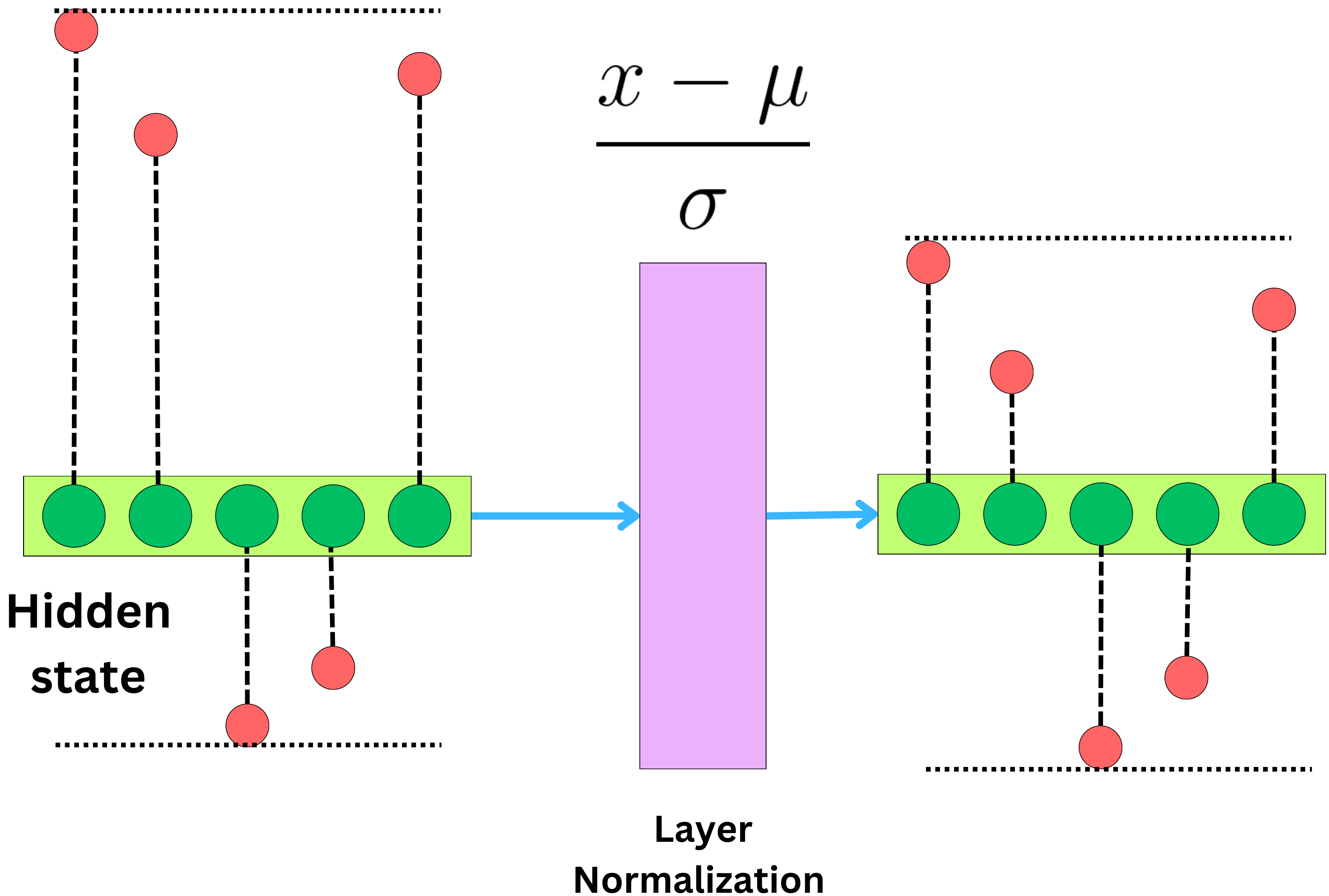
The Transformer Architecture

TheAiEdge.io

