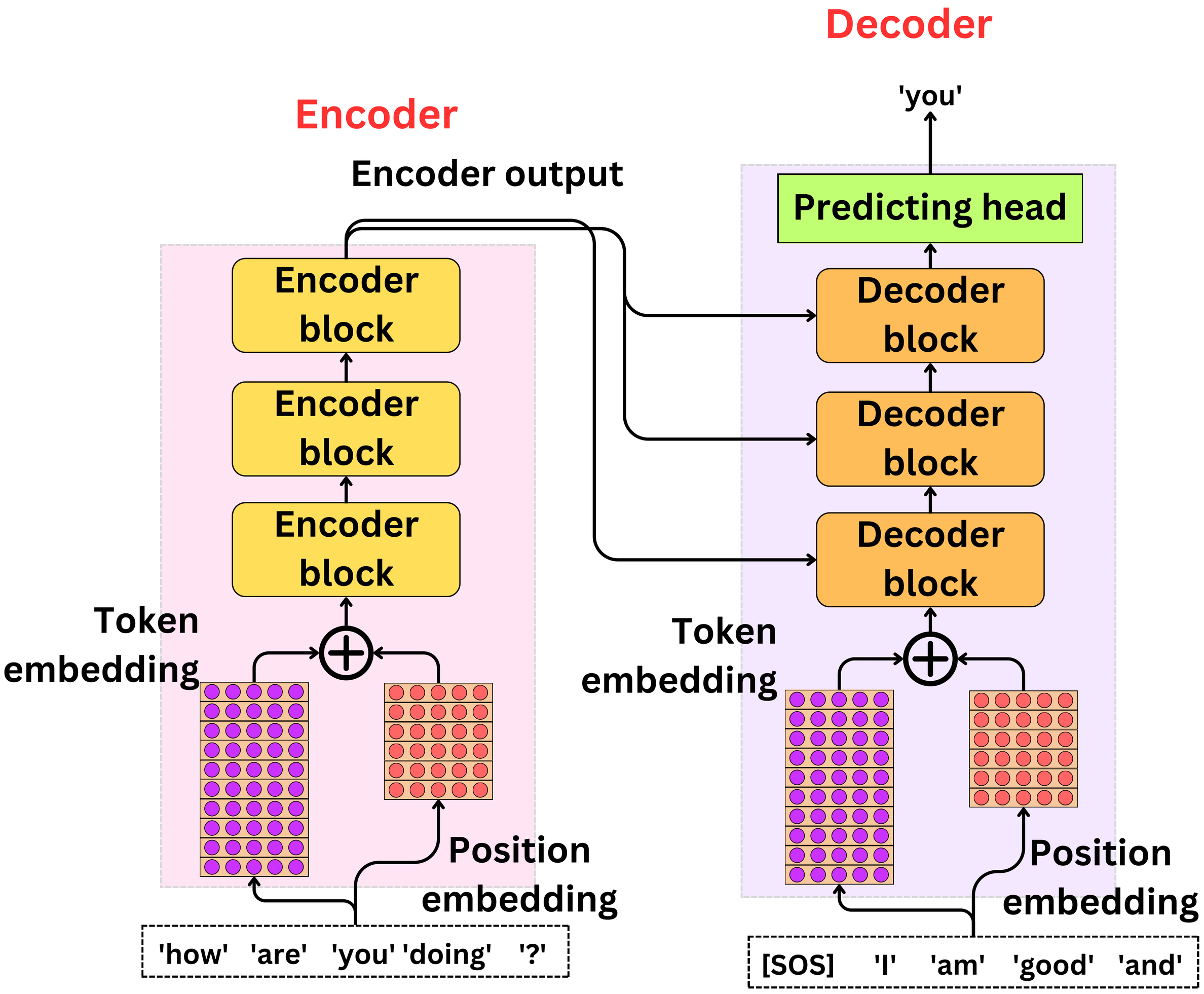
The Transformer Architecture

TheAiEdge.io



The Overall Architecture

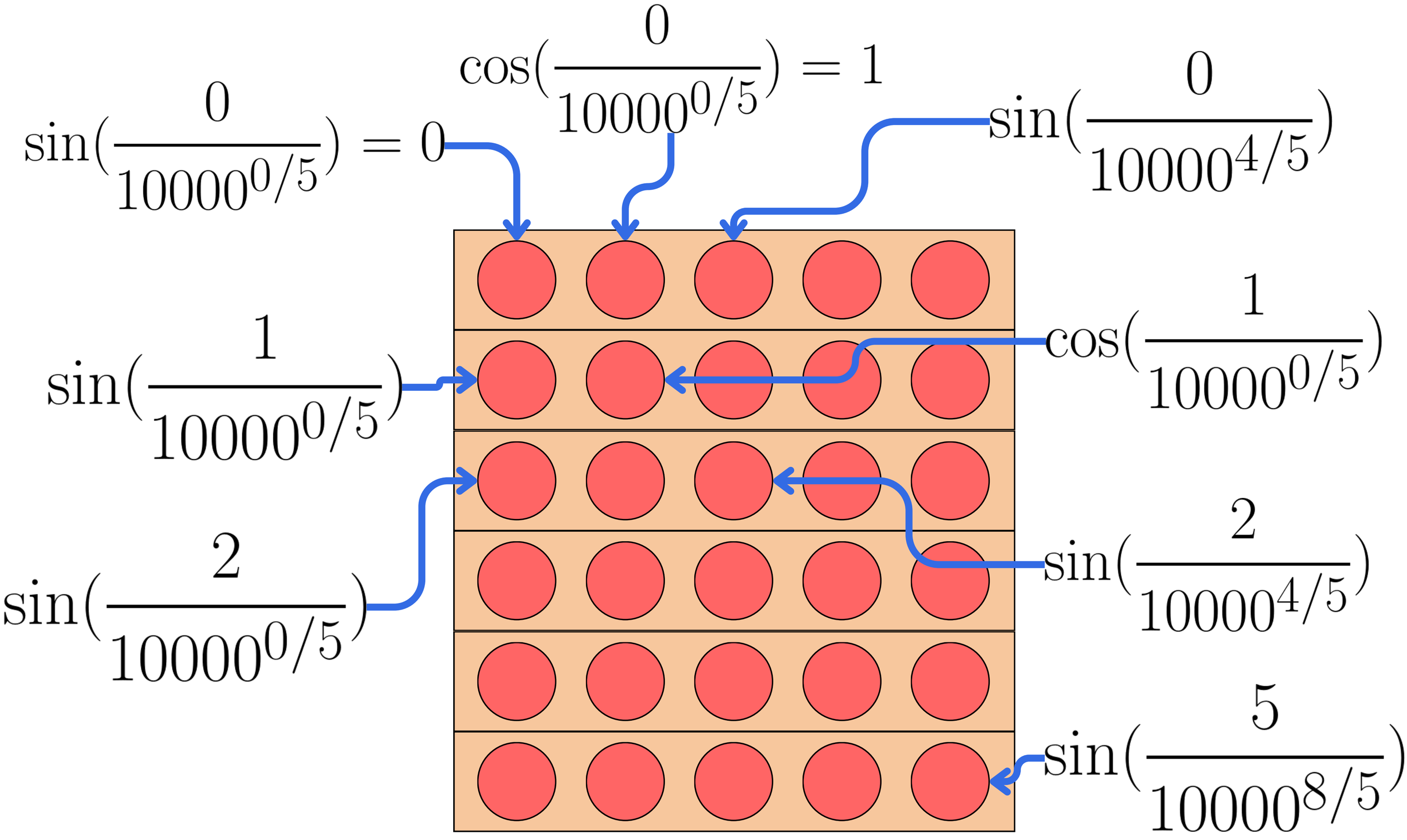
ABNASIA.ORG

