cos similarity heatmap - layer 7 ert-large-uncased_wikitext_(Qx,Ky)_random-params_dense-on_24-layers) 0.75 20 0.50 cosine similarity 0.25 40 token i 0.00 60 -0.25-0.5080 -0.75100 -1.0020 40 60 80 100 token j