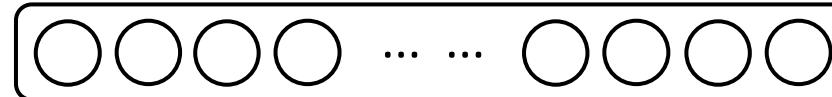


Output Hidden \mathbf{u}_t



$\{\mathbf{o}_{t,i}\}$

Multi-Head Attention (Core Attention)