The Transformer Architecture

TheAiEdge.io

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

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



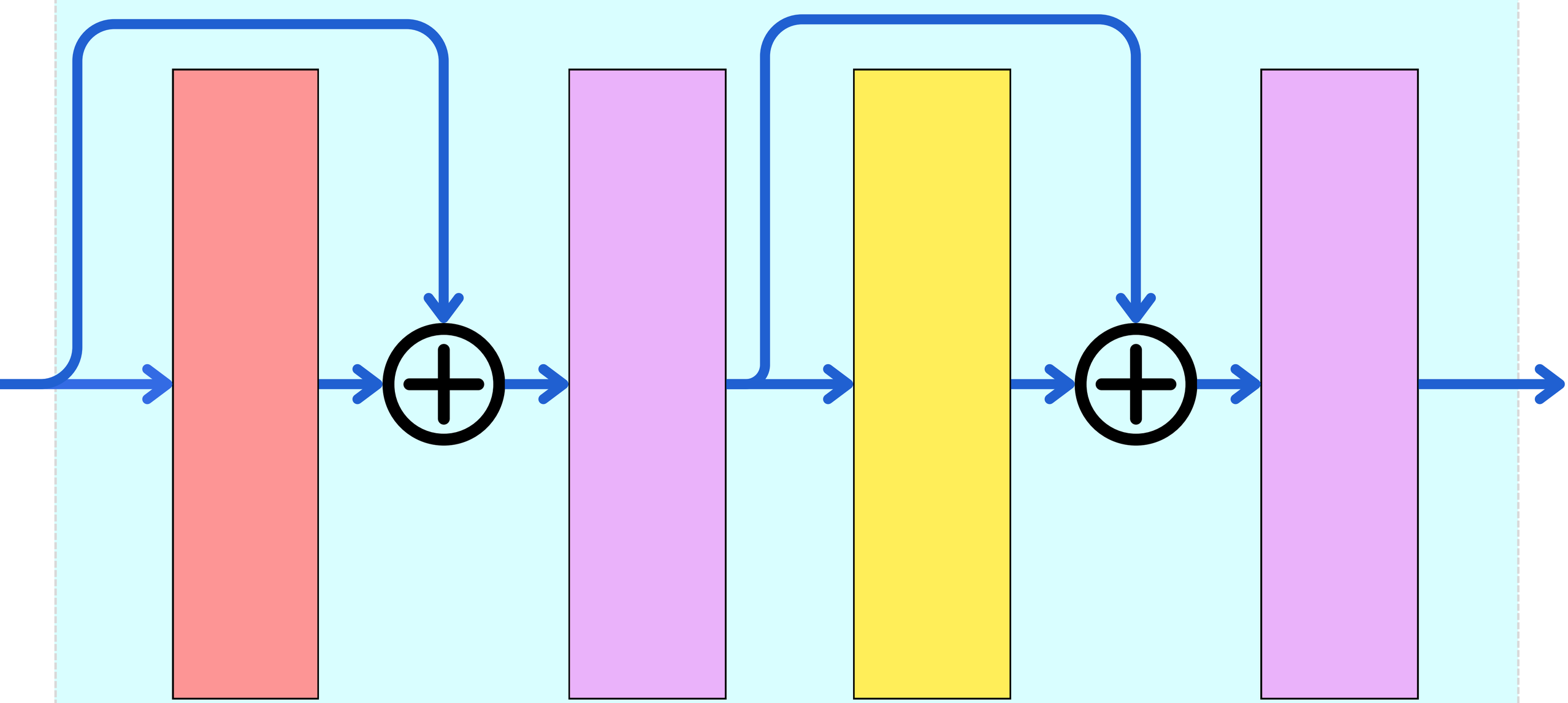
The Position Embedding

ABNASIA.ORG

The Transformer Architecture

TheAiEdge.io

Encoder block



Multihead Attention Layer

Layer Normalization