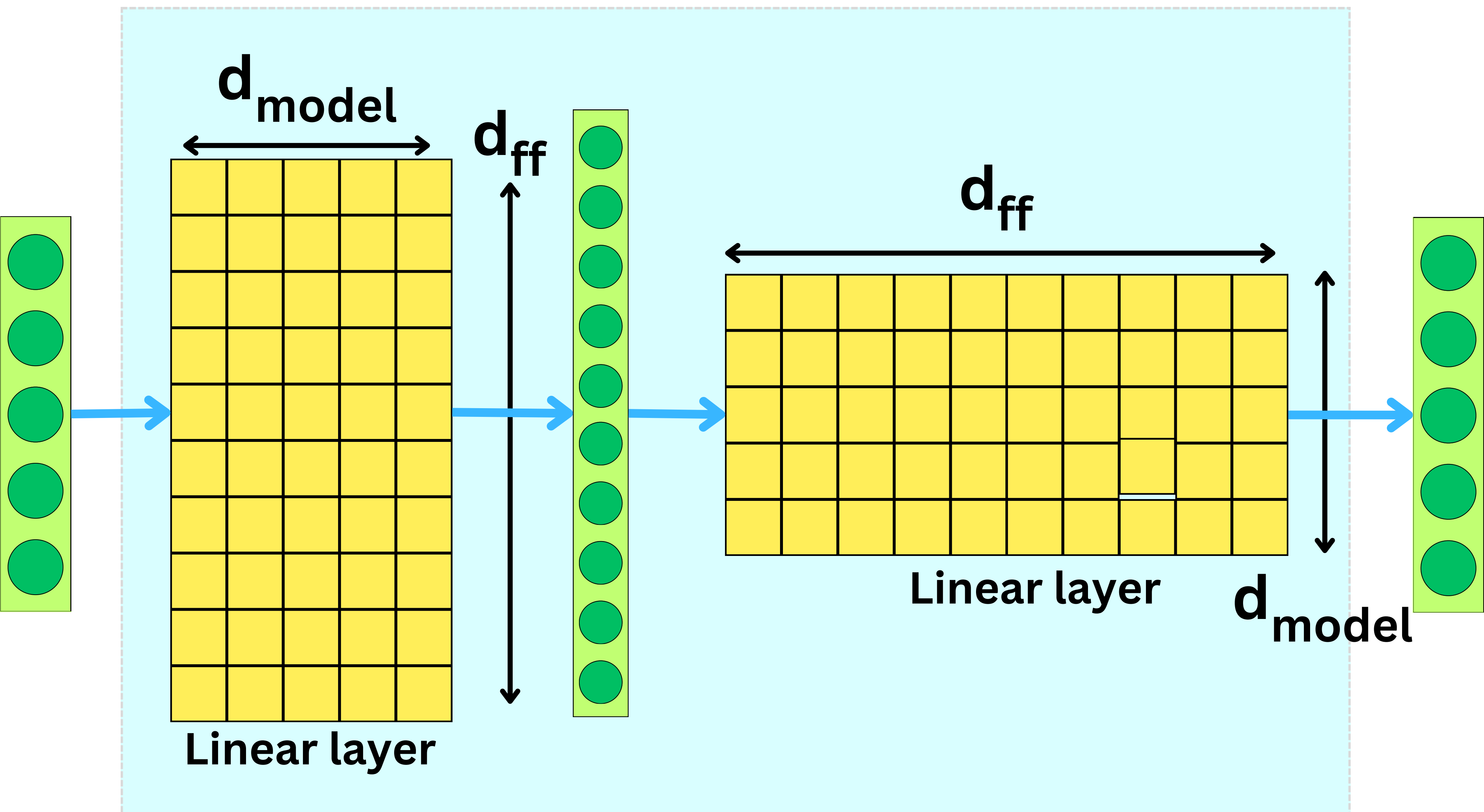


The Self-Attention Layer

The Transformer Architecture

