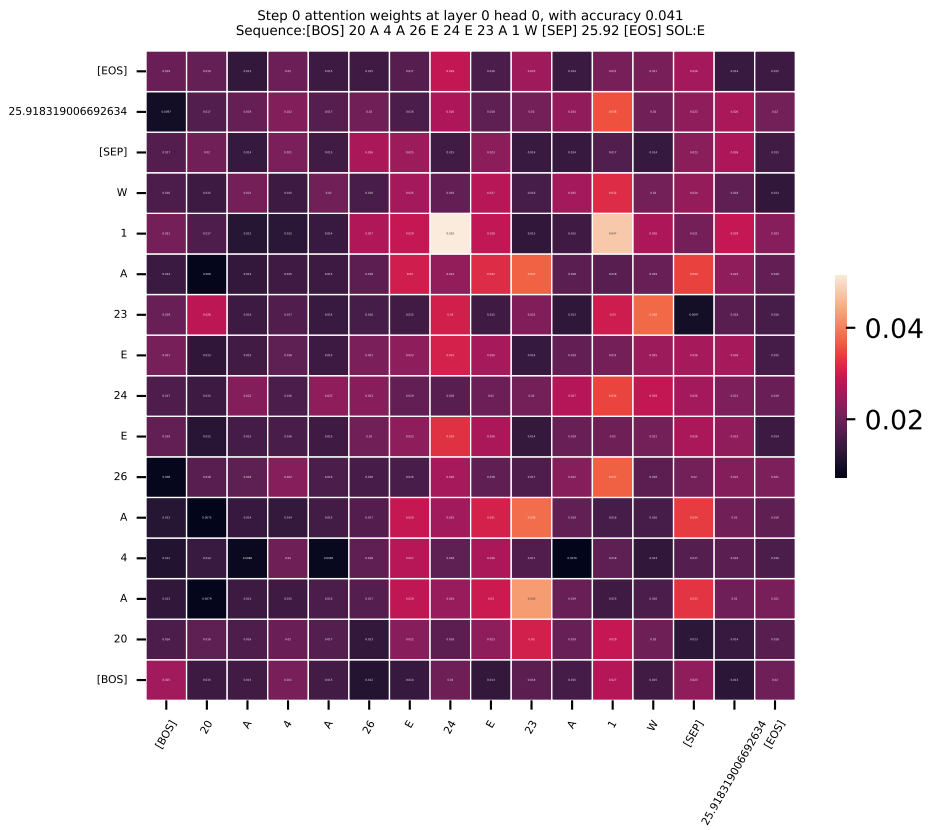
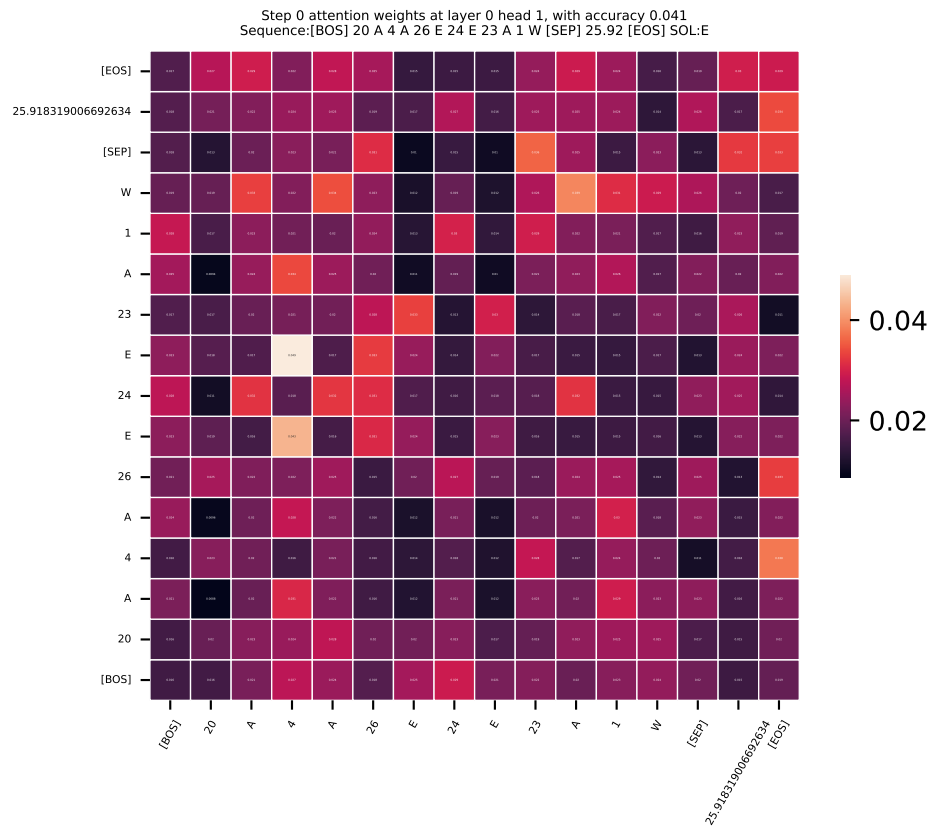
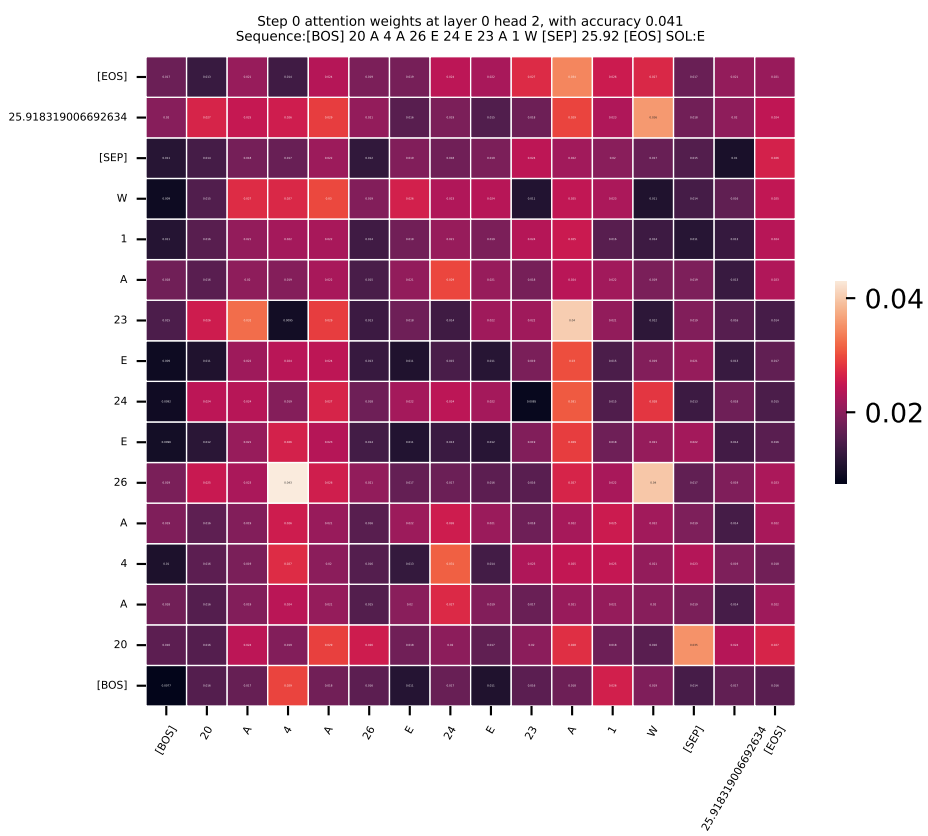
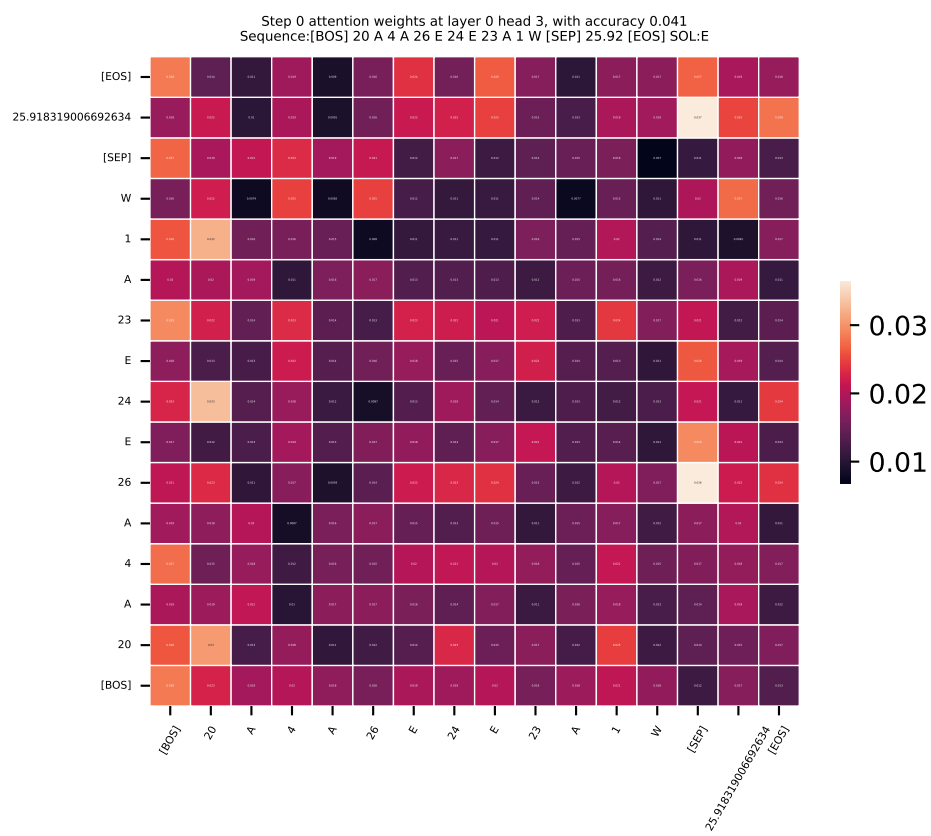
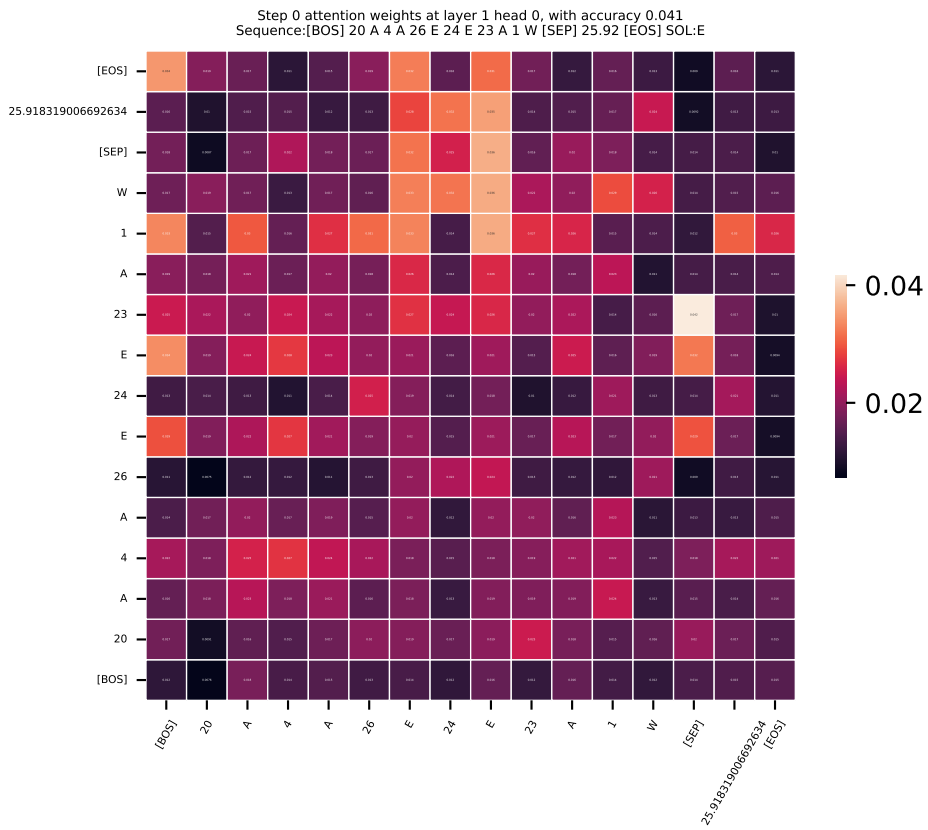
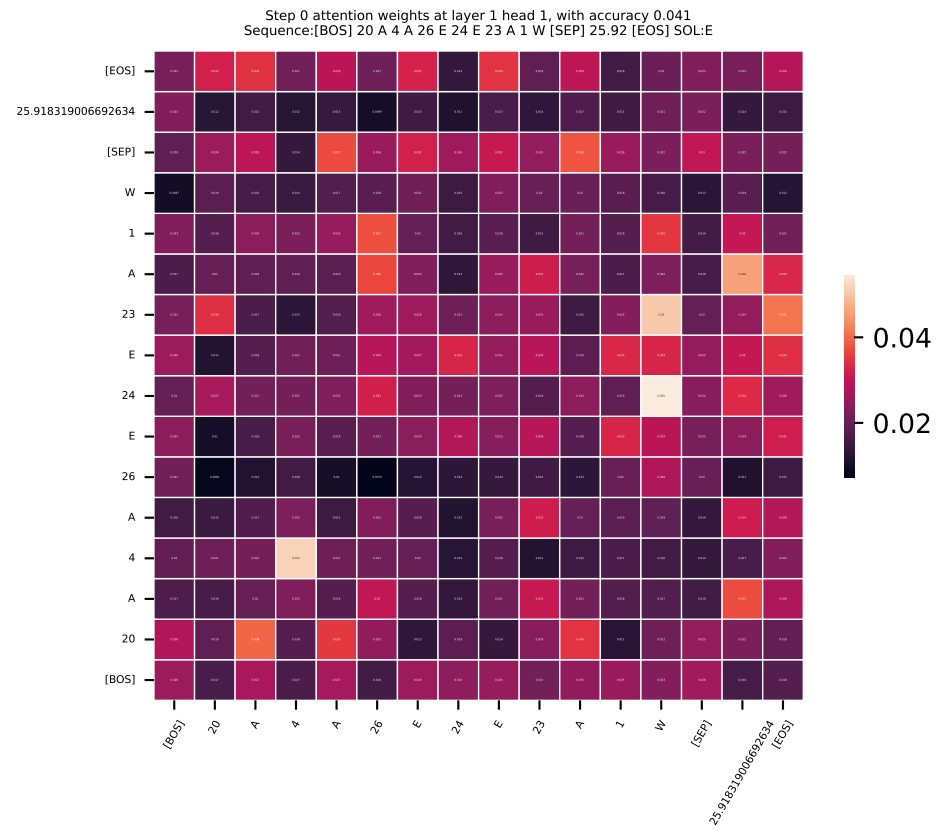

Step 0 attention weights at layer 0 head 0, with accuracy 0.041
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.04
0.02

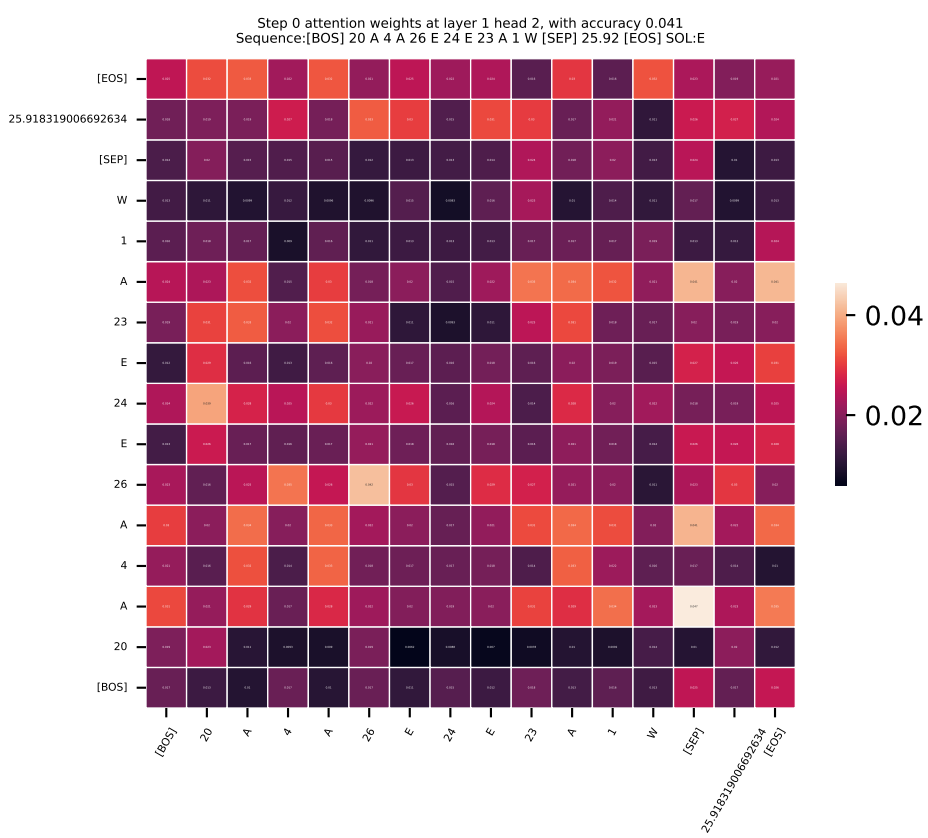

Step 0 attention weights at layer 0 head 1, with accuracy 0.041
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.04
0.02

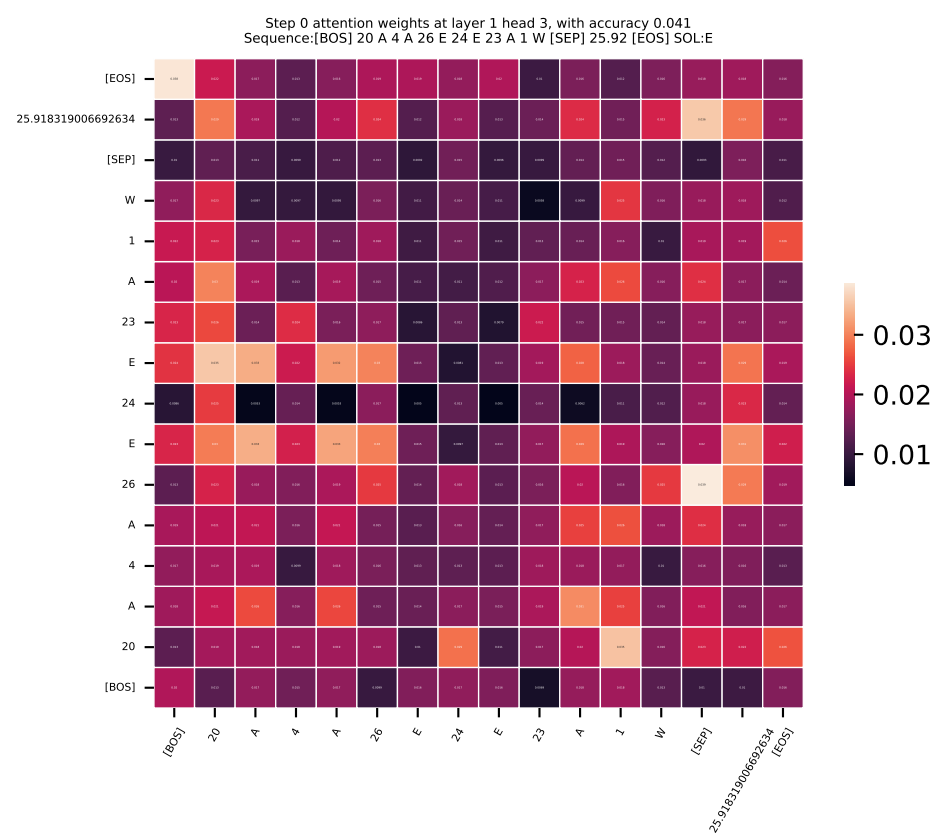

Step 0 attention weights at layer 0 head 2, with accuracy 0.041
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.04
0.02

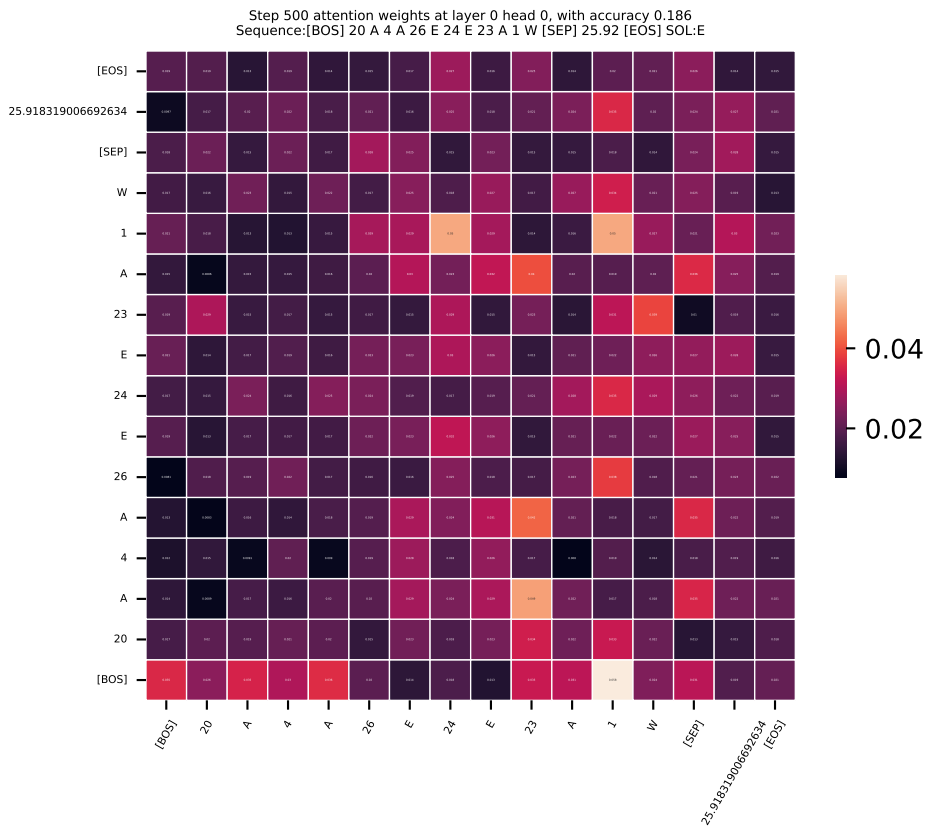

Step 0 attention weights at layer 0 head 3, with accuracy 0.041
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.03
0.02
0.01

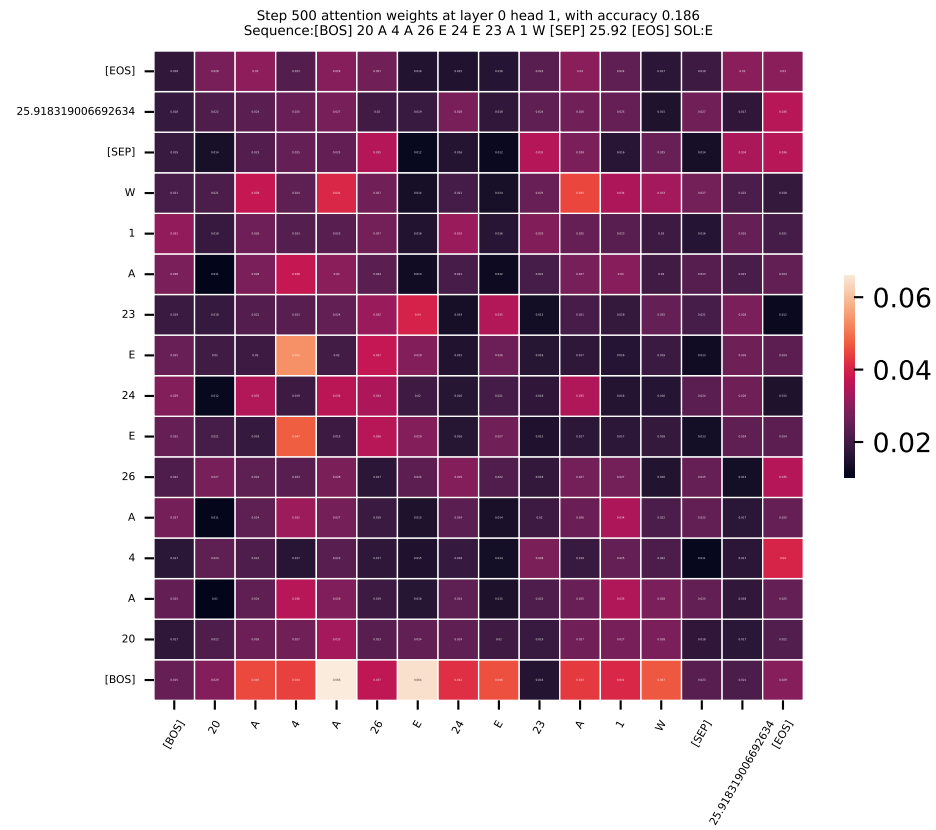

Step 0 attention weights at layer 1 head 0, with accuracy 0.041
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.04
0.02

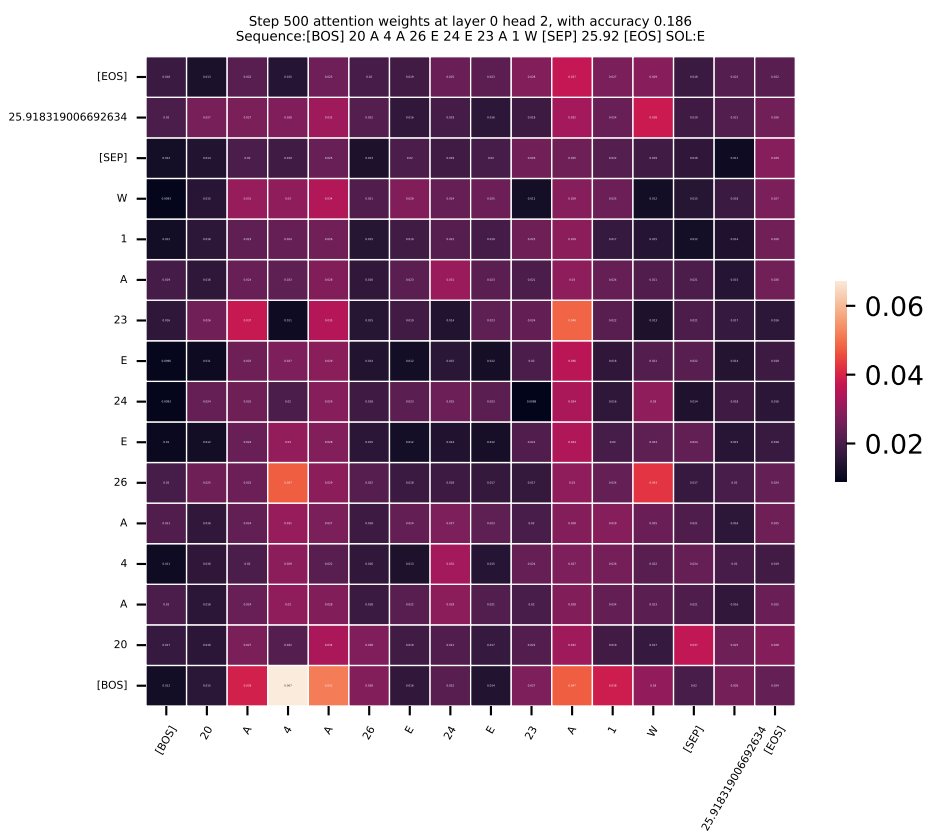

Step 0 attention weights at layer 1 head 1, with accuracy 0.041
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.04
0.02

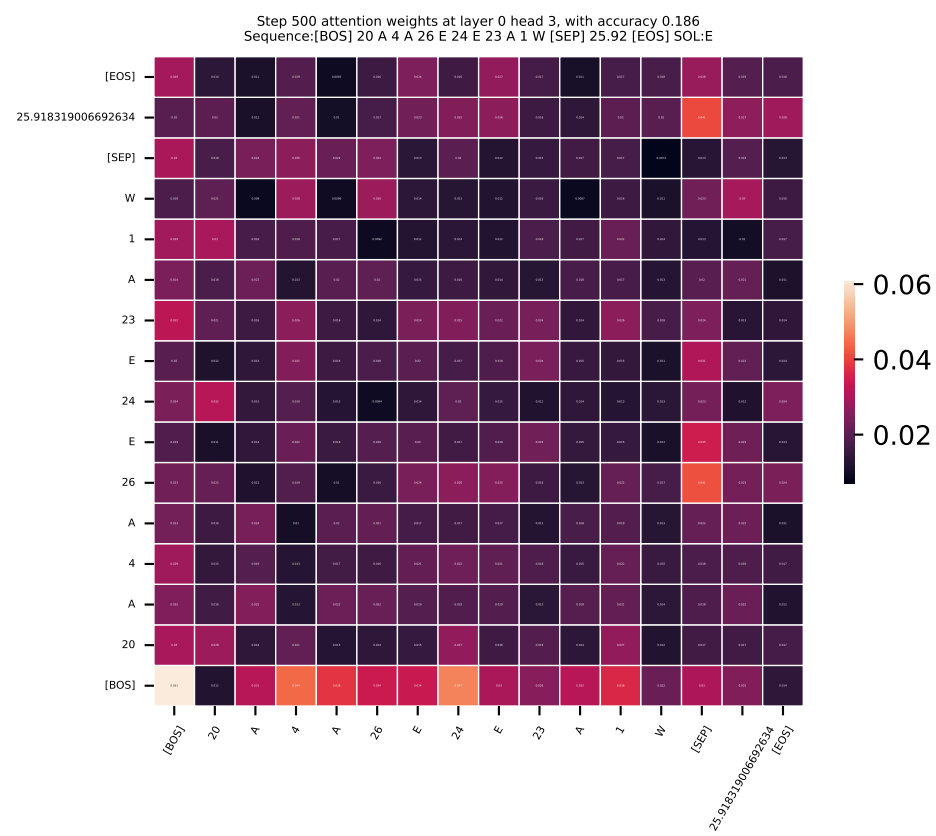

Step 0 attention weights at layer 1 head 2, with accuracy 0.041
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.04
0.02

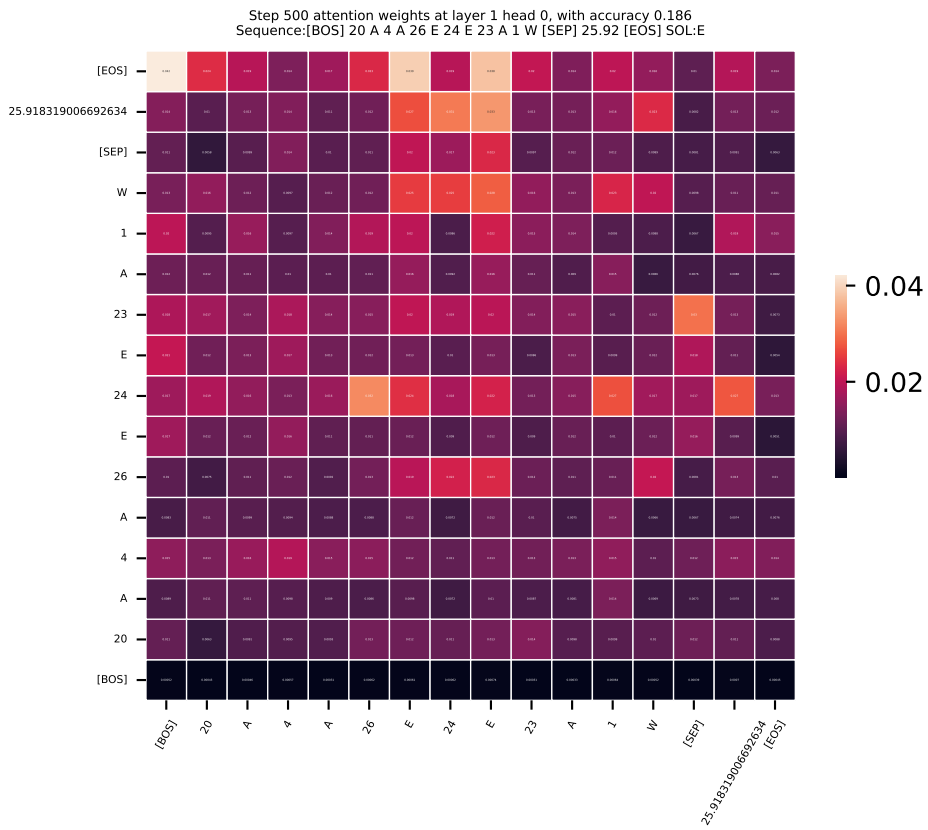

Step 0 attention weights at layer 1 head 3, with accuracy 0.041
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.03
0.02
0.01

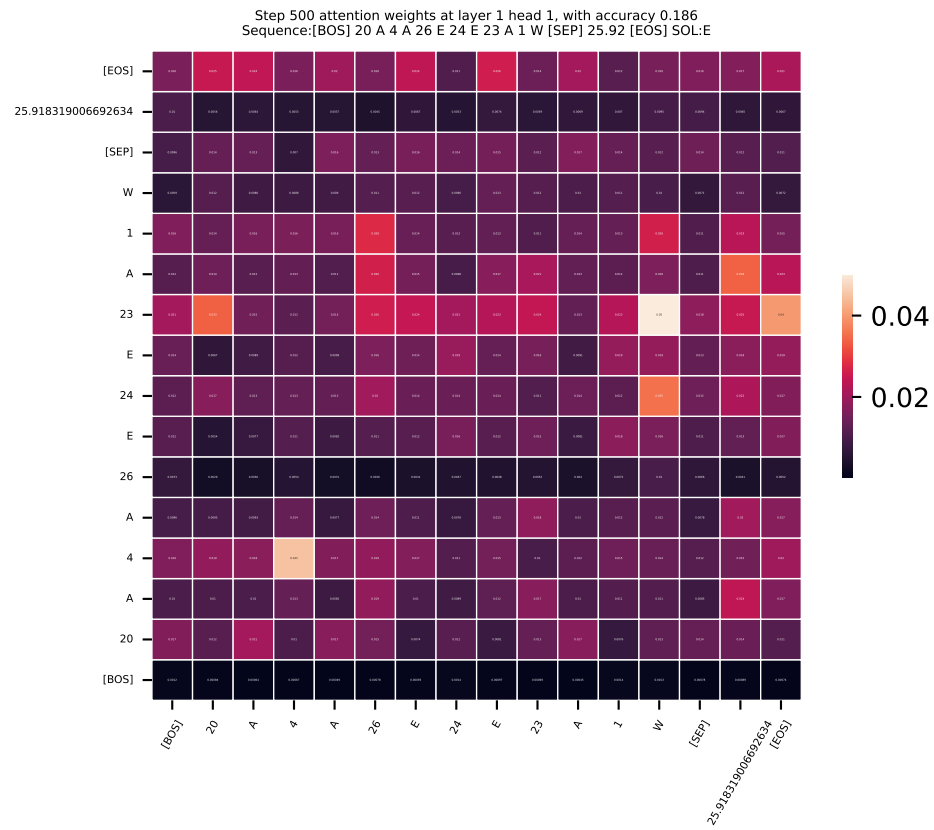

Step 500 attention weights at layer 0 head 0, with accuracy 0.186
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.04
0.02

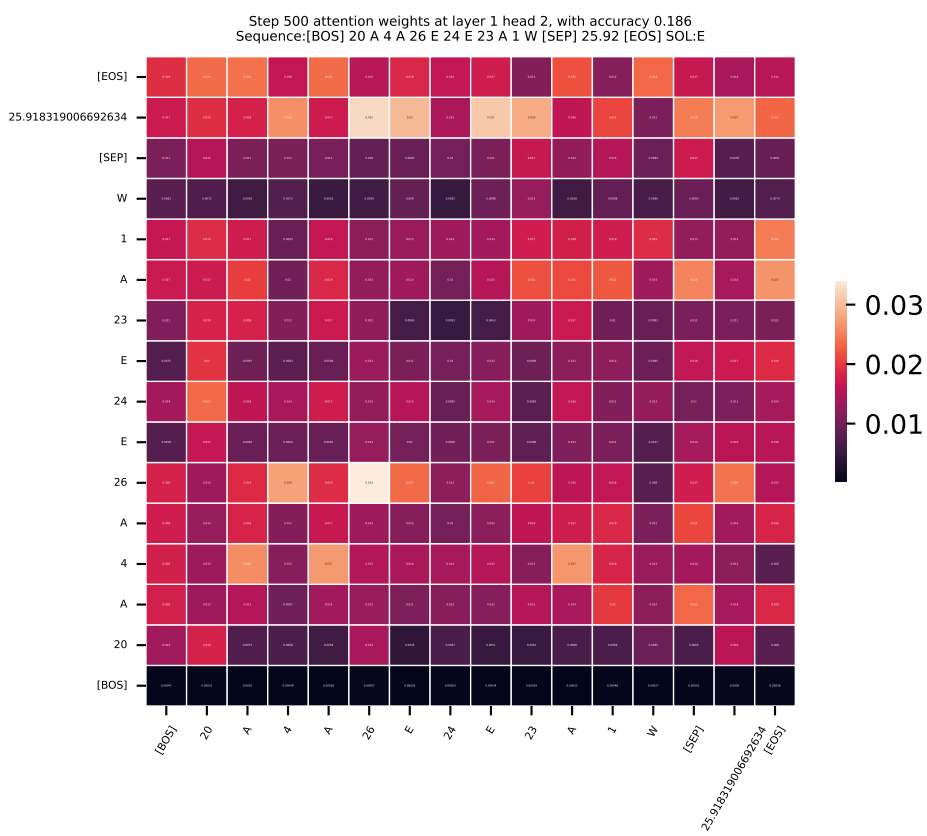

Step 500 attention weights at layer 0 head 1, with accuracy 0.186
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.06
0.04
0.02

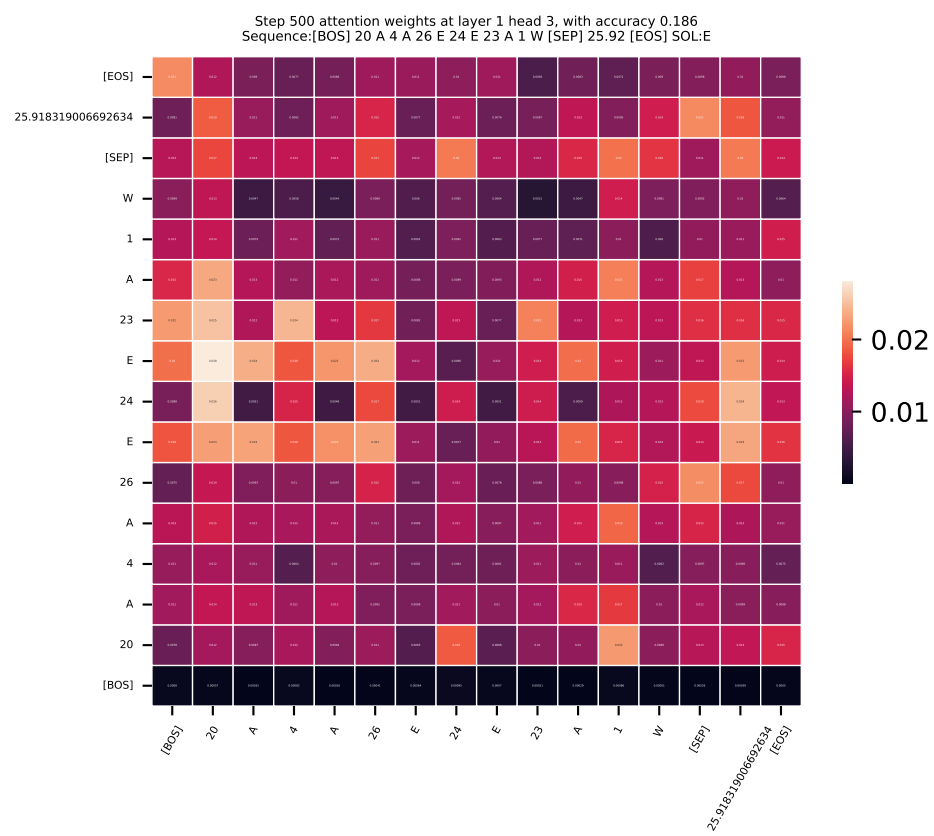

Step 500 attention weights at layer 0 head 2, with accuracy 0.186
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.06
0.04
0.02

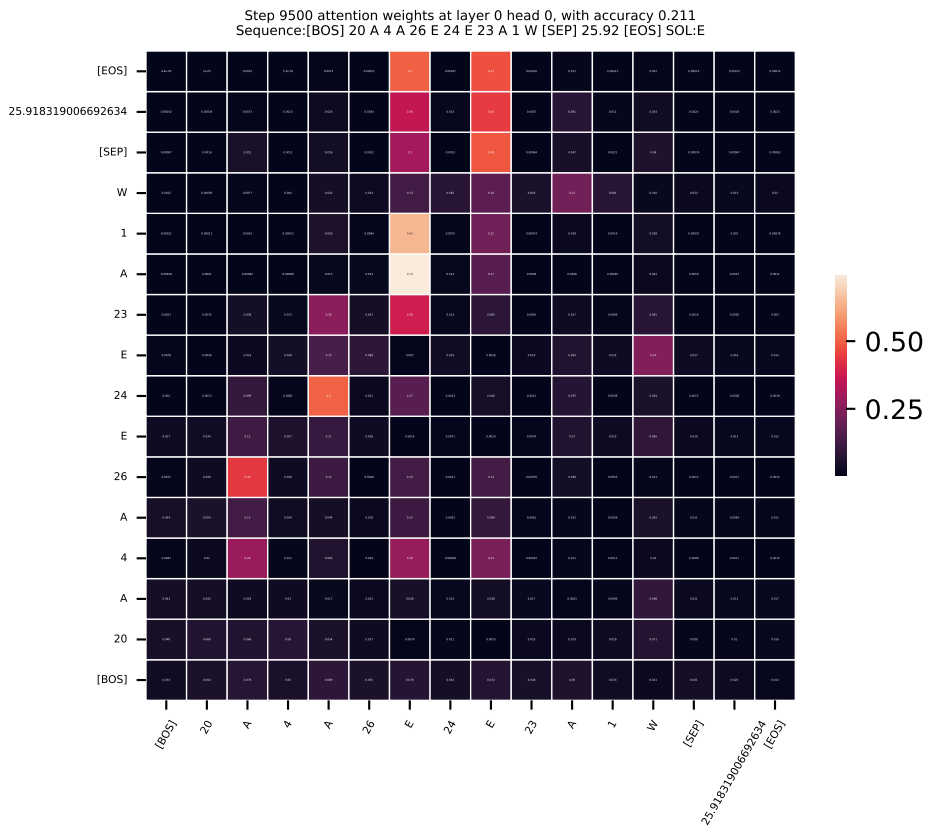

Step 500 attention weights at layer 0 head 3, with accuracy 0.186
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.06
0.04
0.02

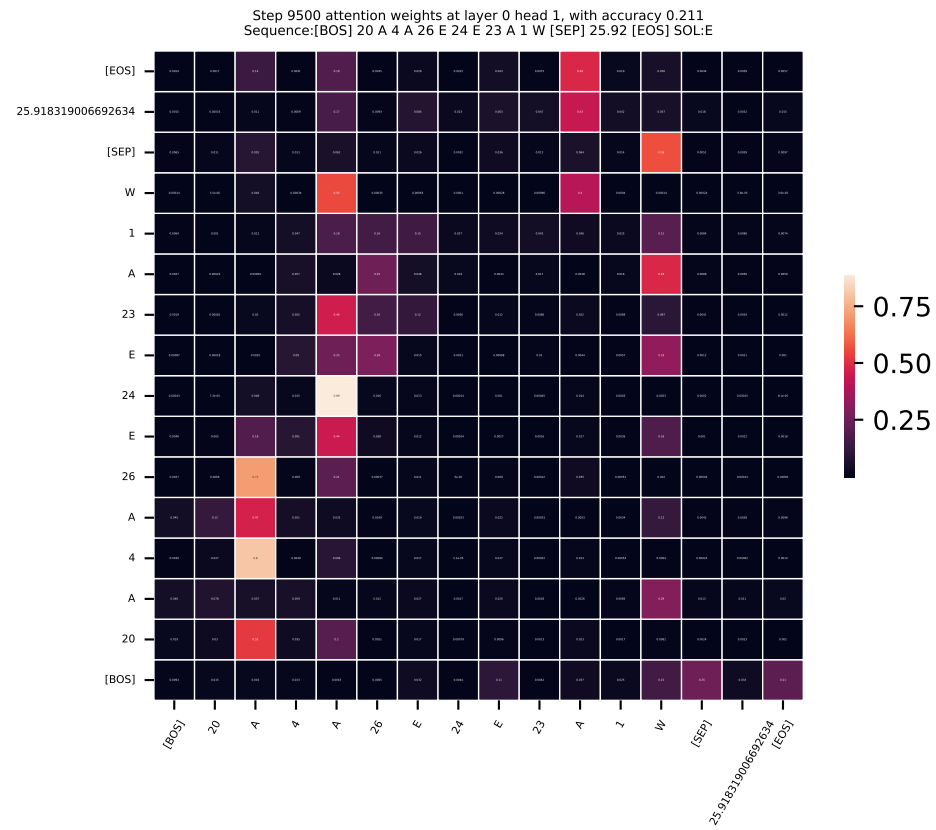

Step 500 attention weights at layer 1 head 0, with accuracy 0.186
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.04
0.02

