Heatmaps of attention logits $\frac{1}{\sqrt{d'}}\langle Qx_i, Kx_j\rangle$ $({\tt wikitext_albert_xlarge_v2_params_trained_dense_on_layers_48})$

