cos similarity heatmap - layer 16 bert-large-uncased\_wikitext\_(x,y)\_trained-params\_dense-off\_24-layers) 0.75 50 0.50 100 0.25 token i 0.00 150 -0.25200 -0.50250 -0.75-1.0050 100 150 200 250 token j