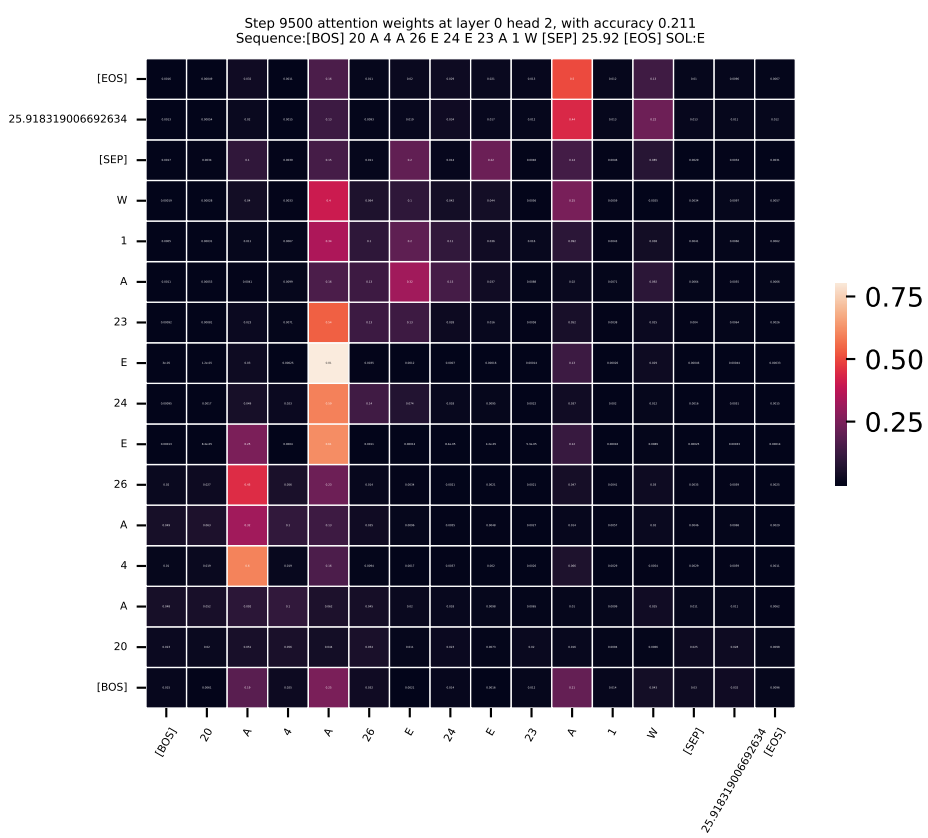

Step 500 attention weights at layer 1 head 1, with accuracy 0.186
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.04
0.02

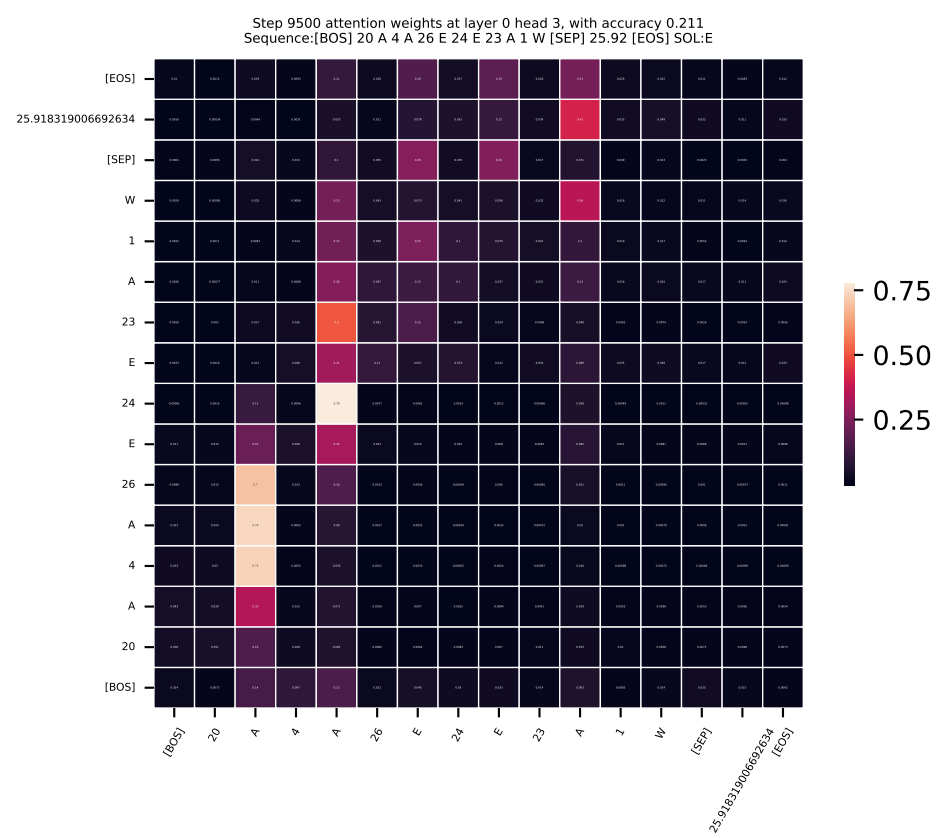

Step 500 attention weights at layer 1 head 2, with accuracy 0.186
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.03
0.02
0.01

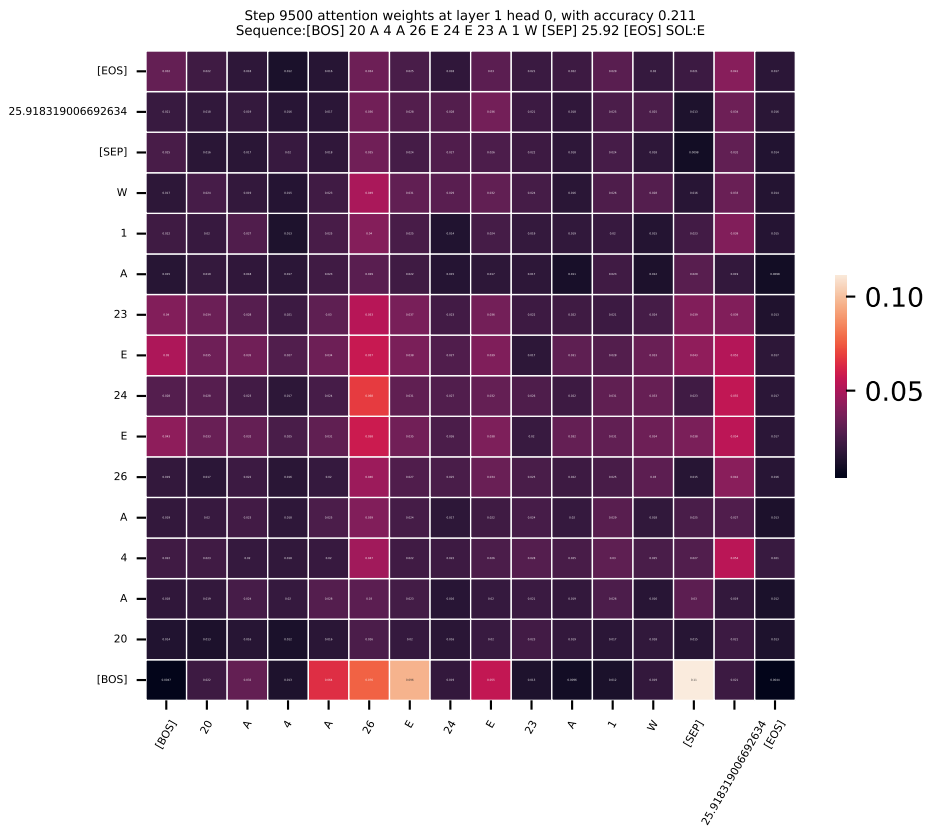

Step 500 attention weights at layer 1 head 3, with accuracy 0.186
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.02
0.01

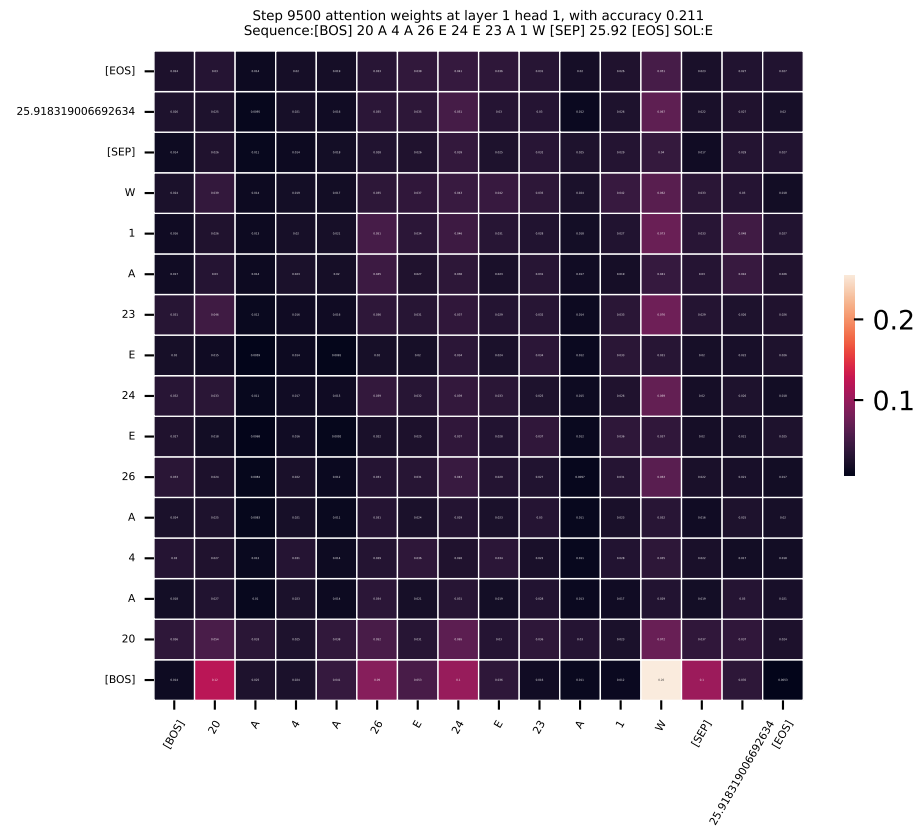

Step 9500 attention weights at layer 0 head 0, with accuracy 0.211
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

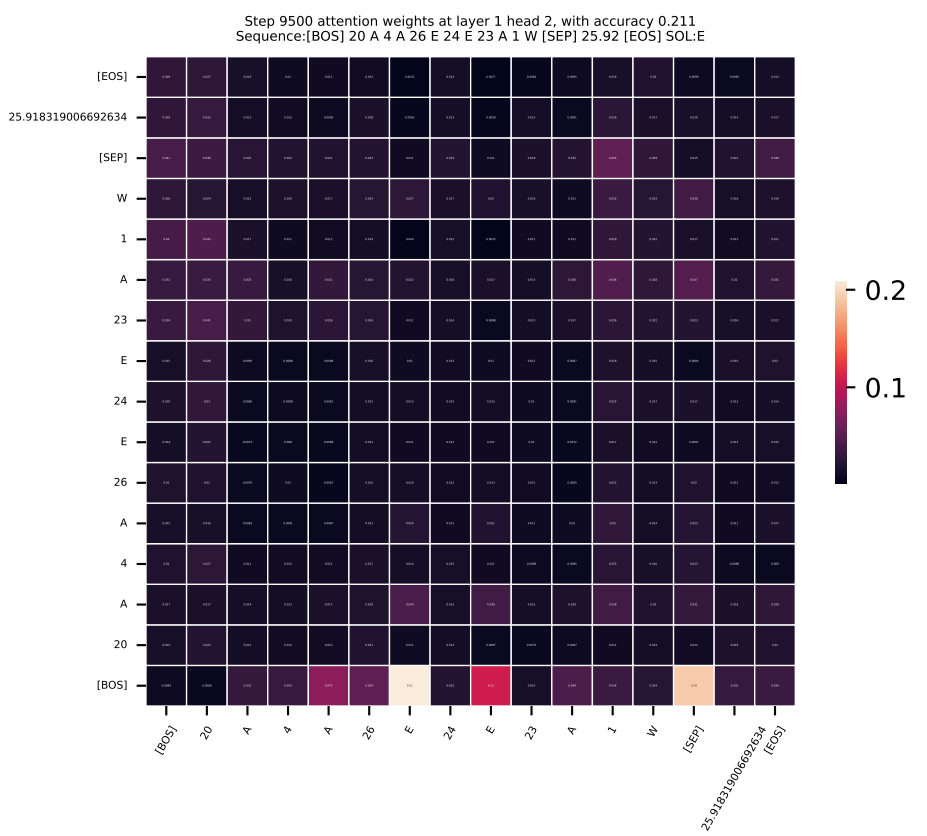

Step 9500 attention weights at layer 0 head 1, with accuracy 0.211
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

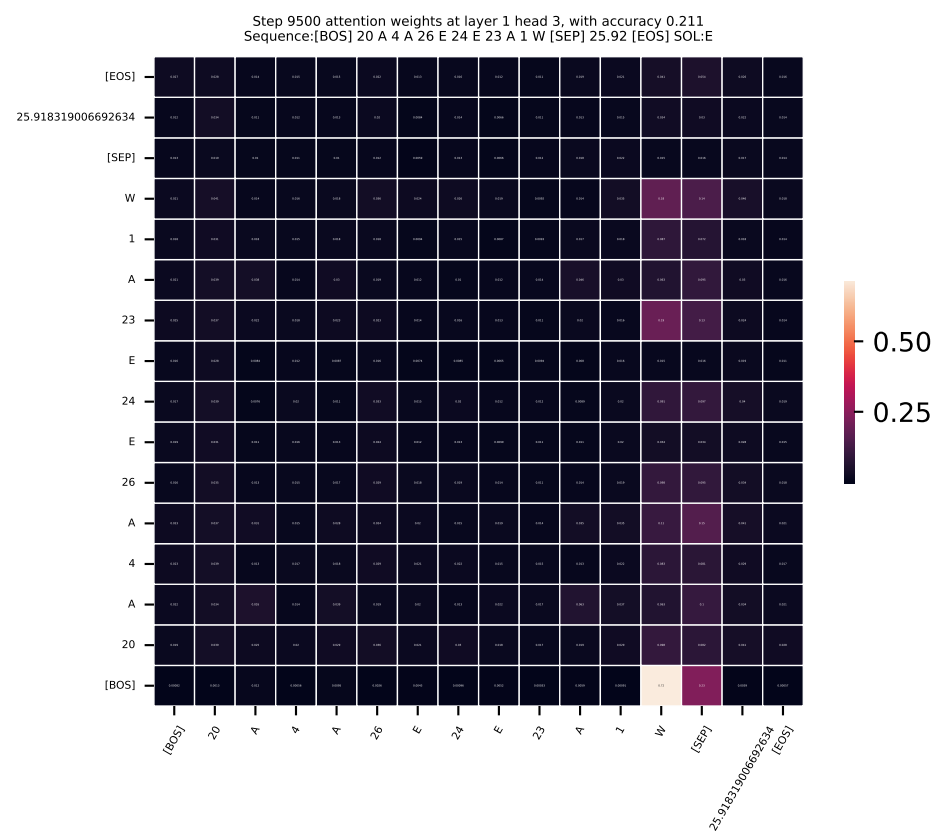

Step 9500 attention weights at layer 0 head 2, with accuracy 0.211
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

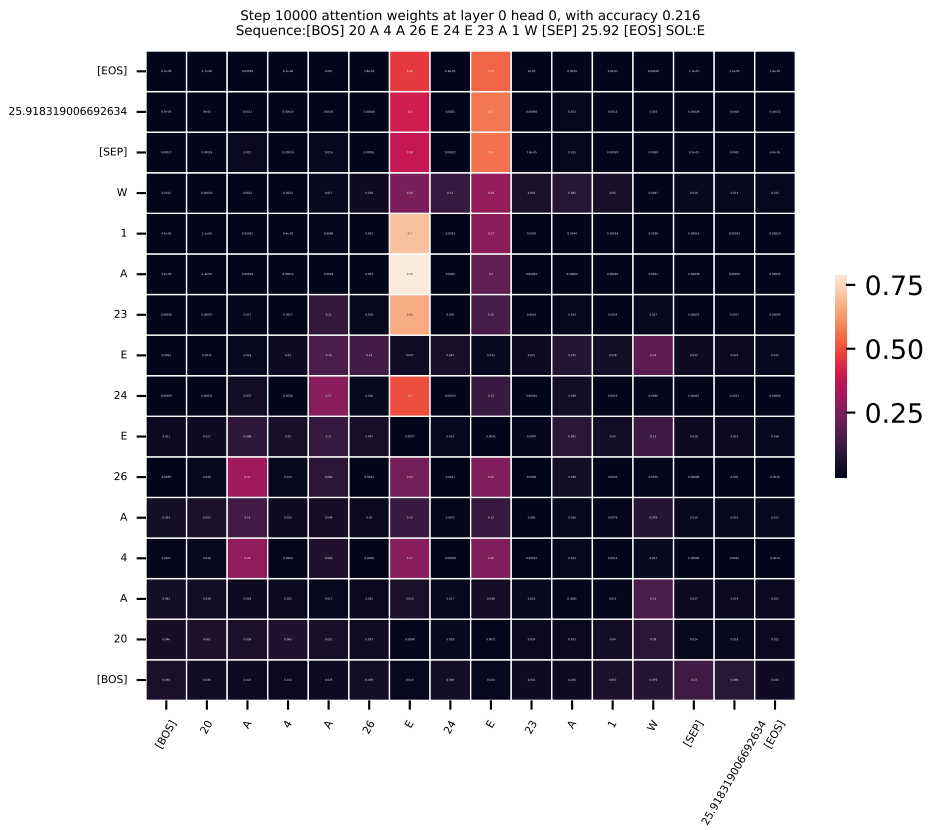

Step 9500 attention weights at layer 0 head 3, with accuracy 0.211
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

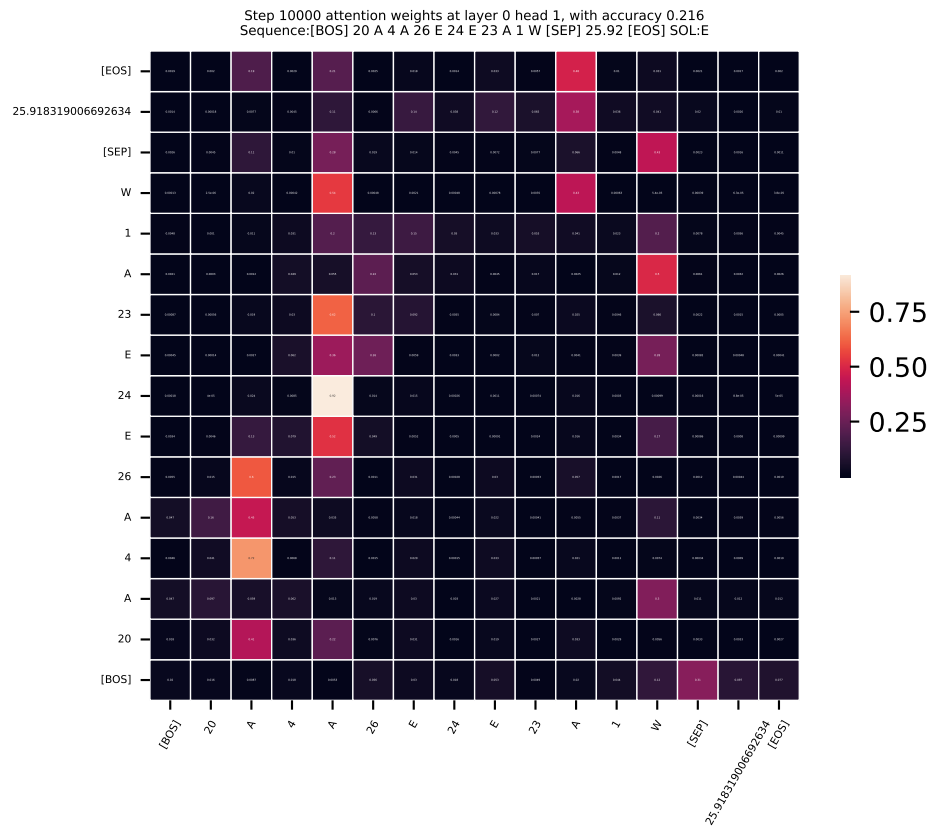

Step 9500 attention weights at layer 1 head 0, with accuracy 0.211
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.10
0.05

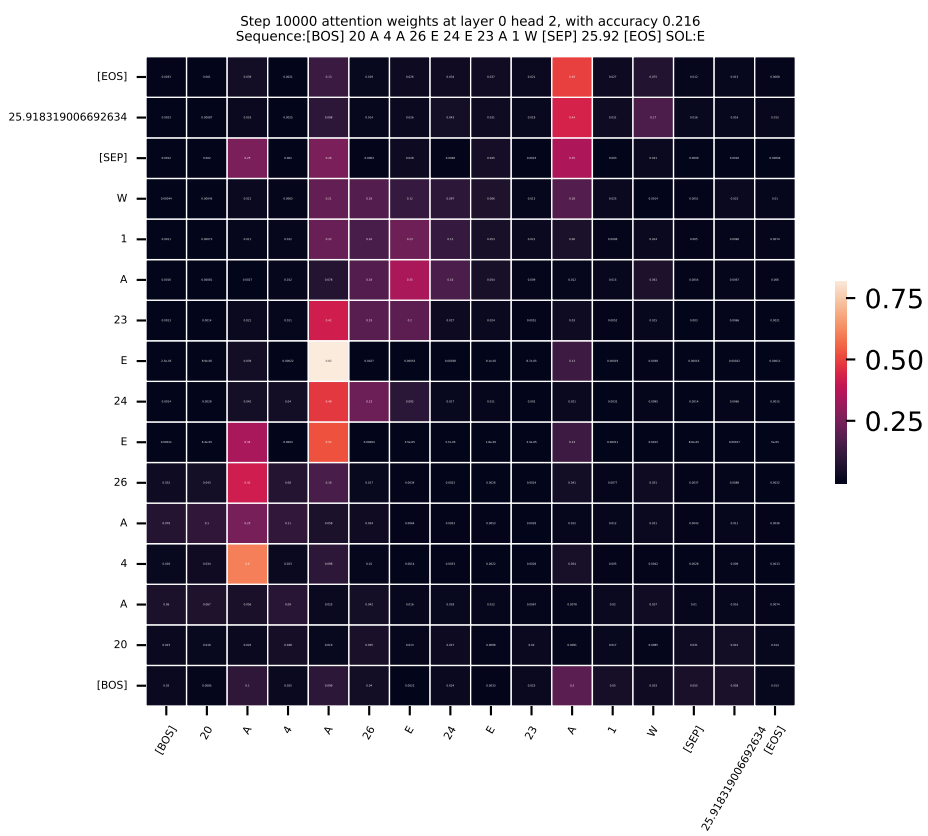

Step 9500 attention weights at layer 1 head 1, with accuracy 0.211
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.2
0.1

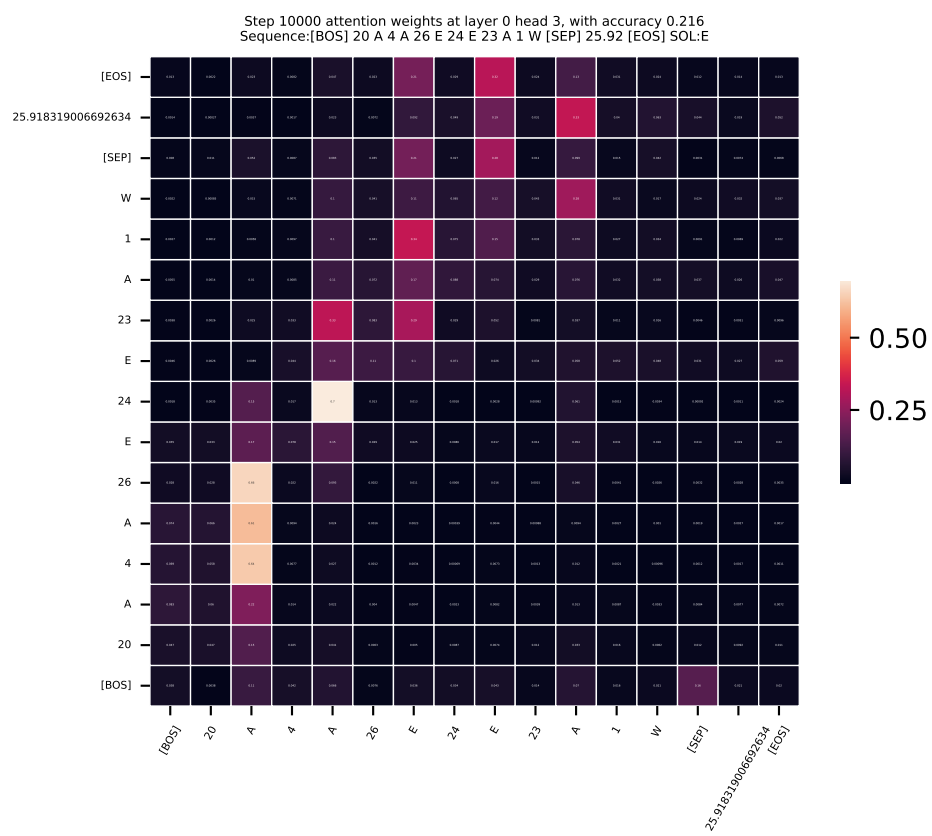

Step 9500 attention weights at layer 1 head 2, with accuracy 0.211
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.2
0.1

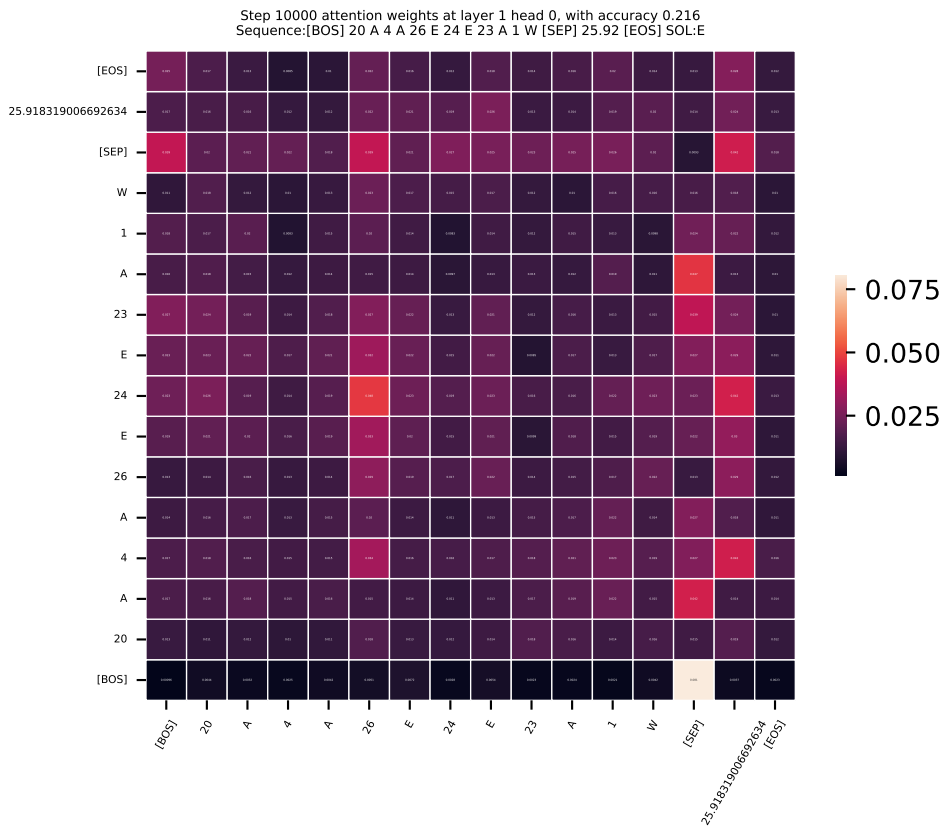

Step 9500 attention weights at layer 1 head 3, with accuracy 0.211
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.50
0.25

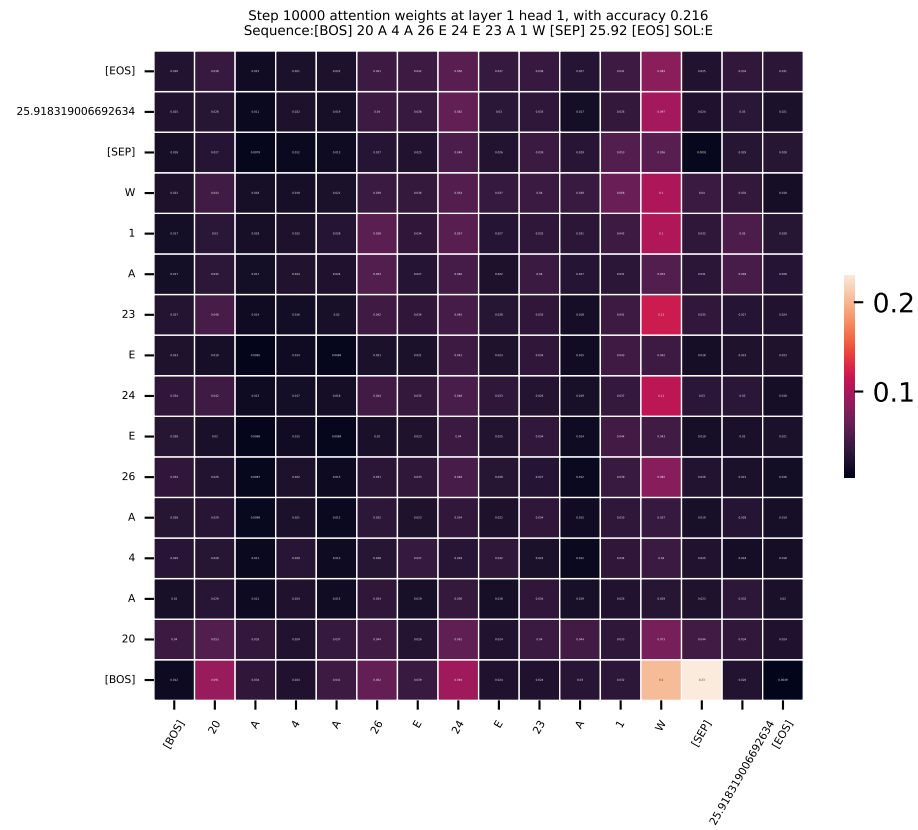

Step 10000 attention weights at layer 0 head 0, with accuracy 0.216
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

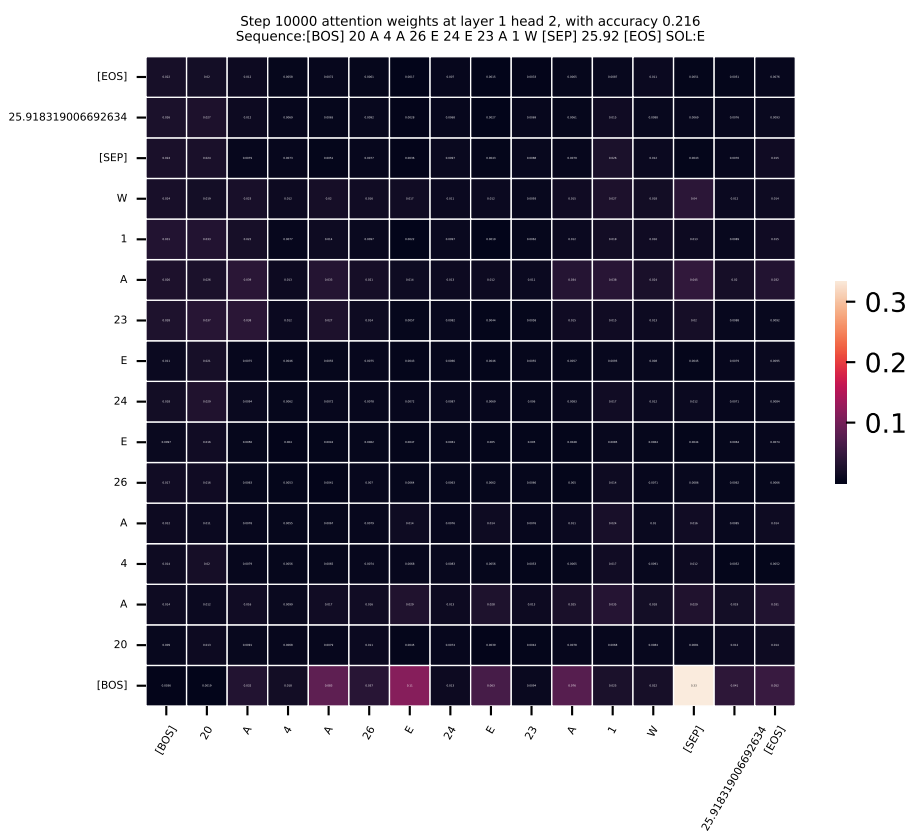

Step 10000 attention weights at layer 0 head 1, with accuracy 0.216
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

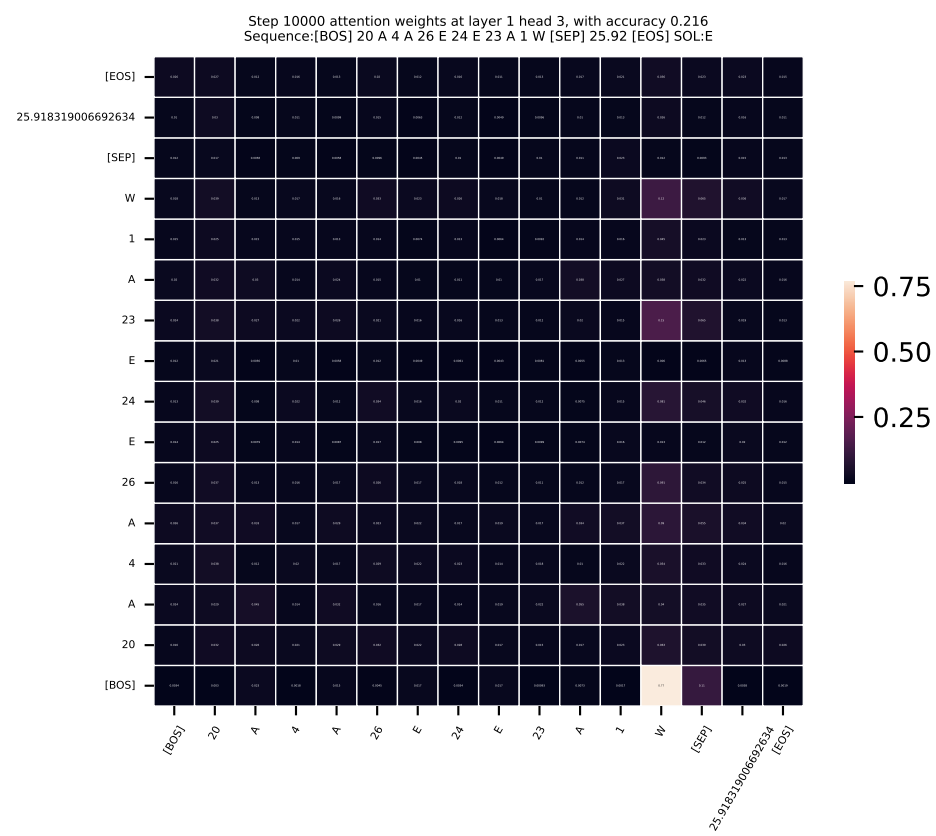

Step 10000 attention weights at layer 0 head 2, with accuracy 0.216
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

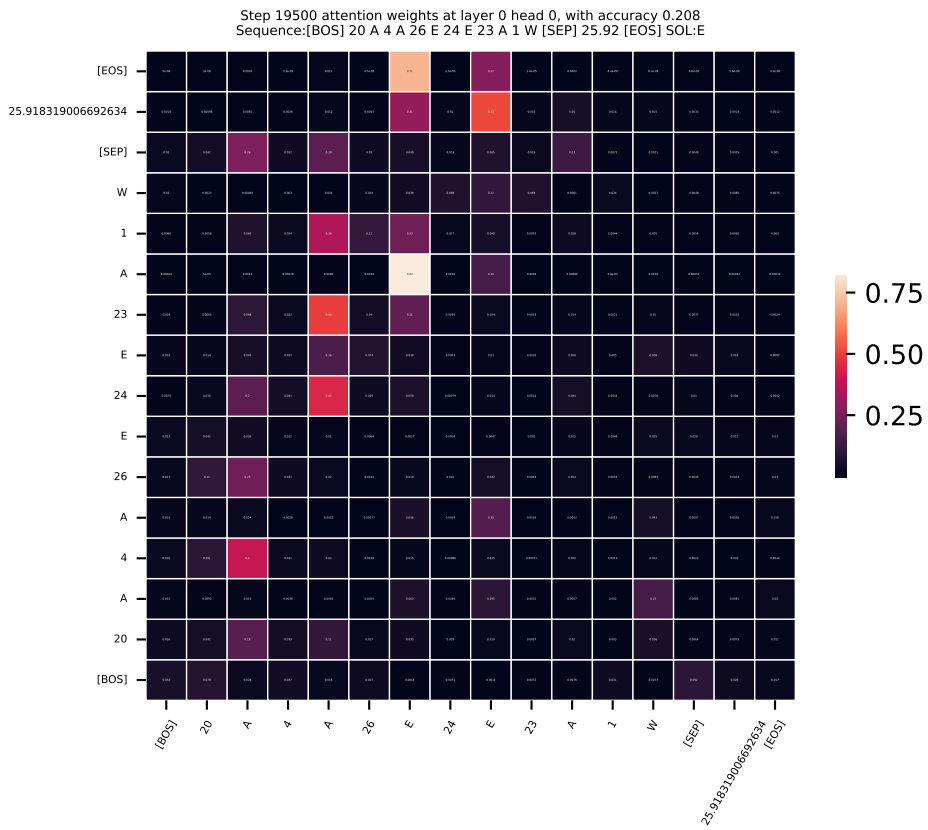

Step 10000 attention weights at layer 0 head 3, with accuracy 0.216
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.50
0.25

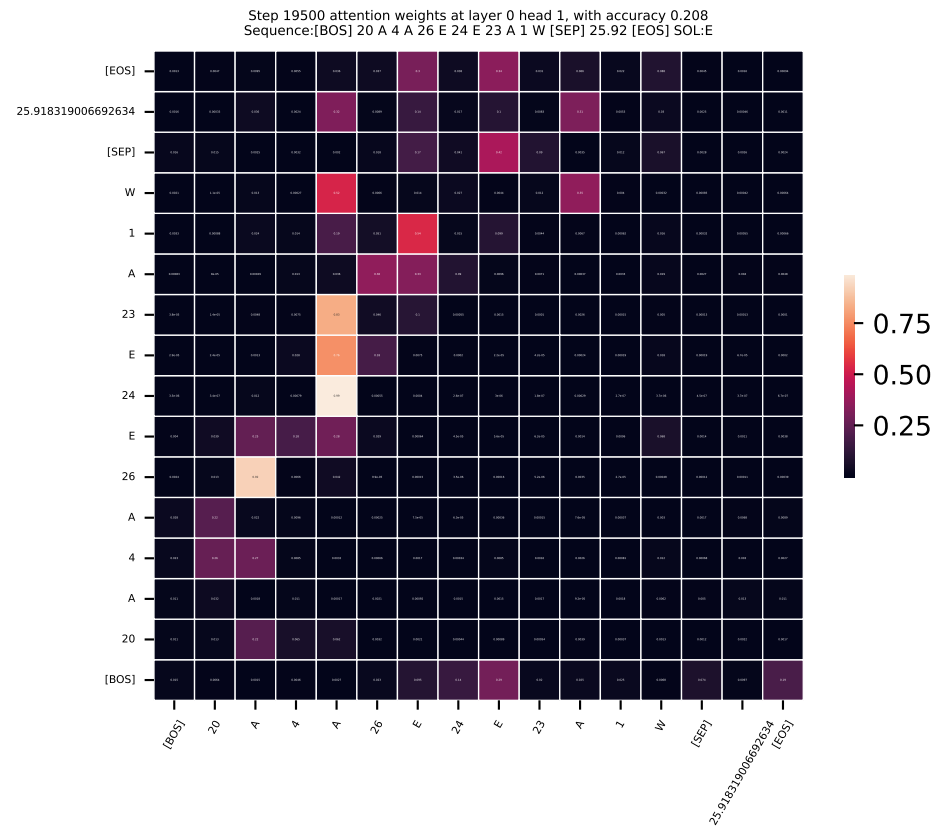

Step 10000 attention weights at layer 1 head 0, with accuracy 0.216
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.075
0.050
0.025

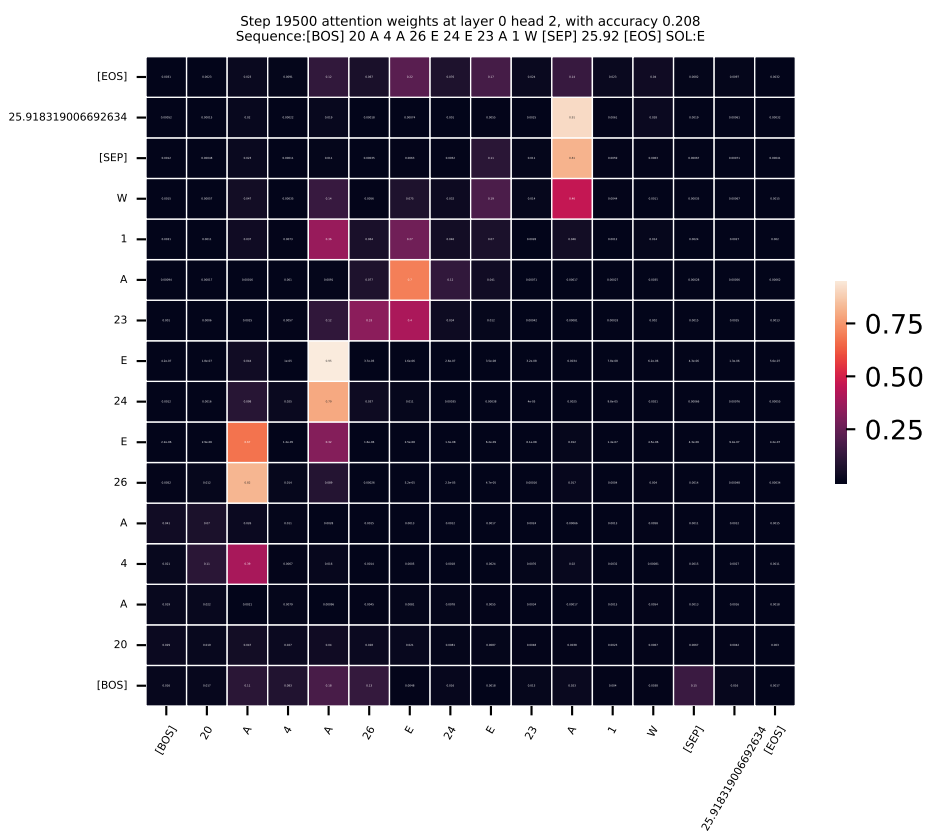

Step 10000 attention weights at layer 1 head 1, with accuracy 0.216
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.2
0.1

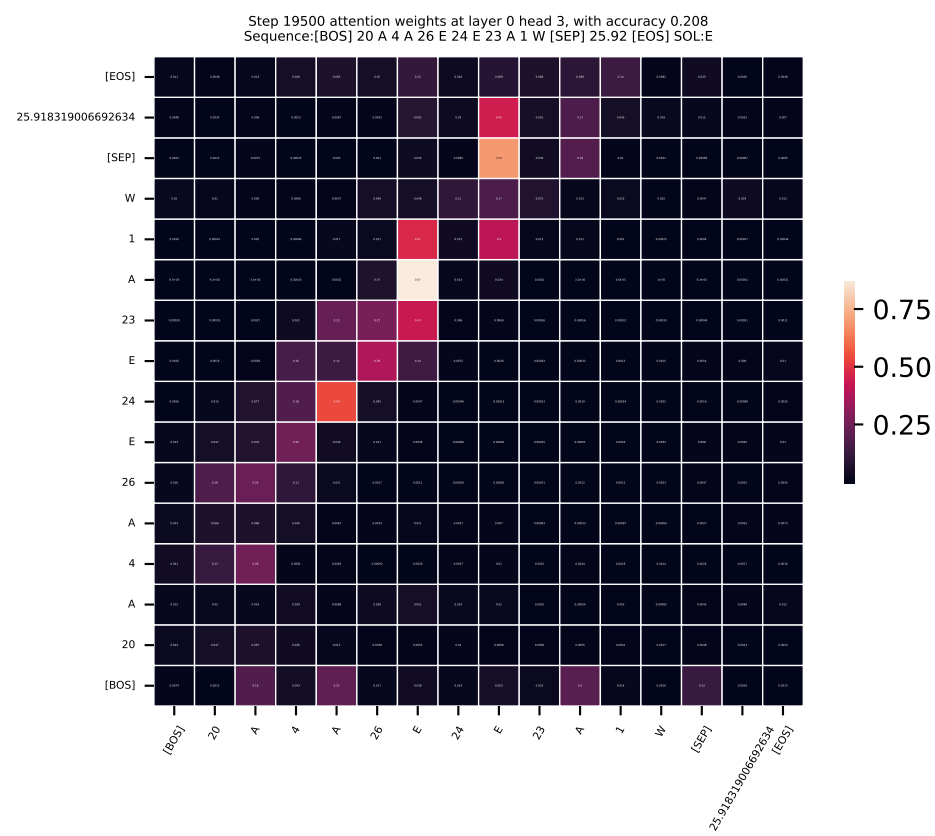

Step 10000 attention weights at layer 1 head 2, with accuracy 0.216
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.3
0.2
0.1

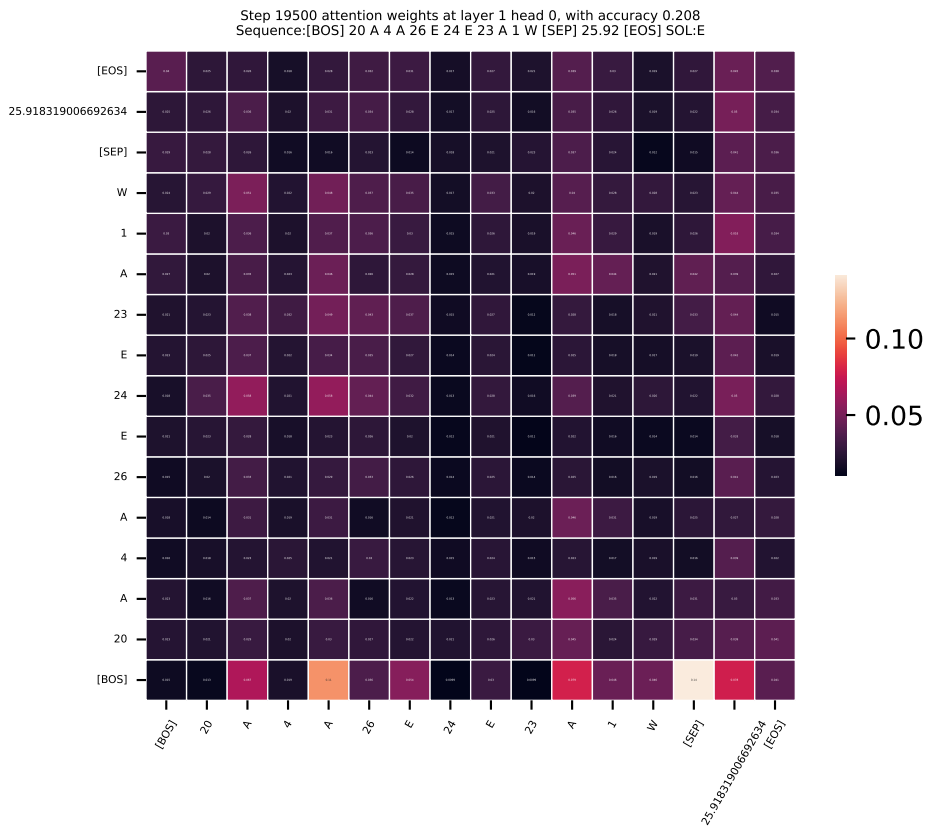

Step 10000 attention weights at layer 1 head 3, with accuracy 0.216
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

