cos similarity heatmap - layer 3 pert-xlarge-v2\_uniform-tokens\_(x,y)\_trained-params\_dense-off\_24-layers) 0.75 100 0.50 0.00 vimilarity cosine similarity cosine similar 200 token i 300 -0.50400 -0.75500 -1.00100 200 300 400 500 token j