cos similarity heatmap - layer 21 ert-xlarge-v2\_uniform-tokens\_(Qx,Ky)\_trained-params\_dense-off\_24-layers) 0.75 100 0.50 0.25 0.00 cosine similarity 0.25 200 token i 0.00 300 -0.50400 -0.75500 -1.00500 100 200 300 400 token j