attention_logits histograms (randomtext_bert-large-uncased_params-trained_dense-on_layers-24) after layer 1 after layer 3 after layer 4 after layer 2 0.4 0.4 0.4 0.4 0.3 0.3 0.3 0.3 0.2 0.2 0.2 0.2 0.1 0.1 0.1 0.1 0.0 0.0 0.0 0.0 0.0 0.0 0.5 0.0 0.5 0.5 -0.50.5 1.0 -0.5-0.5-0.50.0 -1.0-1.01.0 -1.01.0 -1.01.0 after layer 5 after layer 8 after layer 6 after layer 7 0.4 0.4 0.4 0.4 0.3 0.3 0.3 0.3 0.2 0.2 0.2 0.2 0.1 0.1 0.1 0.1 0.0 0.0 0.0 0.0 0.0 0.5 1.0 -0.50.0 0.5 -0.50.0 0.5 1.0 0.0 -0.5-1.01.0 -1.0-1.0-0.51.0 -1.0after layer 9 after layer 10 after layer 11 after layer 12 0.4 0.4 0.4 0.4 0.3 0.3 0.3 0.3 0.2 0.2 0.2 0.2 0.1 0.1 0.1 0.1 0.0 0.0 0.0 0.0 0.0 0.5 1.0 0.0 0.5 -0.50.0 0.5 1.0 0.0 0.5 -1.0-0.5-1.0-0.51.0 -1.0-1.0-0.51.0 after layer 13 after layer 14 after layer 15 after layer 16 0.4 0.4 0.4 0.4 0.3 0.3 0.3 0.3 0.2 0.2 0.2 0.2 0.1 0.1 0.1 0.1 0.0 0.0 0.0 0.0 0.0 0.0 0.5 0.0 0.0 -0.50.5 -0.5-0.50.5 -1.0-0.50.5 -1.01.0 -1.01.0 -1.01.0 1.0 after layer 17 after layer 18 after layer 19 after layer 20 0.4 0.4 0.4 0.4 0.3 0.3 0.3 0.3 0.2 0.2 0.2 0.2 0.1 0.1 0.1 -0.1 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.5 -1.0-0.50.5 1.0 -1.0-0.50.5 1.0 -1.0-0.51.0 -1.0-0.50.0 0.5 1.0 after layer 21 after layer 22 after layer 23 after layer 24 0.4 0.4 0.4 0.4 0.3 0.3 0.3 0.3 0.2 0.2 0.2 0.2 0.1 0.1 0.1 0.1

0.0

-1.0

1.0

-0.5

0.0

0.5

0.0

-1.0

-0.5

0.0

0.5

0.0

-1.0

-0.5

0.0

0.5

1.0

0.0

-1.0

-0.5

0.0

0.5

1.0

1.0