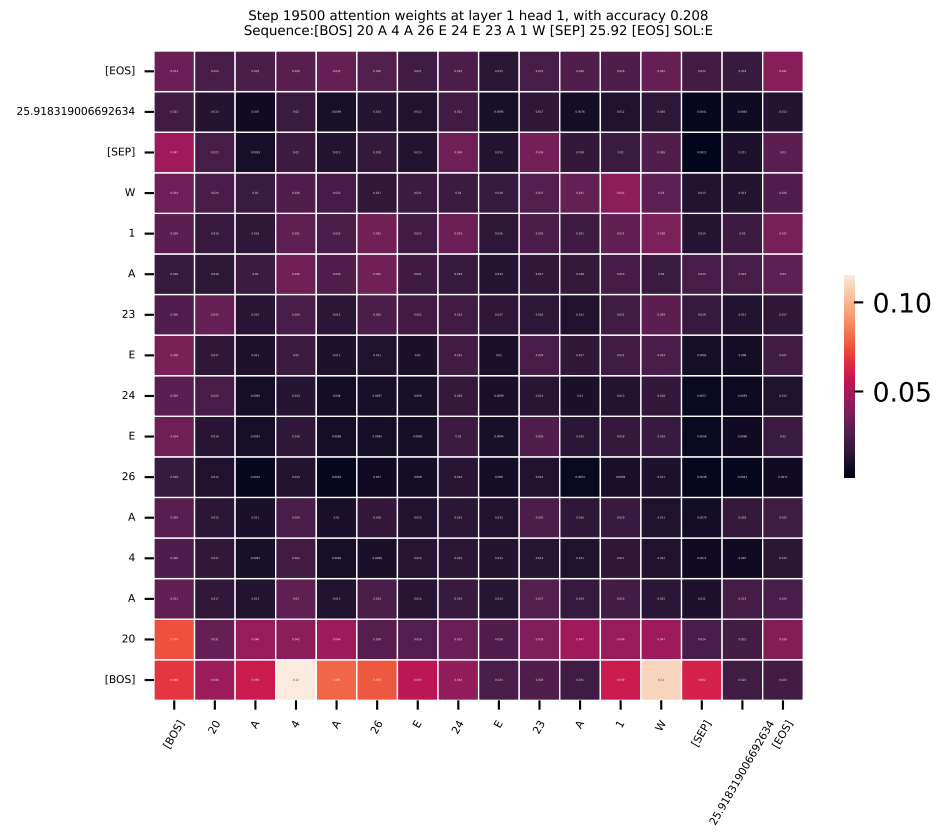

Step 19500 attention weights at layer 0 head 0, with accuracy 0.208
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

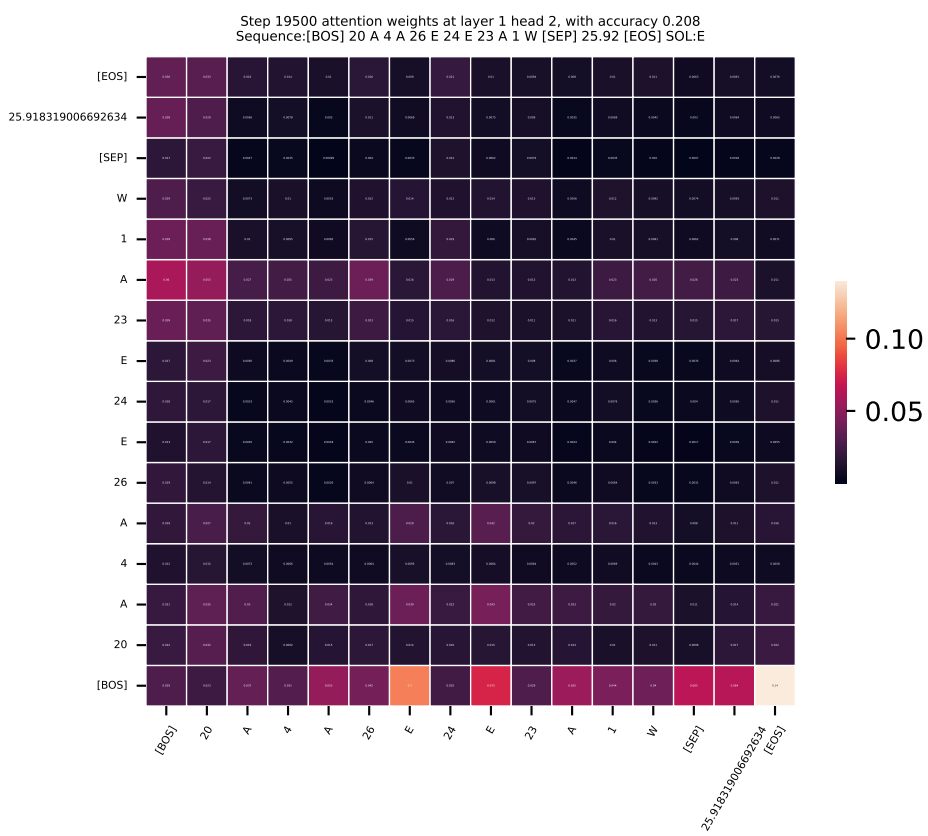

Step 19500 attention weights at layer 0 head 1, with accuracy 0.208
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

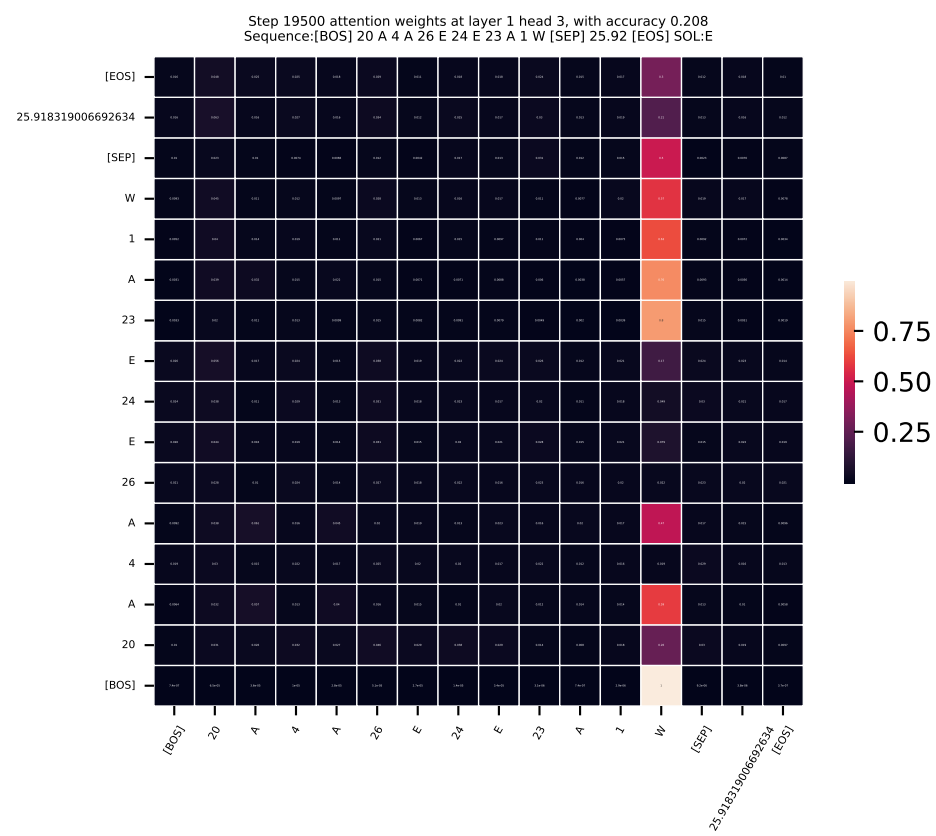

Step 19500 attention weights at layer 0 head 2, with accuracy 0.208
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

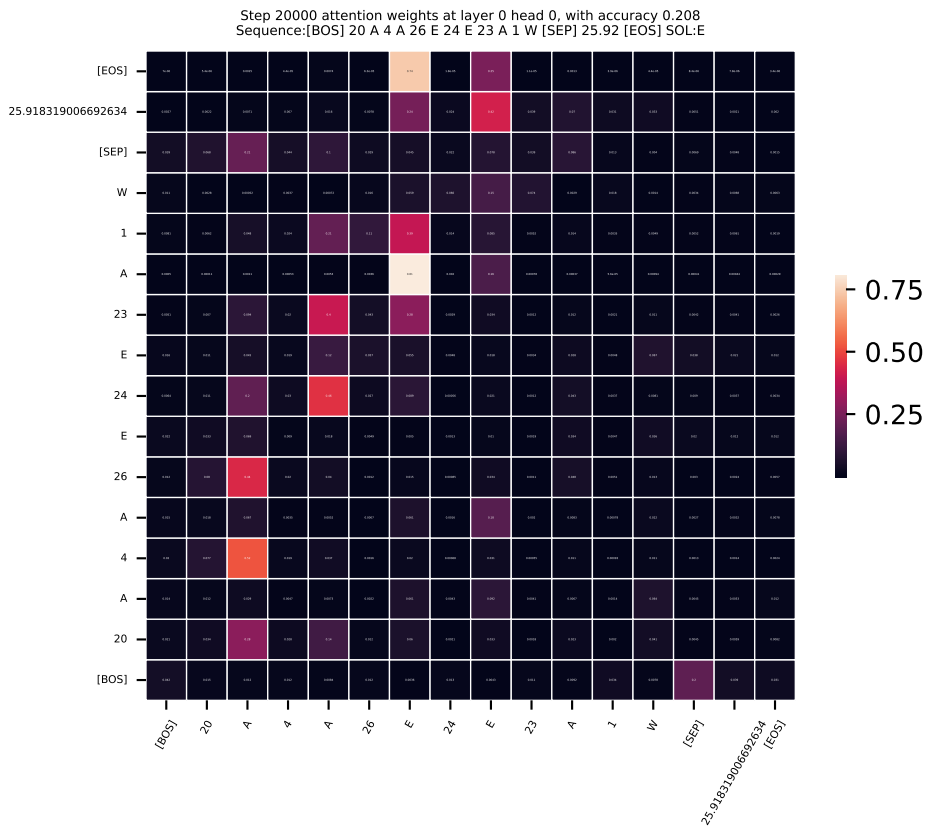

Step 19500 attention weights at layer 0 head 3, with accuracy 0.208
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

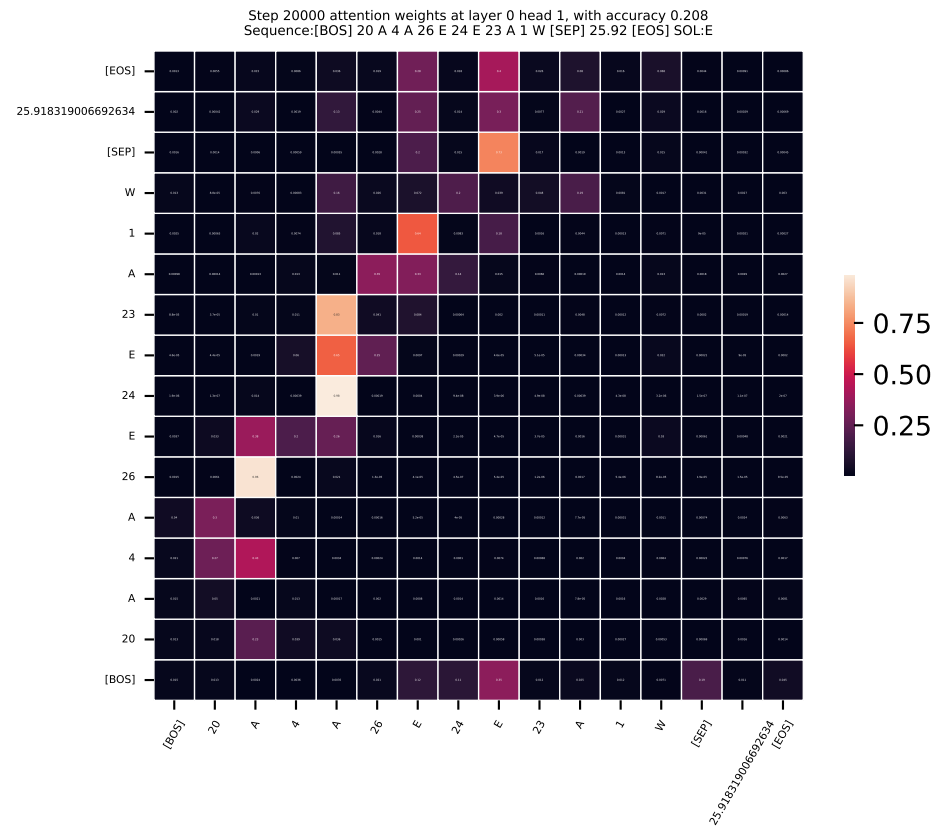

Step 19500 attention weights at layer 1 head 0, with accuracy 0.208
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.10
0.05

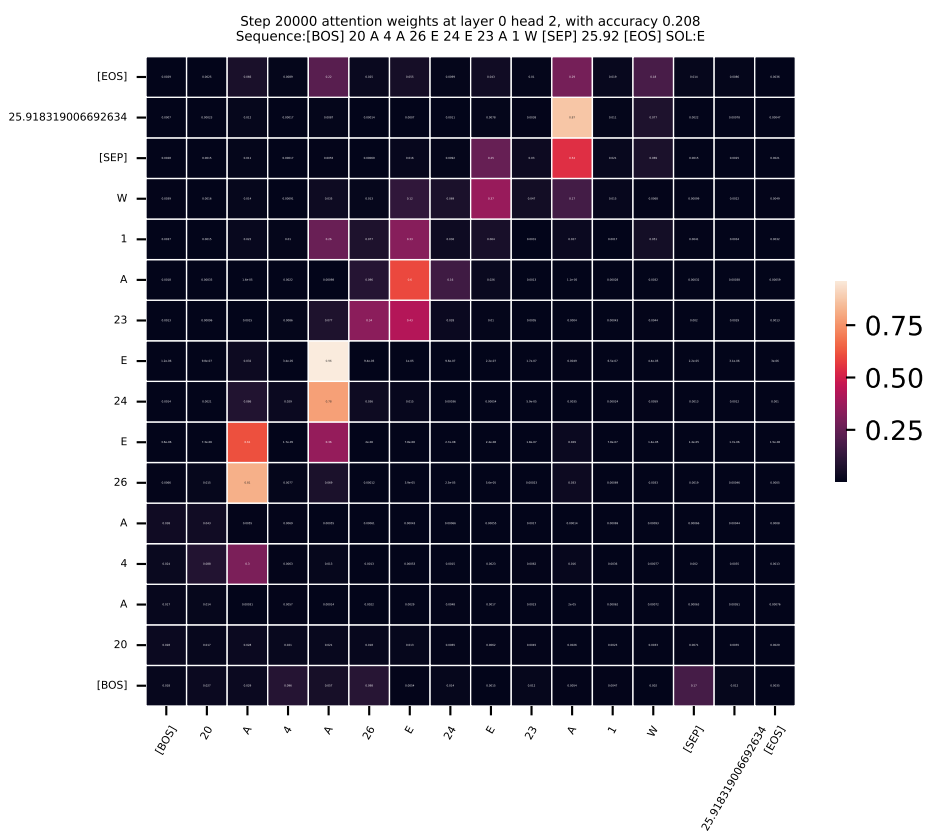

Step 19500 attention weights at layer 1 head 1, with accuracy 0.208
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.10
0.05

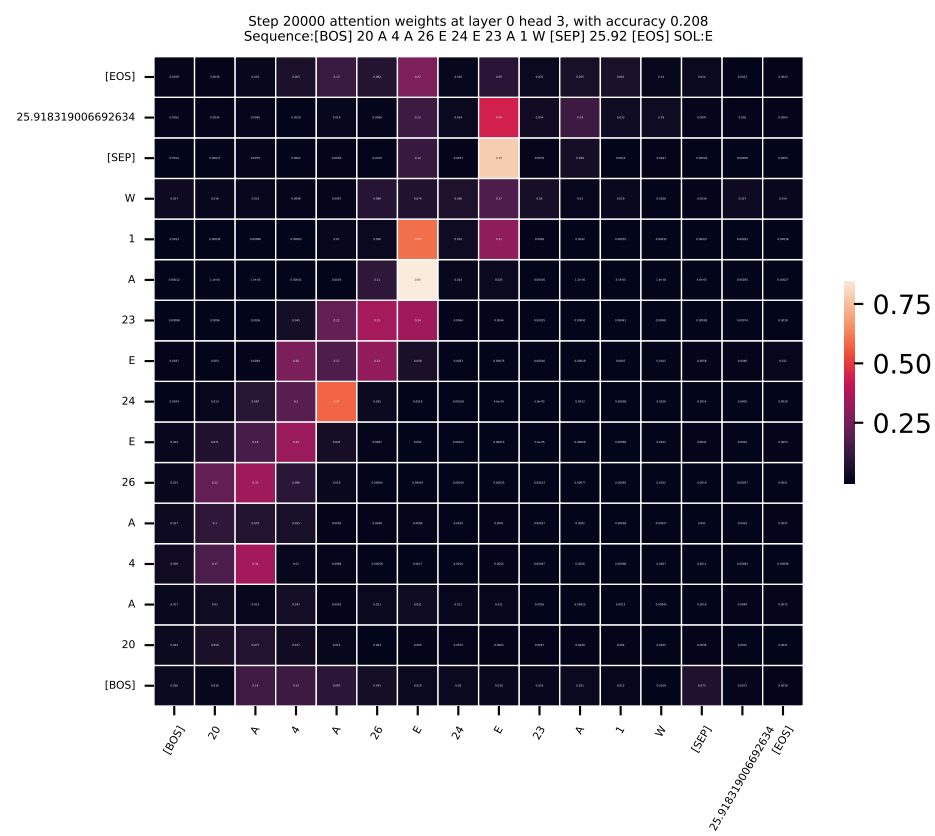

Step 19500 attention weights at layer 1 head 2, with accuracy 0.208
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.10
0.05

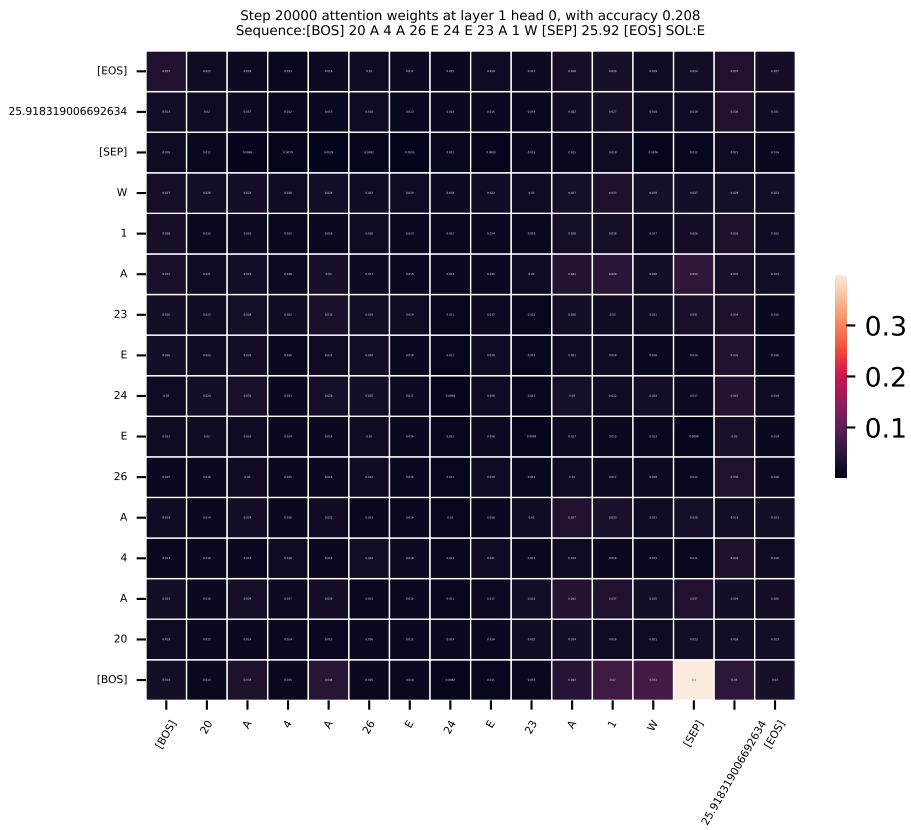

Step 19500 attention weights at layer 1 head 3, with accuracy 0.208
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

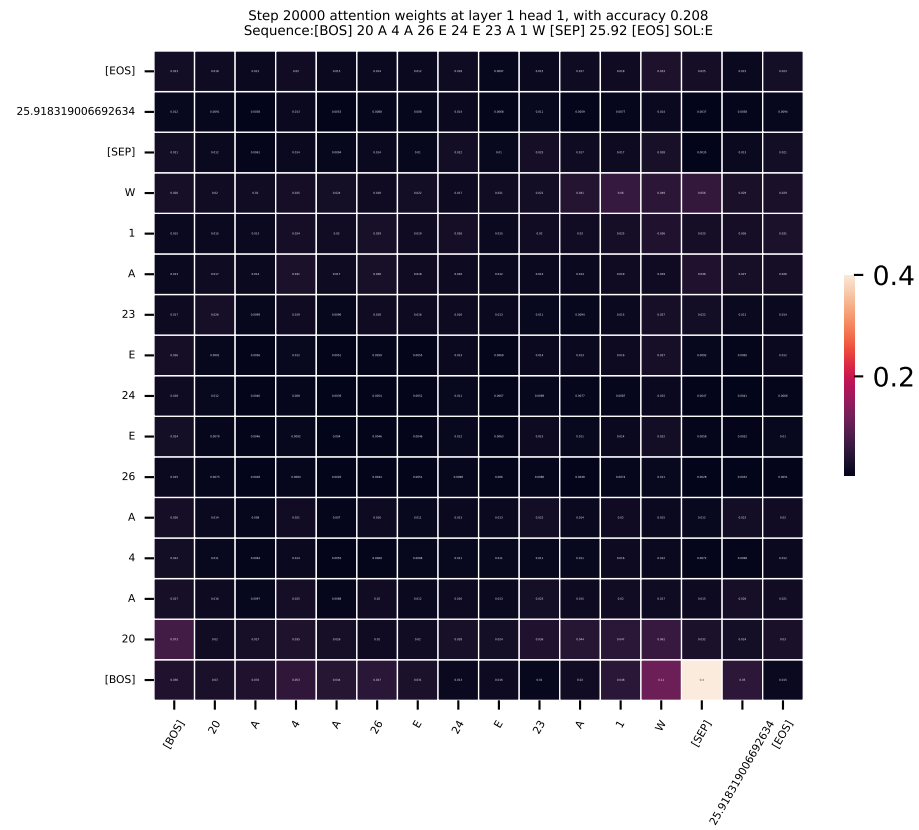

Step 20000 attention weights at layer 0 head 0, with accuracy 0.208
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

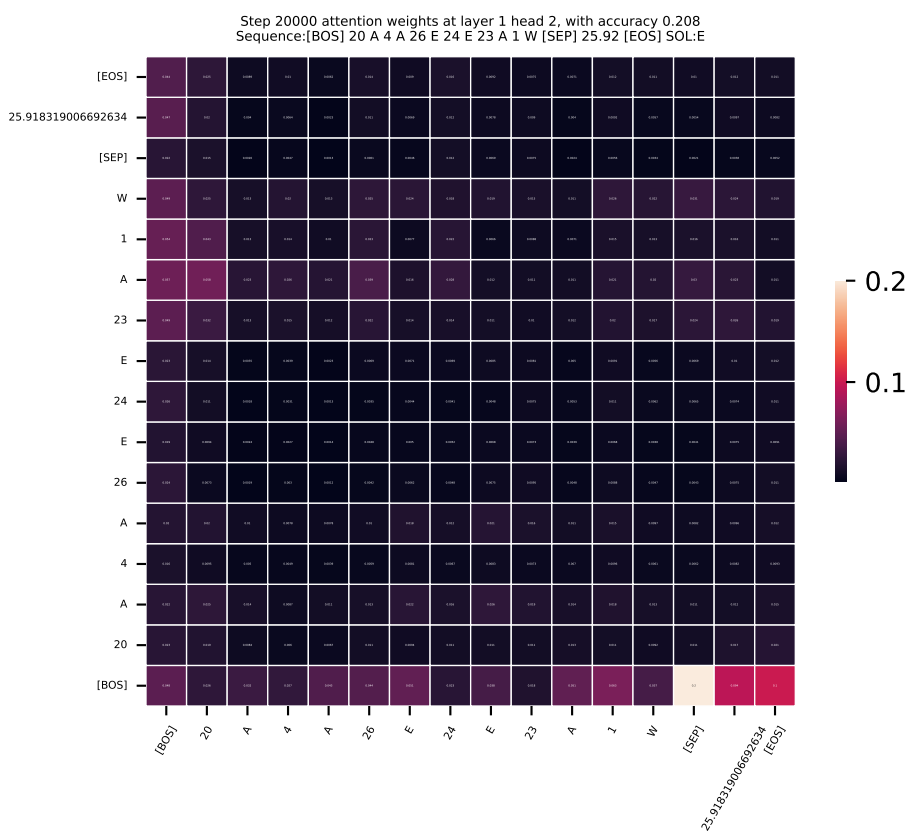

Step 20000 attention weights at layer 0 head 1, with accuracy 0.208
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

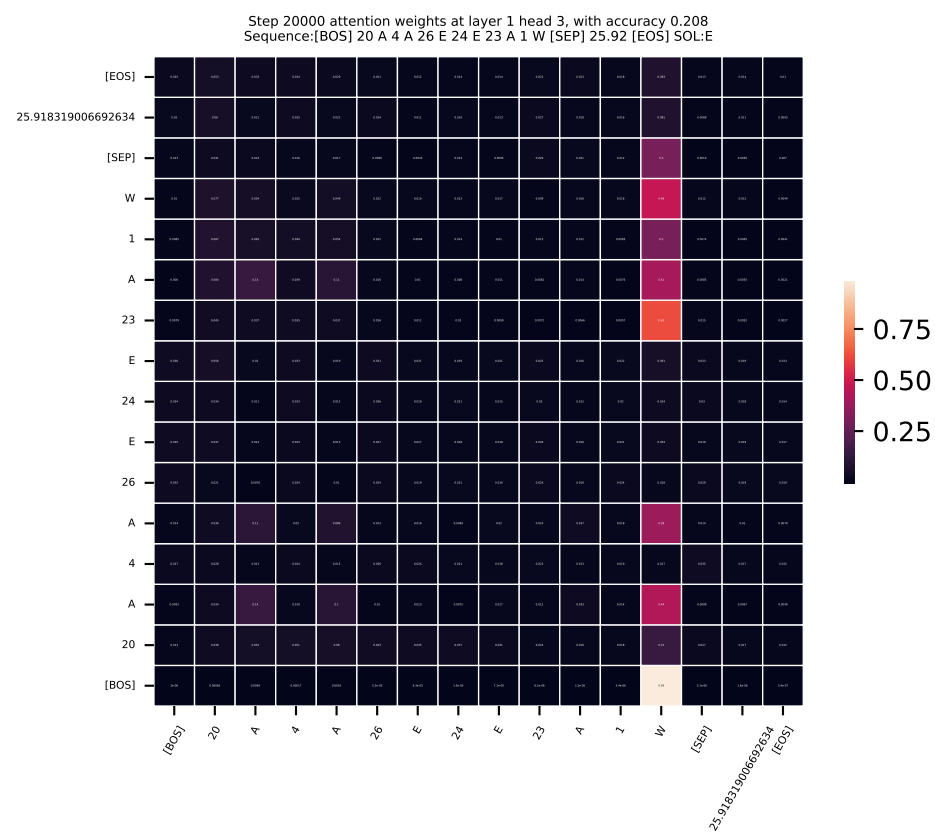

Step 20000 attention weights at layer 0 head 2, with accuracy 0.208
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

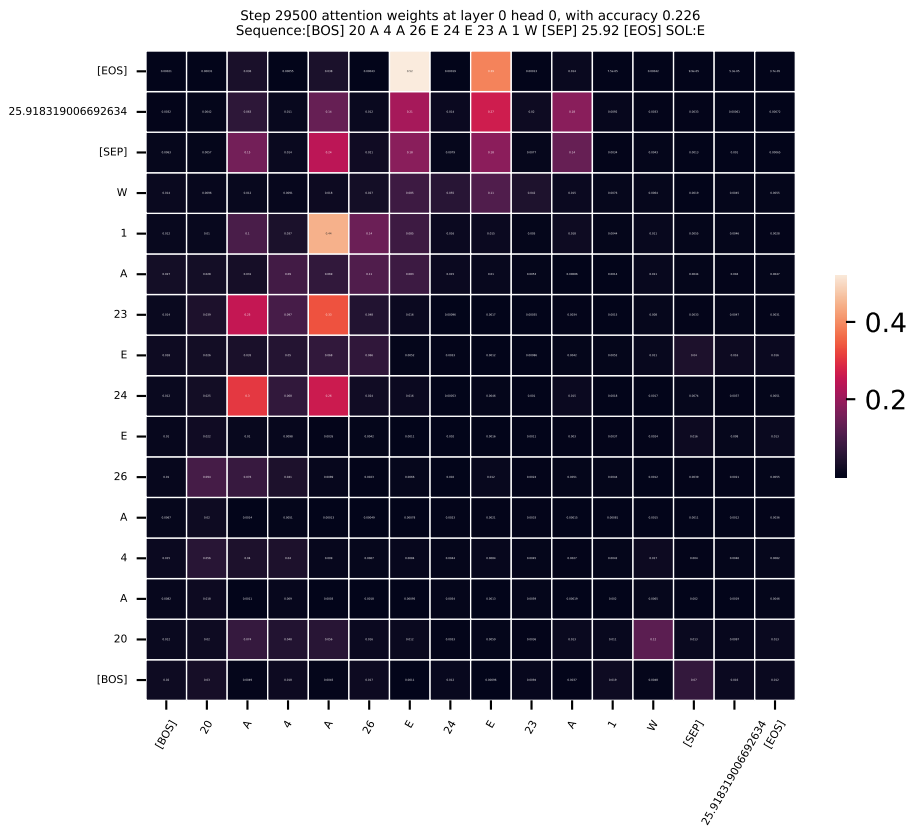

Step 20000 attention weights at layer 0 head 3, with accuracy 0.208
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

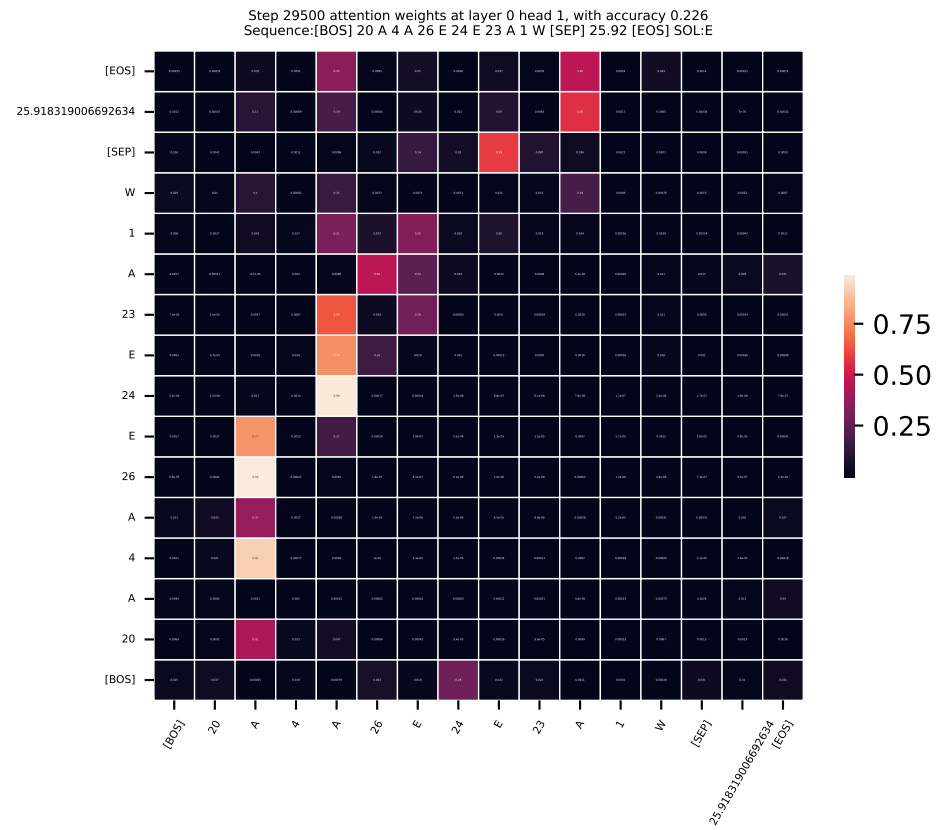

Step 20000 attention weights at layer 1 head 0, with accuracy 0.208
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

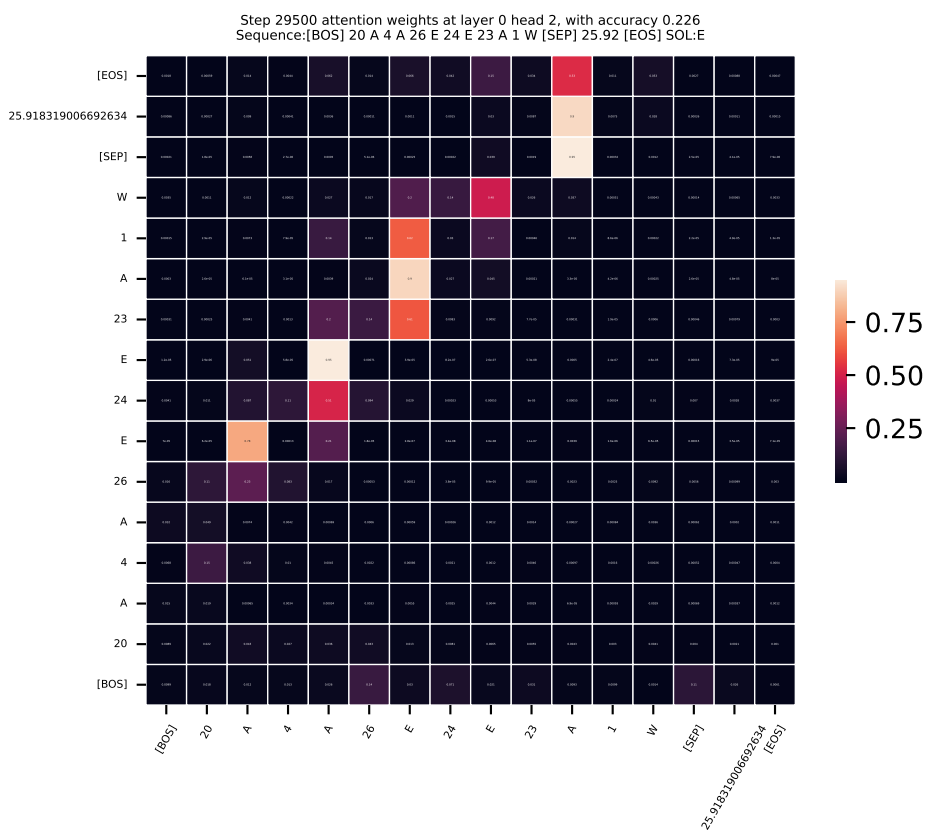

Step 20000 attention weights at layer 1 head 1, with accuracy 0.208
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

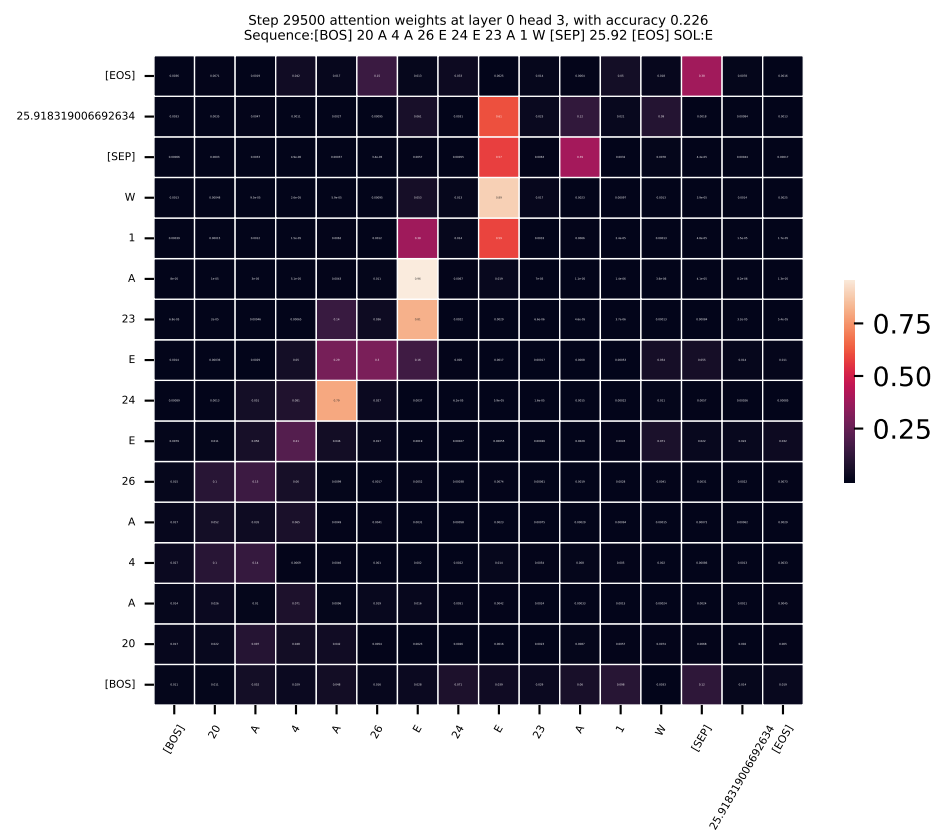

Step 20000 attention weights at layer 1 head 2, with accuracy 0.208
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.2
0.1

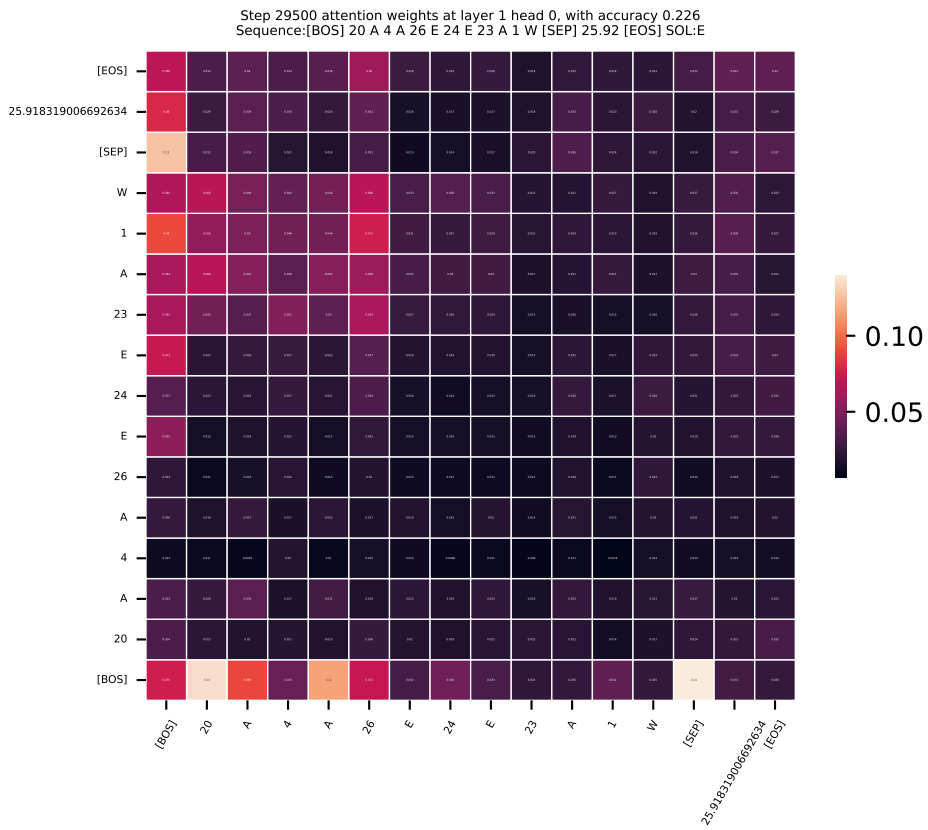

Step 20000 attention weights at layer 1 head 3, with accuracy 0.208
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

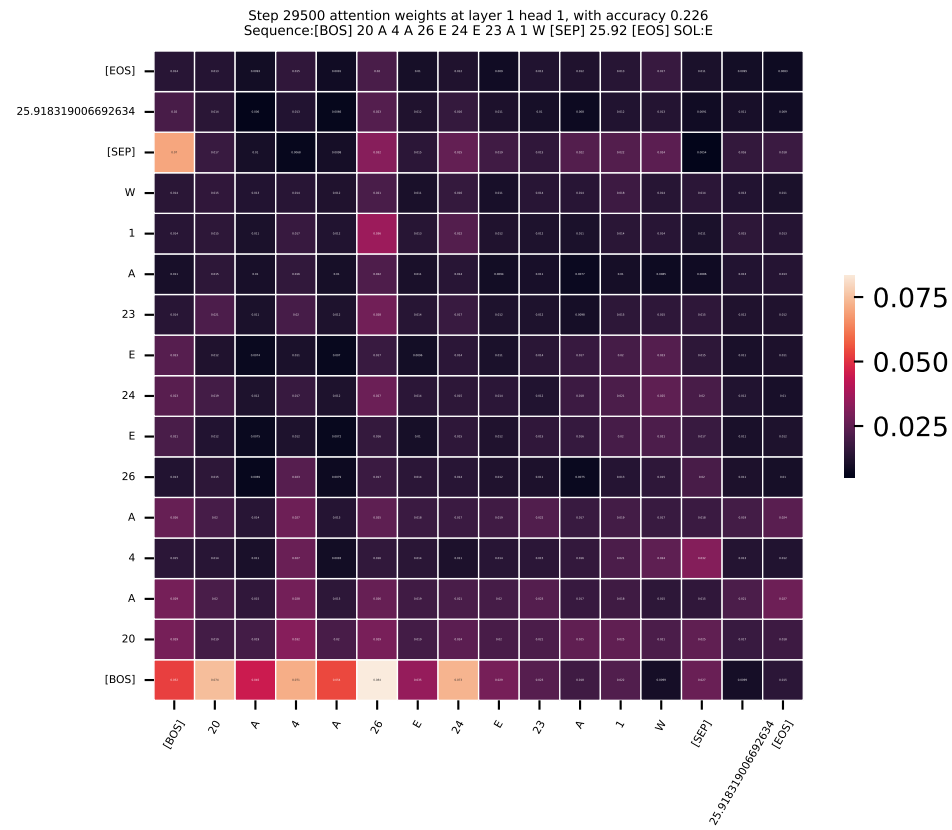

Step 29500 attention weights at layer 0 head 0, with accuracy 0.226
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

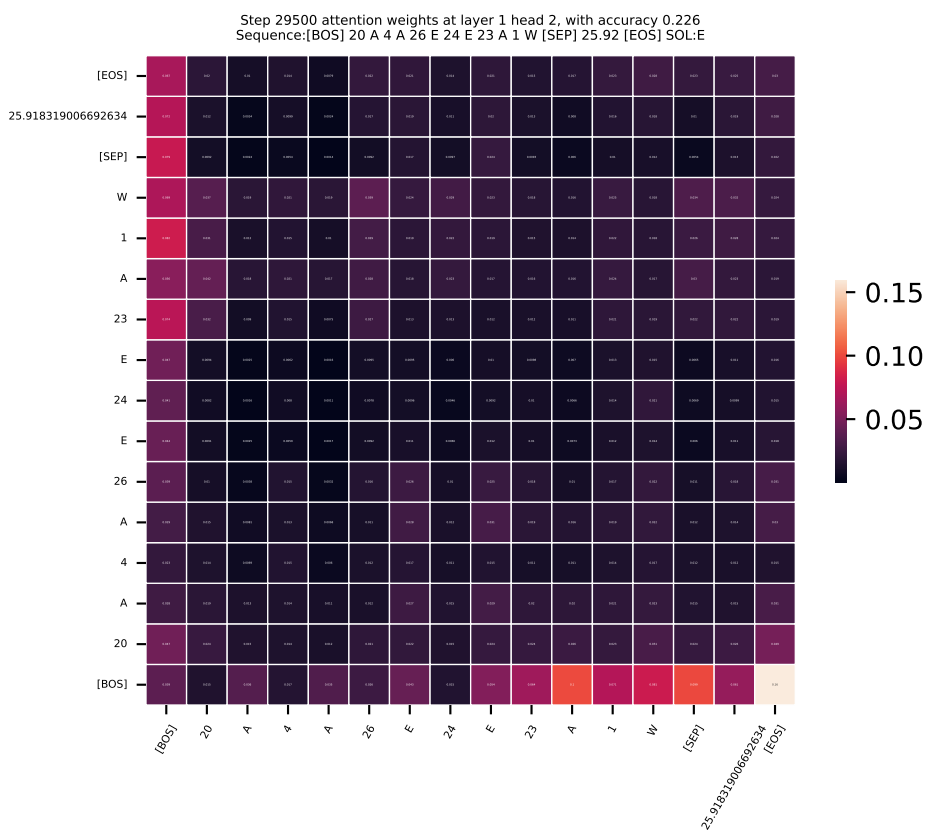

Step 29500 attention weights at layer 0 head 1, with accuracy 0.226
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

