Heatmaps of attention logits $\frac{1}{\sqrt{d'}} \langle Qx_i, Kx_j \rangle$
(`uniform_albert-xlarge-v2_params-trained_dense-off_layers-48`, sample 9)