cos similarity heatmap - layer 21 ert-large-uncased_wikitext_(Qx,Ky)_trained-params_dense-off_24-layers) 0.75 20 0.50 0.25 cosine similarity 40 token i 0.00 60 -0.25-0.5080 -0.75-1.000 80 20 40 60 token j