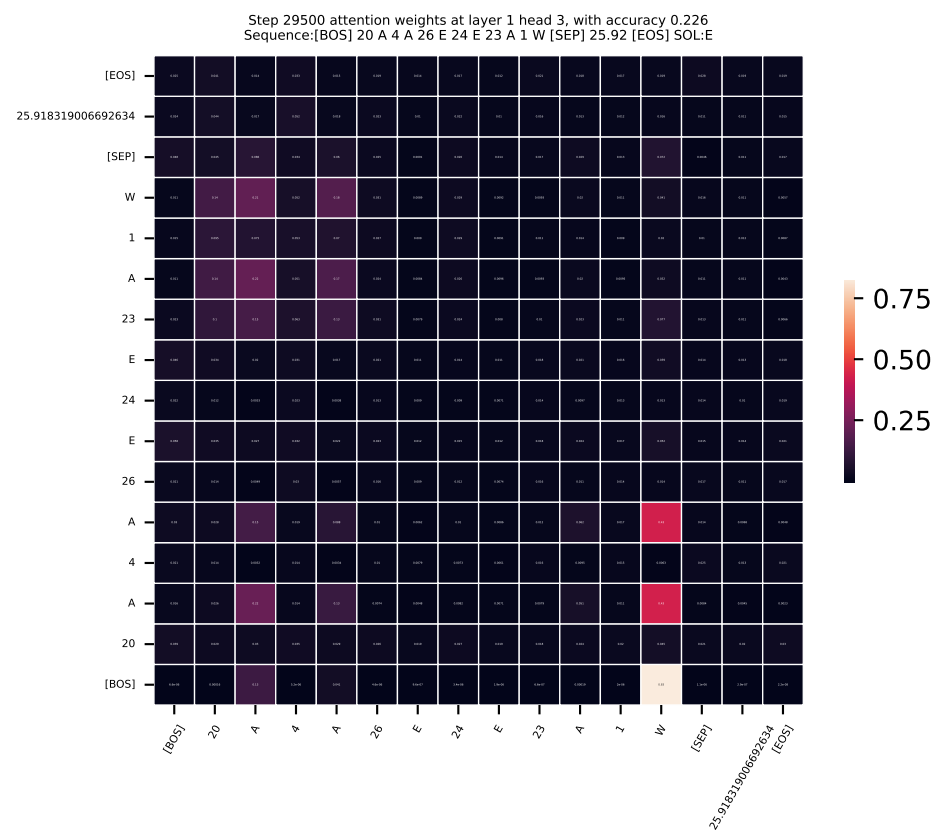

Step 29500 attention weights at layer 0 head 2, with accuracy 0.226
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

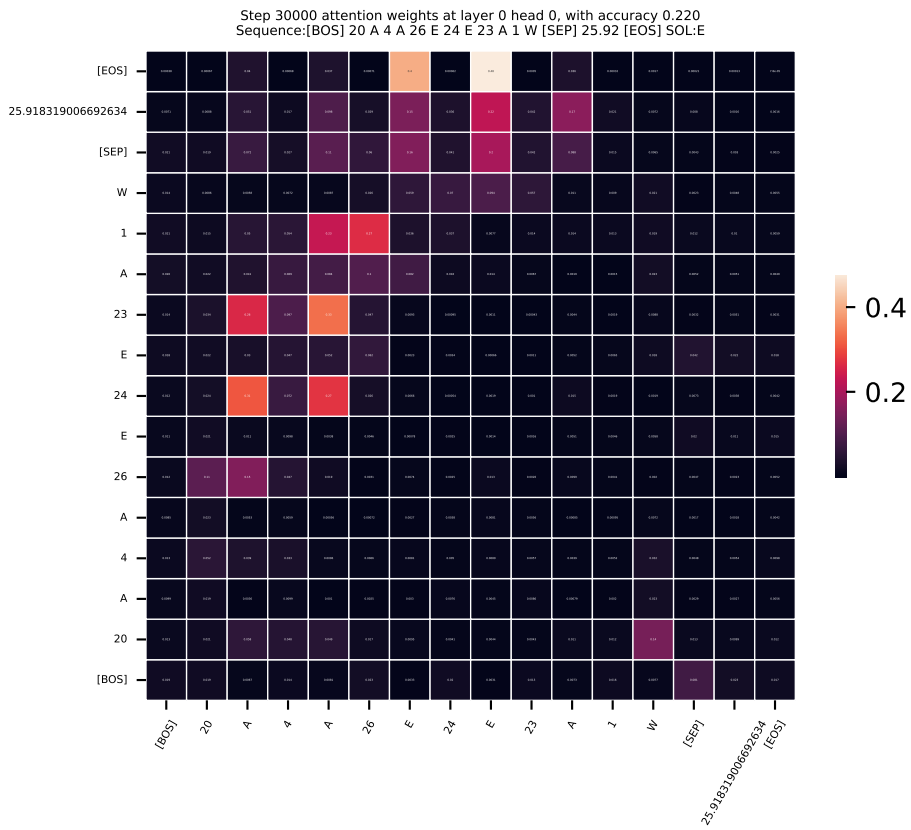

Step 29500 attention weights at layer 0 head 3, with accuracy 0.226
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

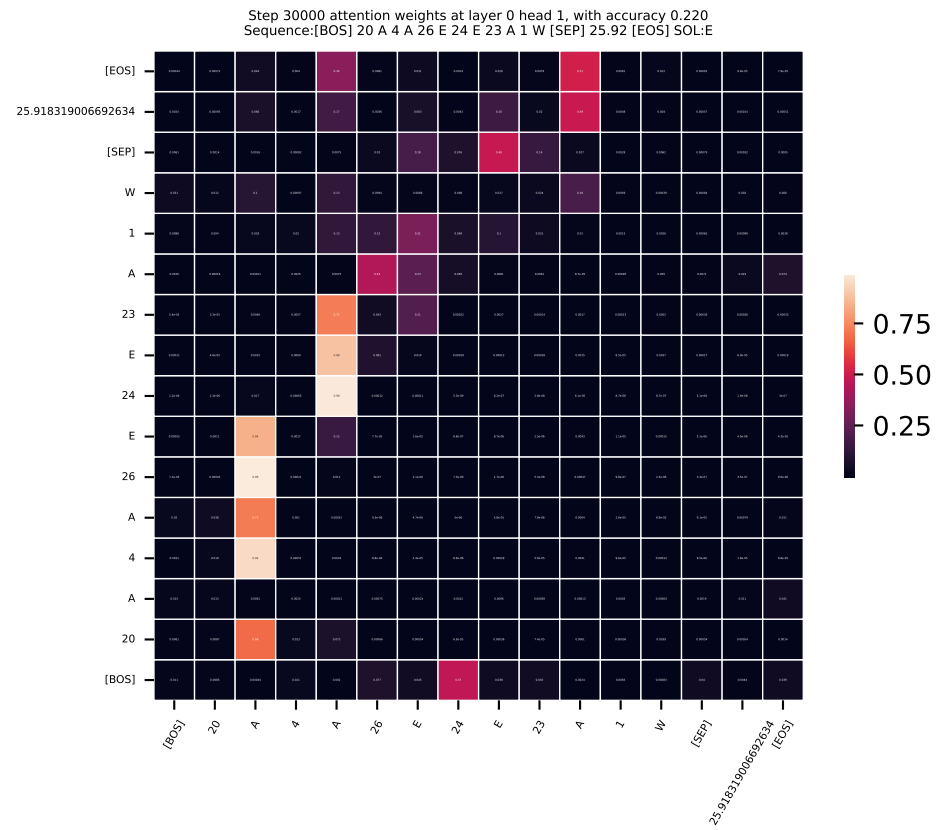

Step 29500 attention weights at layer 1 head 0, with accuracy 0.226
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.10
0.05

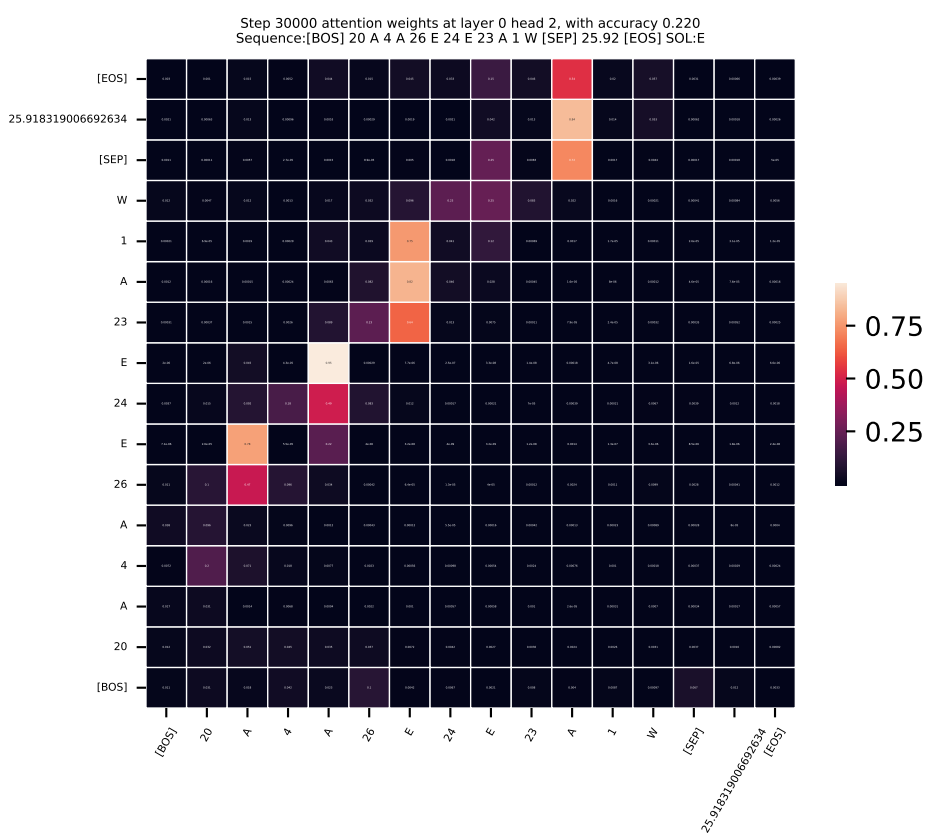

Step 29500 attention weights at layer 1 head 1, with accuracy 0.226
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.075
0.050
0.025

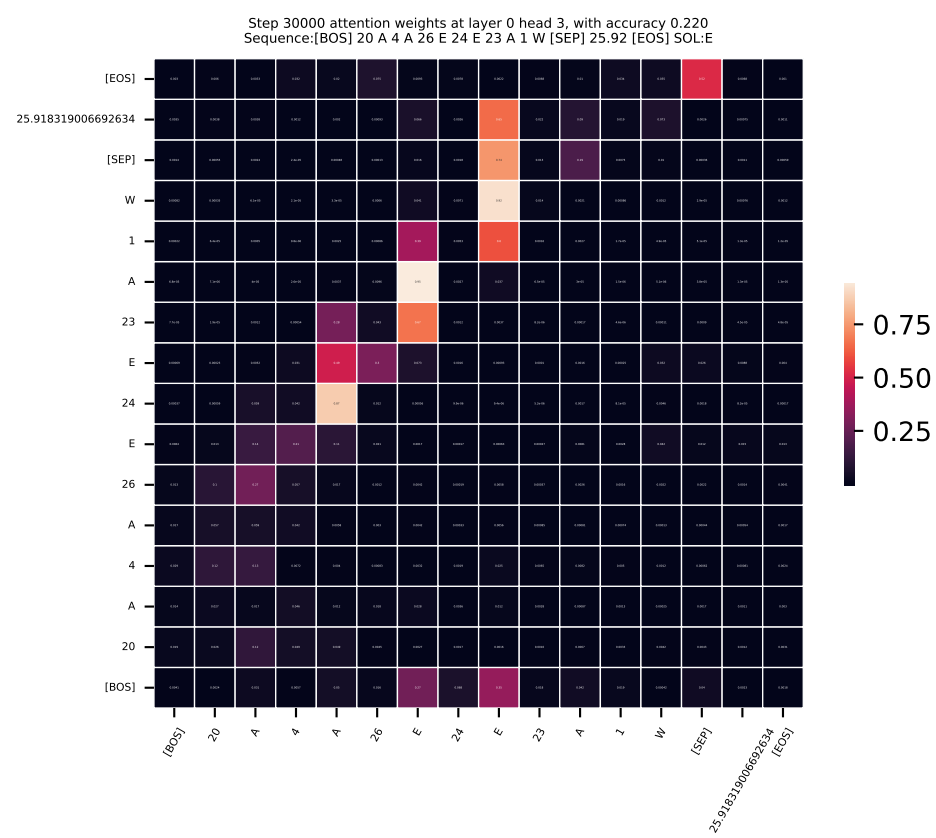

Step 29500 attention weights at layer 1 head 2, with accuracy 0.226
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.15
0.10
0.05

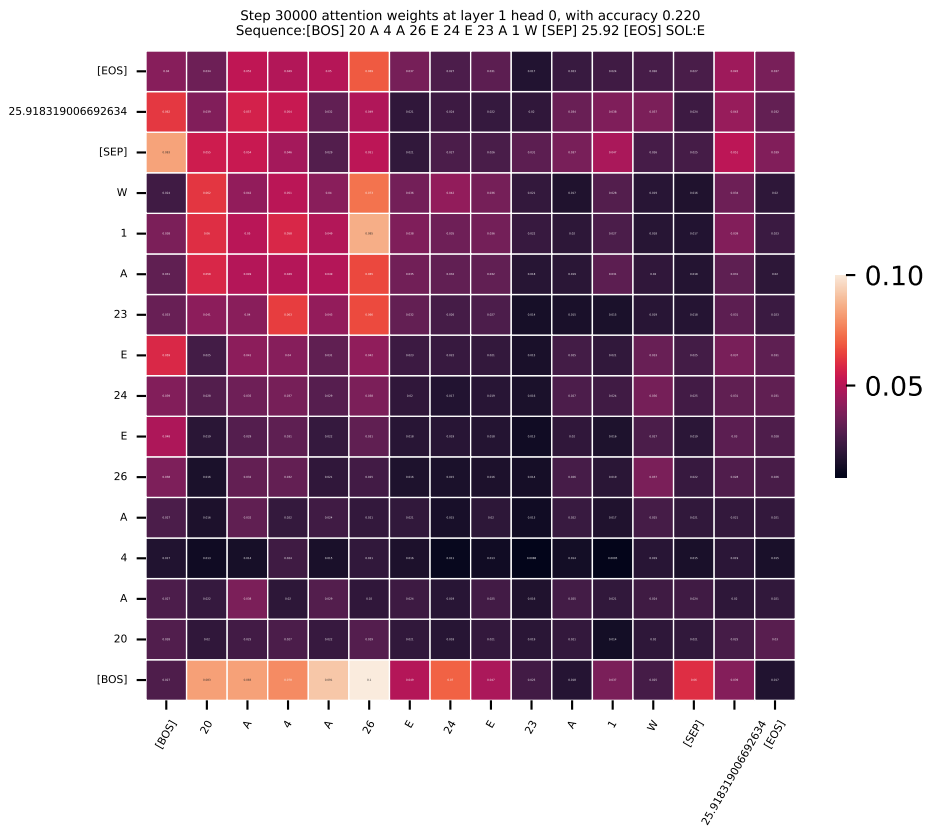

Step 29500 attention weights at layer 1 head 3, with accuracy 0.226
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

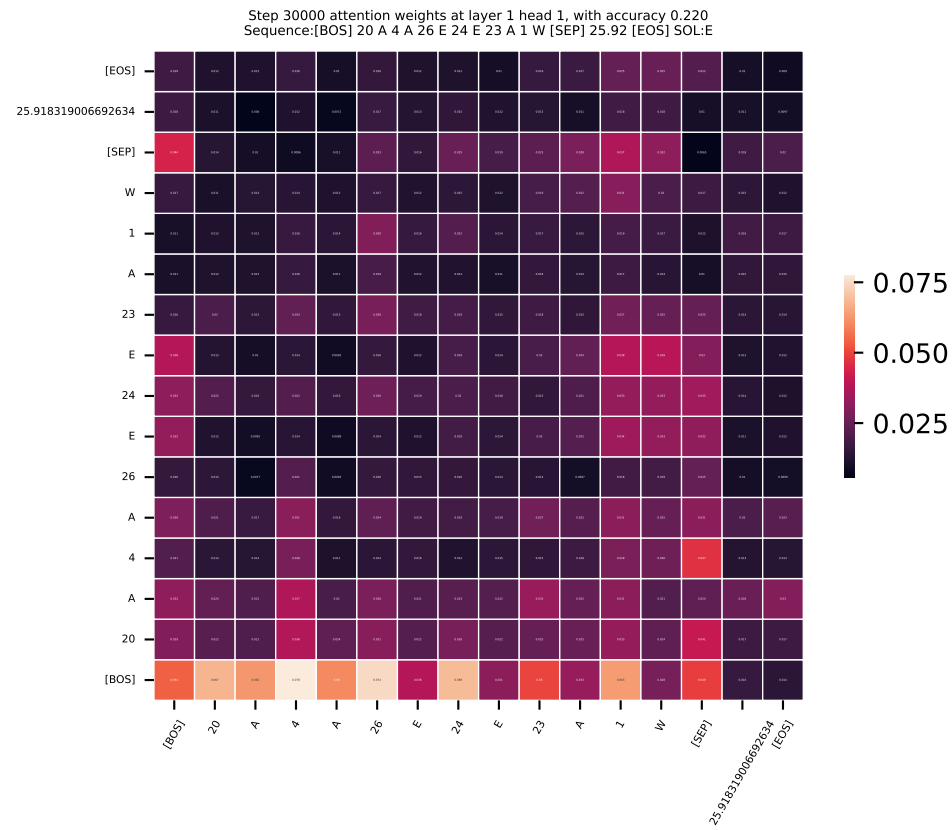

Step 30000 attention weights at layer 0 head 0, with accuracy 0.220
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.4
0.2

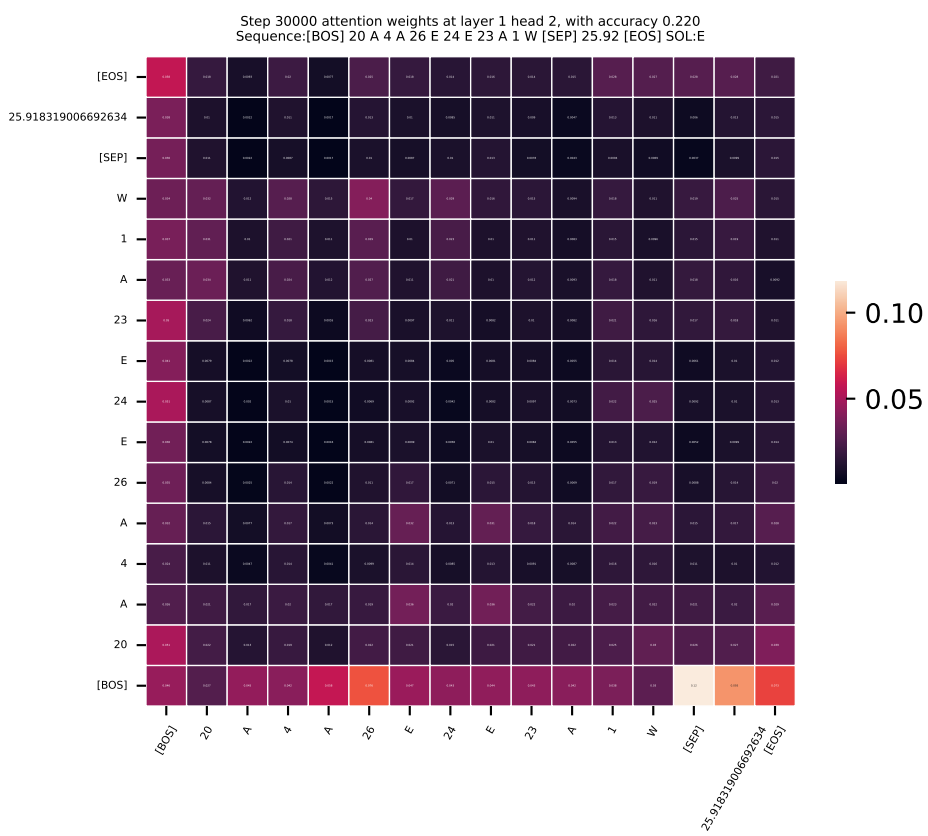

Step 30000 attention weights at layer 0 head 1, with accuracy 0.220
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

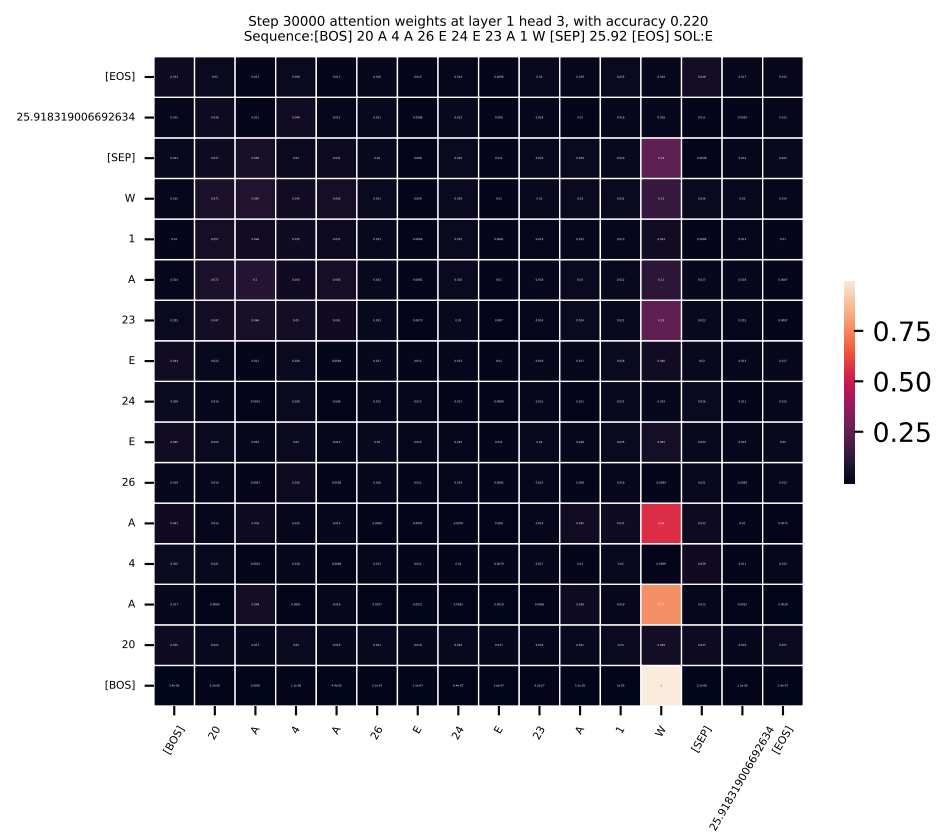

Step 30000 attention weights at layer 0 head 2, with accuracy 0.220
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

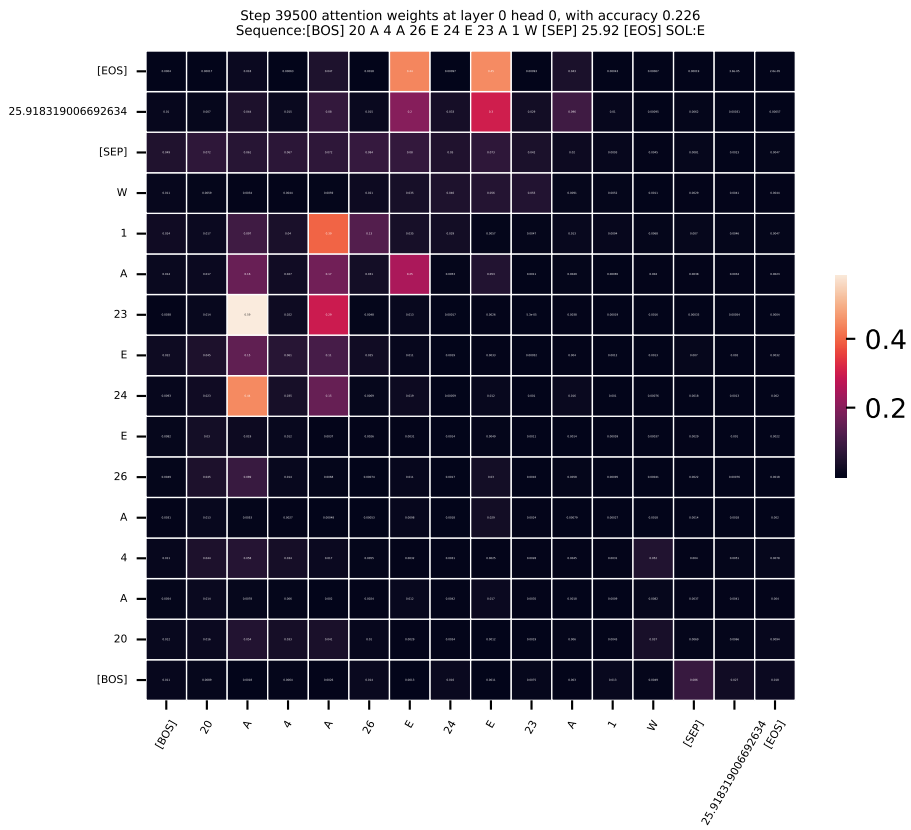

Step 30000 attention weights at layer 0 head 3, with accuracy 0.220
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

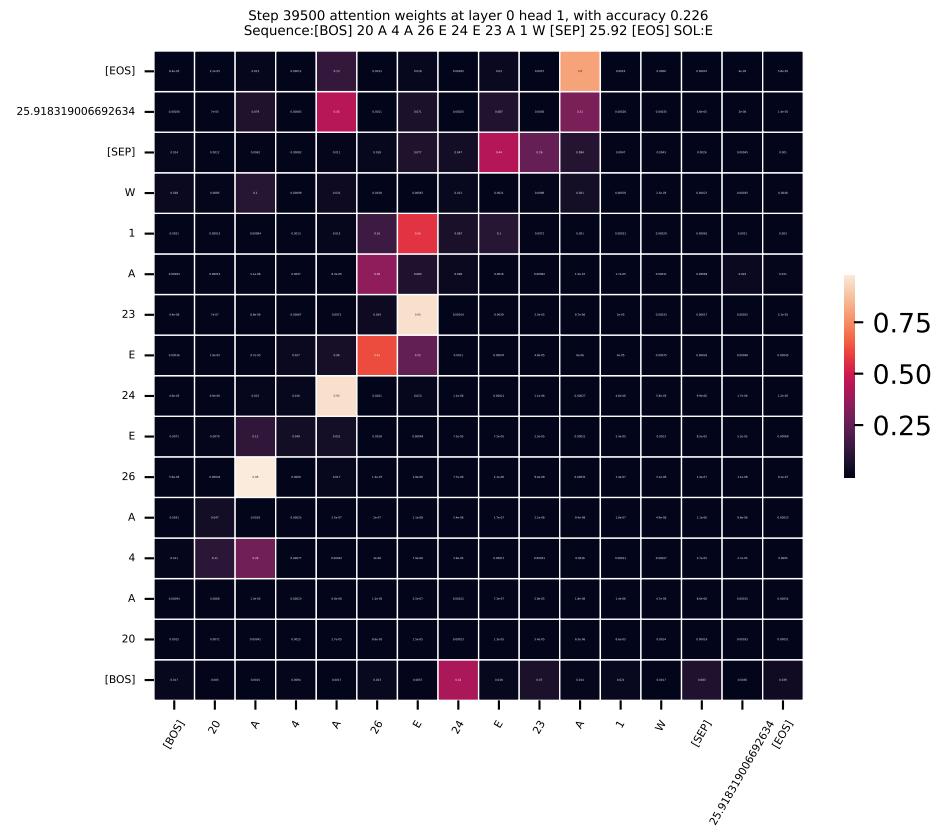

Step 30000 attention weights at layer 1 head 0, with accuracy 0.220
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.10
0.05

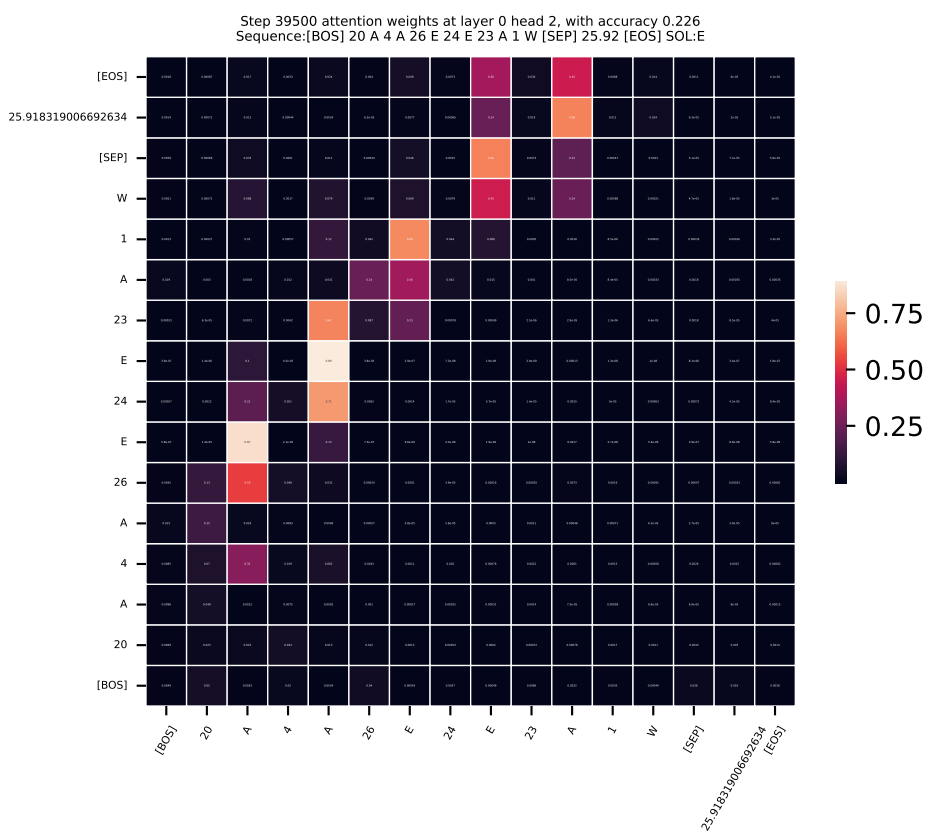

Step 30000 attention weights at layer 1 head 1, with accuracy 0.220
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.075
0.050
0.025

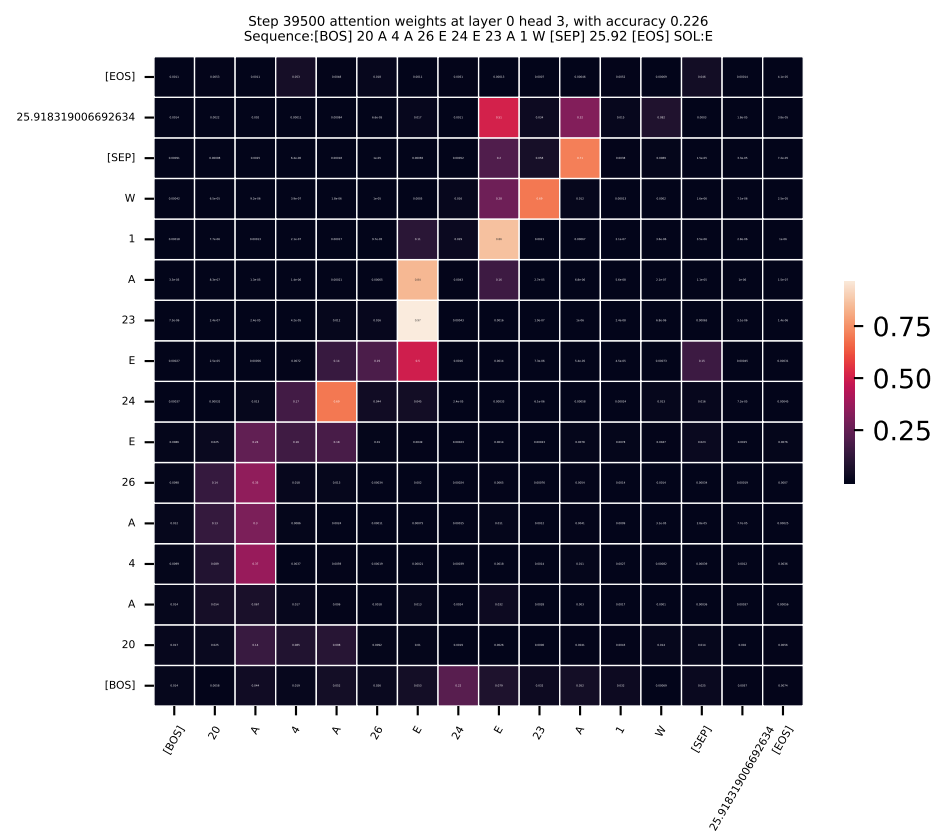

Step 30000 attention weights at layer 1 head 2, with accuracy 0.220
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.10
0.05

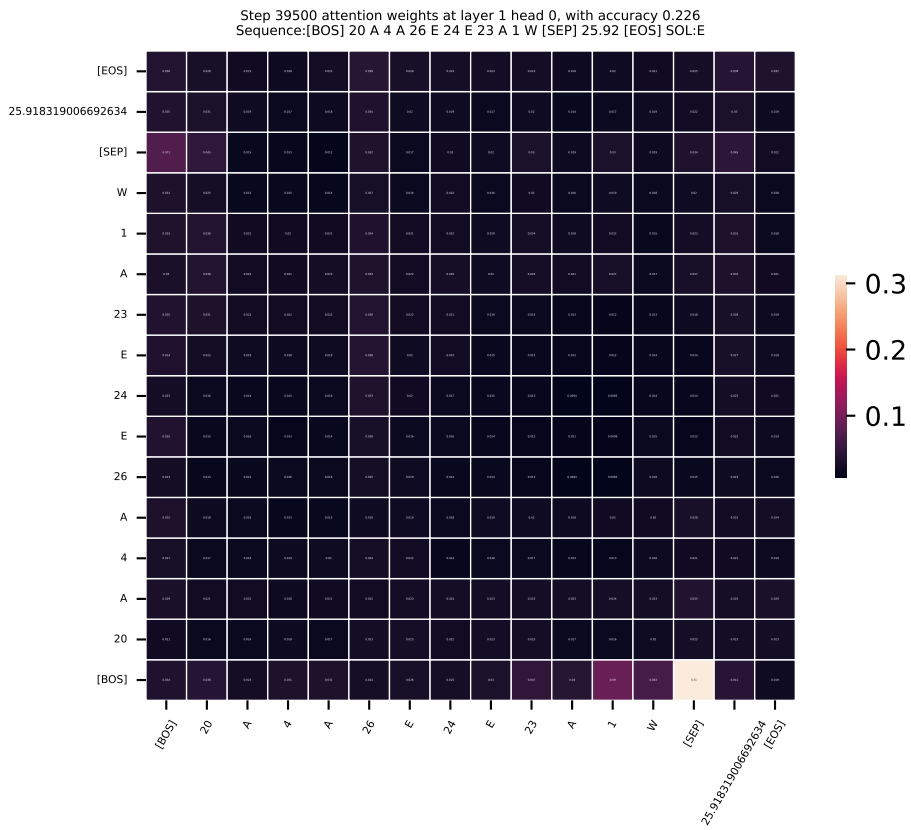

Step 30000 attention weights at layer 1 head 3, with accuracy 0.220
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

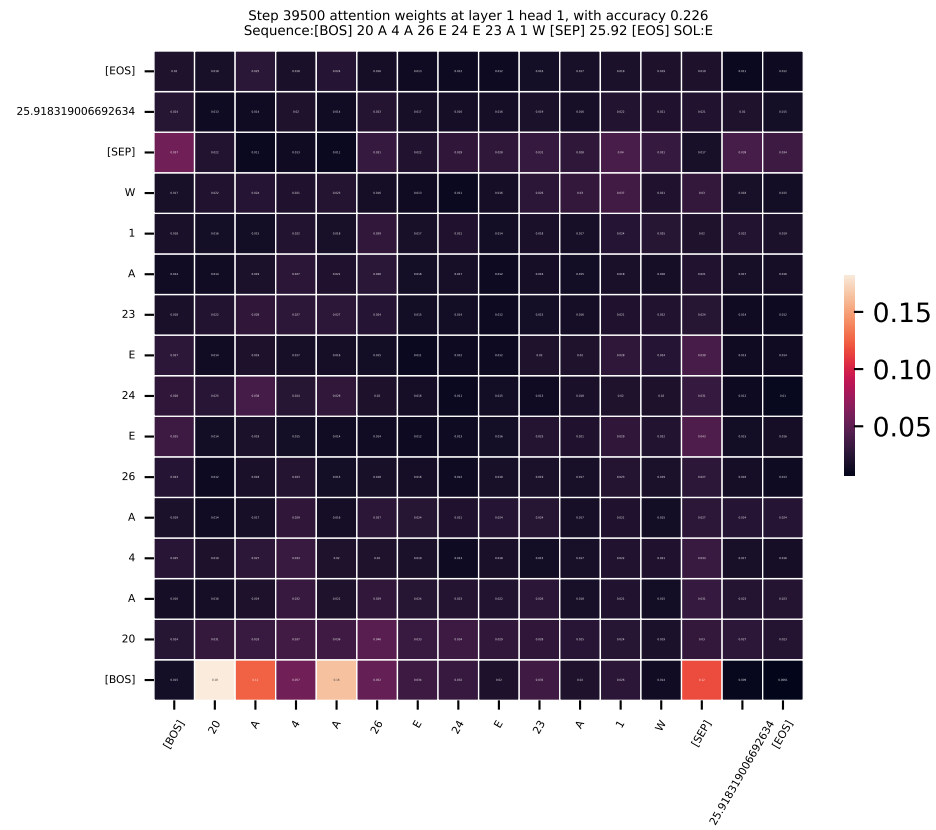

Step 39500 attention weights at layer 0 head 0, with accuracy 0.226
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.4
0.2

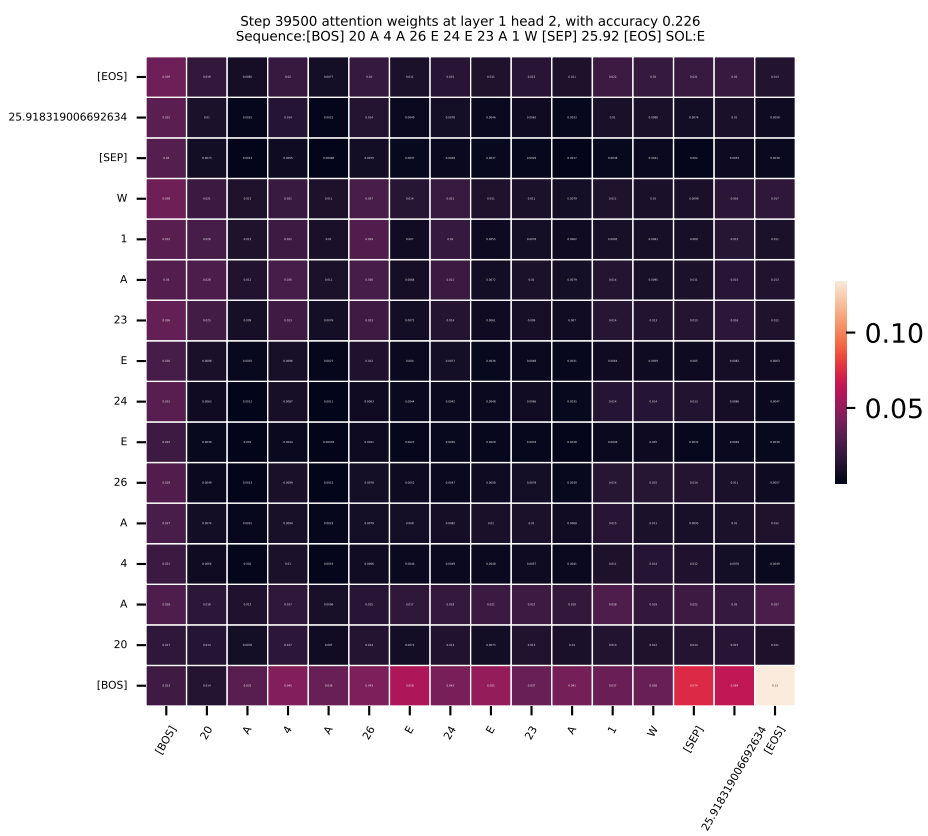

Step 39500 attention weights at layer 0 head 1, with accuracy 0.226
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

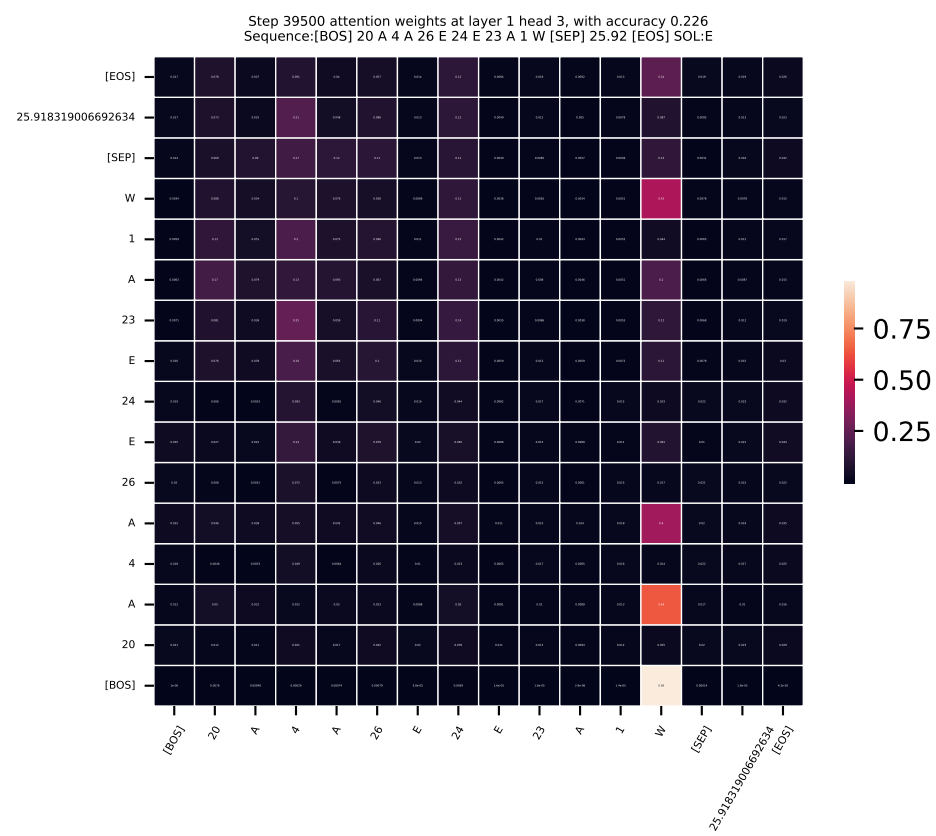

Step 39500 attention weights at layer 0 head 2, with accuracy 0.226
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

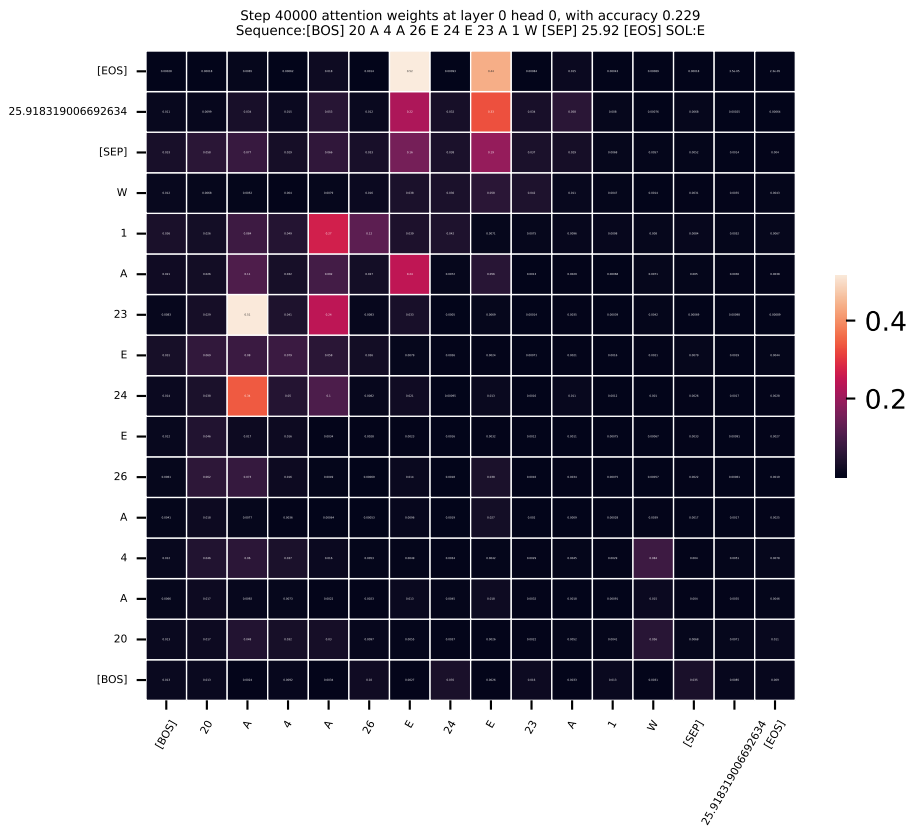

Step 39500 attention weights at layer 0 head 3, with accuracy 0.226
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

