Heatmaps of attention logits $\frac{1}{\sqrt{d'}}\langle Qx_i,Kx_j\rangle$ wikitext_bert-large-uncased_params-trained_dense-off_layers-24)

