cos similarity heatmap - layer 12 bert-large-uncased\_wikitext\_(x,y)\_trained-params\_dense-off\_24-layers) 0.75 50 0.50 100 0.00 vimilarity cosine similarity cosine similar token i 150 200 -0.50250 -0.75-1.000 50 100 150 250 200 token j