cos similarity heatmap - layer 12 ert-xlarge-v2_uniform-tokens_(Qx,Ky)_trained-params_dense-off_24-layers) 0.75 100 0.50 0.00 vimilarity cosine similarity cosine similar 200 token i 300 -0.50400 -0.75500 -1.00500 100 200 300 400 token j