cos similarity histograms (albert-xlarge-v2\_uniform-tokens\_(x,y)\_random-params\_dense-on\_24-layers) layer 2 layer 0 layer 1 layer 3 50 50 50 50 40 40 40 40 30 30 30 30 20 20 20 20 10 10 10 10 0.0 0.5 1.0 0.0 1.0 0.0 0.0 -0.5-0.50.5 -0.50.5 1.0 -0.50.5 1.0 -1.0-1.0-1.0layer 4 layer 5 layer 6 layer 7 50 50 50 50 40 40 40 40 30 30 30 30 20 20 20 20 10 10 10 10 -0.50.0 0.5 -0.50.0 0.5 -0.50.0 0.5 -0.50.0 1.0 -1.01.0 -1.01.0 -1.00.5 1.0 -1.0layer 9 layer 10 layer 8 layer 11 50 50 50 50 40 40 40 40 30 30 30 30 20 20 20 20 10 10 10 10 0.0 0.0 0.5 0.0 0.5 -0.50.0 0.5 -0.5-1.0-0.51.0 -1.0-0.51.0 -1.01.0 -1.00.5 1.0 layer 12 layer 13 layer 14 layer 15 50 50 50 50 40 40 40 40 30 30 30 30 20 20 20 20 10 10 10 10 -0.50.0 -0.50.0 -0.50.0 -0.5 0.0 0.5 1.0 0.5 0.5 -1.0-1.0-1.01.0 -1.01.0 0.5 1.0 layer 16 layer 17 layer 18 layer 19 50 50 50 50 40 40 40 40 30 30 30 30 20 20 20 20 10 -10 10 10 0.0 0.5 0.0 0.5 0.0 0.0 -1.0-0.51.0 -1.0-0.5 1.0 -1.0-0.50.5 1.0 -1.0-0.5 0.5 1.0 layer 20 layer 21 layer 22 layer 23 50 50 50 50 40 40 40 40 30 30 30 30 20 20 20 20 10 10 10 10 0 -0 0 0

-0.5

-1.0

0.0

0.5

1.0

-1.0

-0.5

0.0

0.5

-1.0

1.0

-0.5

0.0

0.5

-1.0

1.0

-0.5

0.0

0.5

1.0