attention_logits histograms (randomtext_albert-large-v2_params-random_dense-off_layers-48) after layer 4 after layer 1 after layer 2 after layer 3 20 20 20 20 15 15 15 15 10 10 10 10 5 5 5 5 0.0 0.0 0.0 0.5 -0.50.0 0.5 0.5 -0.50.5 1.0 -0.5-0.5-1.0-1.01.0 -1.01.0 -1.01.0 after layer 6 after layer 5 after layer 7 after layer 8 20 20 20 20 15 15 15 15 10 10 10 10 5 5 5 5 -0.50.0 0.5 -0.50.0 -0.50.0 0.5 -0.50.0 1.0 -1.00.5 1.0 -1.01.0 -1.00.5 1.0 -1.0after layer 9 after layer 10 after layer 11 after layer 12 20 20 20 20 15 15 15 15 10 10 10 10 5 5 5 5 0.0 0.0 0.0 0.0 -1.0-0.50.5 1.0 -1.0-0.50.5 1.0 -1.0-0.50.5 1.0 -1.0-0.50.5 1.0 after layer 13 after layer 14 after layer 15 after layer 16 20 20 20 20 15 15 15 15 10 10 10 10 5 5 5 5 0.0 0.0 0.0 0.0 -0.50.5 -0.50.5 -0.50.5 -0.51.0 -1.01.0 -1.01.0 -1.00.5 1.0 -1.0after layer 17 after layer 18 after layer 19 after layer 20 20 20 20 20 15 15 15 15 10 10 10 10 5 · 5 5 5 -0.50.0 0.5 -0.50.0 0.5 -0.50.0 -0.50.0 -1.01.0 -1.01.0 -1.00.5 1.0 -1.00.5 1.0 after layer 21 after layer 22 after layer 23 after layer 24 20 -20 20 20 15 15 15 15 10 -10 10 10 5 5 5 5

0 -

-1.0

1.0

-0.5

0.0

0.5

1.0

-0.5

0.0

0.5

-1.0

1.0

0 -

-1.0

-0.5

0.0

0.5

1.0

-0.5

0.0

0.5

-1.0