Heatmaps of attention logits $\frac{1}{\sqrt{d'}}\langle Qx_i, Kx_j\rangle$ $({\tt randomtext_albert_xlarge_v2_params_trained_dense_off_layers_48})$ layer 25 layer 26 layer 27 0 100 100

