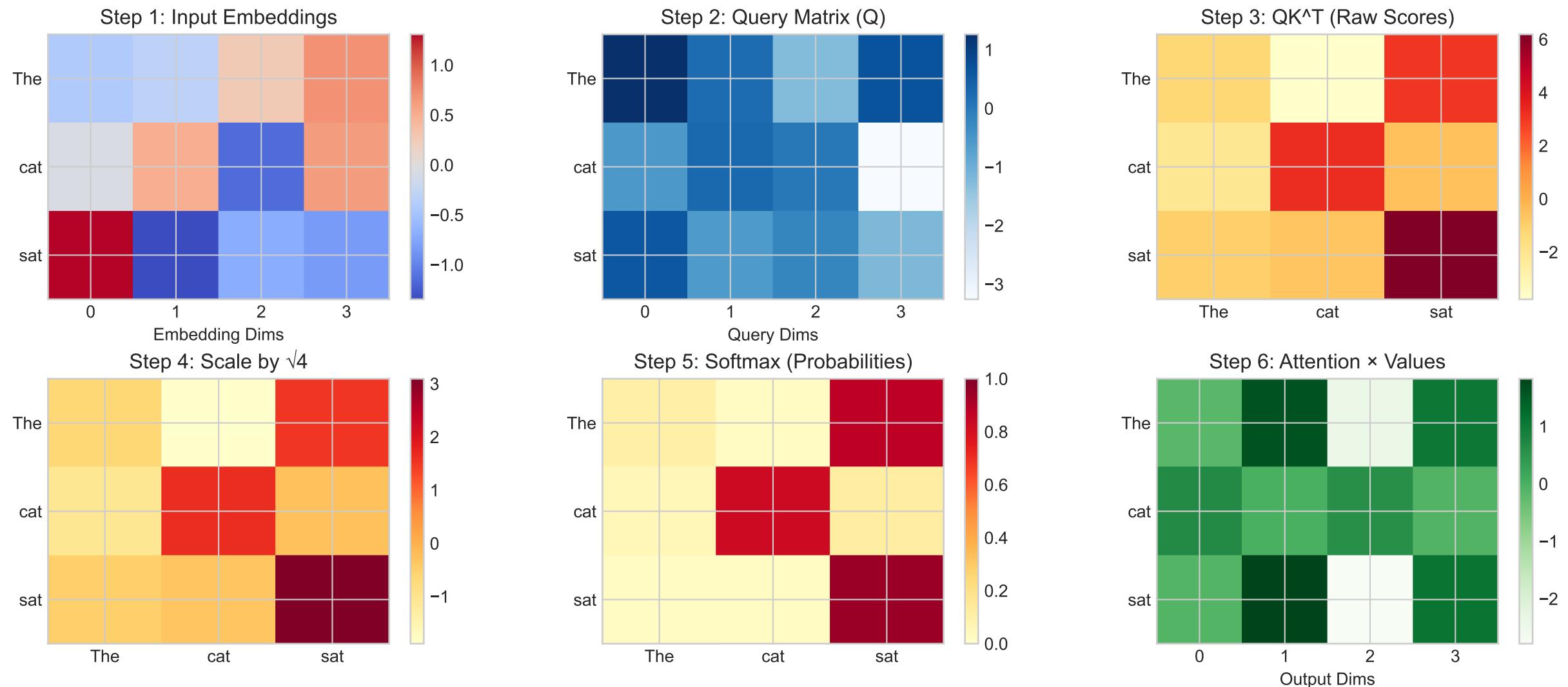


Self-Attention Mechanism: Step-by-Step Computation



$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Where each word decides what to attend to based on Query-Key similarity