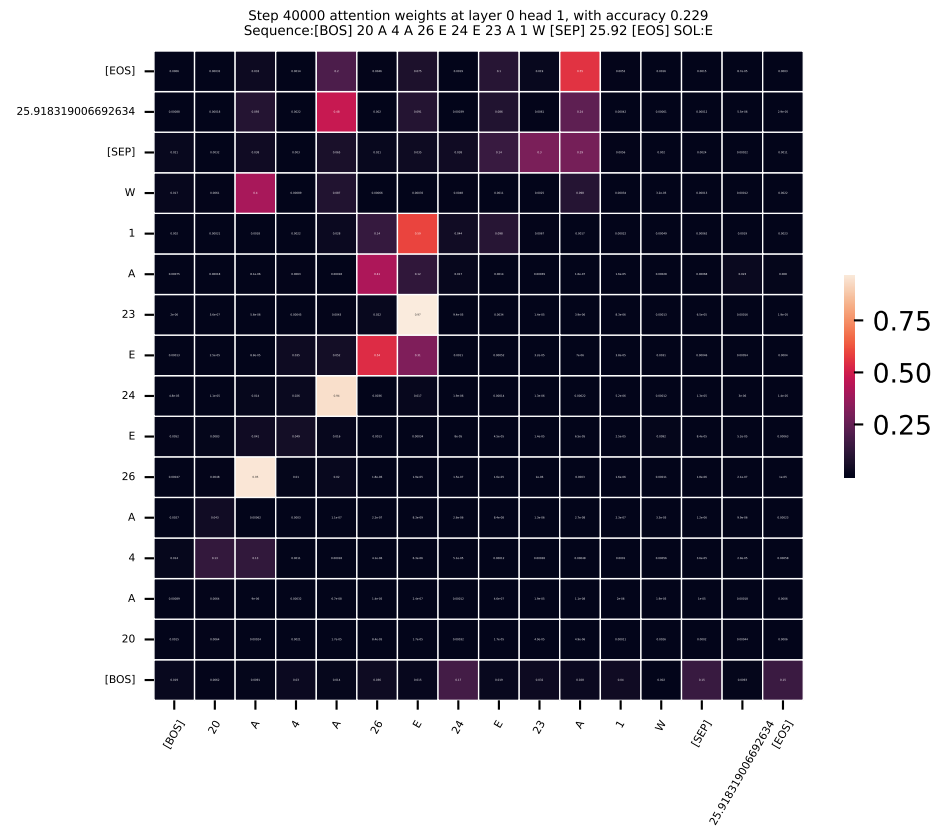

Step 39500 attention weights at layer 1 head 0, with accuracy 0.226
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.3
0.2
0.1

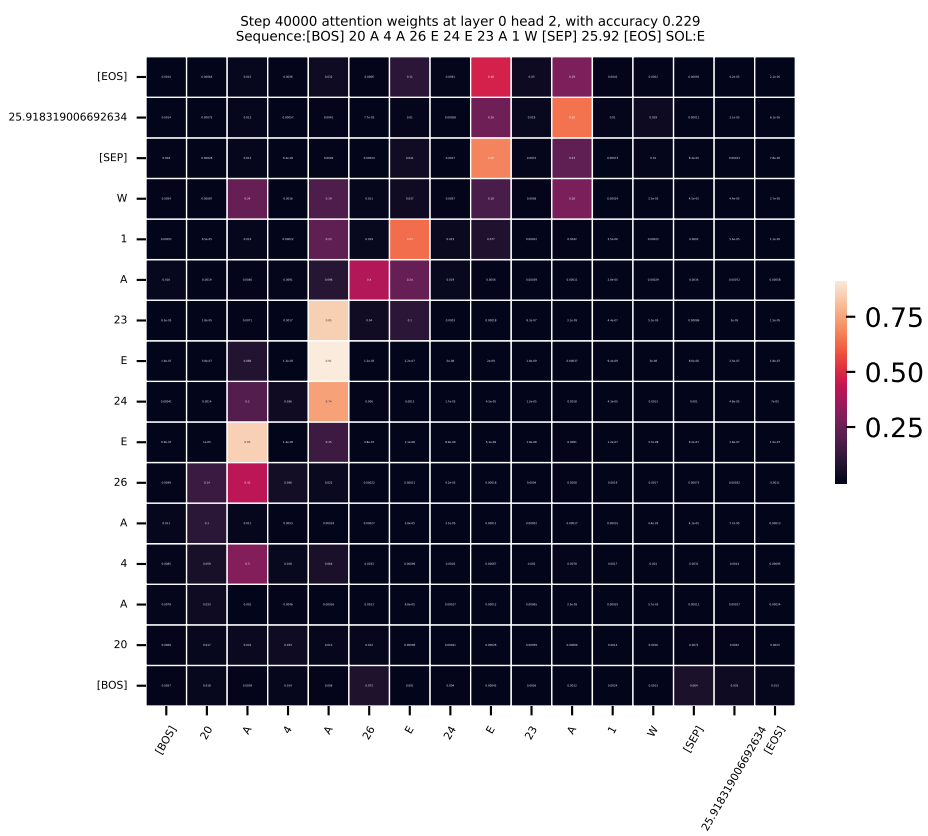

Step 39500 attention weights at layer 1 head 1, with accuracy 0.226
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.15
0.10
0.05

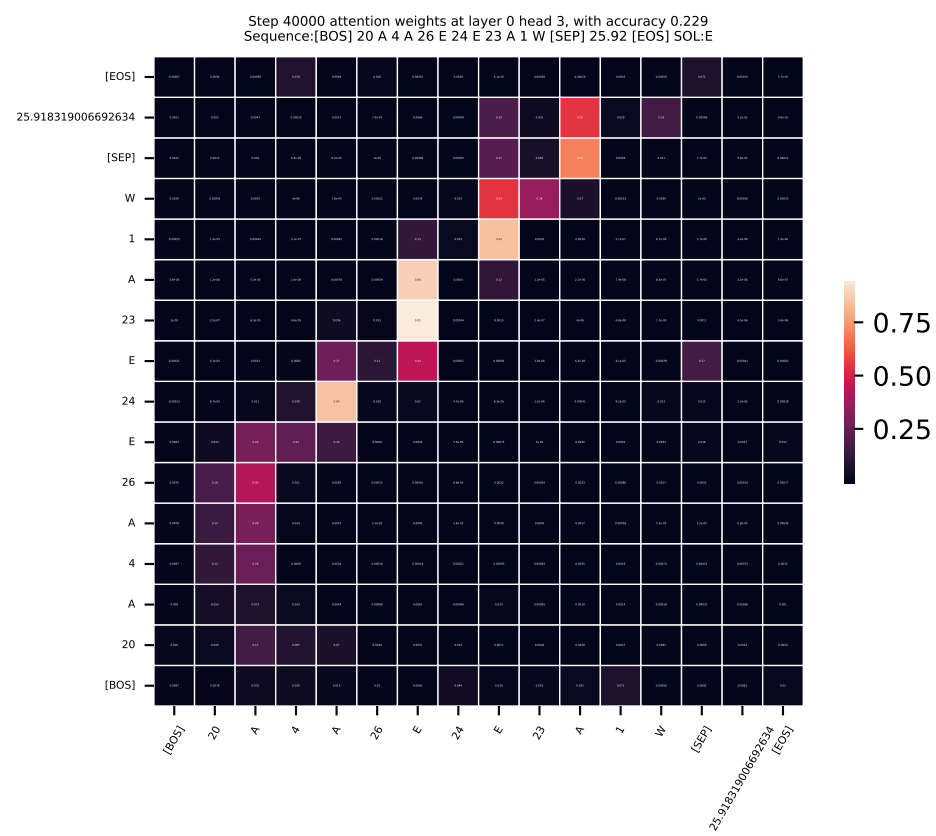

Step 39500 attention weights at layer 1 head 2, with accuracy 0.226
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.10
0.05

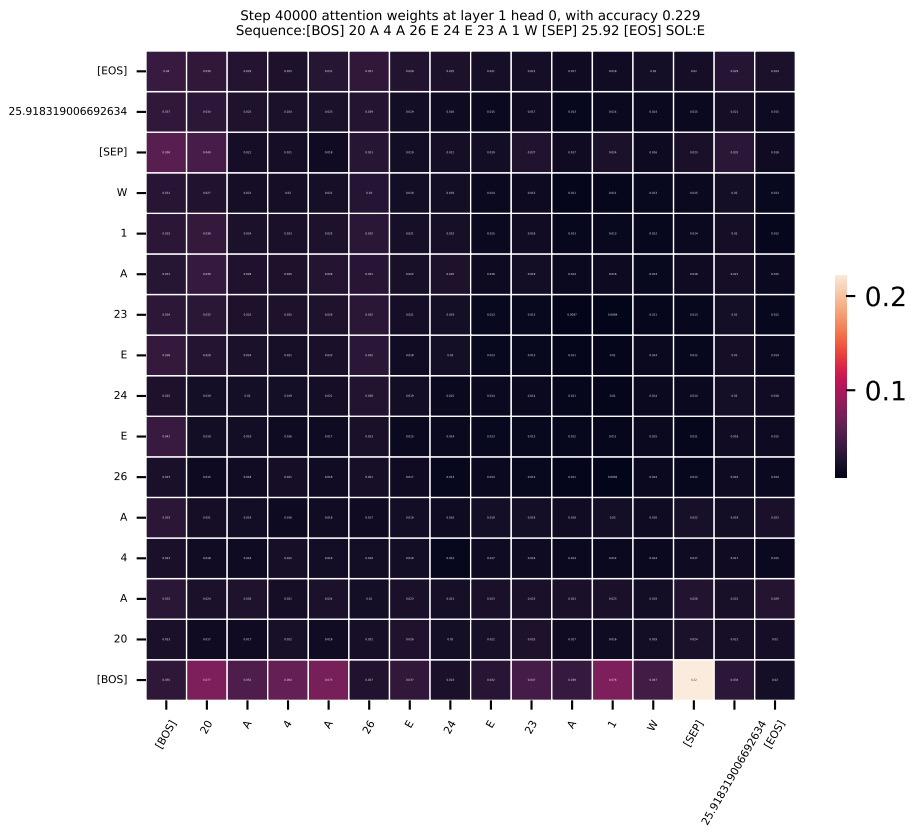

Step 39500 attention weights at layer 1 head 3, with accuracy 0.226
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

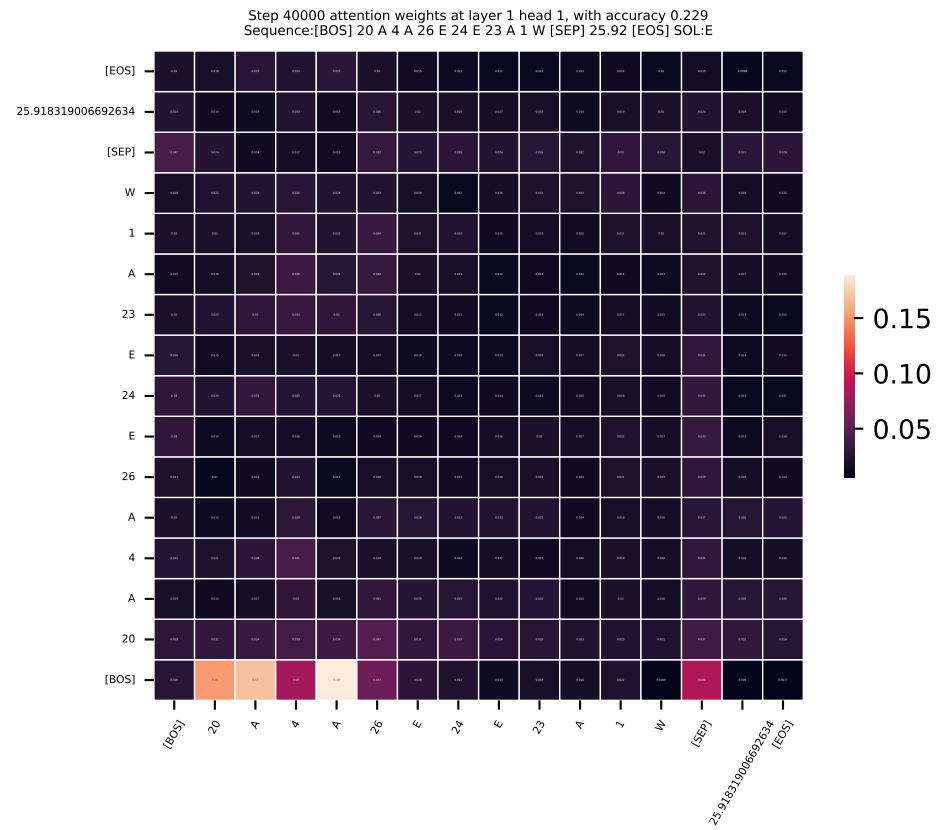

Step 40000 attention weights at layer 0 head 0, with accuracy 0.229
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

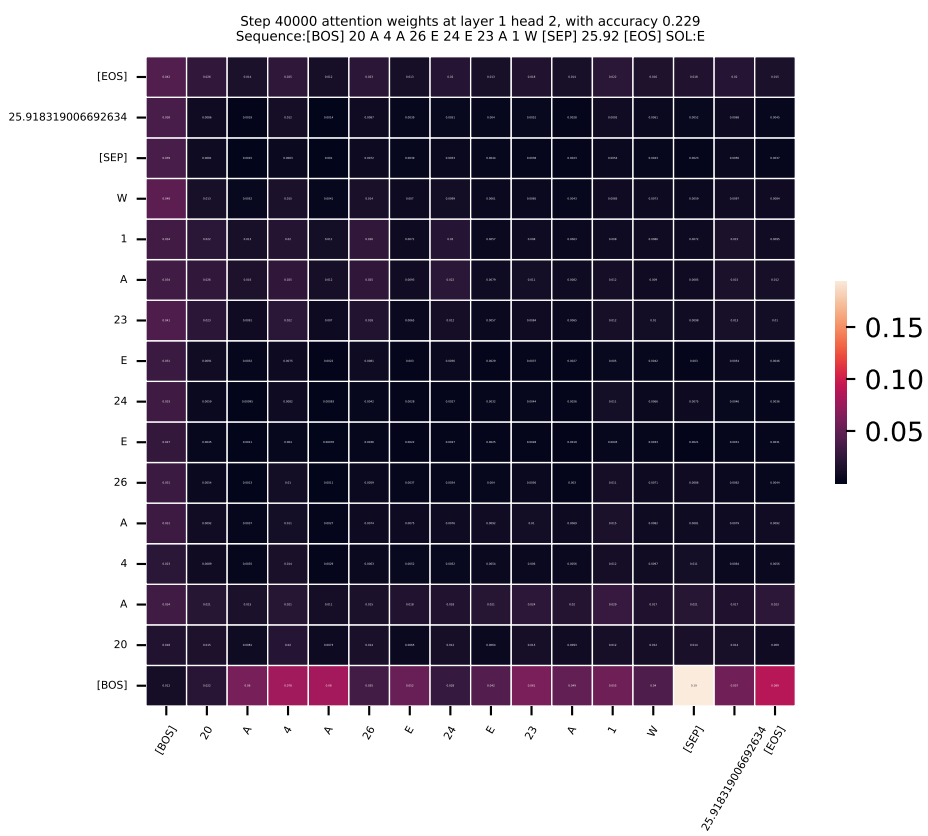

Step 40000 attention weights at layer 0 head 1, with accuracy 0.229
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

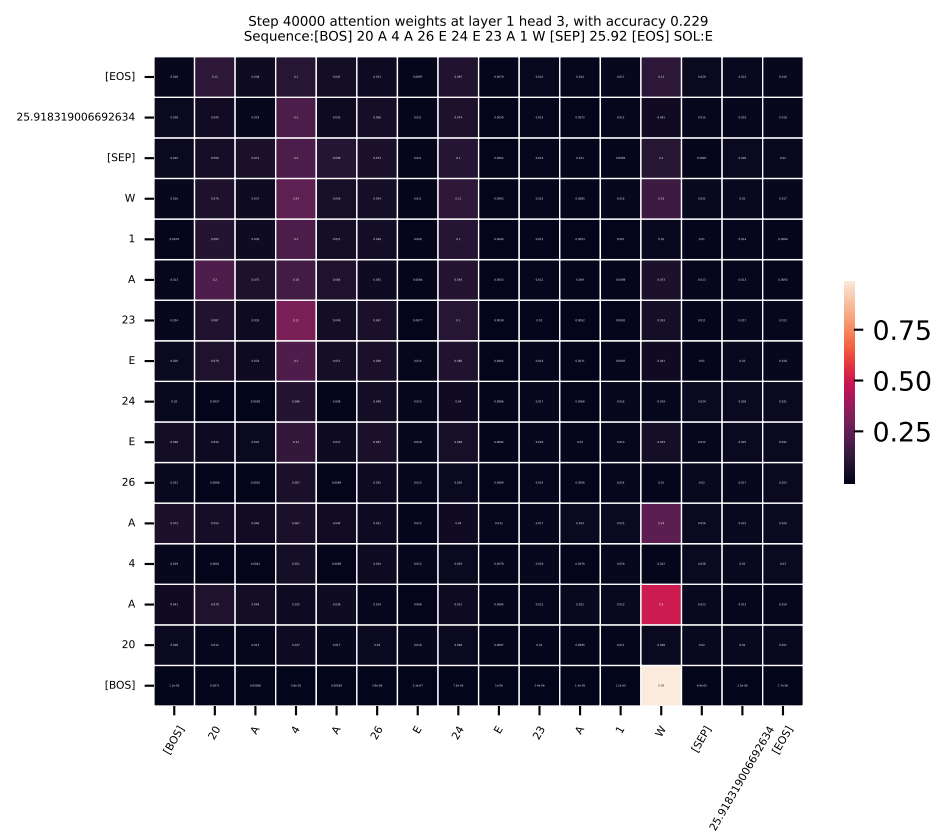

Step 40000 attention weights at layer 0 head 2, with accuracy 0.229
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

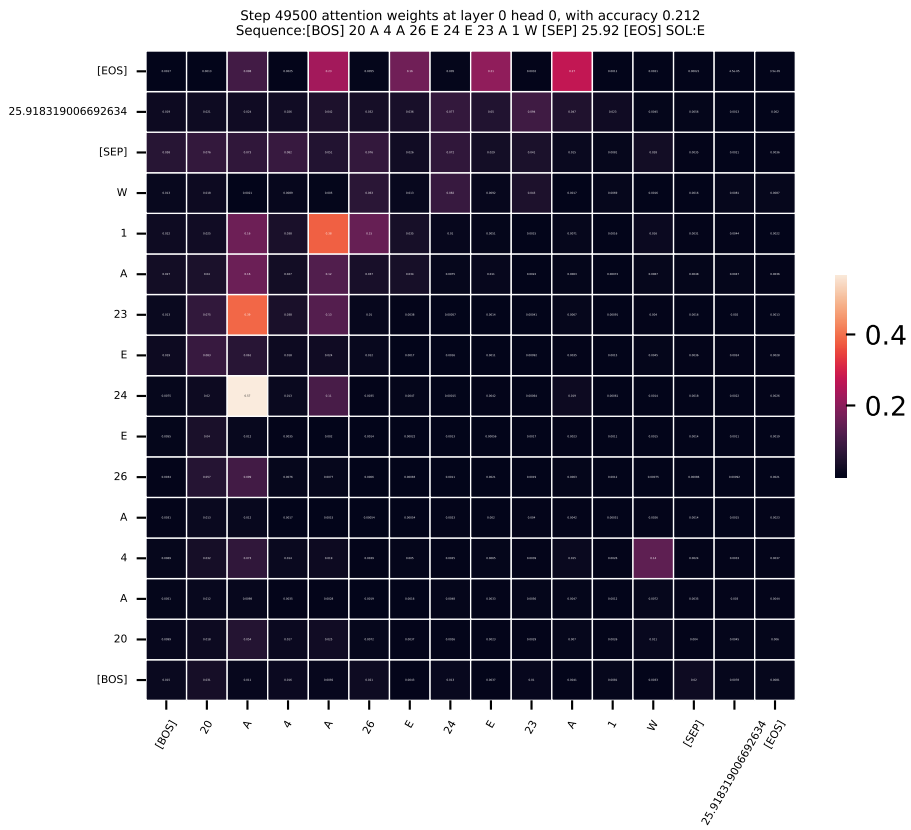

Step 40000 attention weights at layer 0 head 3, with accuracy 0.229
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

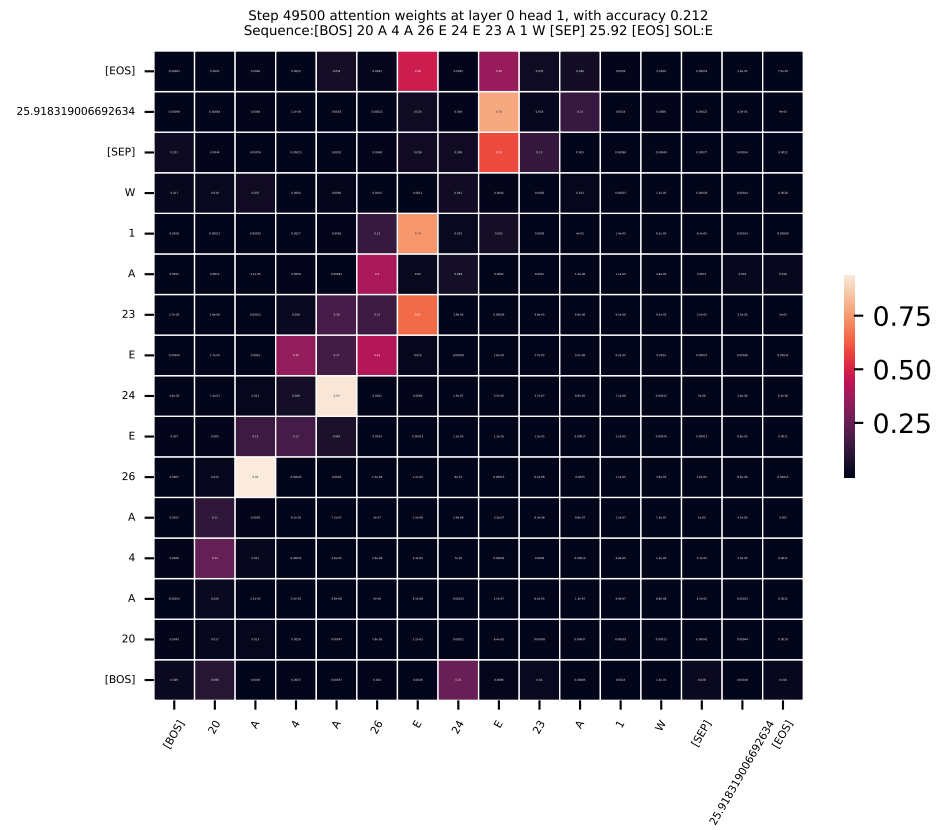

Step 40000 attention weights at layer 1 head 0, with accuracy 0.229
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.2
0.1

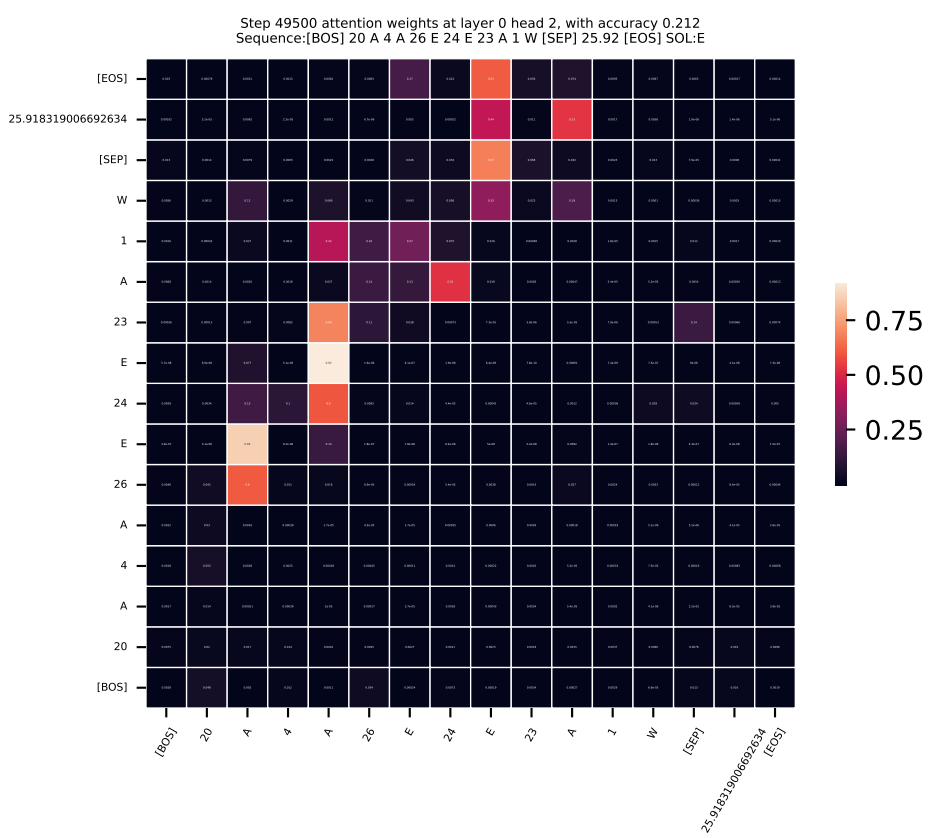

Step 40000 attention weights at layer 1 head 1, with accuracy 0.229
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.15
0.10
0.05

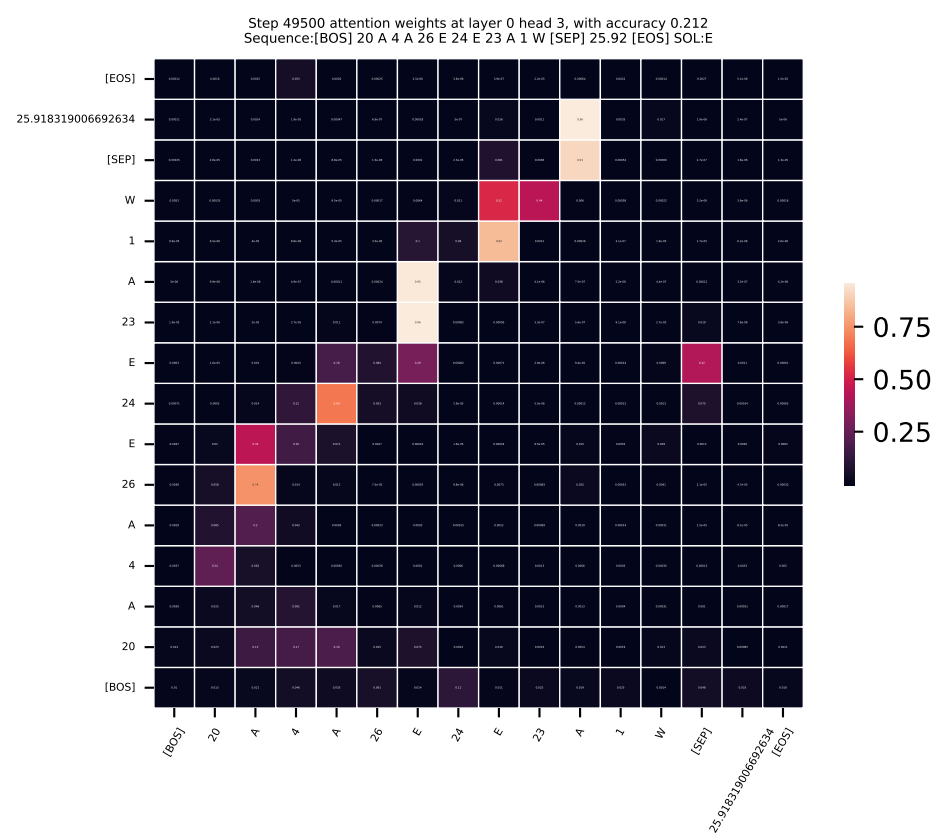

Step 40000 attention weights at layer 1 head 2, with accuracy 0.229
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.15
0.10
0.05

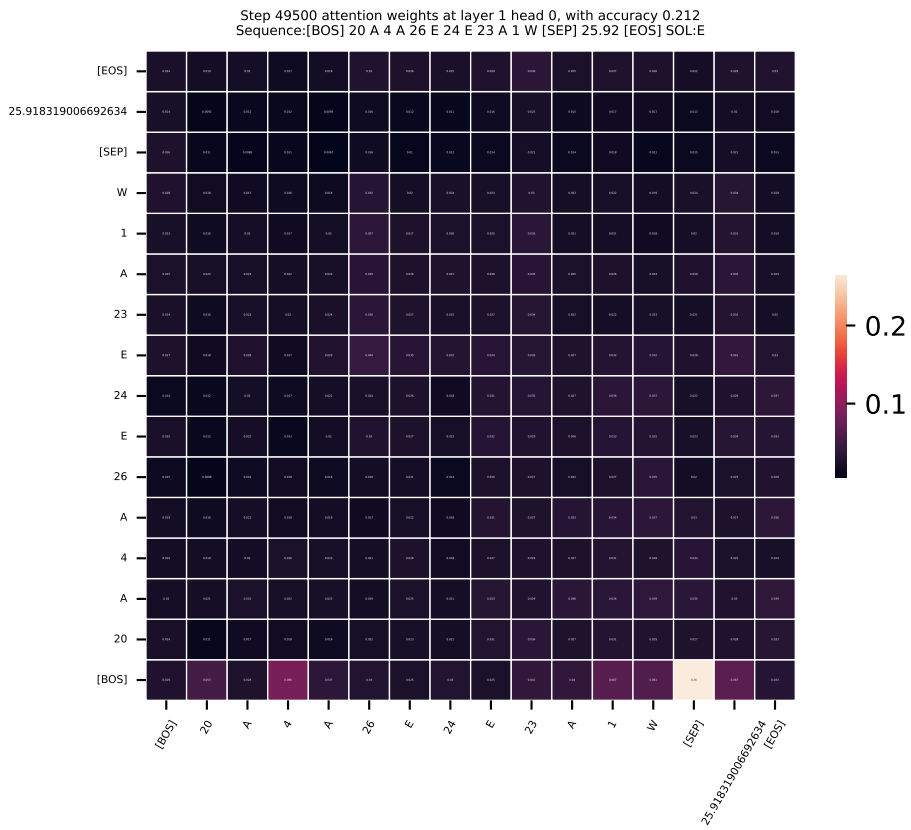

Step 40000 attention weights at layer 1 head 3, with accuracy 0.229
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

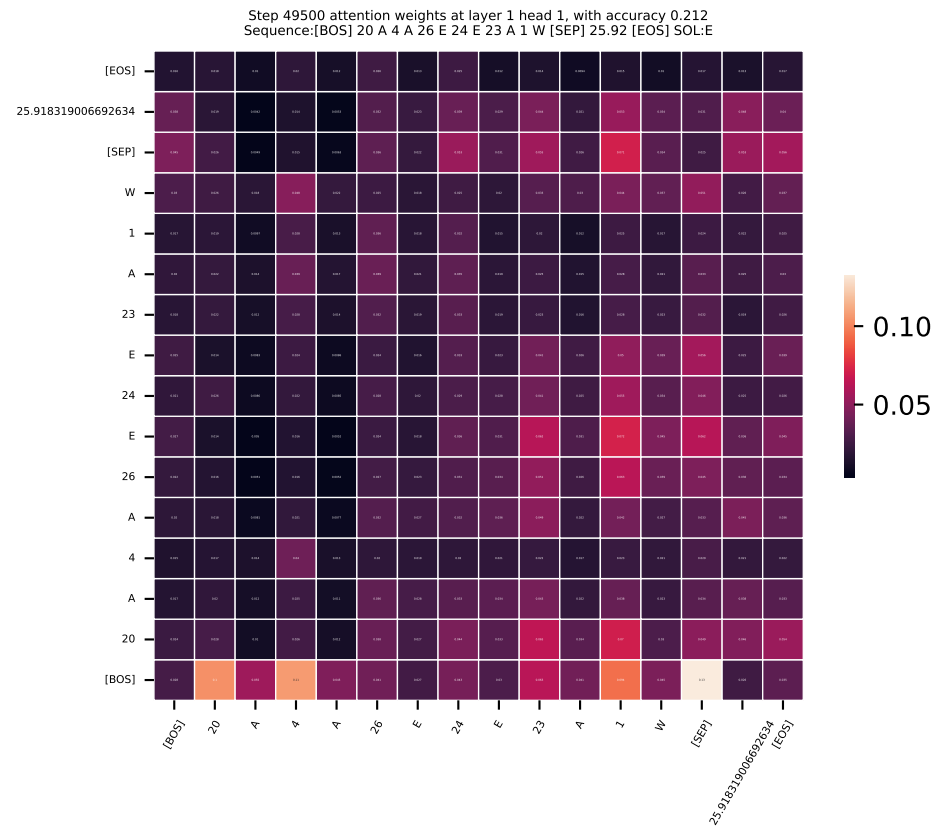

Step 49500 attention weights at layer 0 head 0, with accuracy 0.212
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

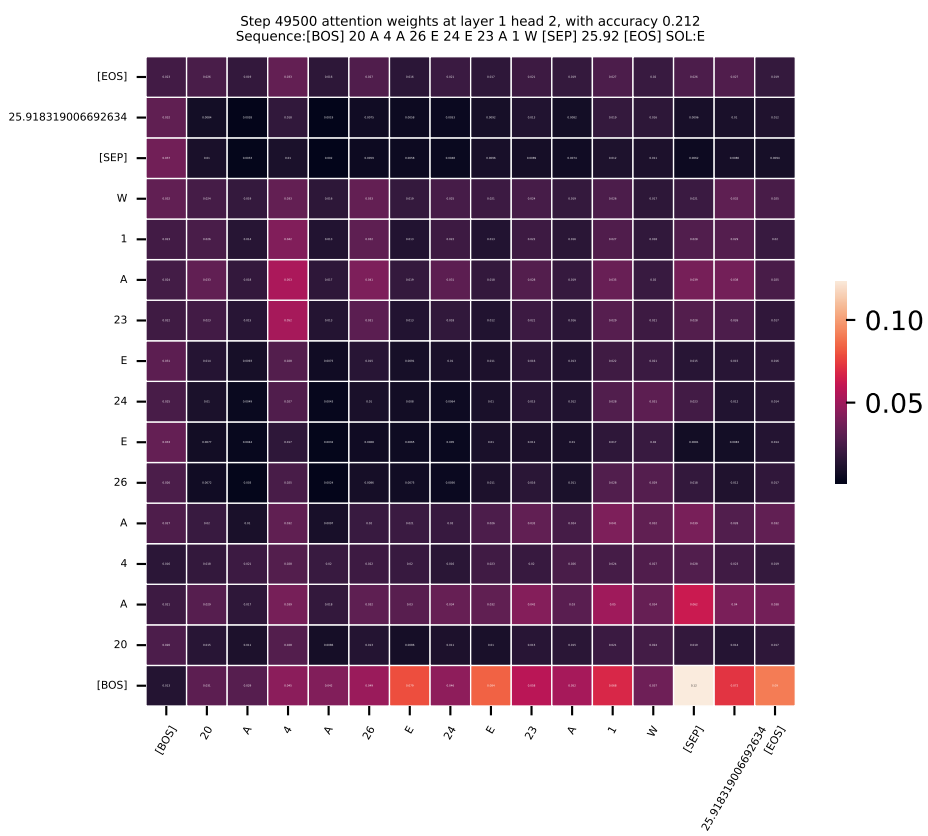

Step 49500 attention weights at layer 0 head 1, with accuracy 0.212
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

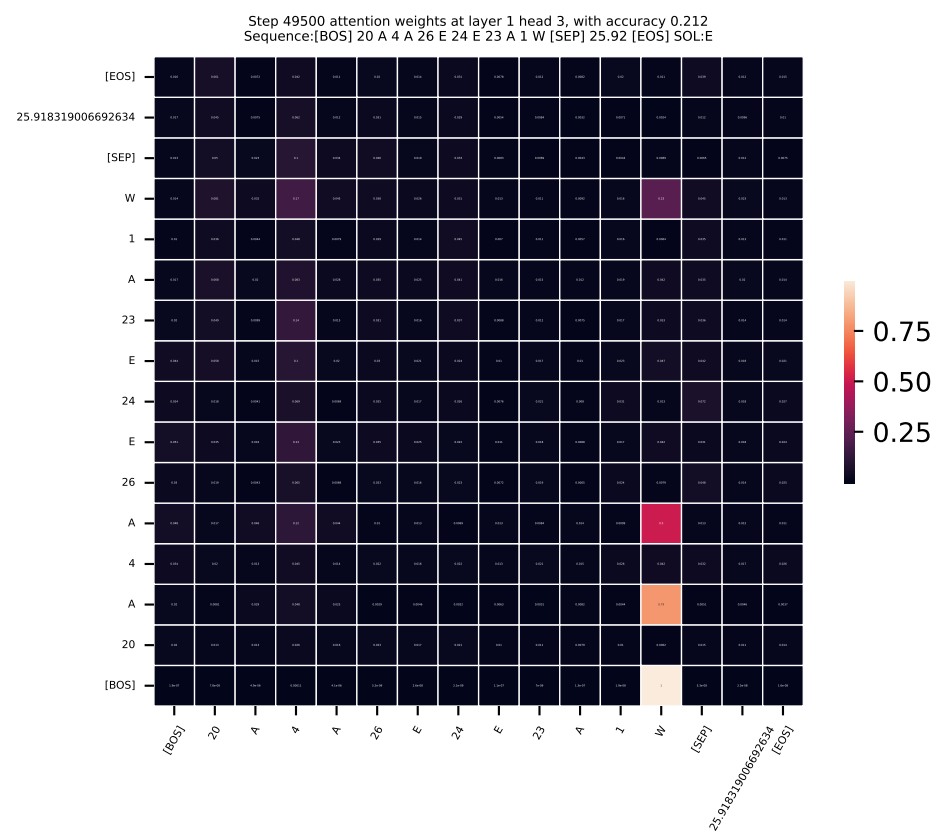

Step 49500 attention weights at layer 0 head 2, with accuracy 0.212
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

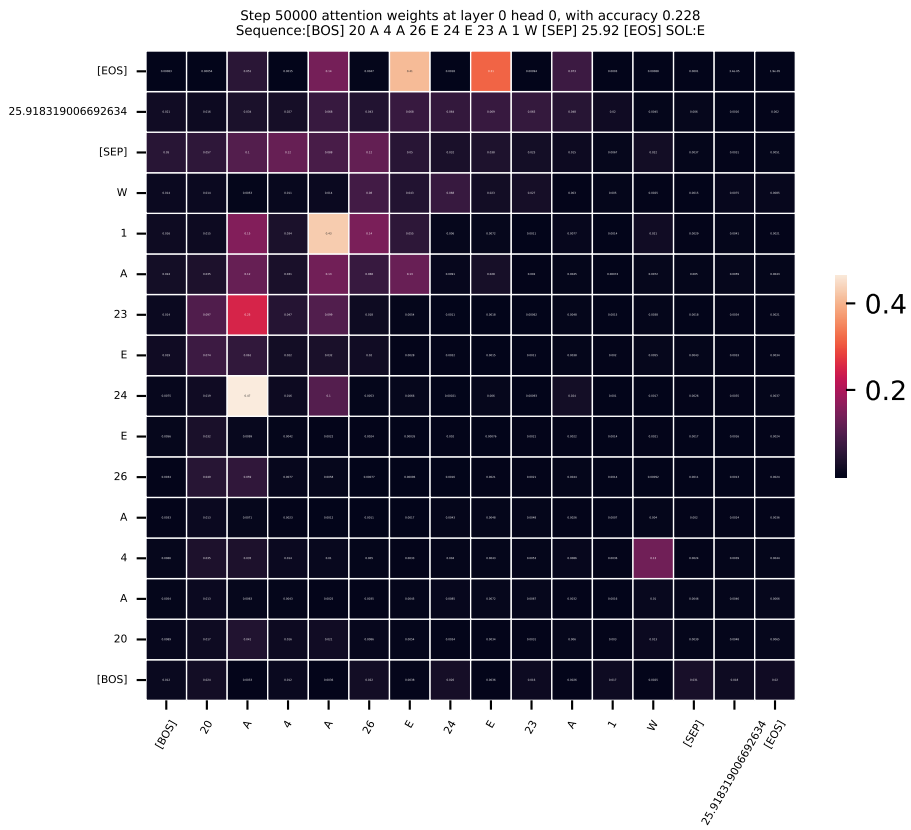

Step 49500 attention weights at layer 0 head 3, with accuracy 0.212
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

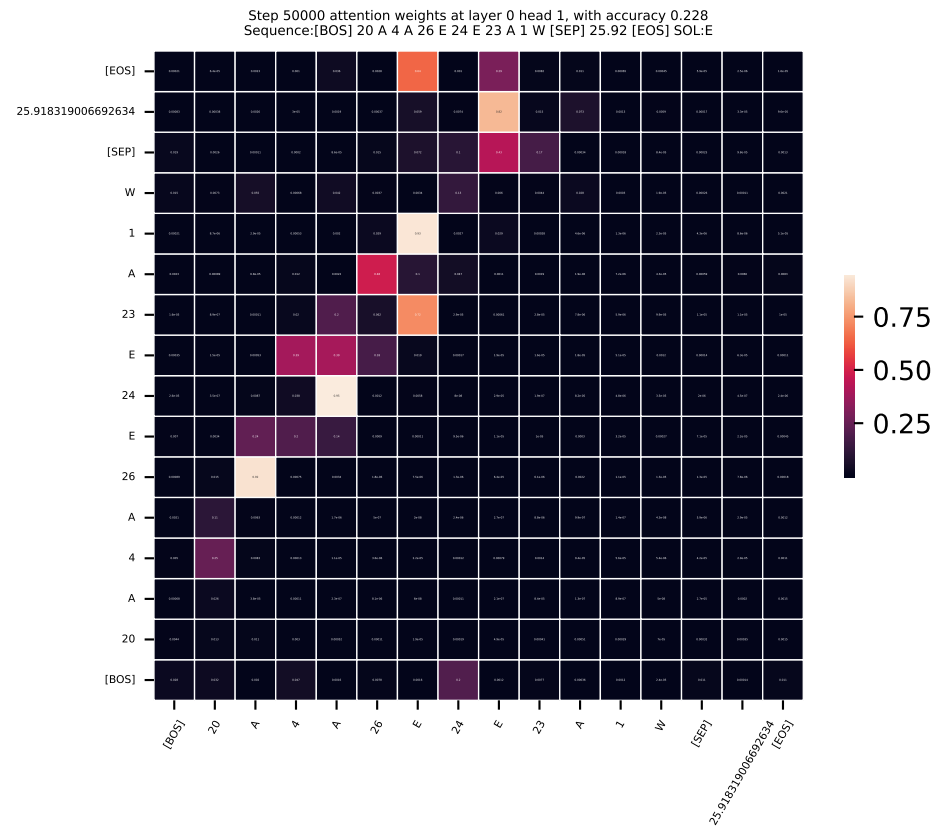

Step 49500 attention weights at layer 1 head 0, with accuracy 0.212
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.2
0.1

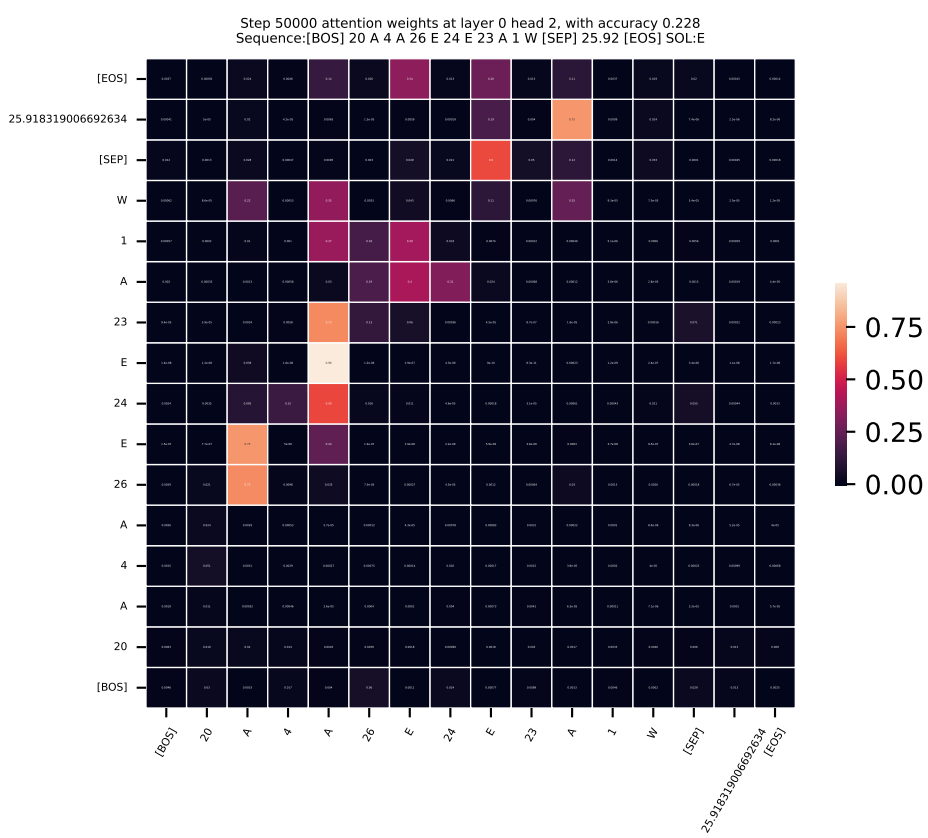

Step 49500 attention weights at layer 1 head 1, with accuracy 0.212
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.10
0.05