Feed Forward network

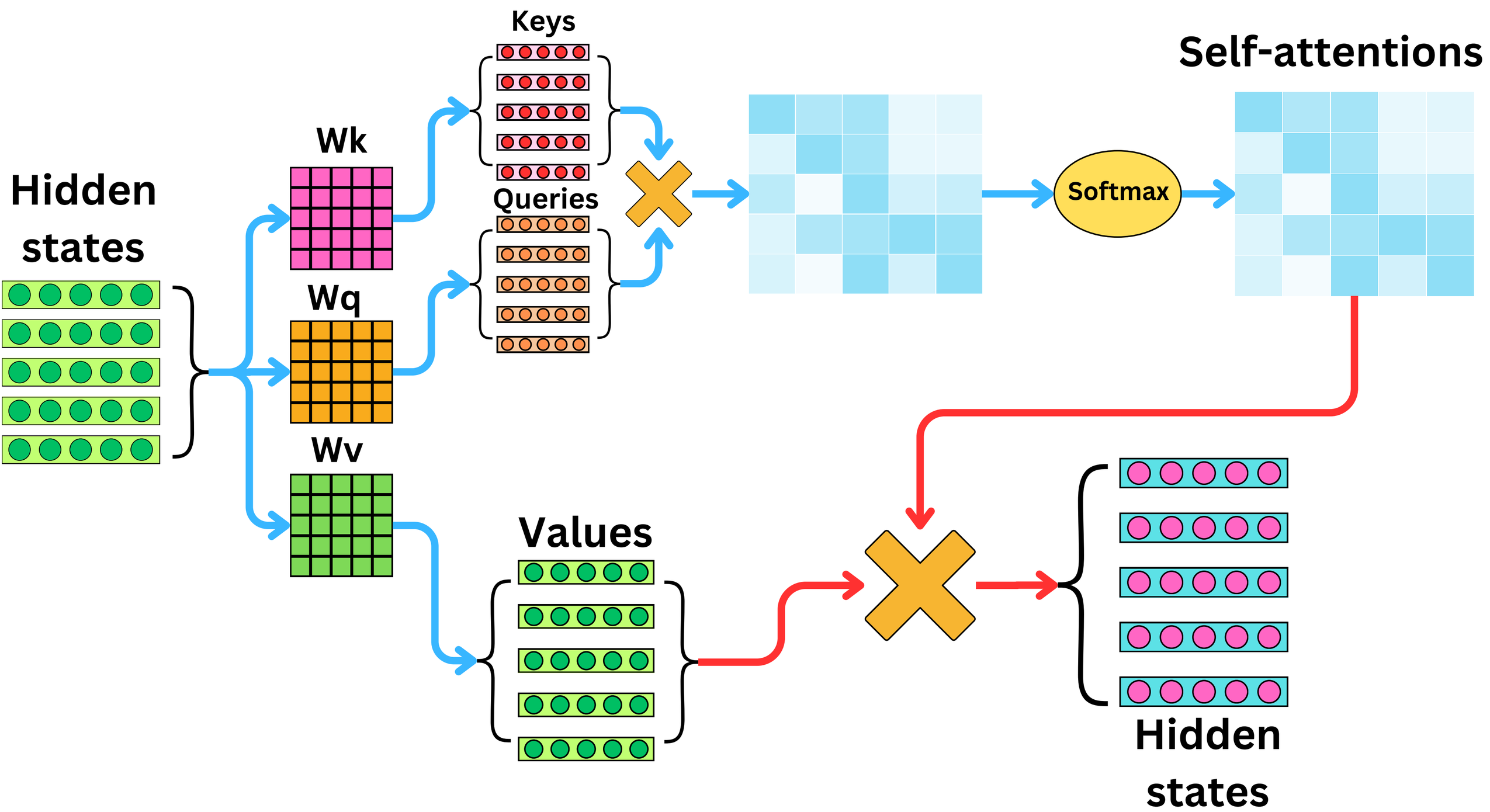
Layer Normalization

The Encoder Block

ABNASIA.ORG

The Transformer Architecture

TheAiEdge.io

