

attention  $\left(\mathbf{q}, \{\mathbf{k}_0, \mathbf{k}_1, \dots\}, \{\mathbf{v}_0, \mathbf{v}_1, \dots\}\right)$

$$= \frac{\sum_t \mathbf{v}_t \exp\left(\mathbf{k}_t^\top \mathbf{q} / \sqrt{d}\right)}{\sum_t \exp\left(\mathbf{k}_t^\top \mathbf{q} / \sqrt{d}\right)}$$