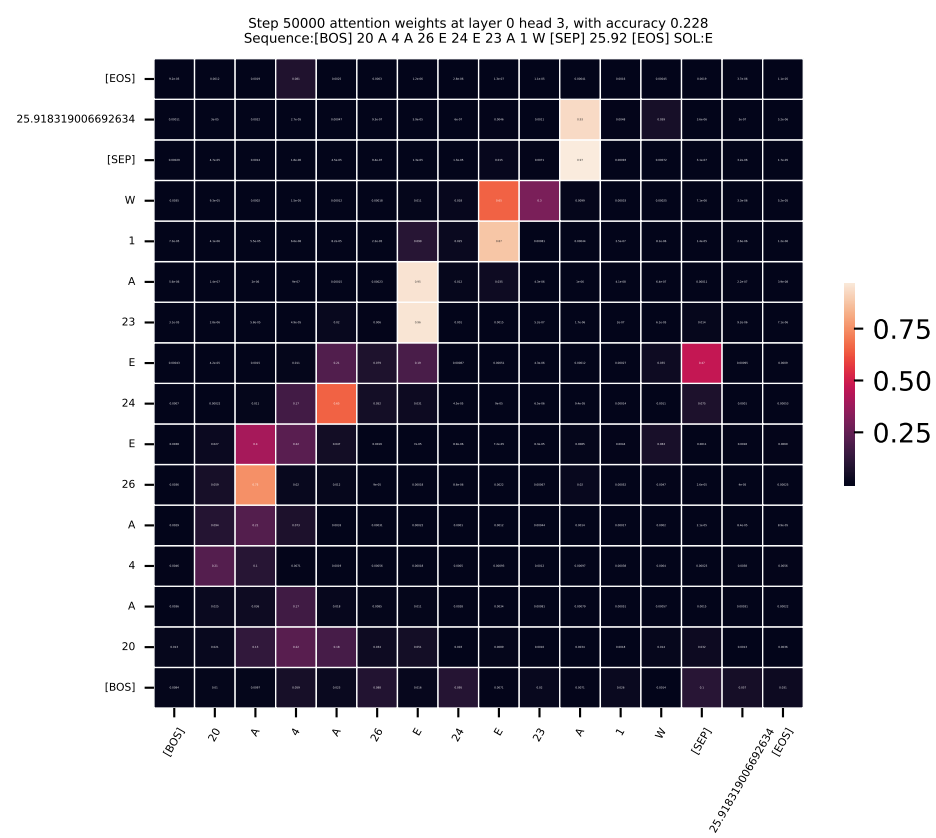

Step 49500 attention weights at layer 1 head 2, with accuracy 0.212
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.10
0.05

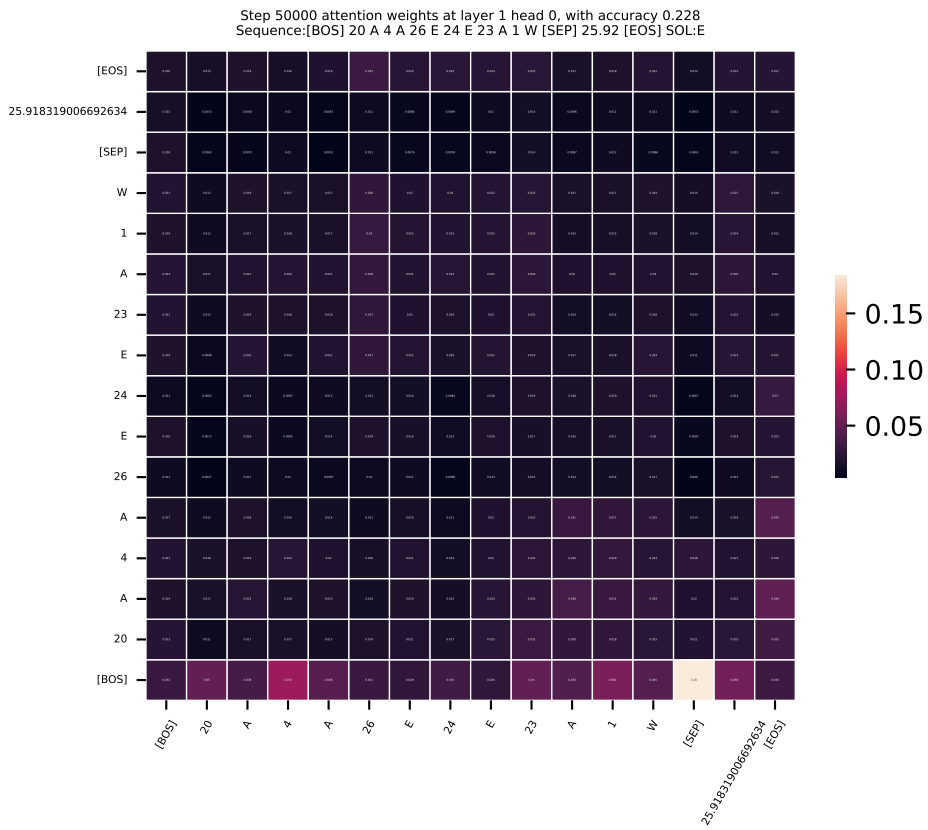

Step 49500 attention weights at layer 1 head 3, with accuracy 0.212
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

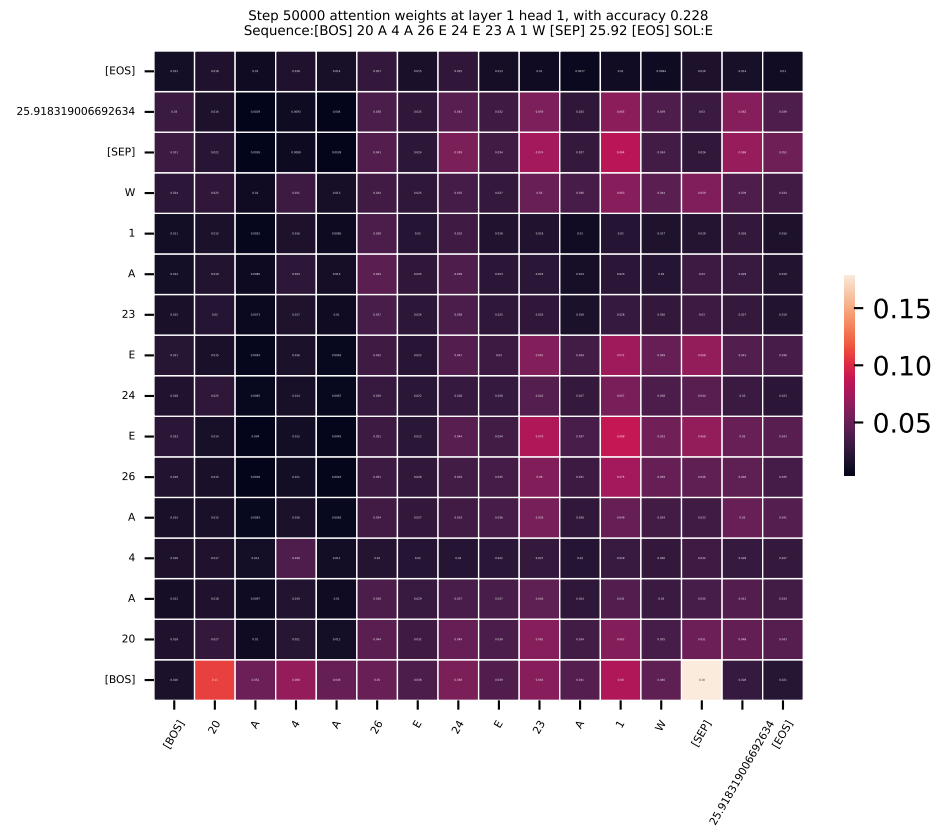

Step 50000 attention weights at layer 0 head 0, with accuracy 0.228
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.4
0.2

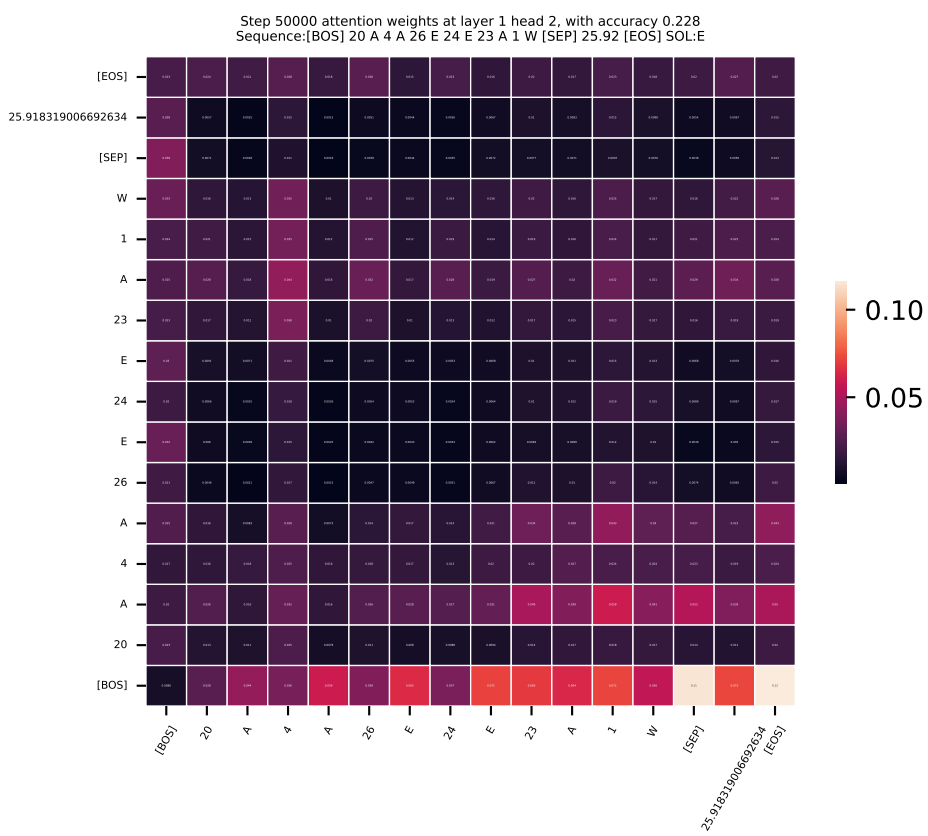

Step 50000 attention weights at layer 0 head 1, with accuracy 0.228
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

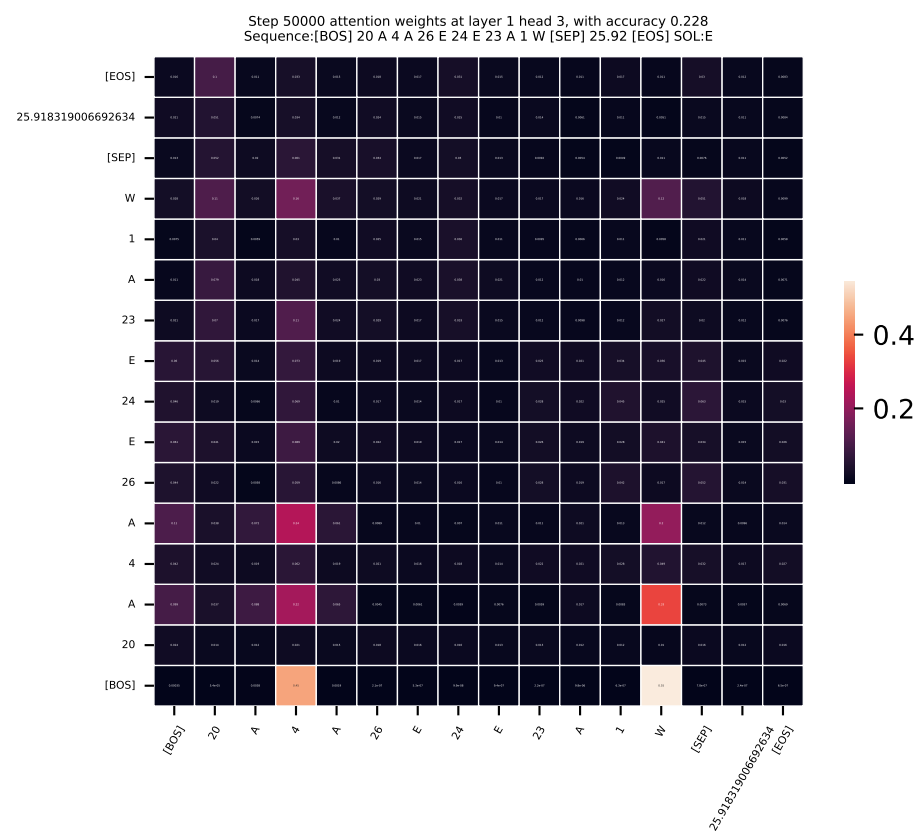

Step 50000 attention weights at layer 0 head 2, with accuracy 0.228
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

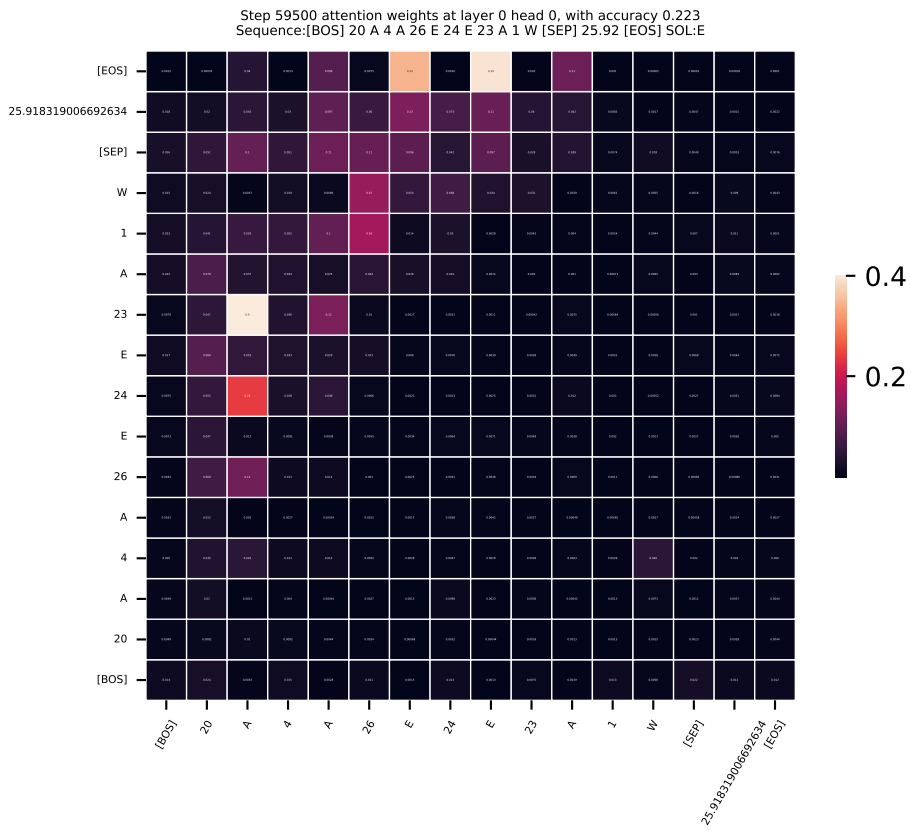

Step 50000 attention weights at layer 0 head 3, with accuracy 0.228
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

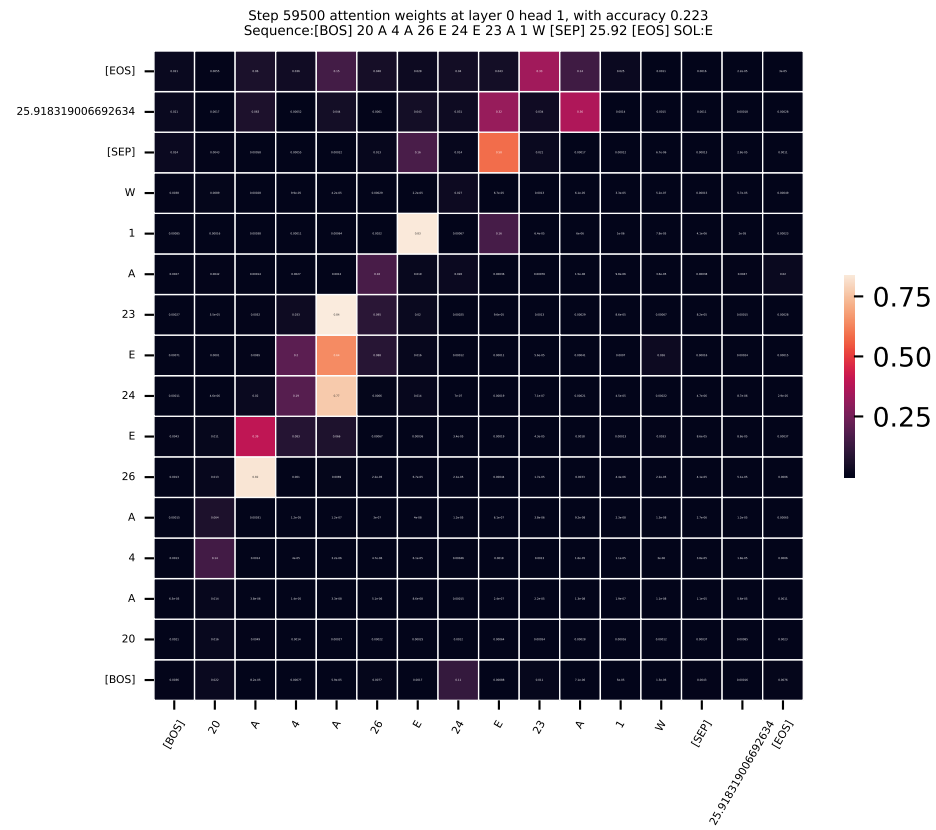

Step 50000 attention weights at layer 1 head 0, with accuracy 0.228
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.15
0.10
0.05

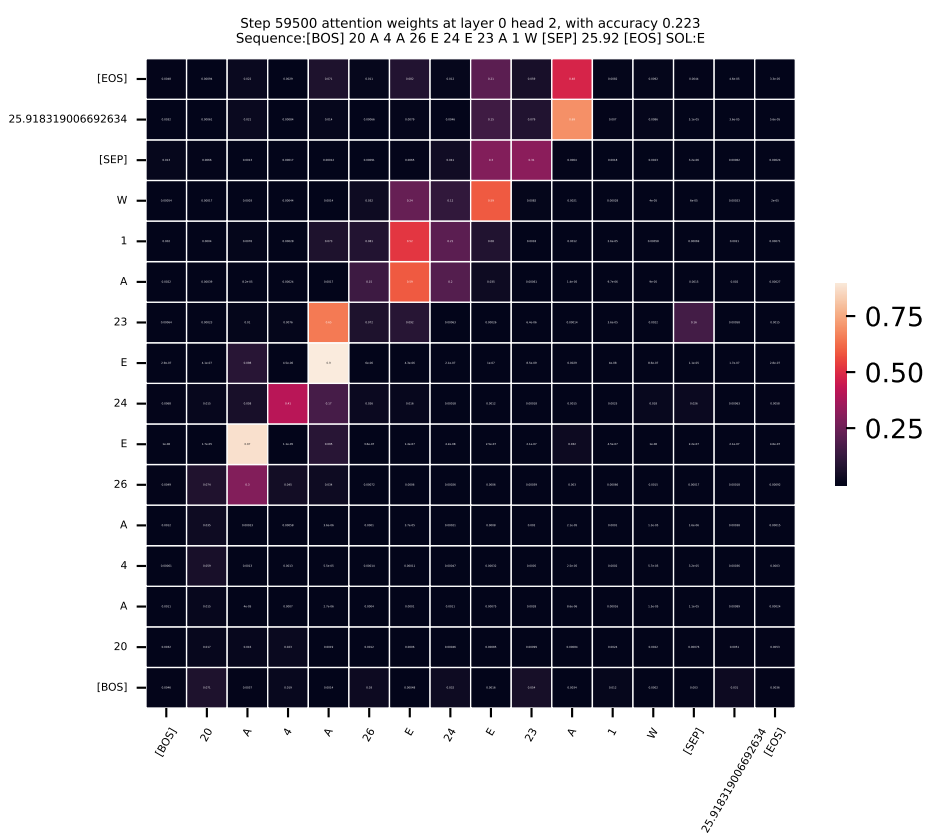

Step 50000 attention weights at layer 1 head 1, with accuracy 0.228
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.15
0.10
0.05

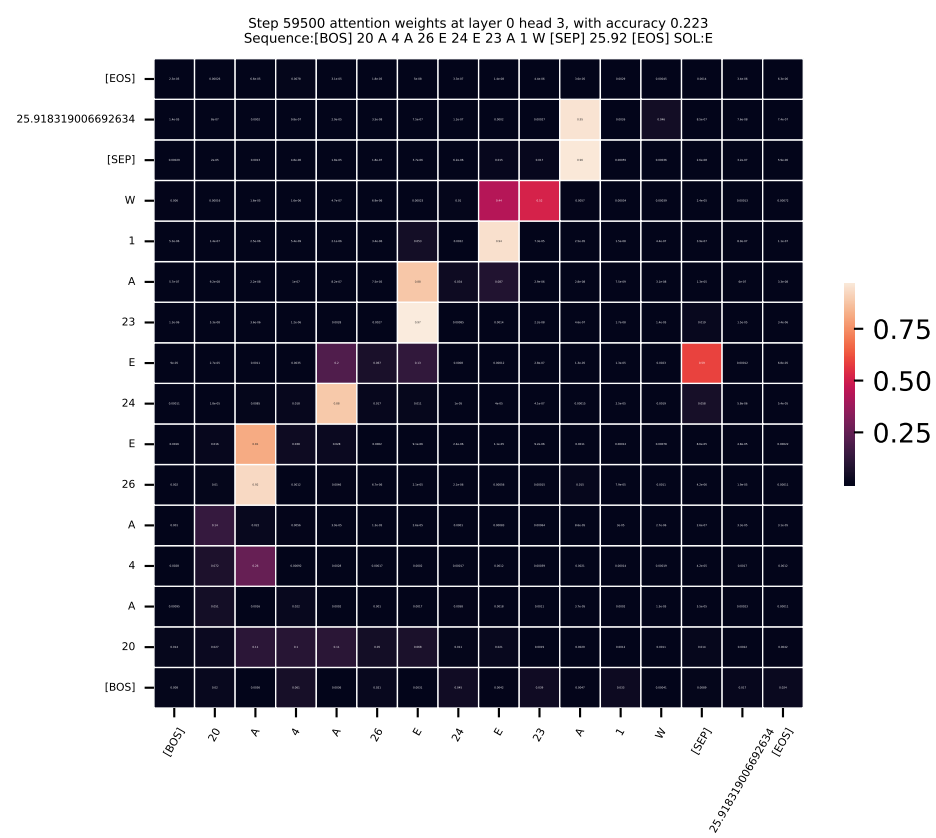

Step 50000 attention weights at layer 1 head 2, with accuracy 0.228
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.10
0.05

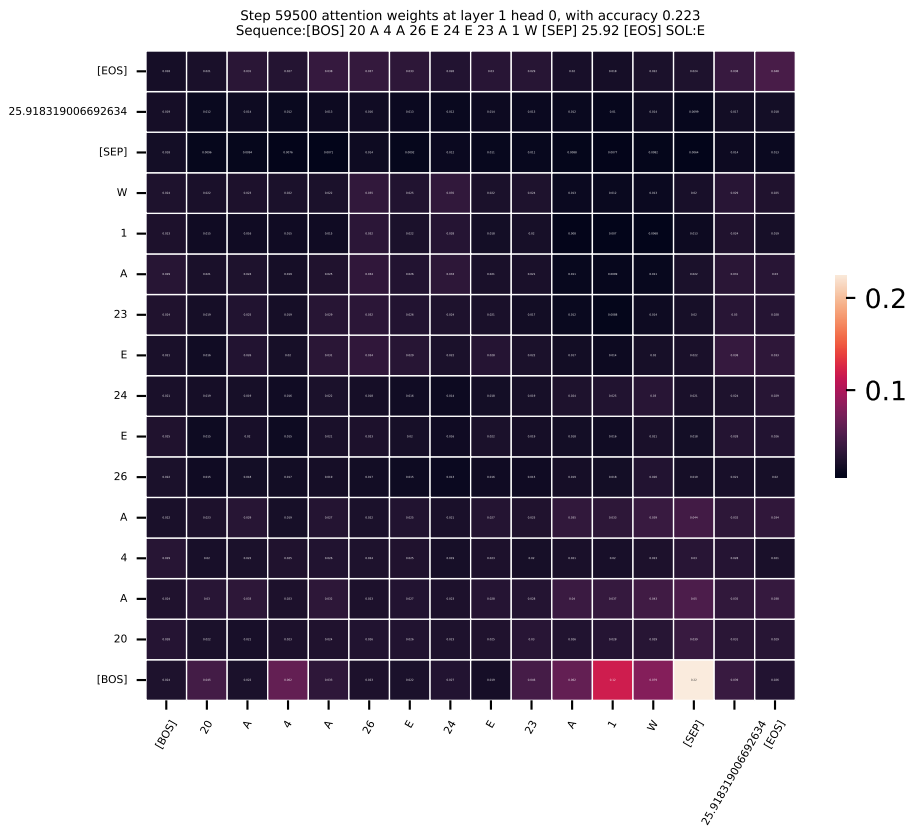

Step 50000 attention weights at layer 1 head 3, with accuracy 0.228
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

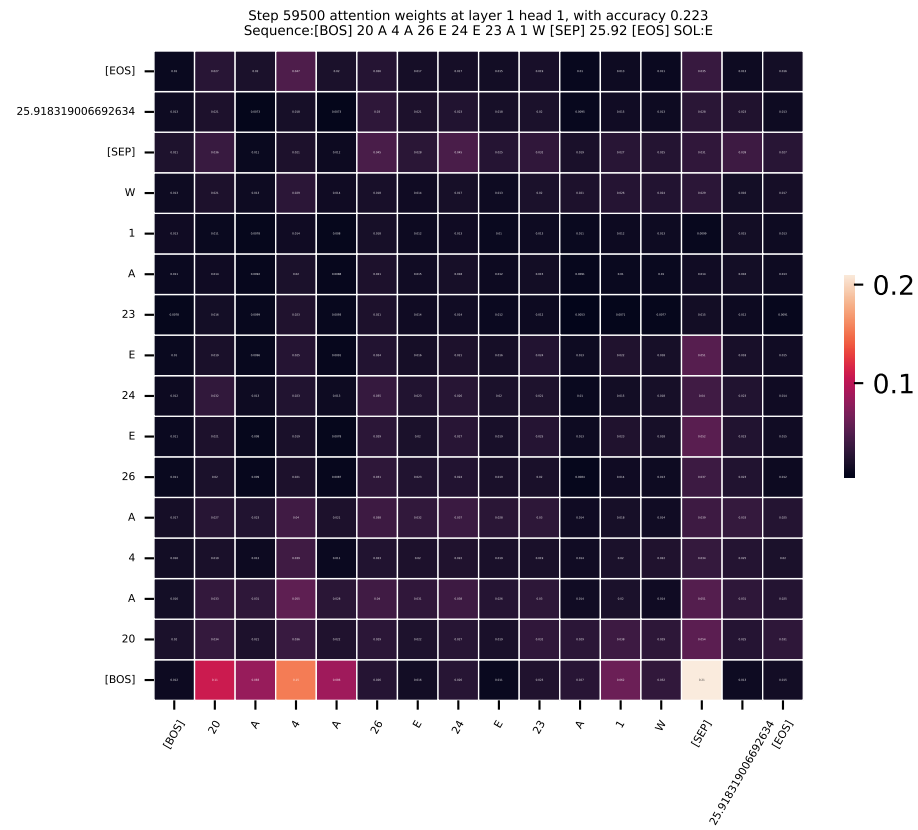

Step 59500 attention weights at layer 0 head 0, with accuracy 0.223
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

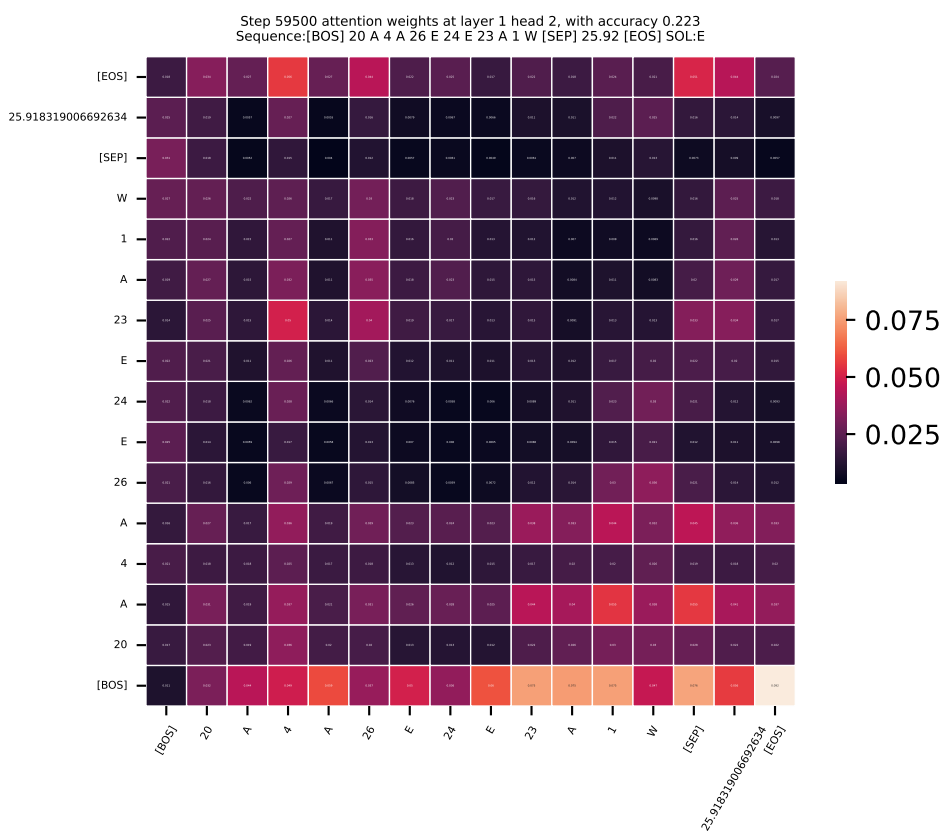

Step 59500 attention weights at layer 0 head 1, with accuracy 0.223
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

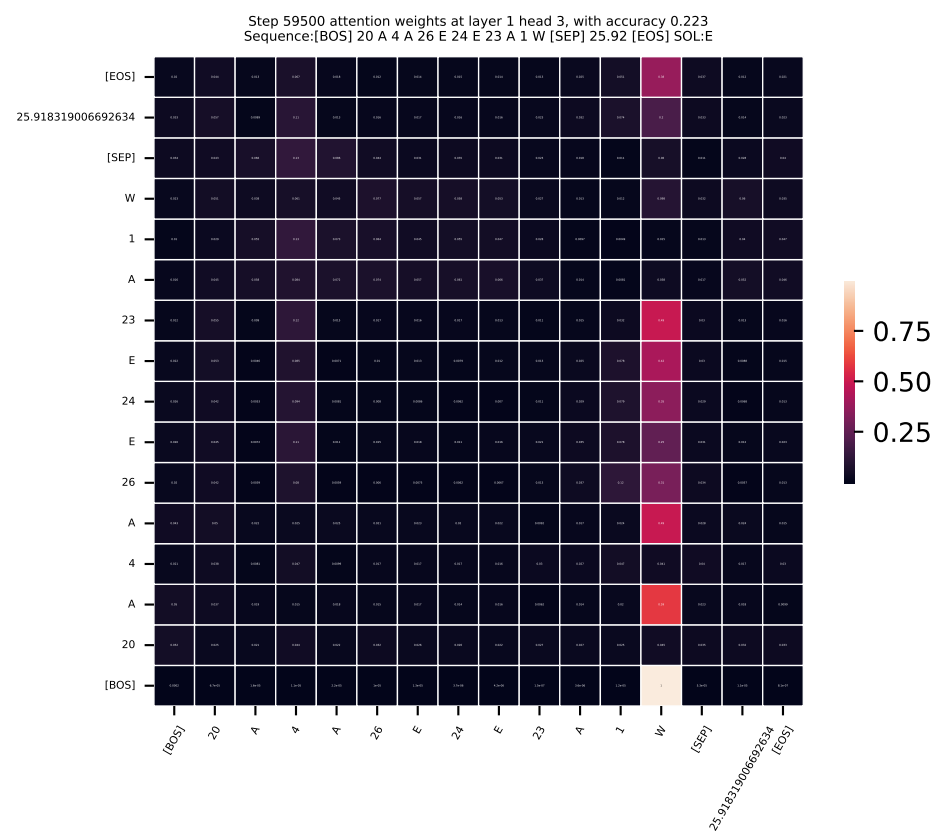

Step 59500 attention weights at layer 0 head 2, with accuracy 0.223
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

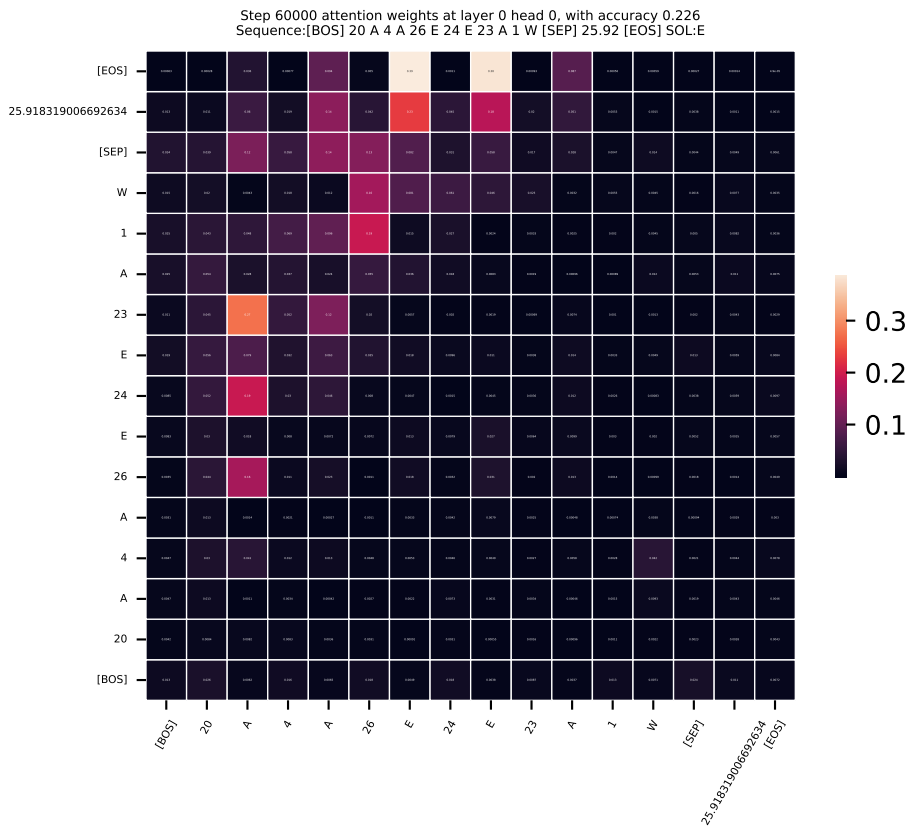

Step 59500 attention weights at layer 0 head 3, with accuracy 0.223
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

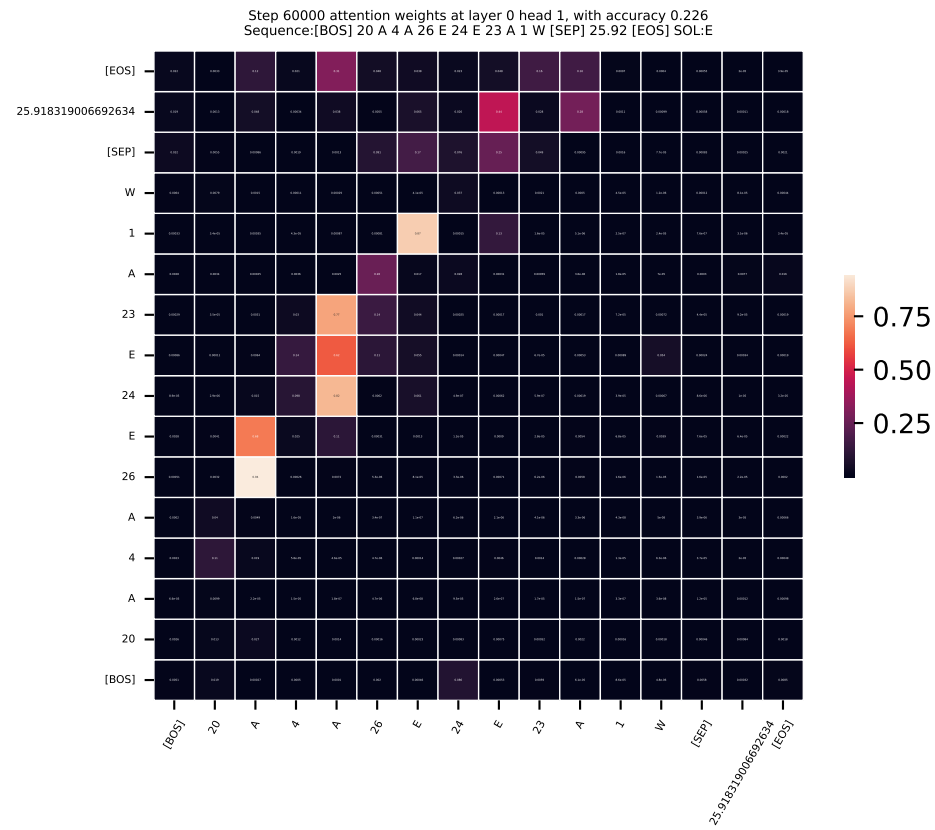

Step 59500 attention weights at layer 1 head 0, with accuracy 0.223
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.2
0.1

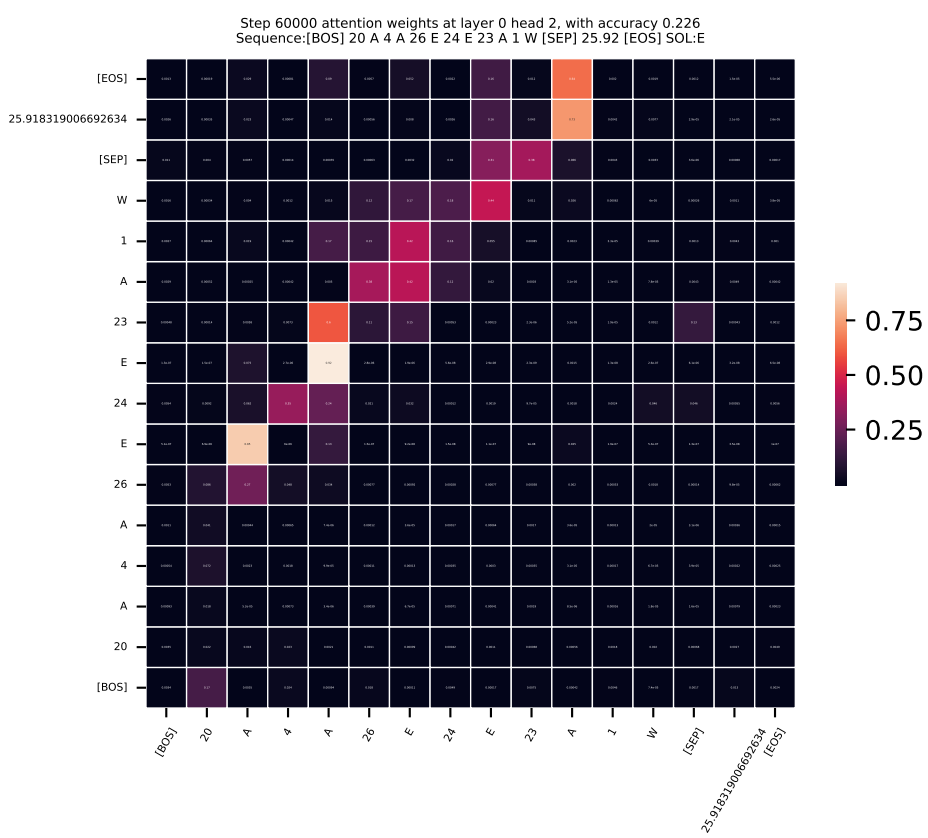

Step 59500 attention weights at layer 1 head 1, with accuracy 0.223
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.2
0.1

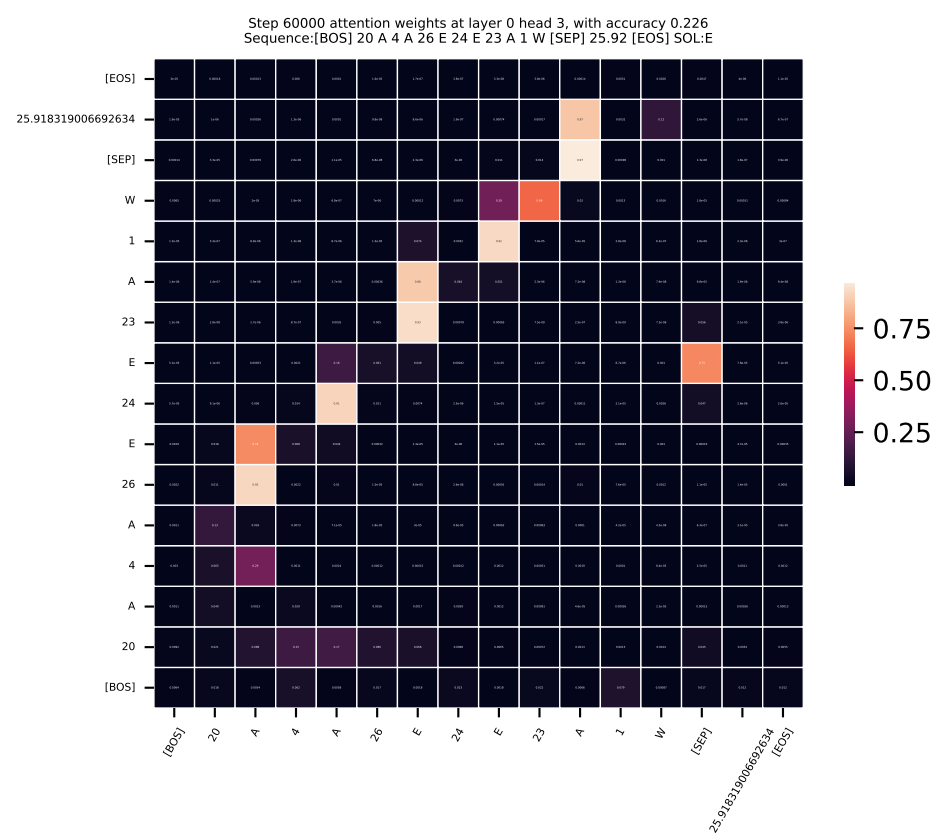

Step 59500 attention weights at layer 1 head 2, with accuracy 0.223
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.075
0.050
0.025

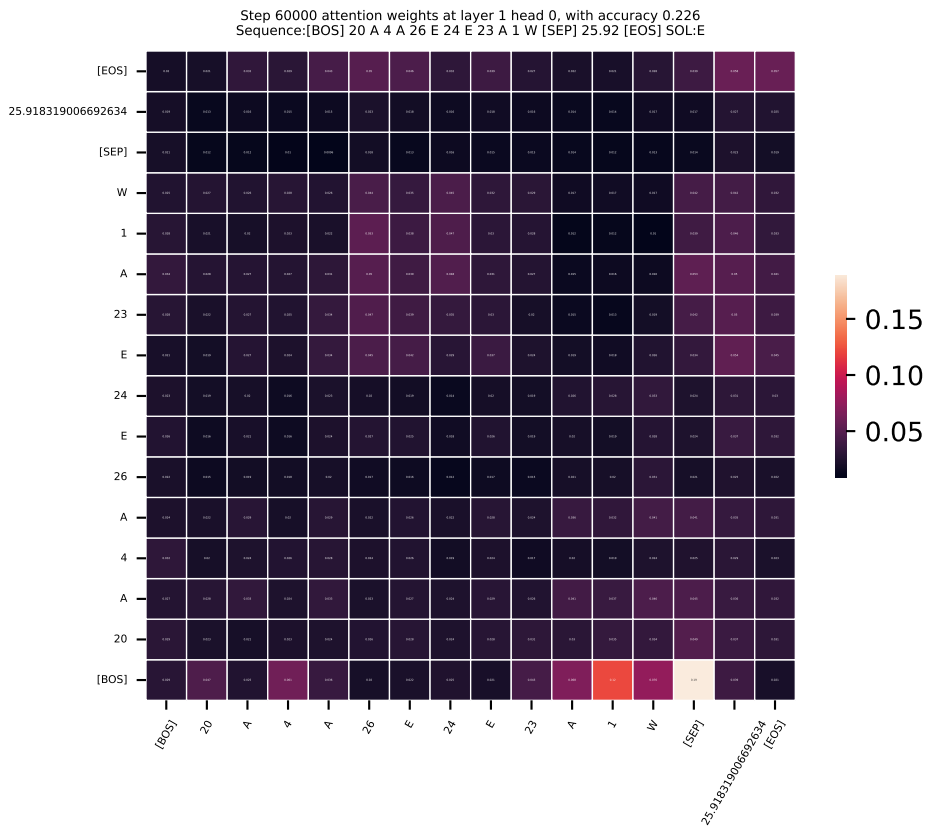

Step 59500 attention weights at layer 1 head 3, with accuracy 0.223
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

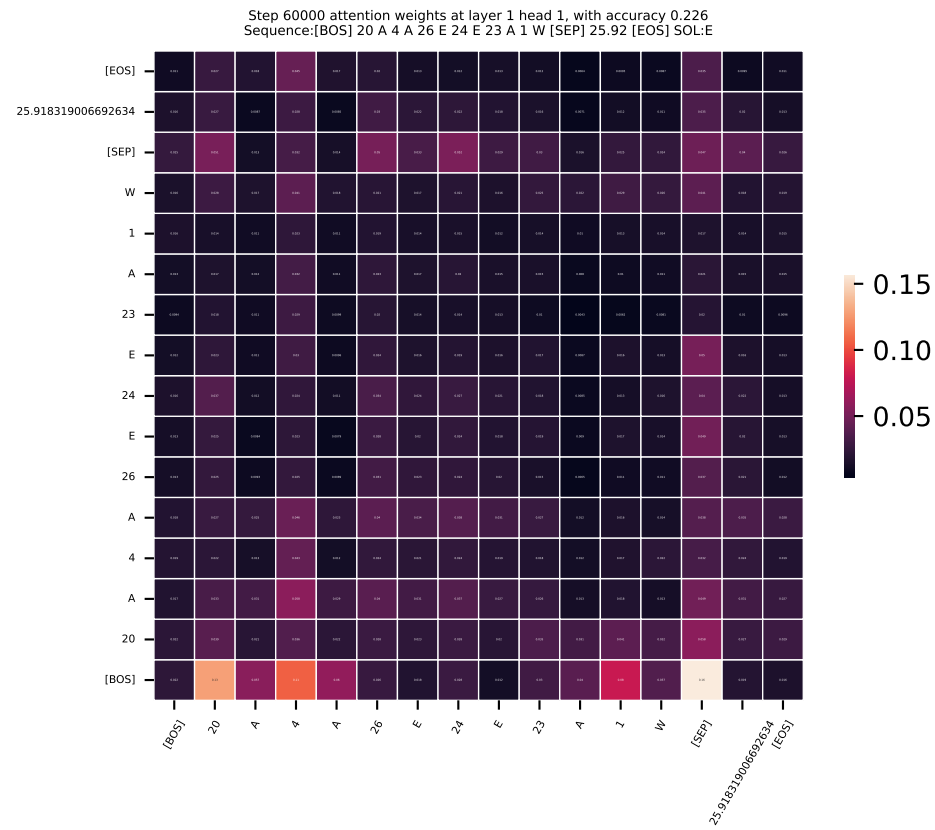

Step 60000 attention weights at layer 0 head 0, with accuracy 0.226
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.3
0.2
0.1