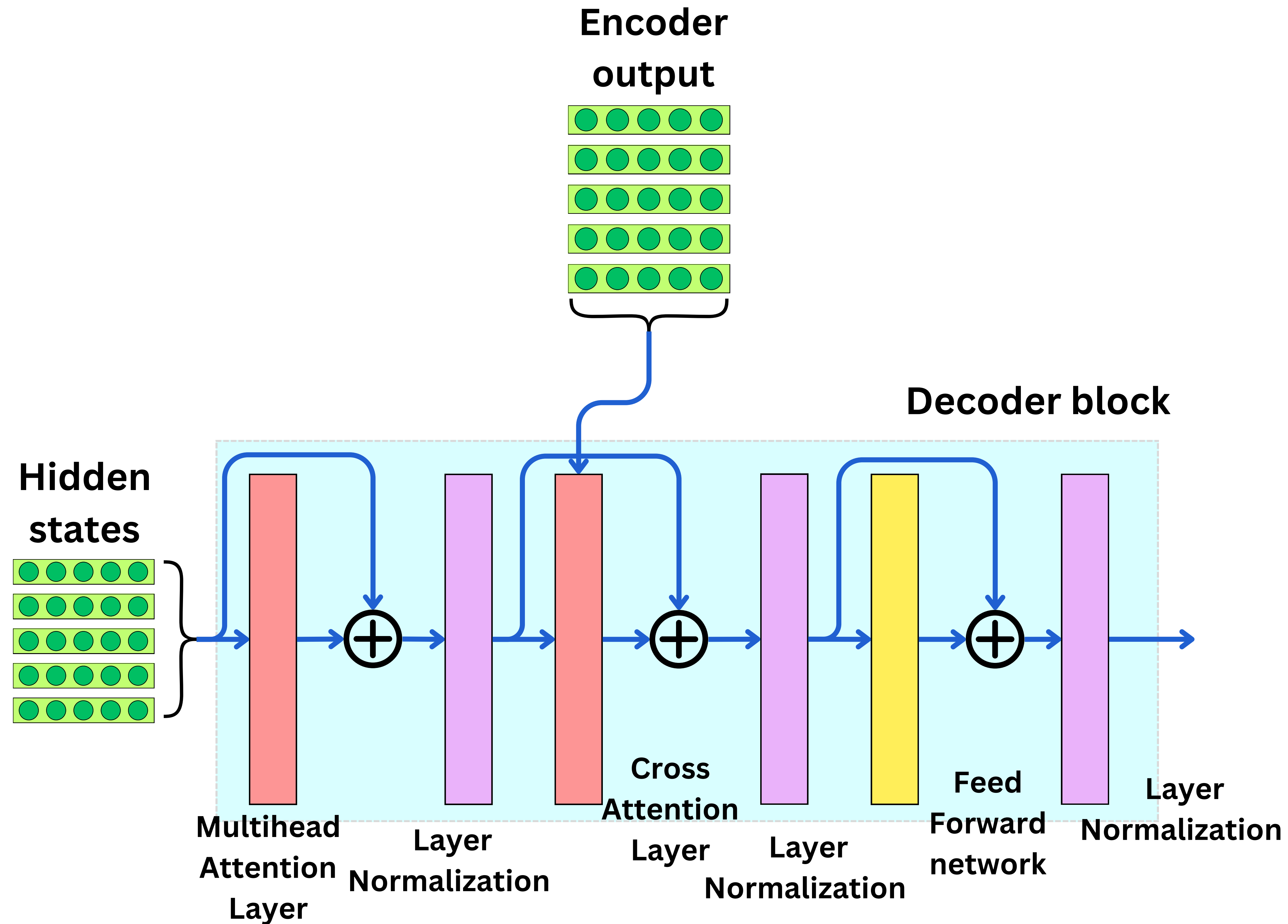


The Layer Normalization

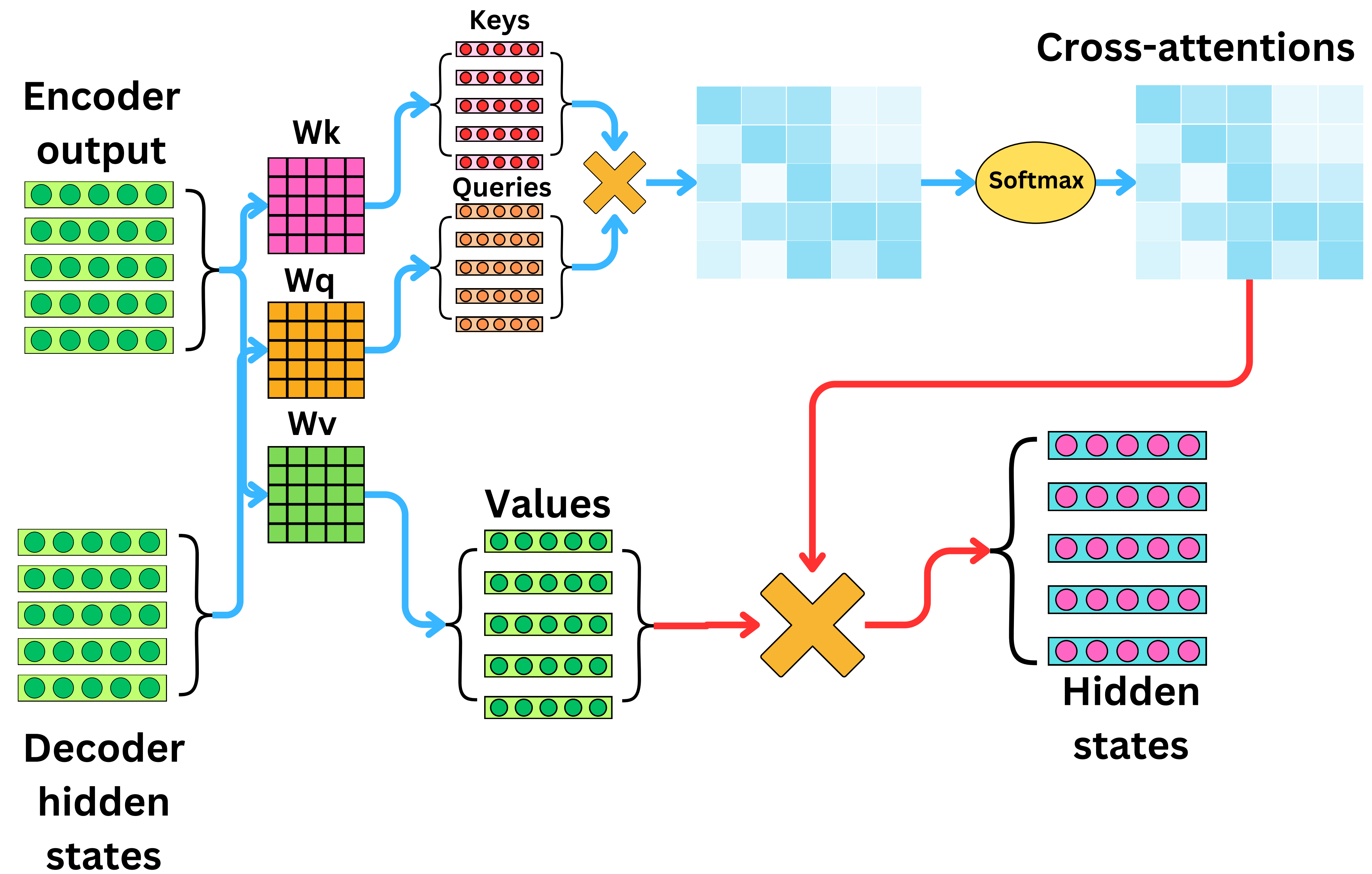