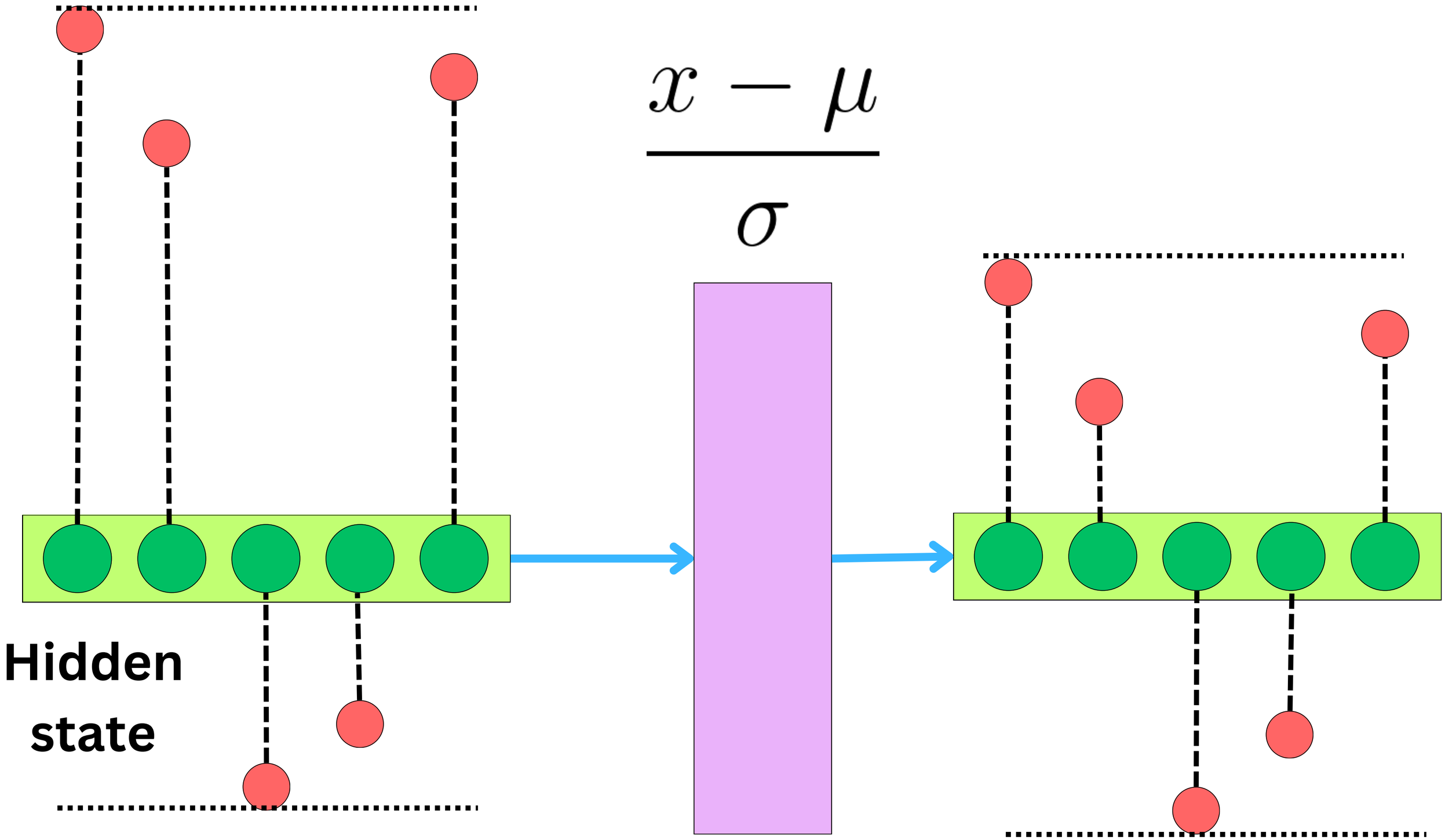


The Self-Attention Layer

ABNASIA.ORG

The Transformer Architecture

TheAiEdge.io

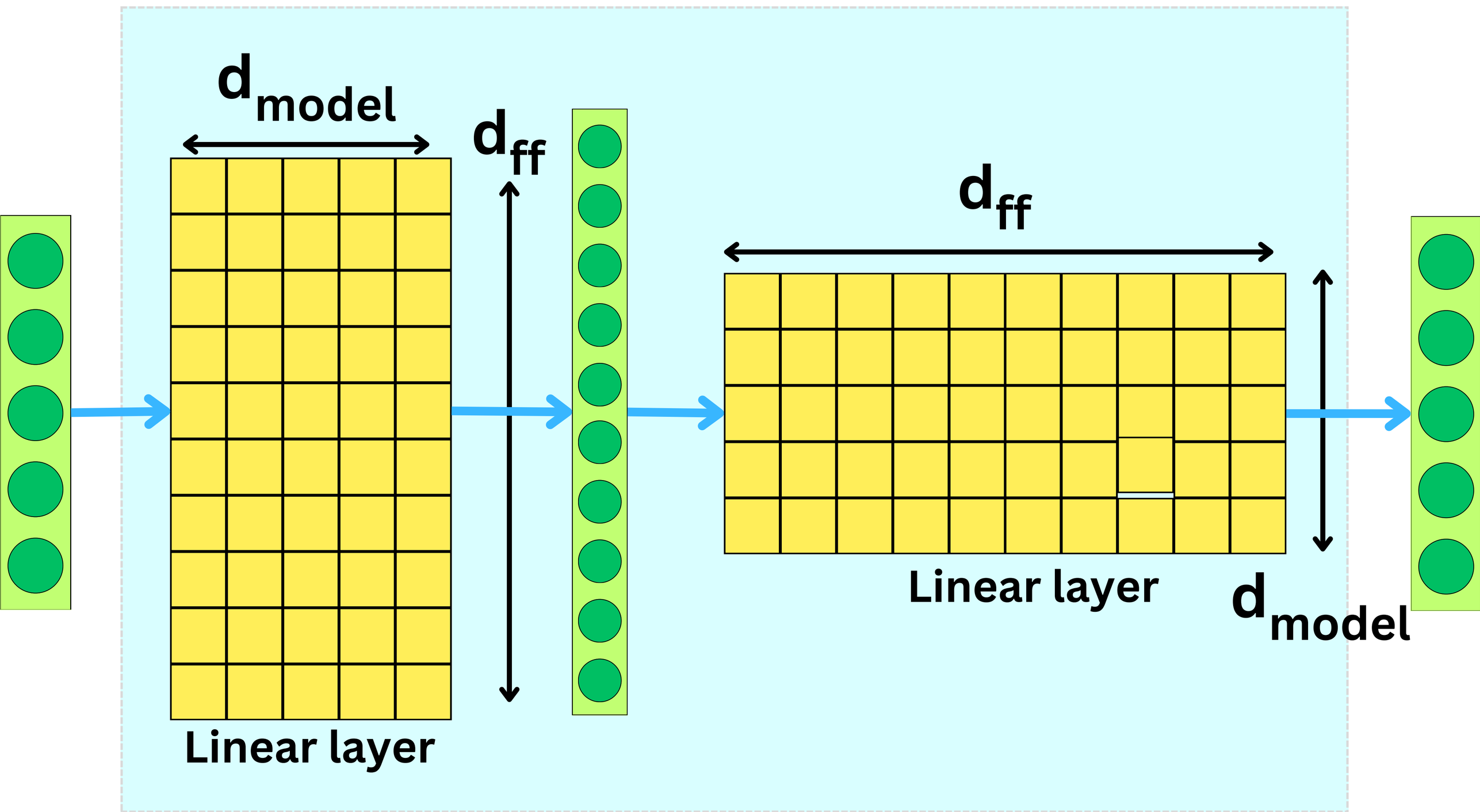