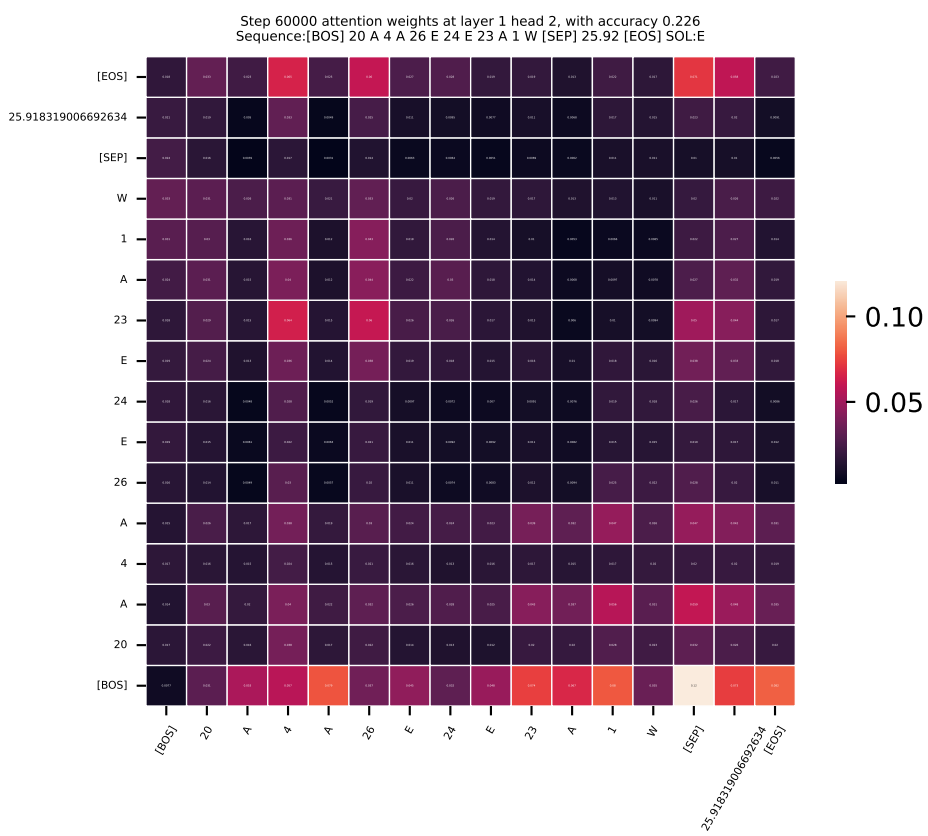

Step 60000 attention weights at layer 0 head 1, with accuracy 0.226
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

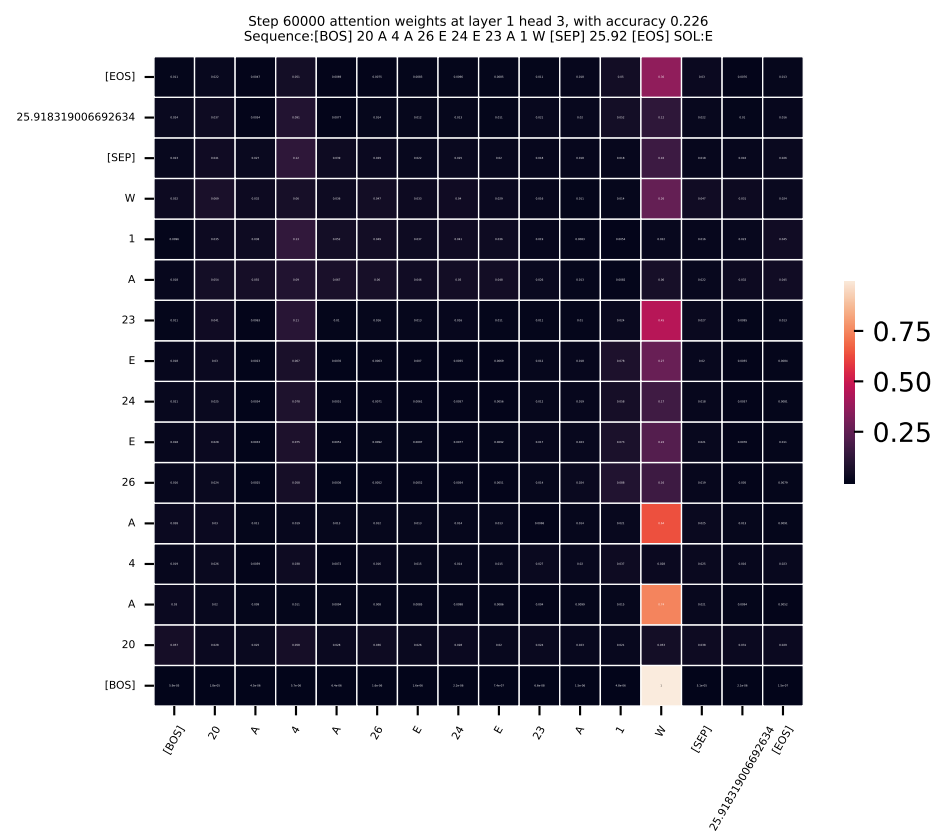

Step 60000 attention weights at layer 0 head 2, with accuracy 0.226
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

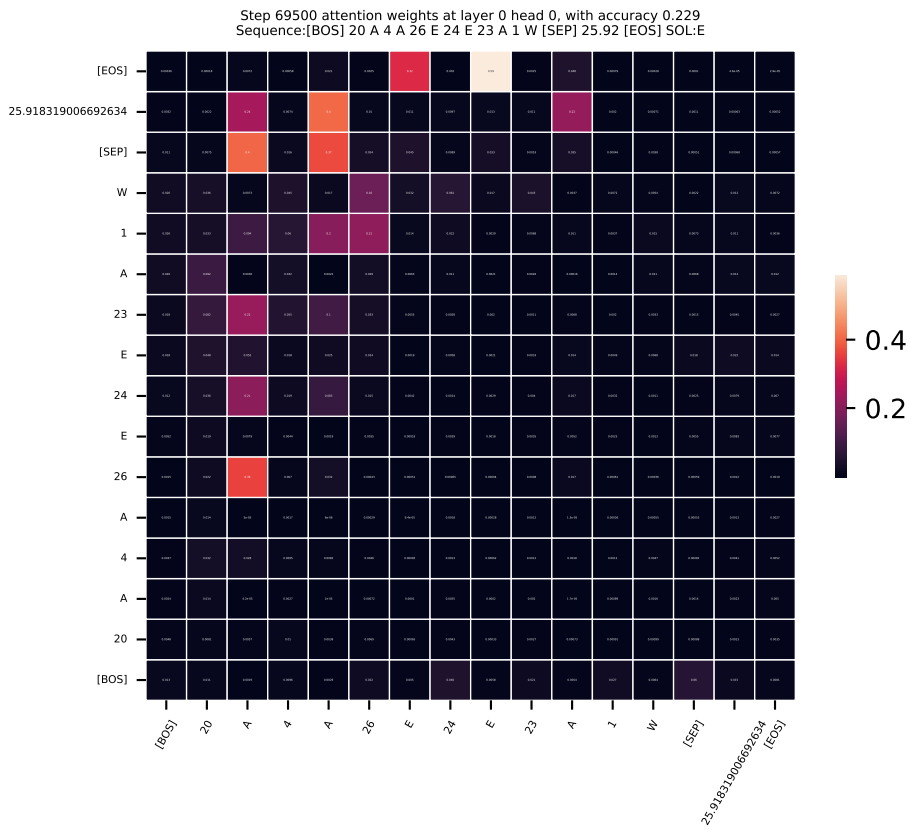

Step 60000 attention weights at layer 0 head 3, with accuracy 0.226
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

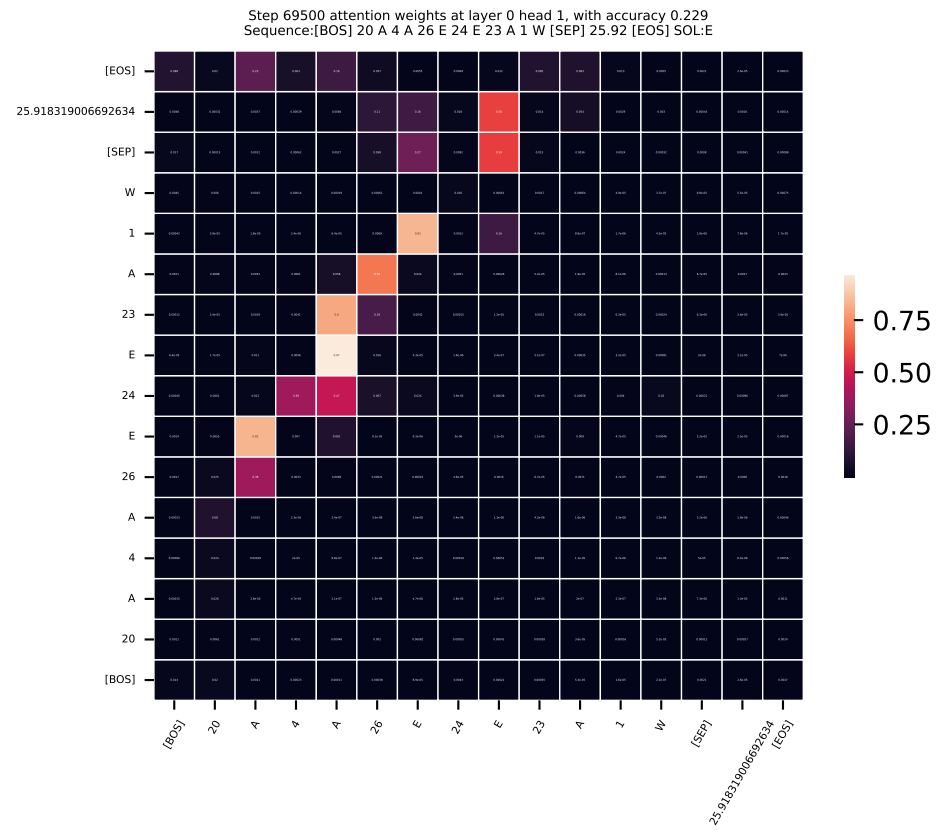

Step 60000 attention weights at layer 1 head 0, with accuracy 0.226
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.15
0.10
0.05

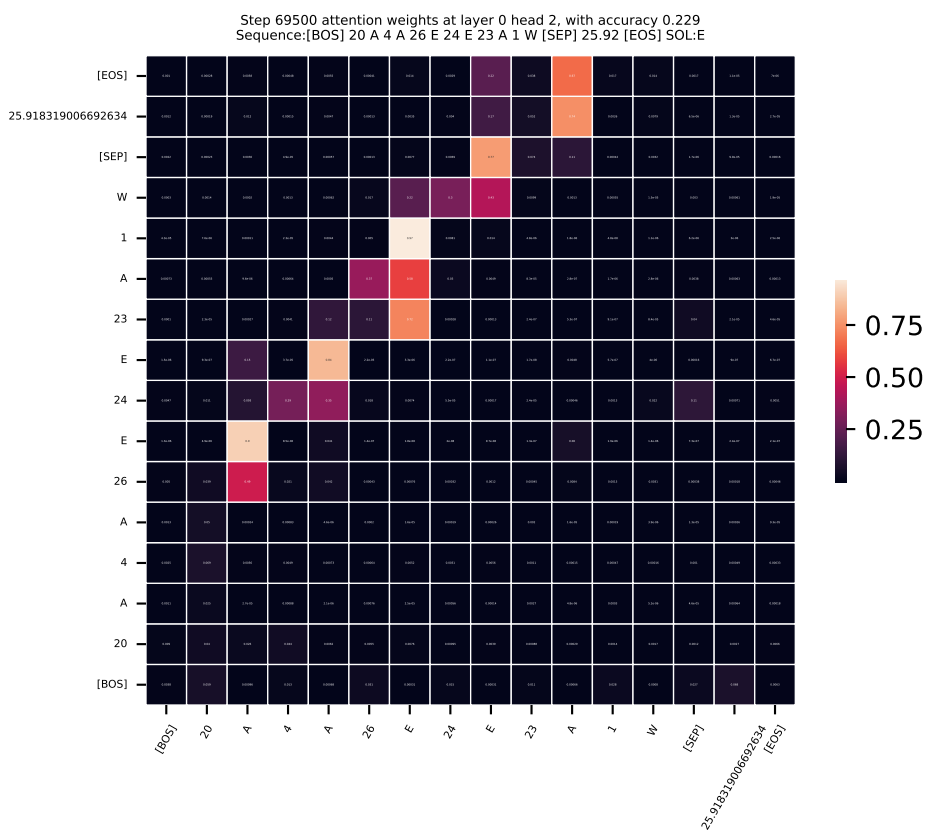

Step 60000 attention weights at layer 1 head 1, with accuracy 0.226
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.15
0.10
0.05

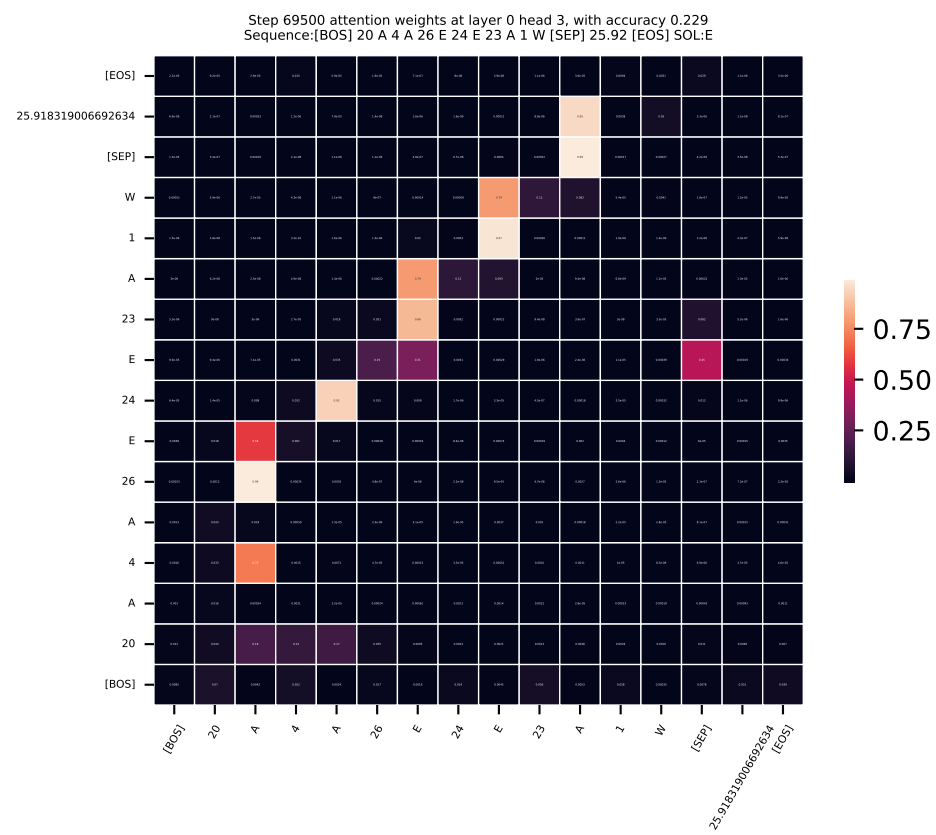

Step 60000 attention weights at layer 1 head 2, with accuracy 0.226
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.10
0.05

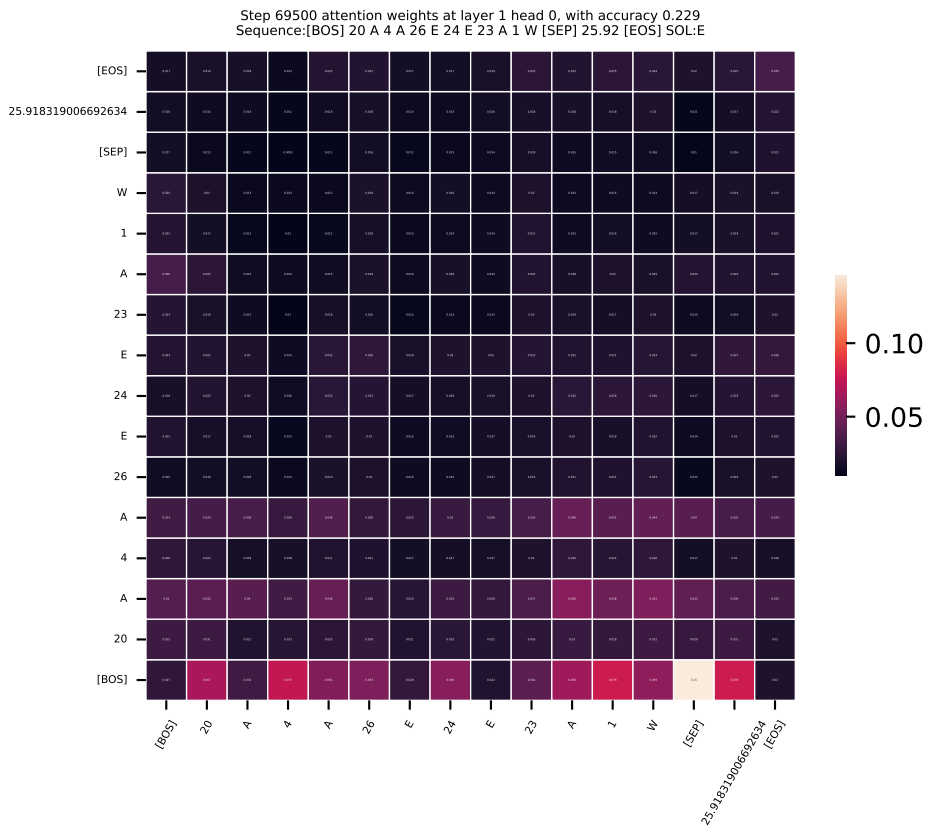

Step 60000 attention weights at layer 1 head 3, with accuracy 0.226
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

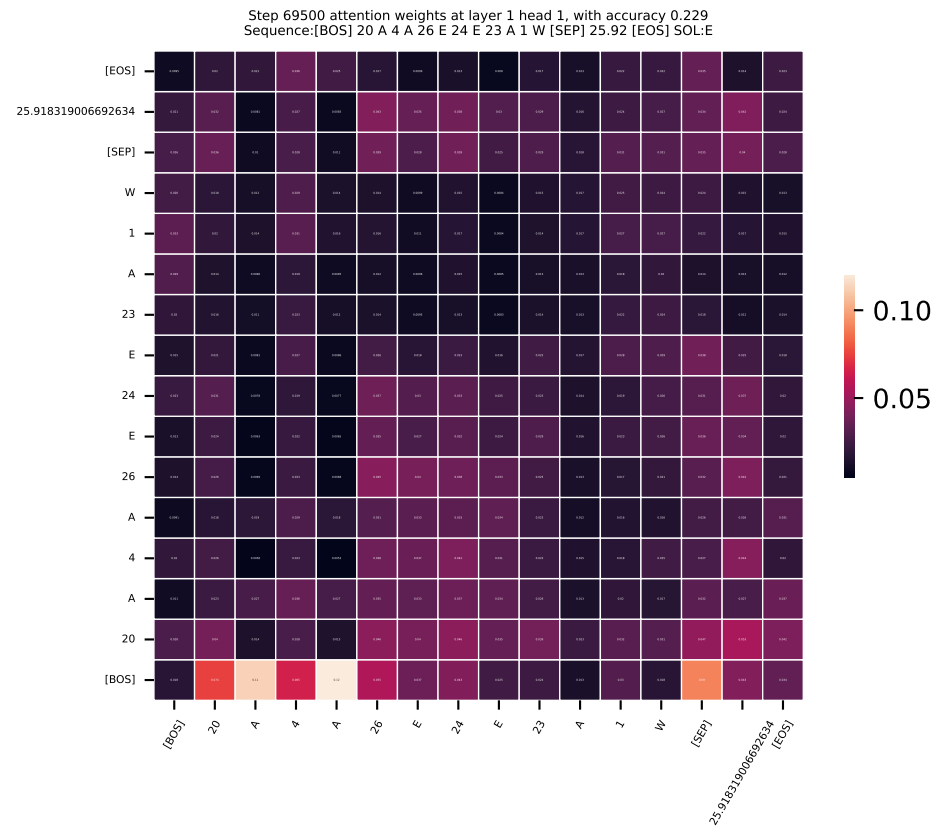

Step 69500 attention weights at layer 0 head 0, with accuracy 0.229
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

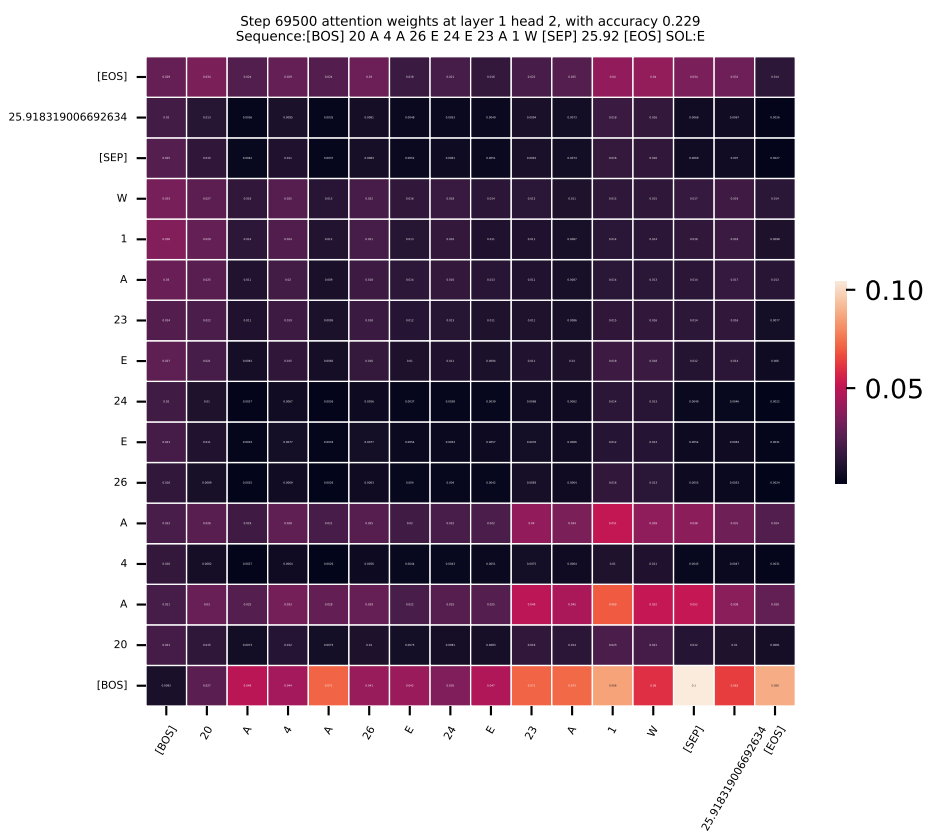

Step 69500 attention weights at layer 0 head 1, with accuracy 0.229
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

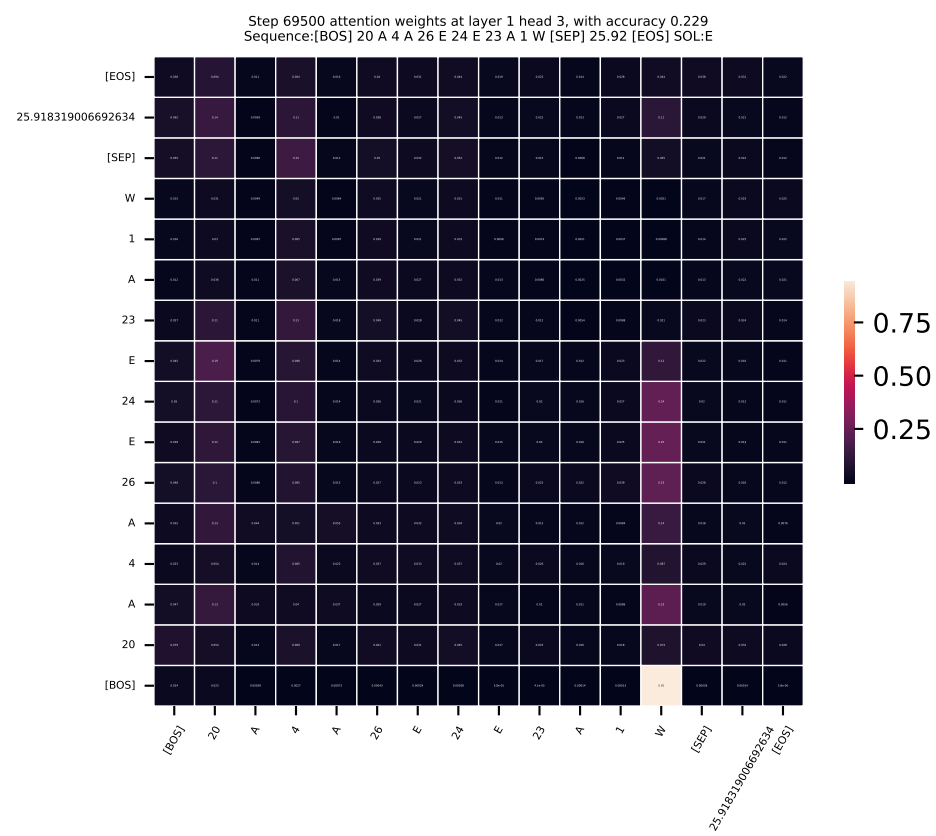

Step 69500 attention weights at layer 0 head 2, with accuracy 0.229
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

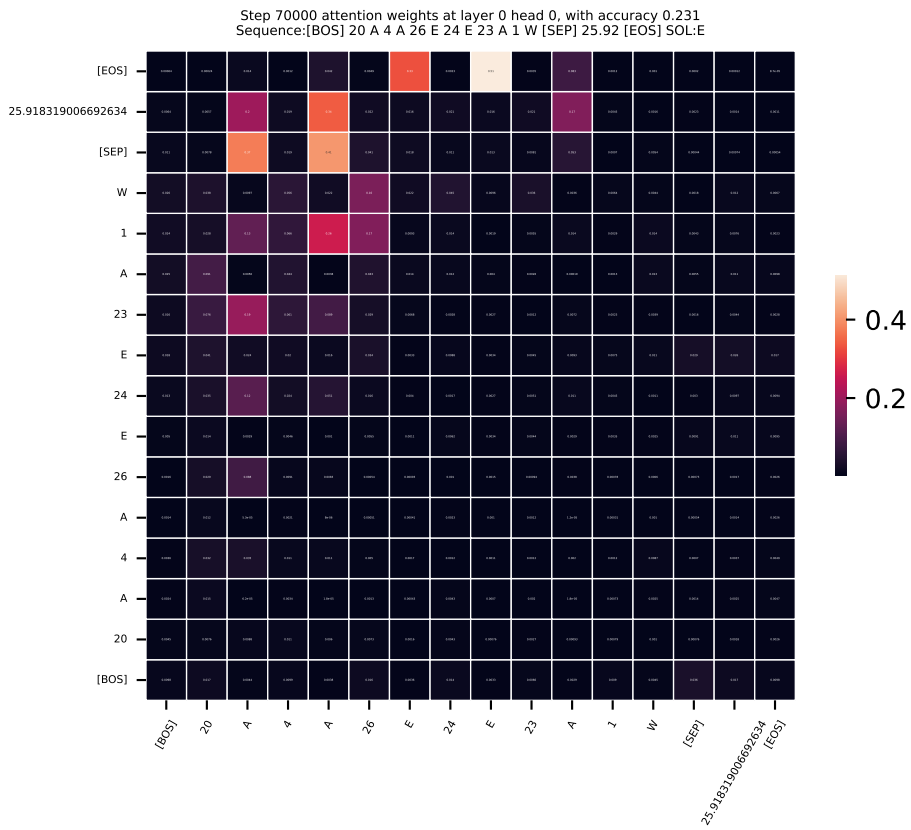

Step 69500 attention weights at layer 0 head 3, with accuracy 0.229
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

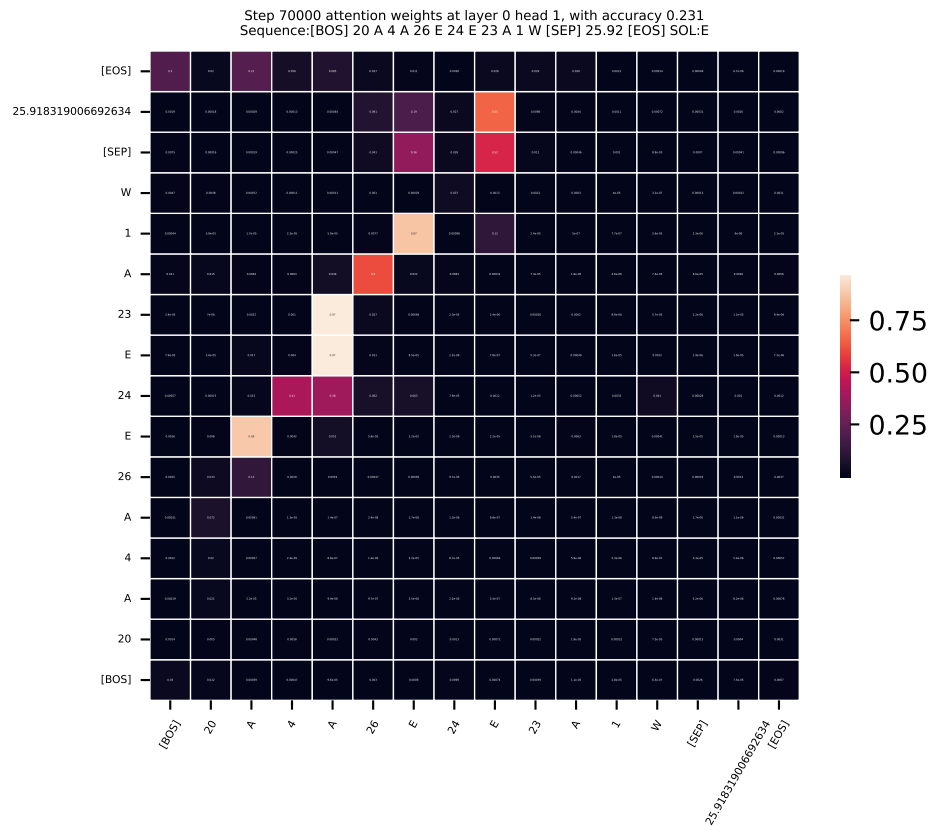

Step 69500 attention weights at layer 1 head 0, with accuracy 0.229
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.10
0.05

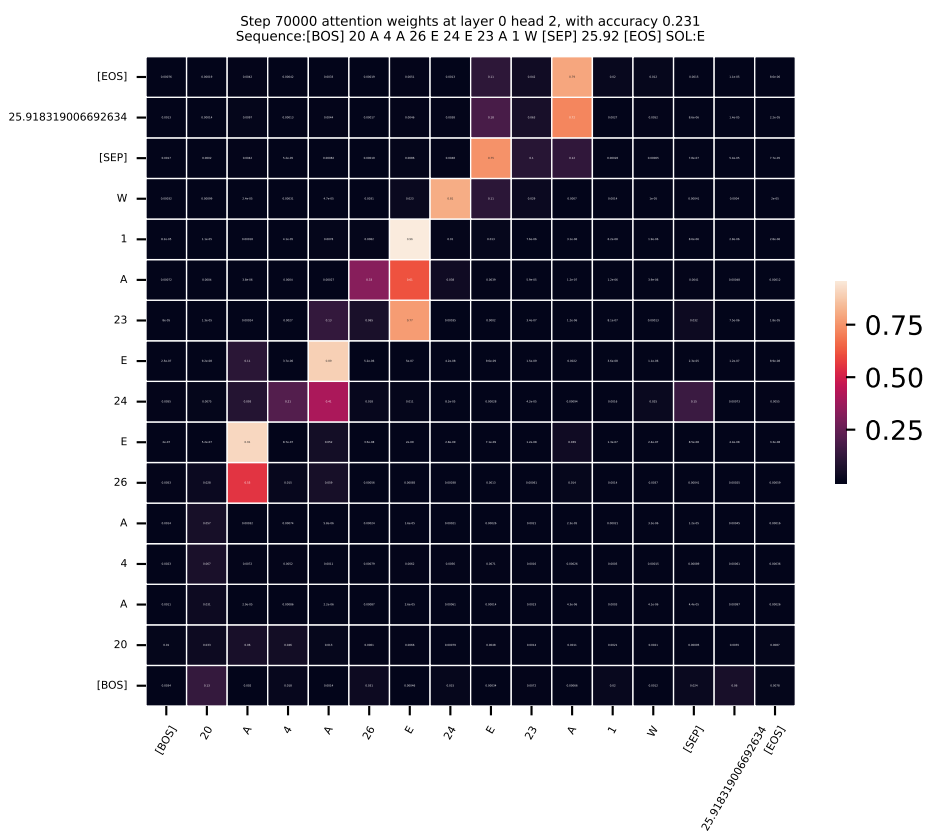

Step 69500 attention weights at layer 1 head 1, with accuracy 0.229
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.10
0.05

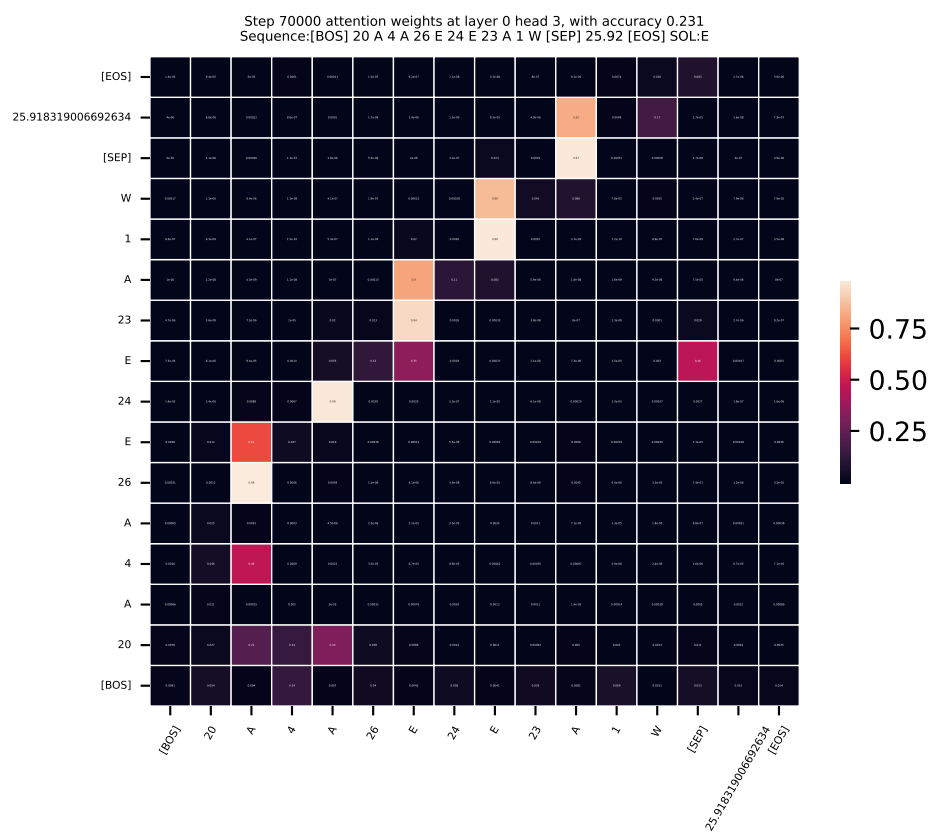

Step 69500 attention weights at layer 1 head 2, with accuracy 0.229
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.10
0.05

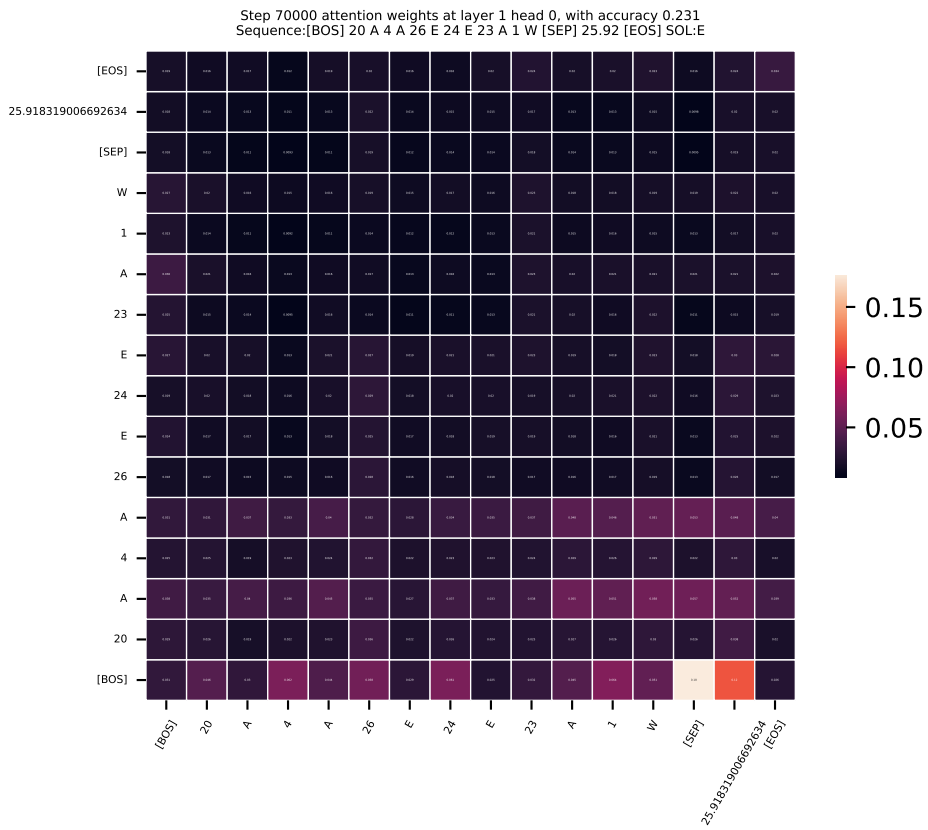

Step 69500 attention weights at layer 1 head 3, with accuracy 0.229
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

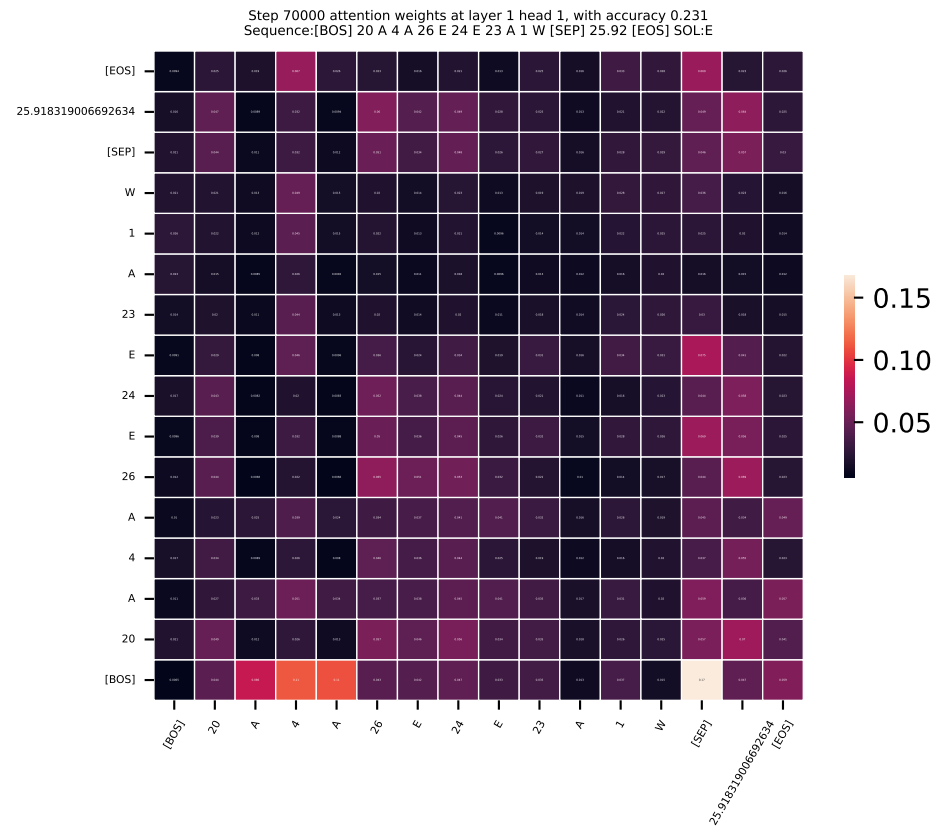

Step 70000 attention weights at layer 0 head 0, with accuracy 0.231
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.4
0.2

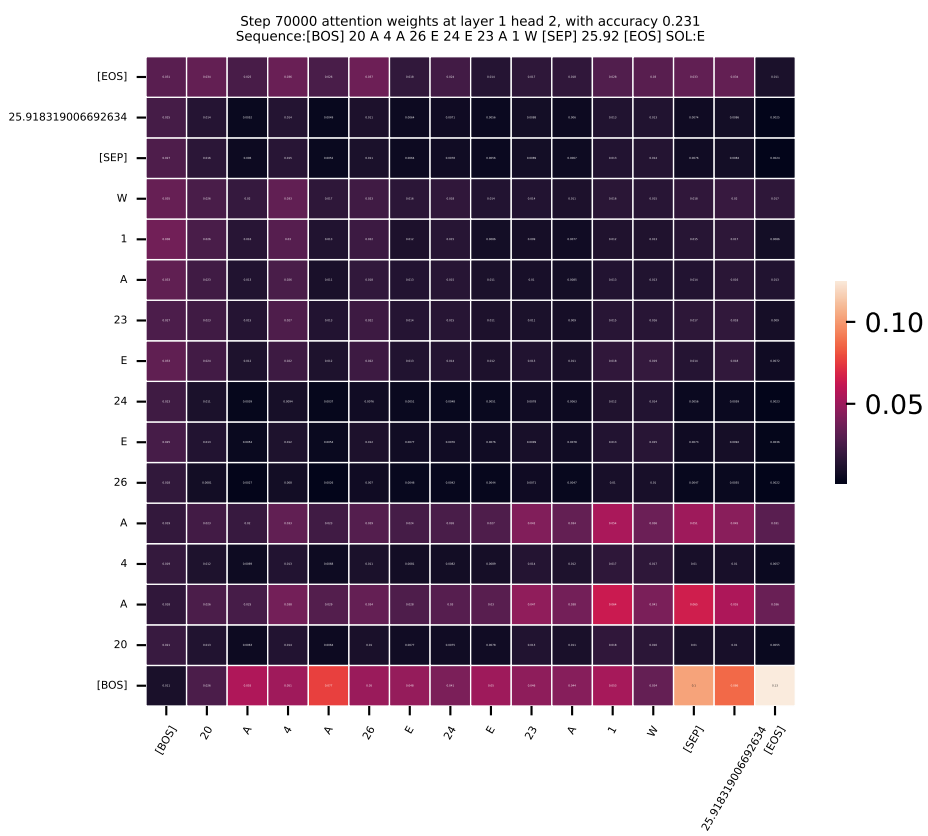

Step 70000 attention weights at layer 0 head 1, with accuracy 0.231
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

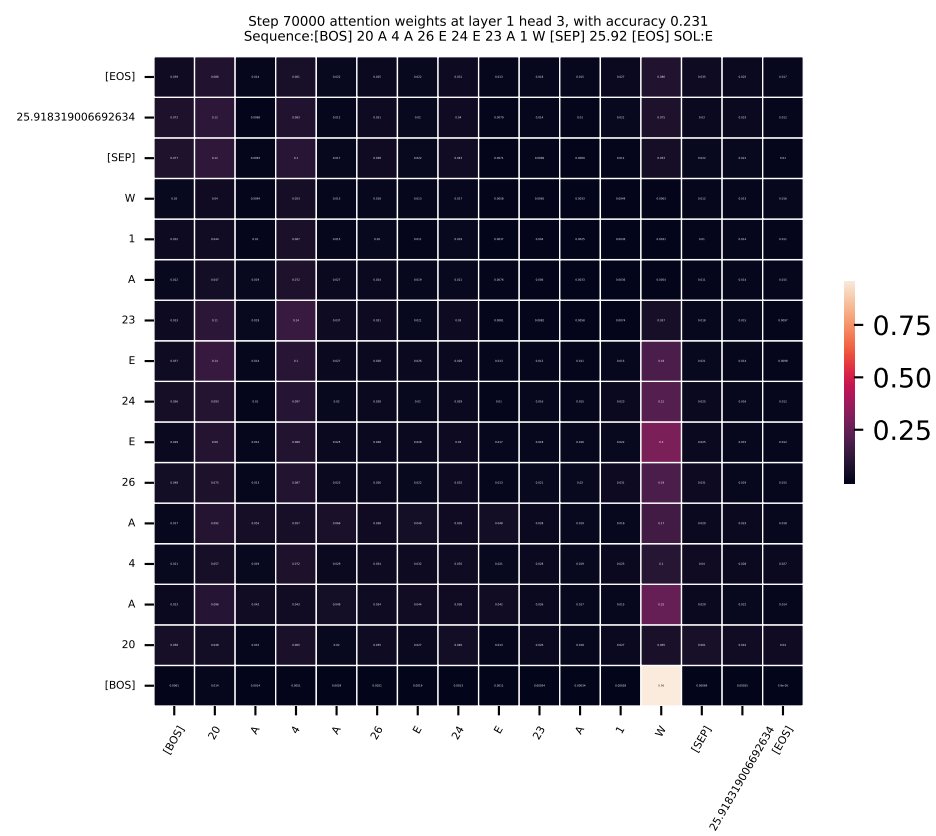

Step 70000 attention weights at layer 0 head 2, with accuracy 0.231
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

