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

(`wikitext_albert-xlarge-v2_params-random_dense-off_layers-48`)