The Layer Normalization

ABNASIA.ORG

The Transformer Architecture

TheAiEdge.io

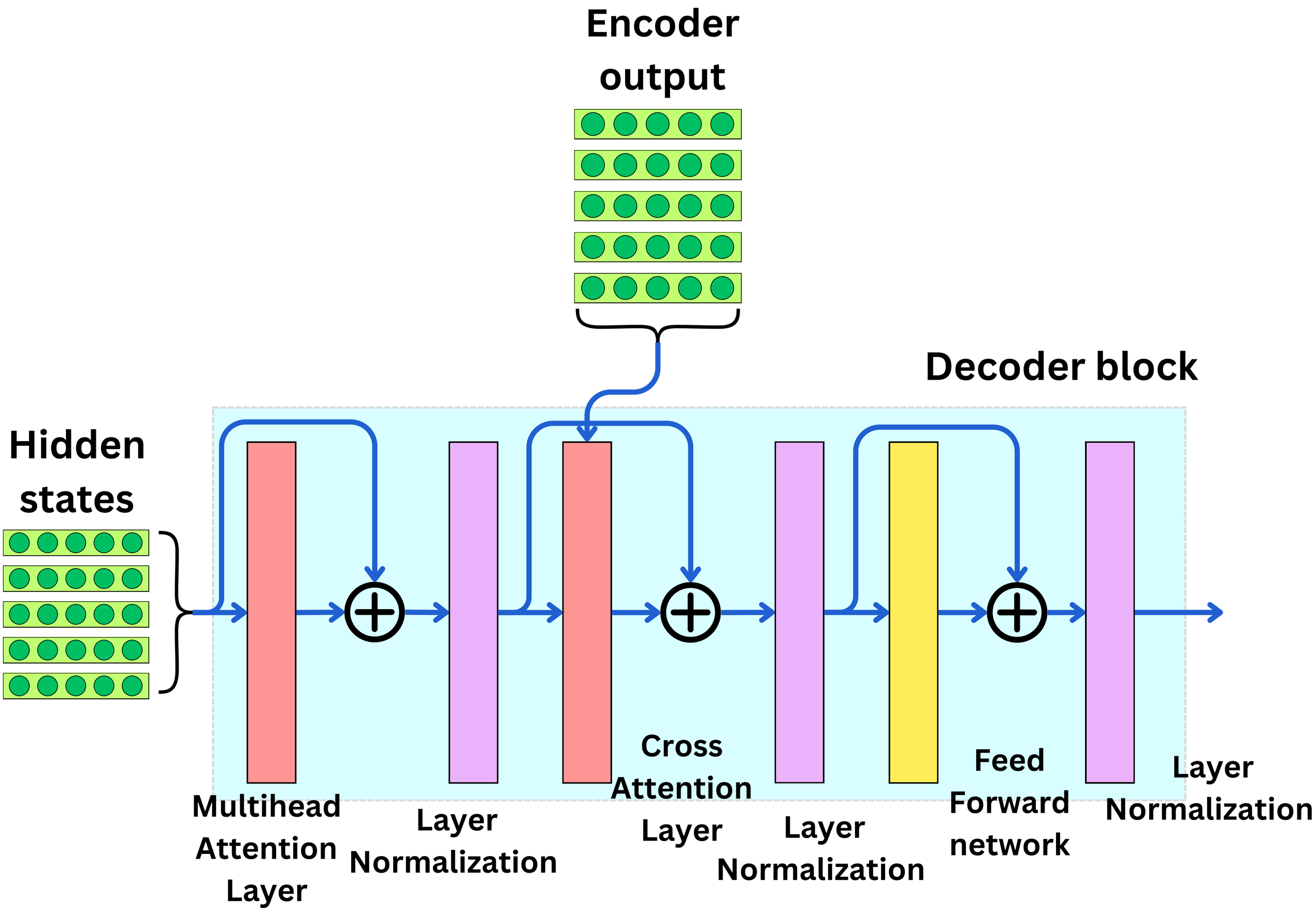


The Position-wise Feed-forward Network

