cos similarity heatmap - layer 1 (albert-xlarge-v2_wikitext_(x,y)_random-params_dense-on_24-layers) 0.75 50 0.50 0.25 x 0.00 cosine similarity 0.25 cosine similarity 0.25 100 token i 0.00 150 -0.50200 -0.75-1.000 50 100 150 200 token j