Heatmaps of attention logits $\frac{1}{\sqrt{d'}}\langle Qx_i,Kx_j\rangle$ (wikitext_albert-xlarge-v2_params-trained_dense-on_layers-48)