ABNASIA.ORG

The Transformer Architecture

TheAiEdge.io

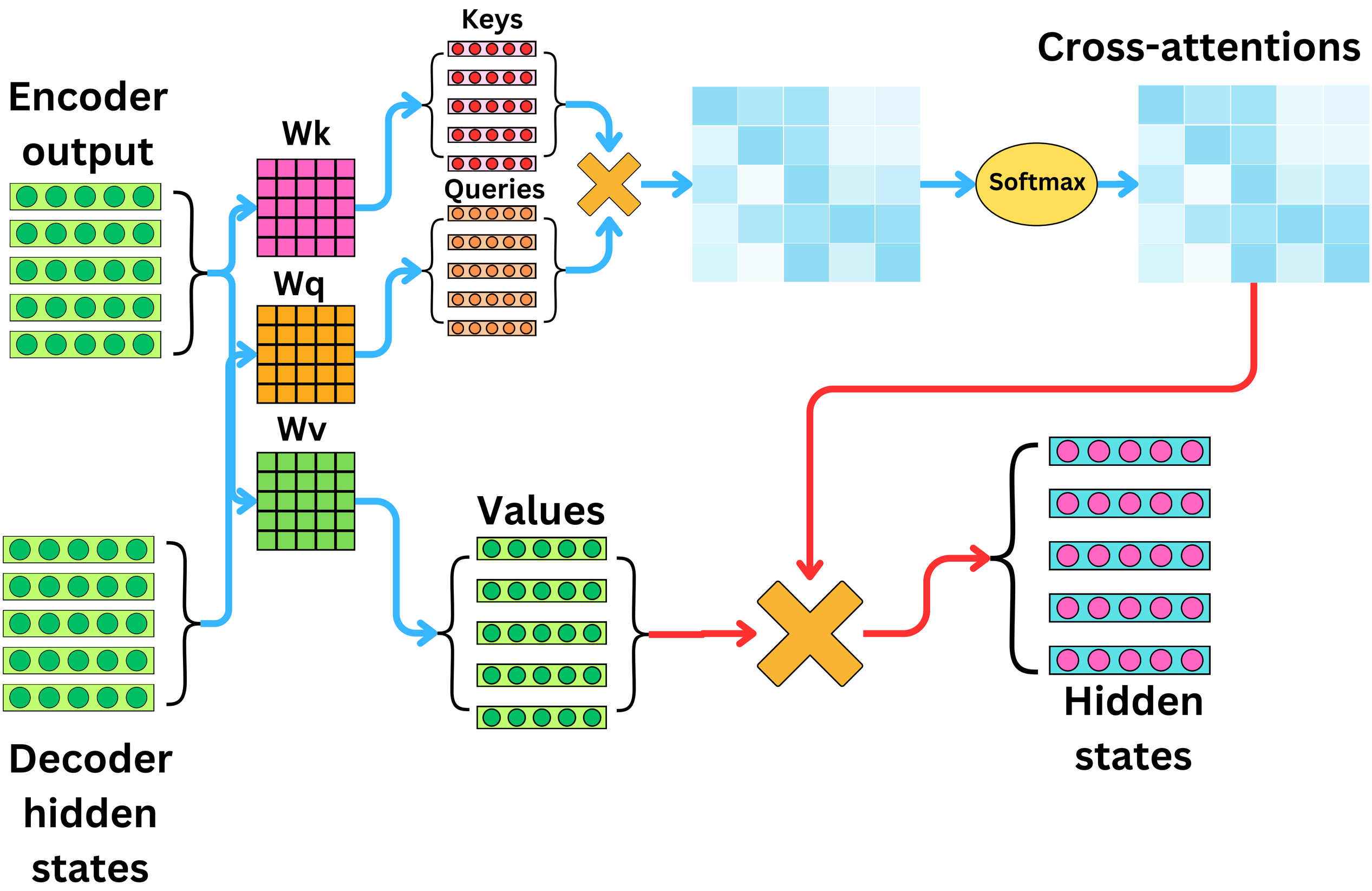


The Decoder Block

ABNASIA.ORG

The Transformer Architecture

TheAiEdge.io

