

Heatmaps of attention logits $\frac{1}{\sqrt{d}} \langle Qx_i, Kx_j \rangle$
(imdb_albert-xlarge-v2_params-random_dense-on_layers-48, sample 3)

