cos similarity heatmap - layer 16 ert-xlarge-v2\_uniform-tokens\_(Qx,Ky)\_trained-params\_dense-on\_24-layers) 0.75 100 0.50 cosine similarity 0.25 200 token i 0.00 300 -0.25-0.50400 -0.75500 -1.00200 100 300 400 500 token j