The Transformer Architecture

TheAiEdge.io



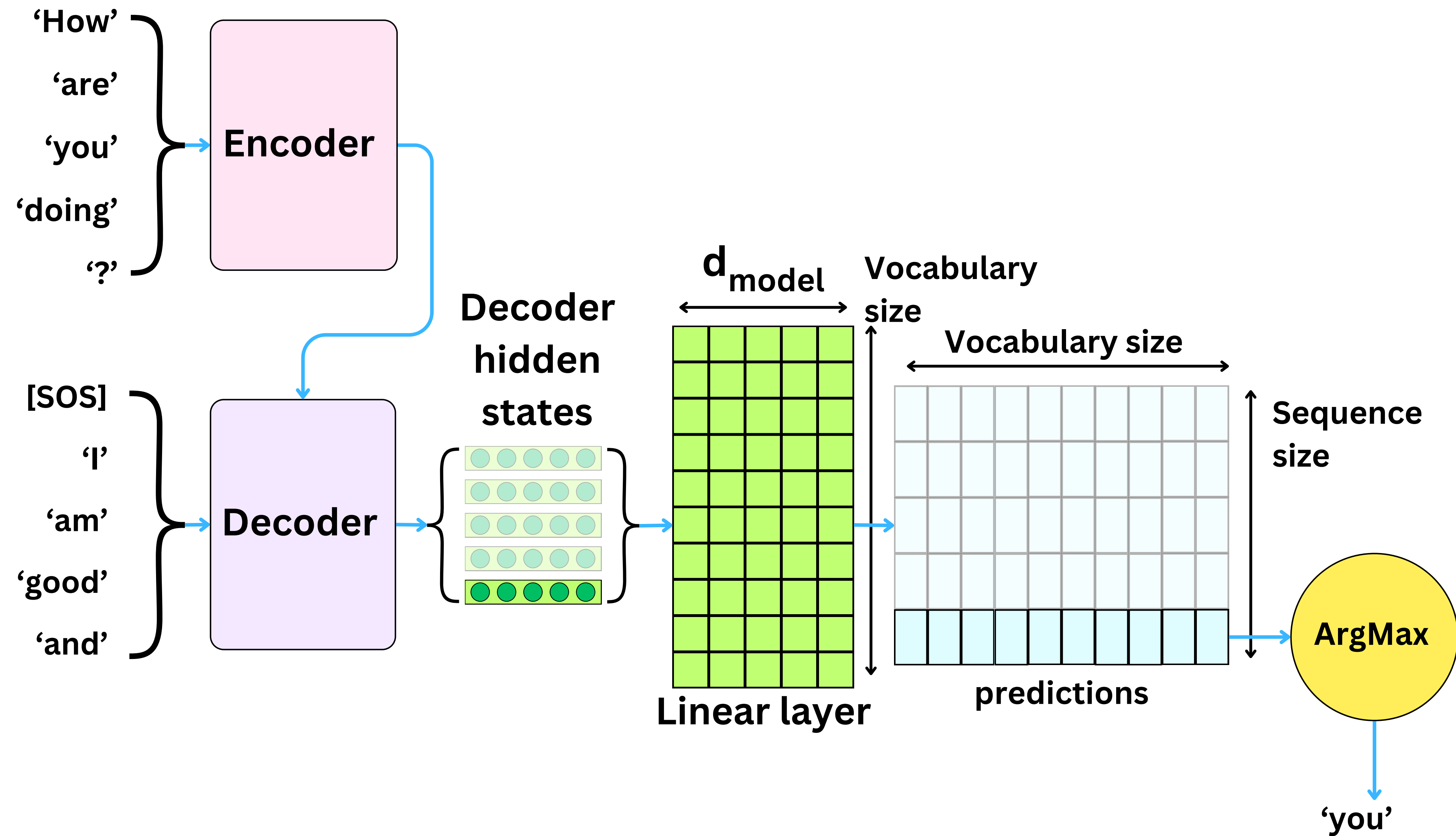
The Decoder Block



The Cross-Attention Layer

The Transformer Architecture

TheAiEdge.io



The Predicting Head