cos similarity heatmap - layer 21 bert-large-uncased_wikitext_(x,y)_trained-params_dense-off_24-layers) 0.75 50 0.50 100 0.25 0.00 cosine similarity 0.25 token i 0.00 150 200 -0.50250 -0.75-1.0050 100 150 250 0 200 token j