Heatmaps of attention logits  $\frac{1}{\sqrt{d'}}\langle Qx_i,Kx_j\rangle$  wikitext\_bert-large-uncased\_params-trained\_dense-on\_layers-24)

