Heatmaps of attention logits $\frac{1}{\sqrt{d'}}\langle Qx_i, Kx_j \rangle$

(uniform_bert-large-uncased_params-trained_dense-on-layers-24, sample 1)