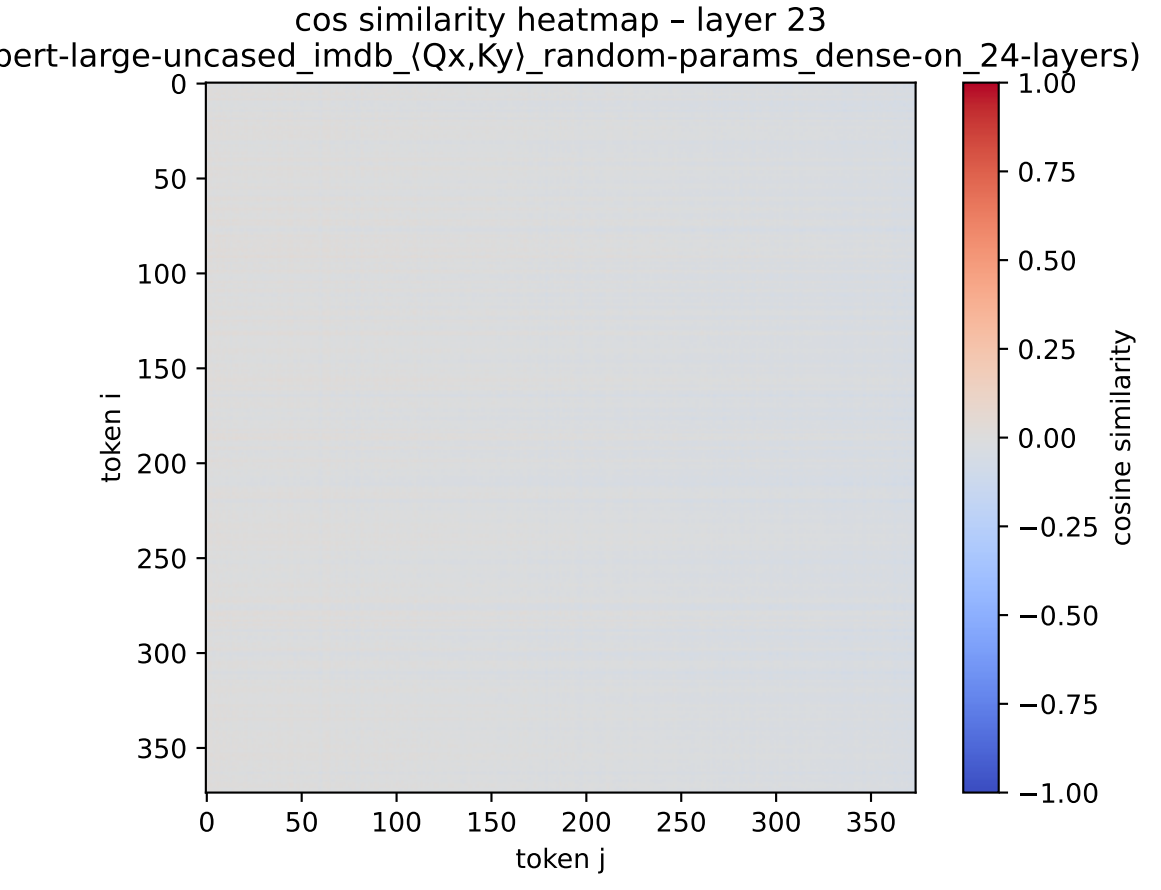
cos similarity heatmap – layer 23
pert-large-uncased_imdb_⟨Qx,Ky⟩_random-params_dense-on_24-layers)
token i
token j
0
50
100
150
200
250
300
350
cosine similarity
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00