

$$W_{\text{att}} = \frac{e^{QK + \alpha M}}{\sum e^{QK + \alpha M}}$$

--- masked    — visible

