cos similarity histogram - layer 1 (albert-xlarge-v2 uniform-tokens (x,y) trained-params dense-off 24-layers 6 5 4 3 2 1

0.00

0.25

0.50

0.75

1.00

-0.50

-1.00

-0.75

-0.25