cos similarity heatmap - layer 7 ert-large-uncased_wikitext_(Qx,Ky)_random-params_dense-off_24-layers) 0.75 20 0.50 40 cosine similarity 0.25 token i 0.00 60 -0.2580 -0.50100 -0.75-1.000 20 40 60 80 100 token j