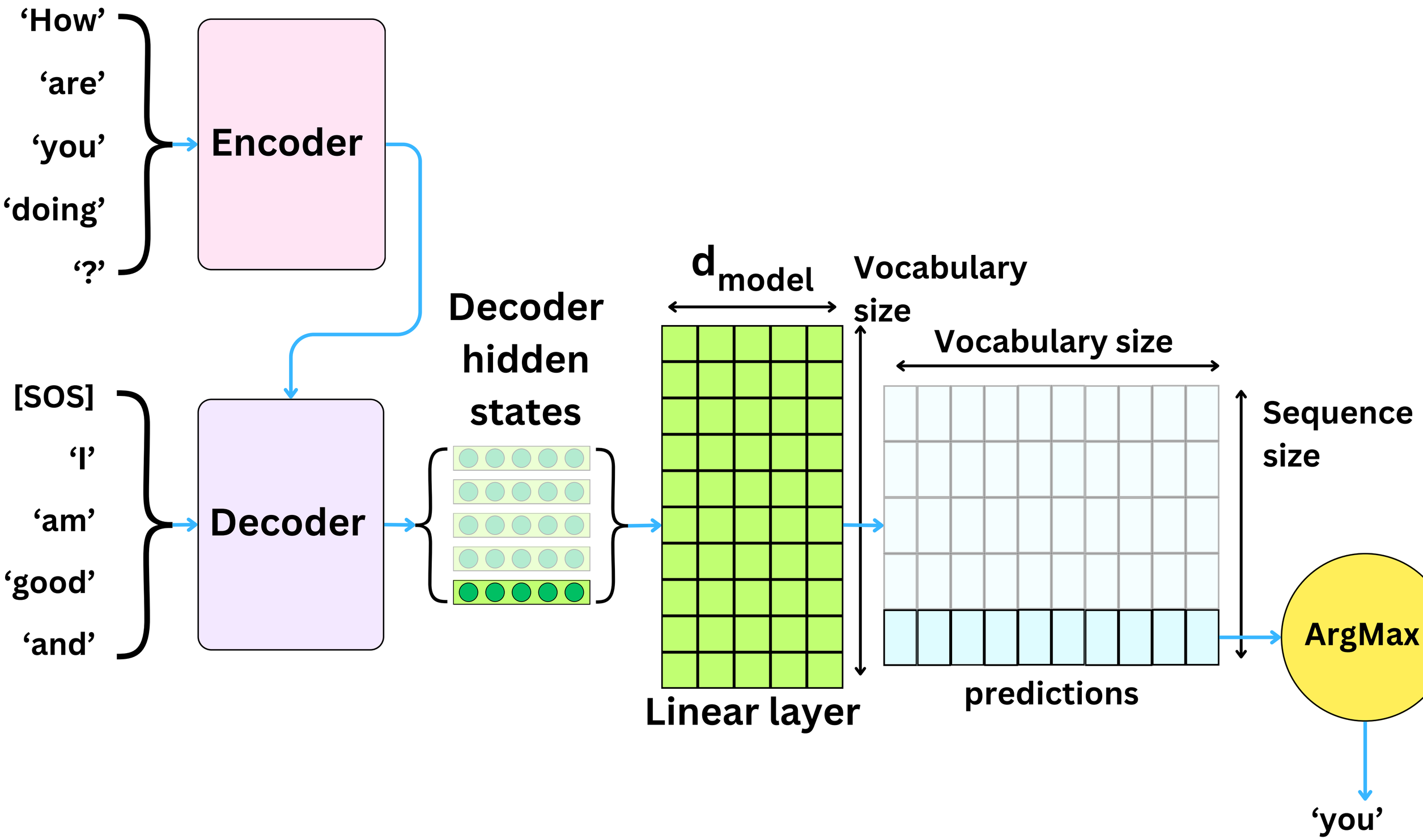


The Cross-Attention Layer

ABNASIA.ORG

The Transformer Architecture

TheAiEdge.io



The Predicting Head

ABNASIA.ORG