Heatmaps of attention logits  $\frac{1}{\sqrt{d'}}\langle Qx_i, Kx_j\rangle$  $(\verb|imdb_albert-x|| arge-v2\_params-trained_dense-on_layers-48)$ layer 3 layer 1 layer 2 0

