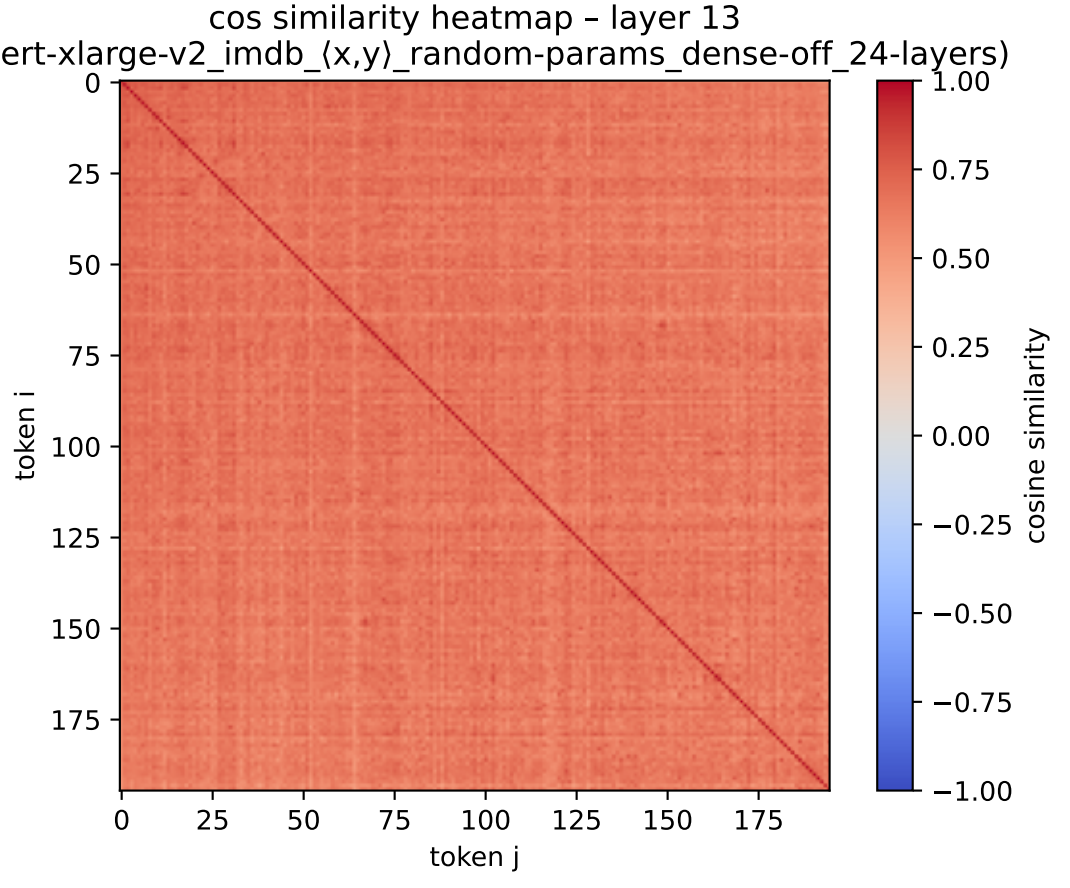
cos similarity heatmap – layer 13
(albert-xlarge-v2_imdb_⟨x,y⟩_random-params_dense-off_24-layers)
token i
token j
0
25
50
75
100
125
150
175
cosine similarity
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00