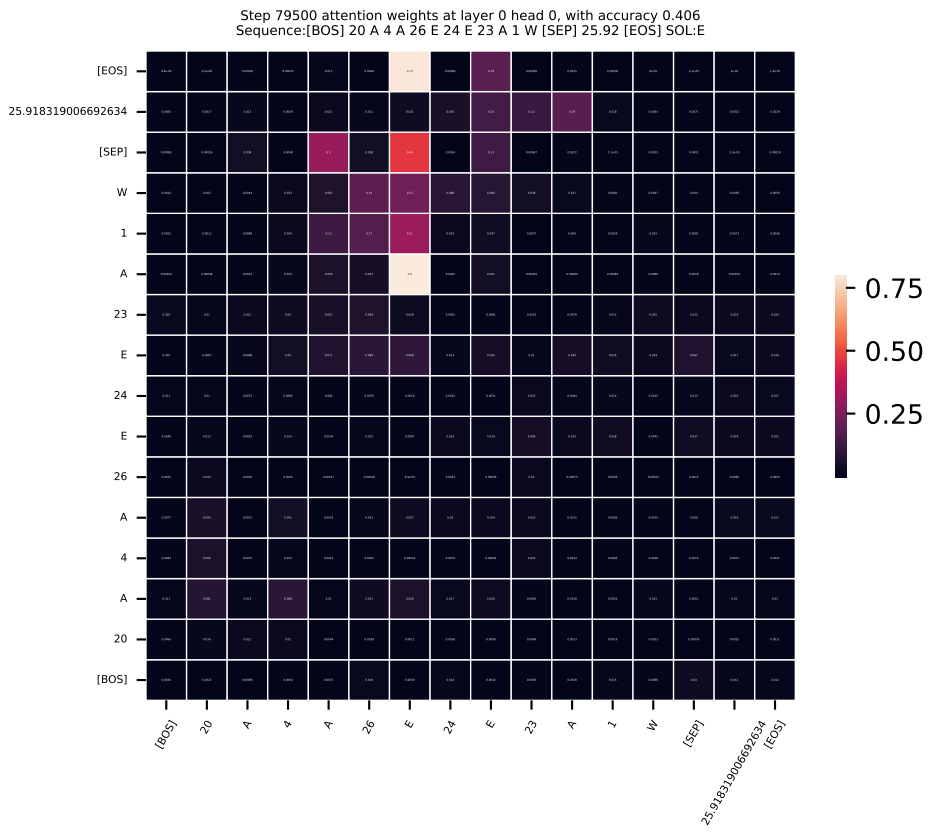

Step 70000 attention weights at layer 0 head 3, with accuracy 0.231
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

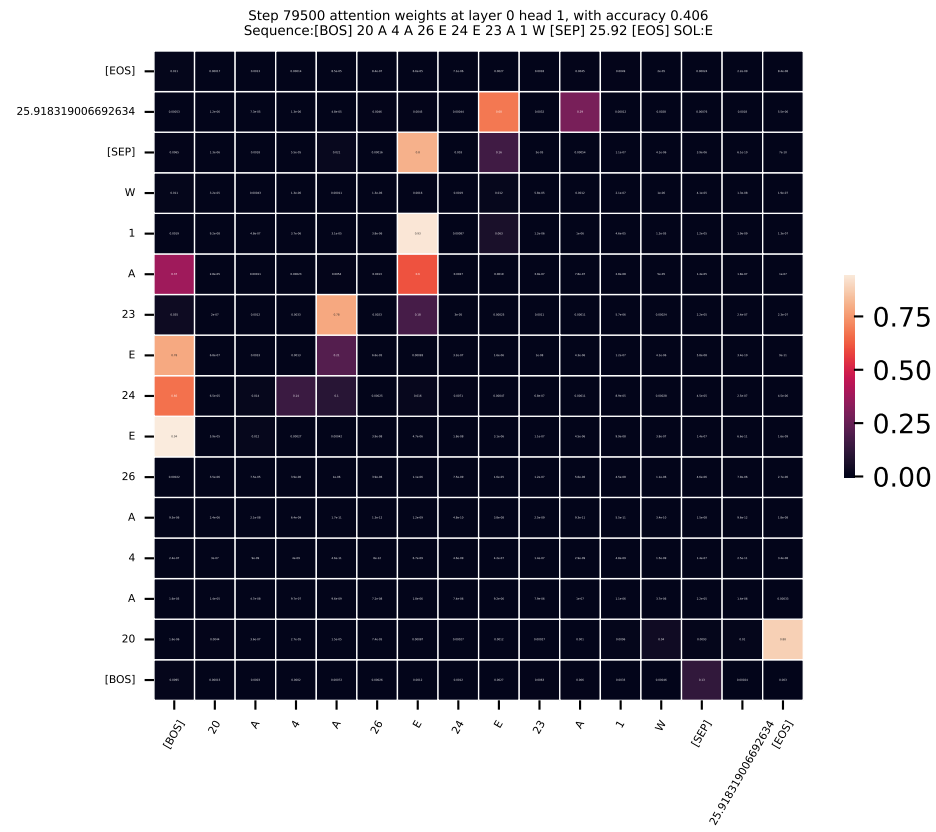

Step 70000 attention weights at layer 1 head 0, with accuracy 0.231
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.15
0.10
0.05

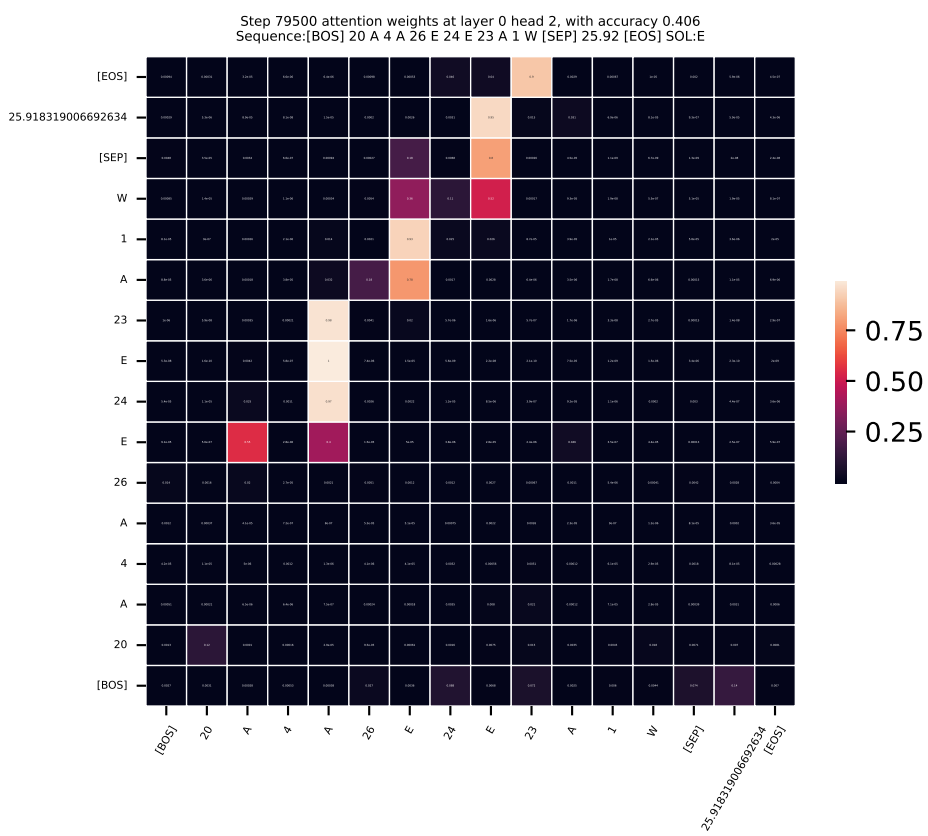

Step 70000 attention weights at layer 1 head 1, with accuracy 0.231
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.15
0.10
0.05

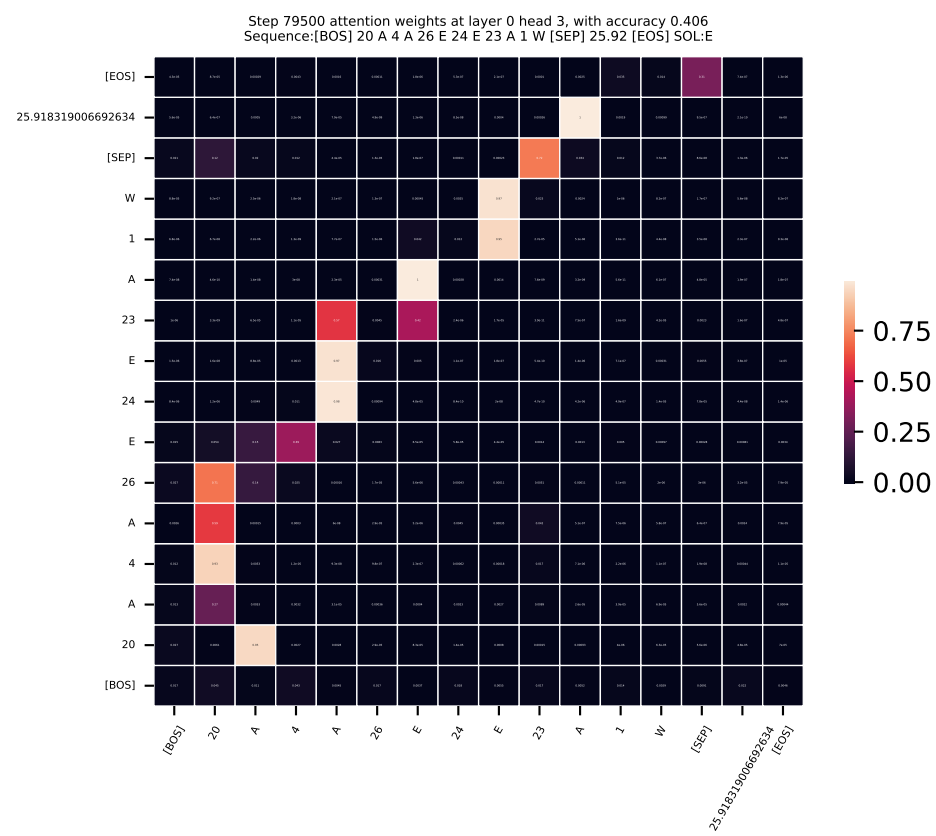

Step 70000 attention weights at layer 1 head 2, with accuracy 0.231
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.10
0.05

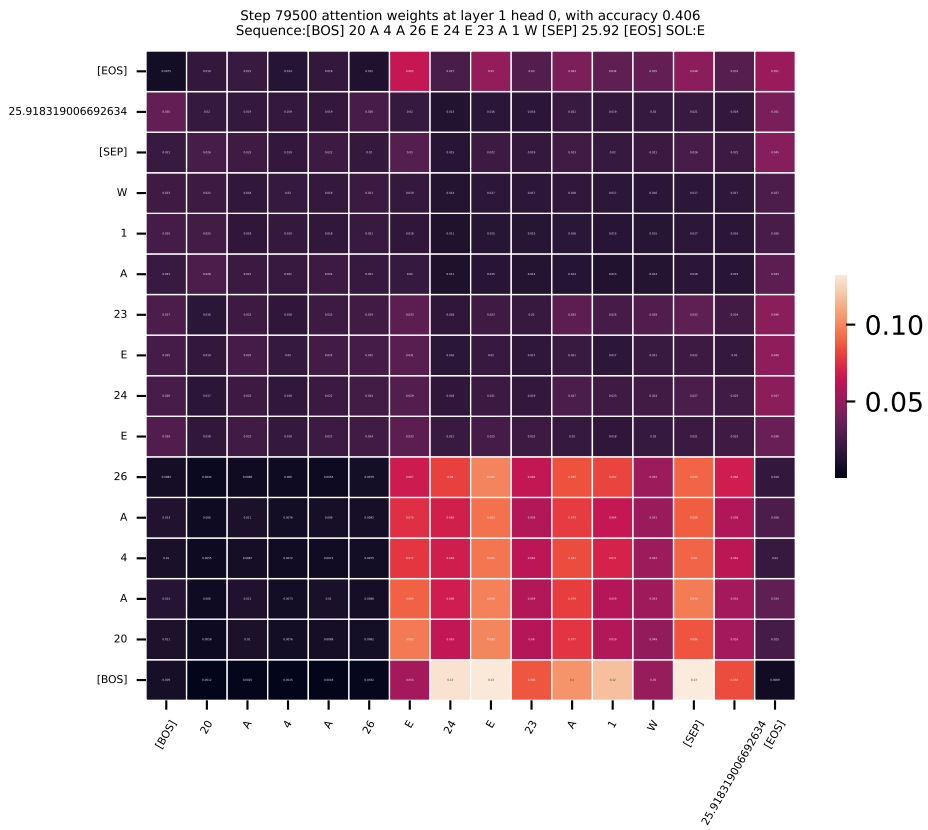

Step 70000 attention weights at layer 1 head 3, with accuracy 0.231
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

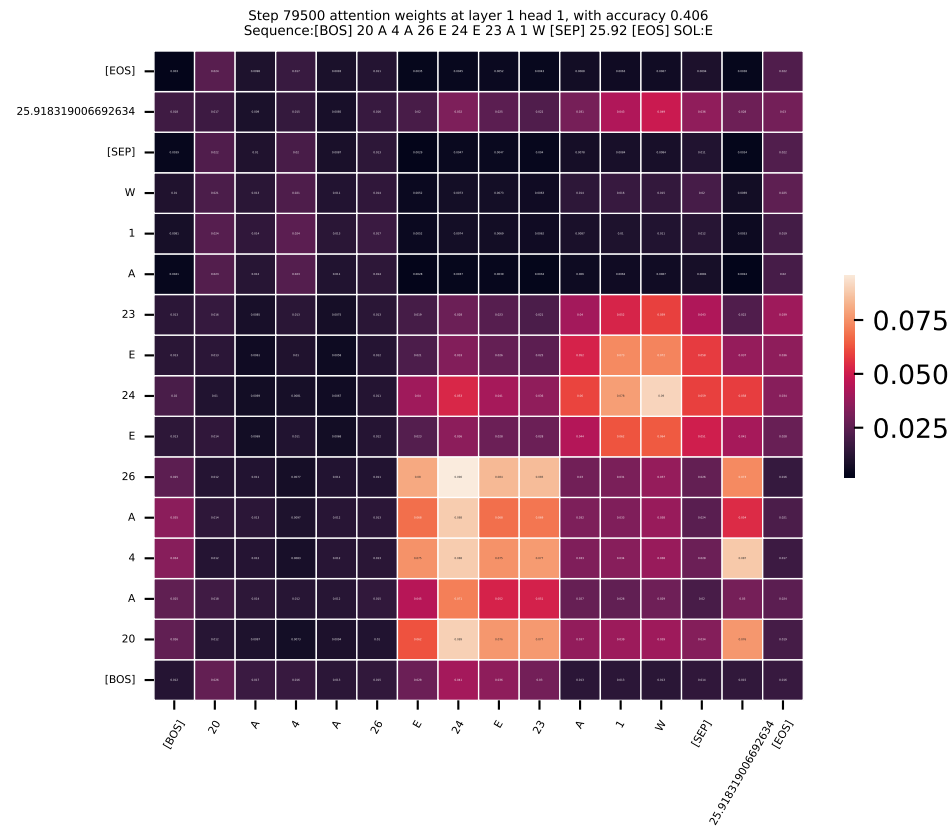

Step 79500 attention weights at layer 0 head 0, with accuracy 0.406
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

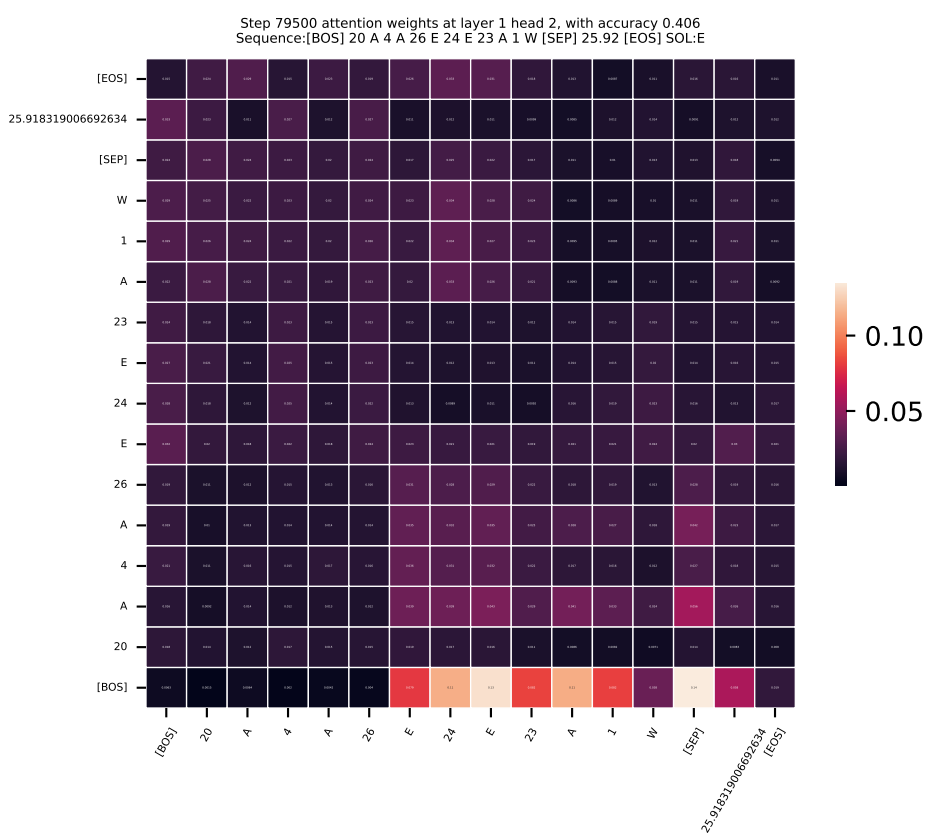

Step 79500 attention weights at layer 0 head 1, with accuracy 0.406
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

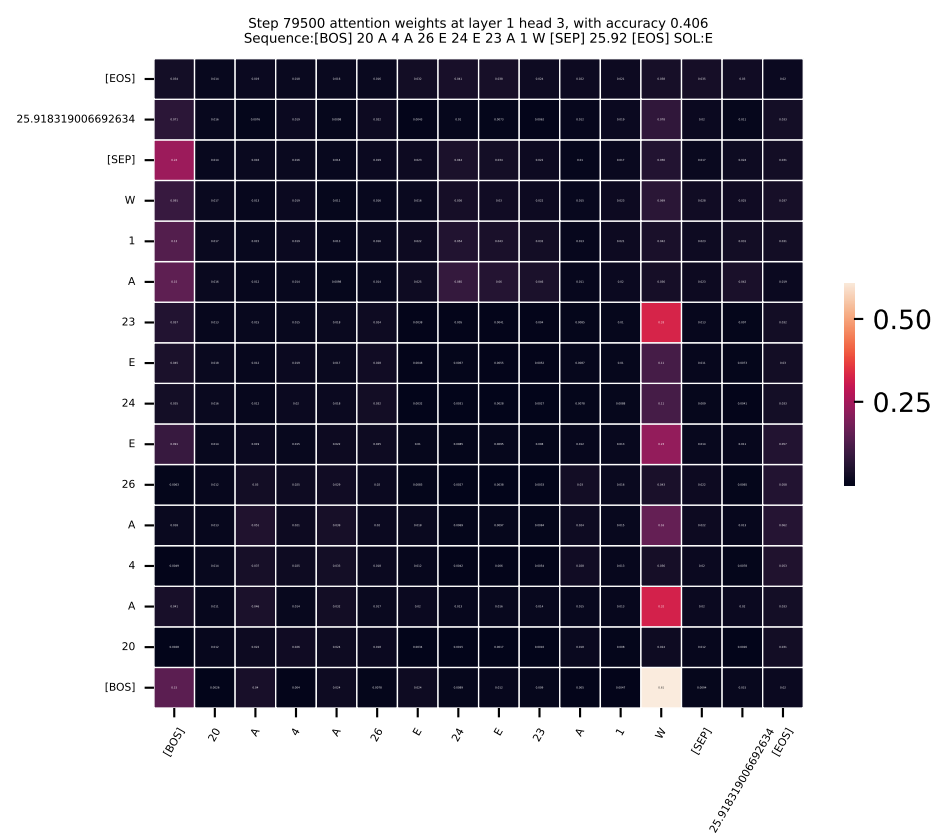

Step 79500 attention weights at layer 0 head 2, with accuracy 0.406
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

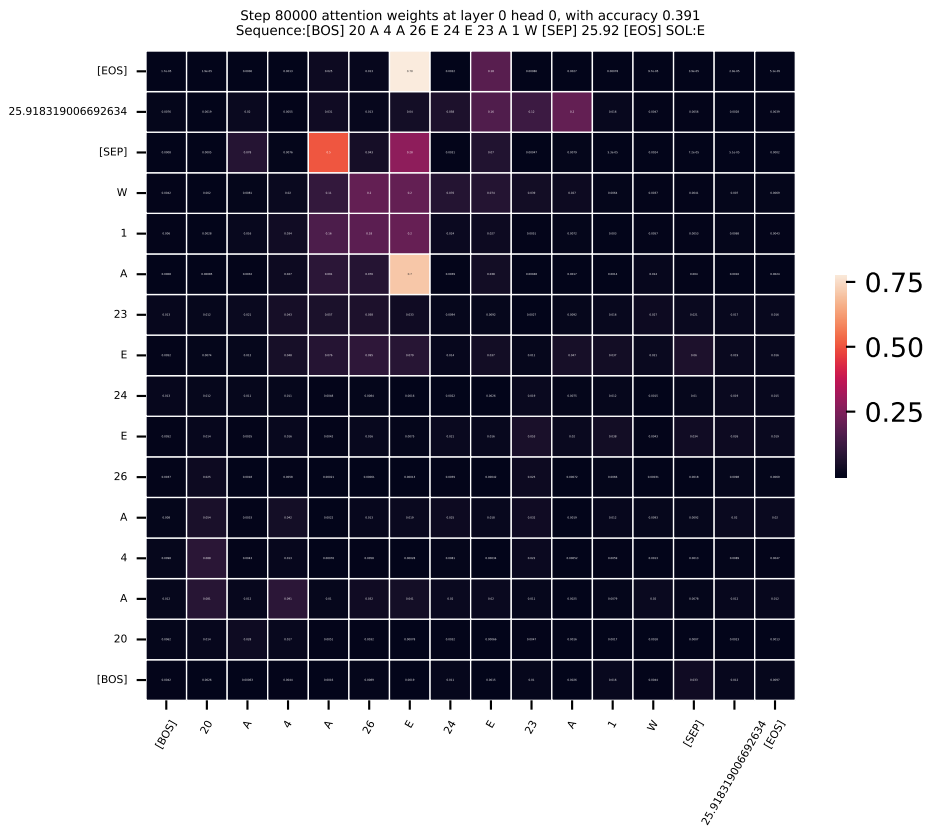

Step 79500 attention weights at layer 0 head 3, with accuracy 0.406
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

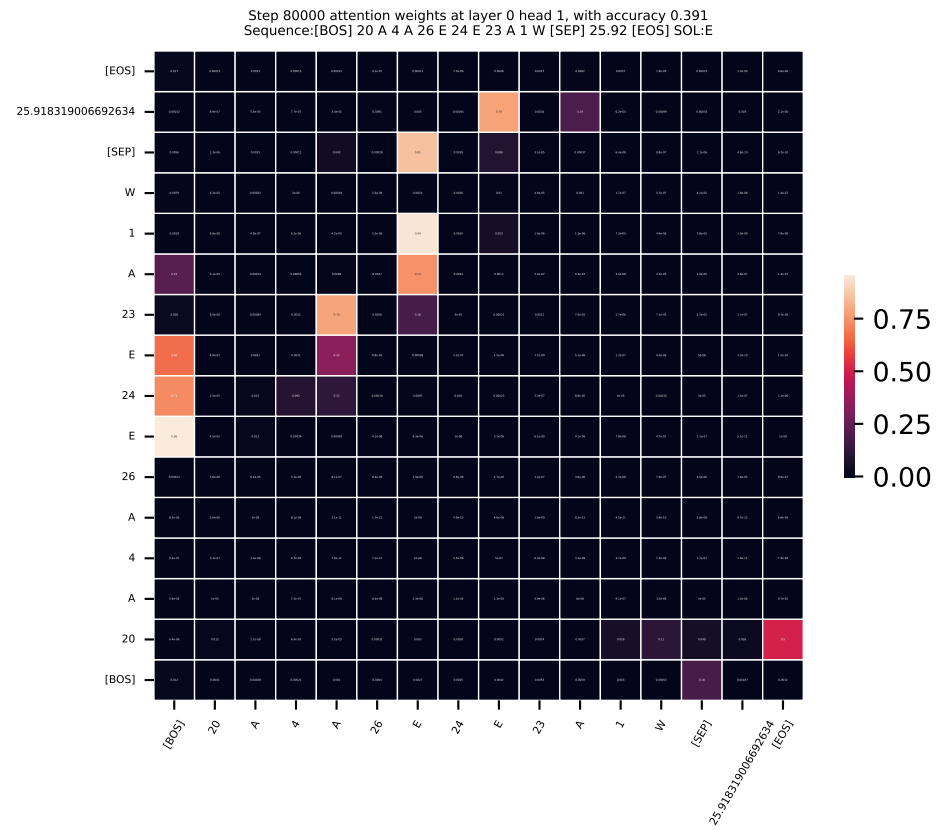

Step 79500 attention weights at layer 1 head 0, with accuracy 0.406
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.10
0.05

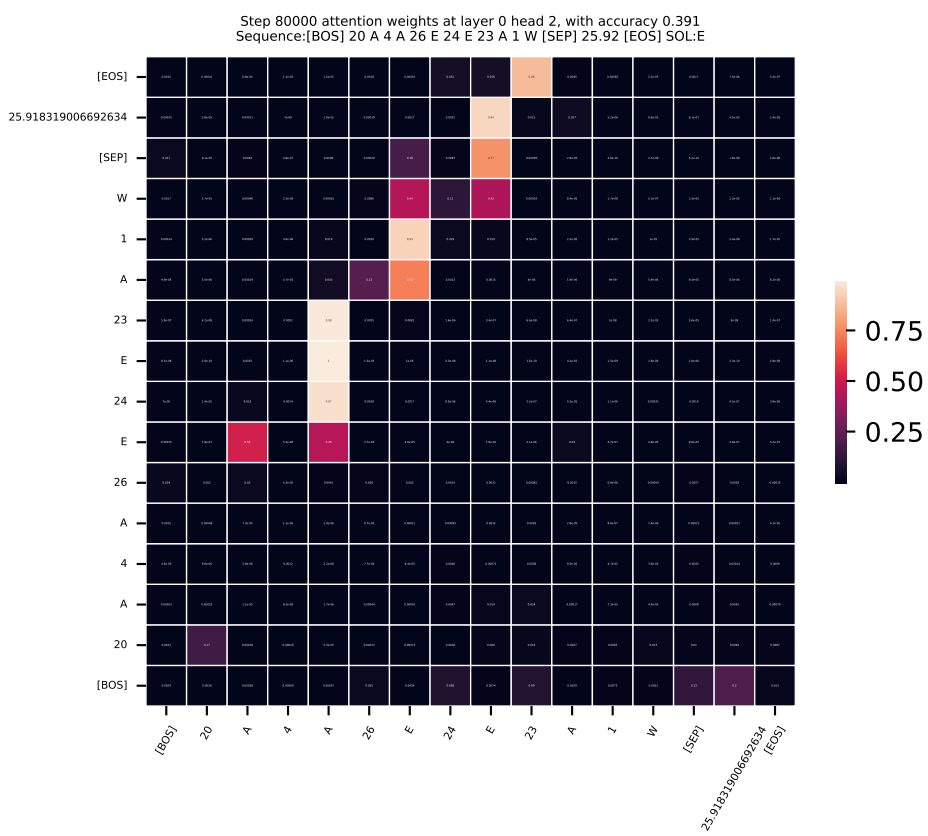

Step 79500 attention weights at layer 1 head 1, with accuracy 0.406
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.075
0.050
0.025

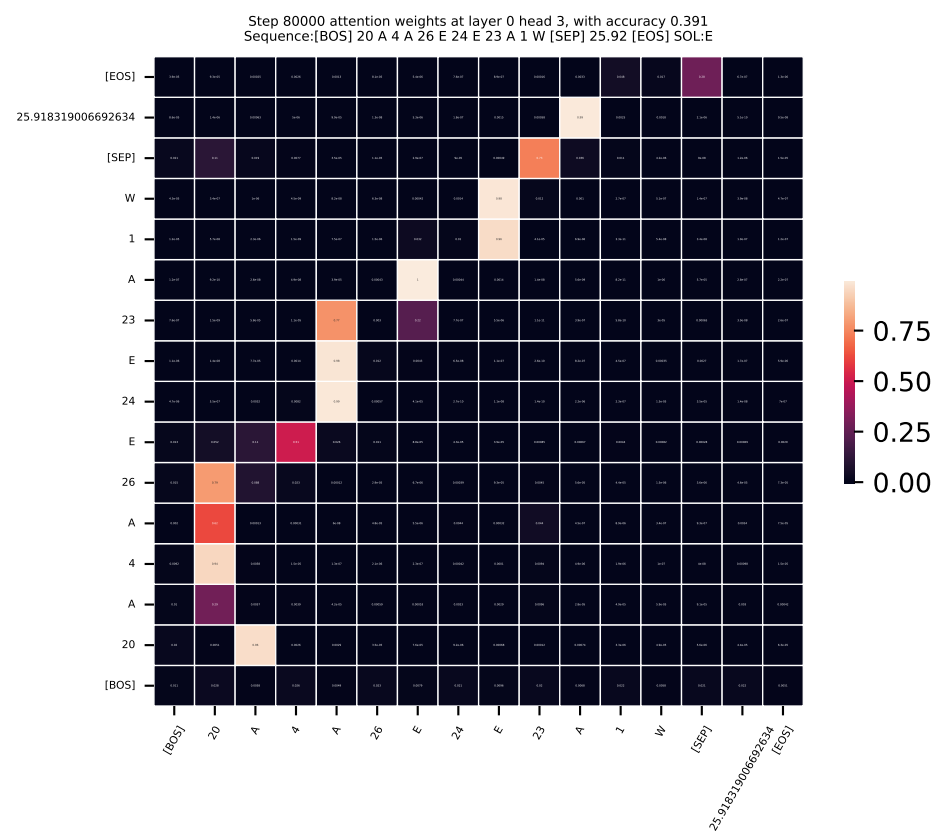

Step 79500 attention weights at layer 1 head 2, with accuracy 0.406
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.10
0.05

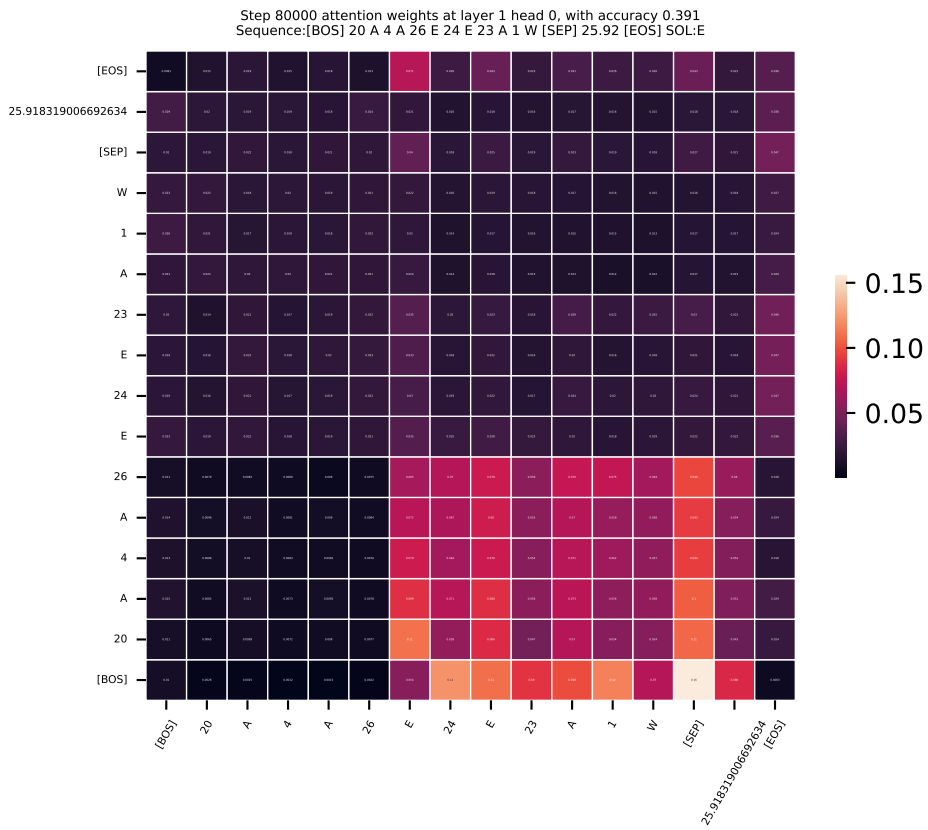

Step 79500 attention weights at layer 1 head 3, with accuracy 0.406
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

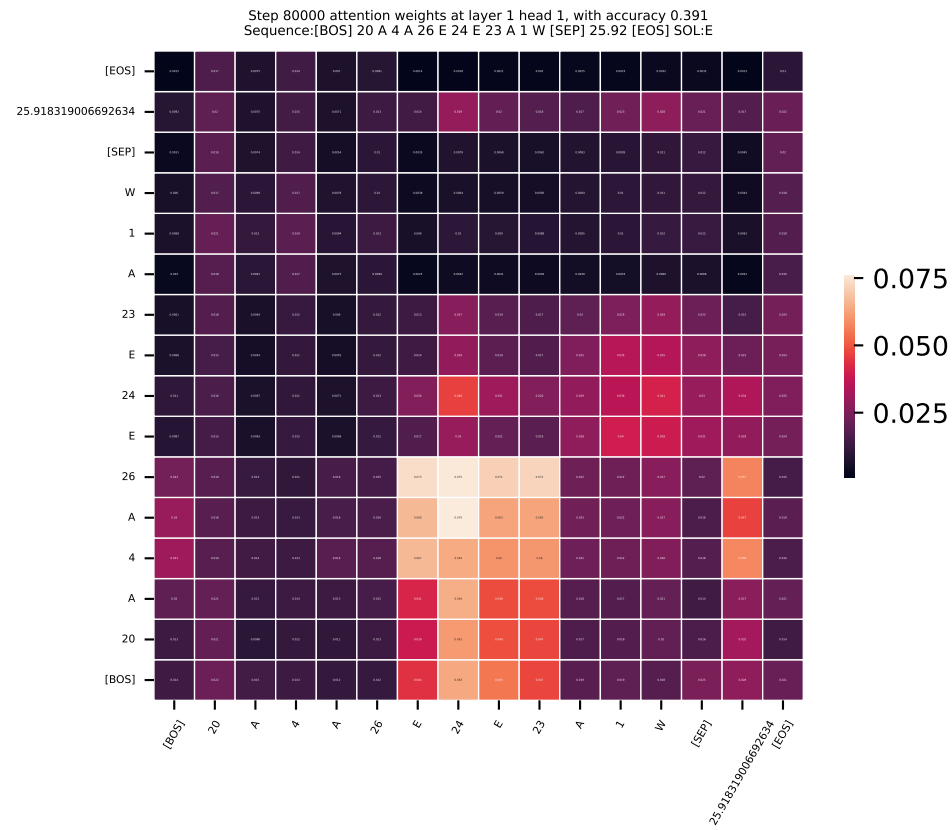

Step 80000 attention weights at layer 0 head 0, with accuracy 0.391
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

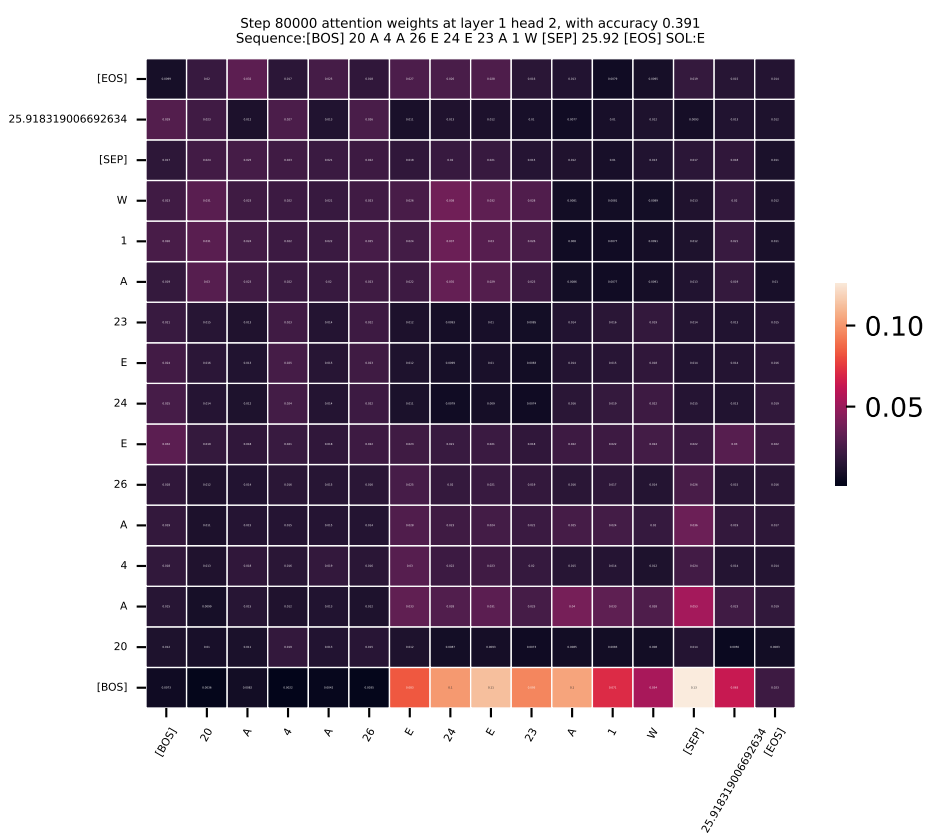

Step 80000 attention weights at layer 0 head 1, with accuracy 0.391
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

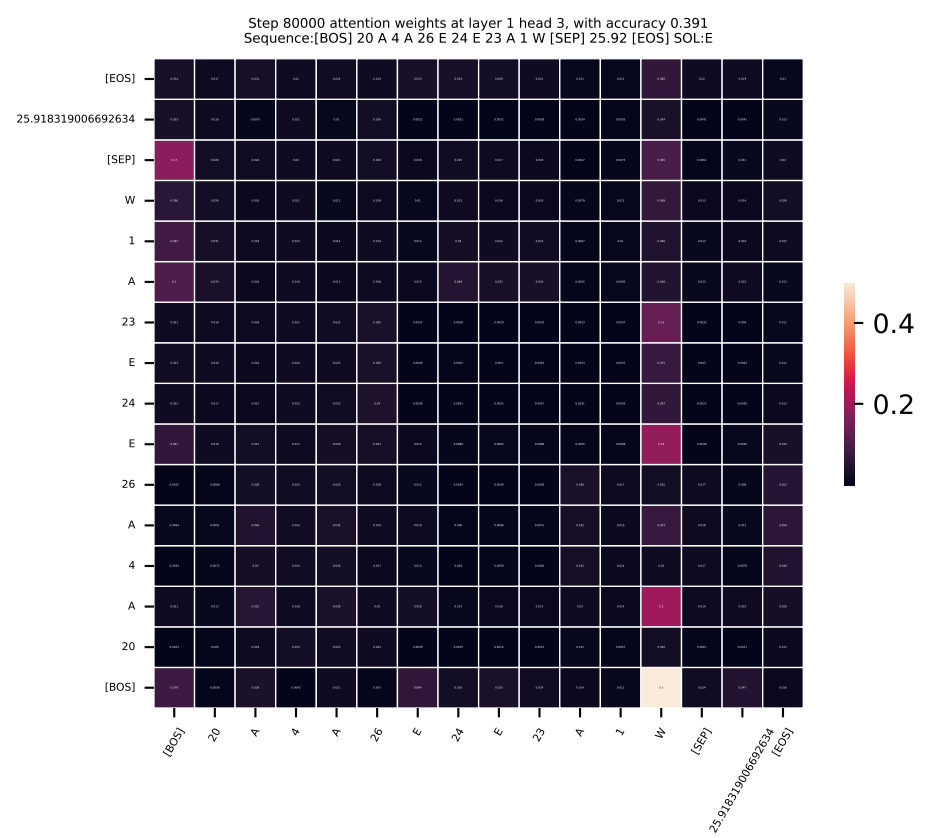

Step 80000 attention weights at layer 0 head 2, with accuracy 0.391
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

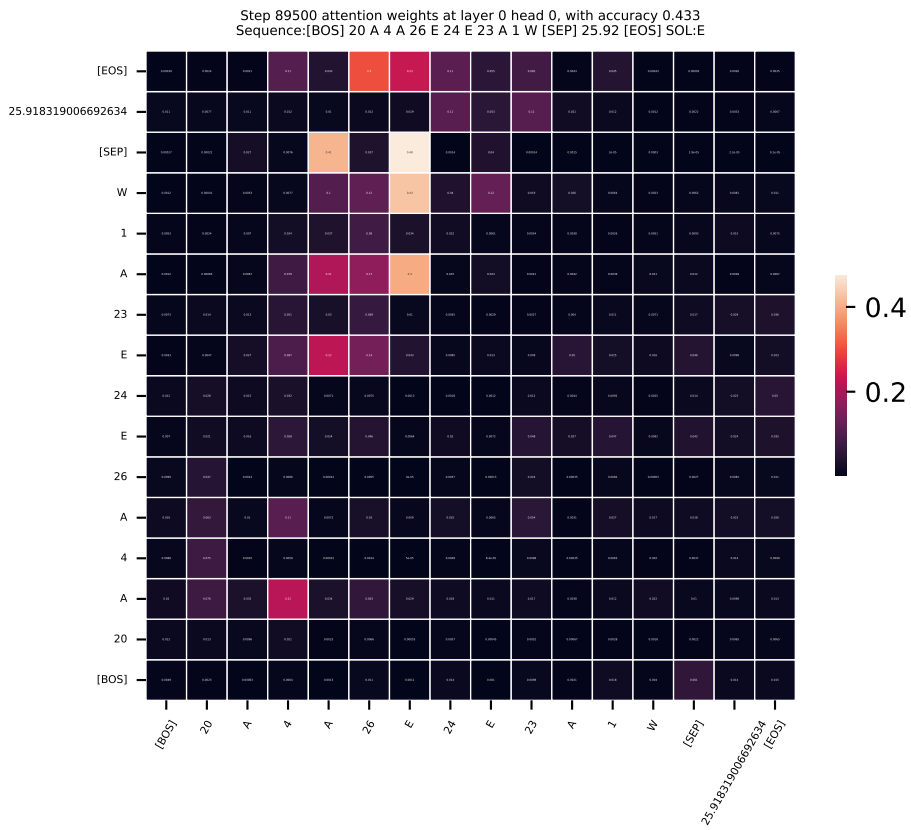

Step 80000 attention weights at layer 0 head 3, with accuracy 0.391
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

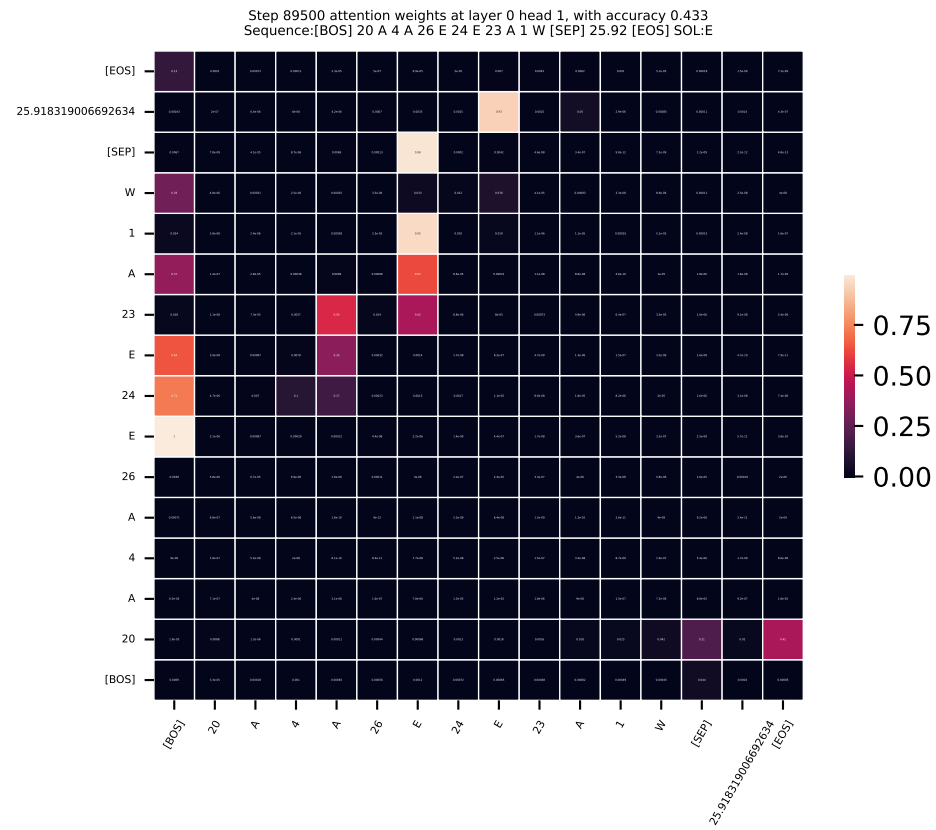

Step 80000 attention weights at layer 1 head 0, with accuracy 0.391
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.15
0.10
0.05

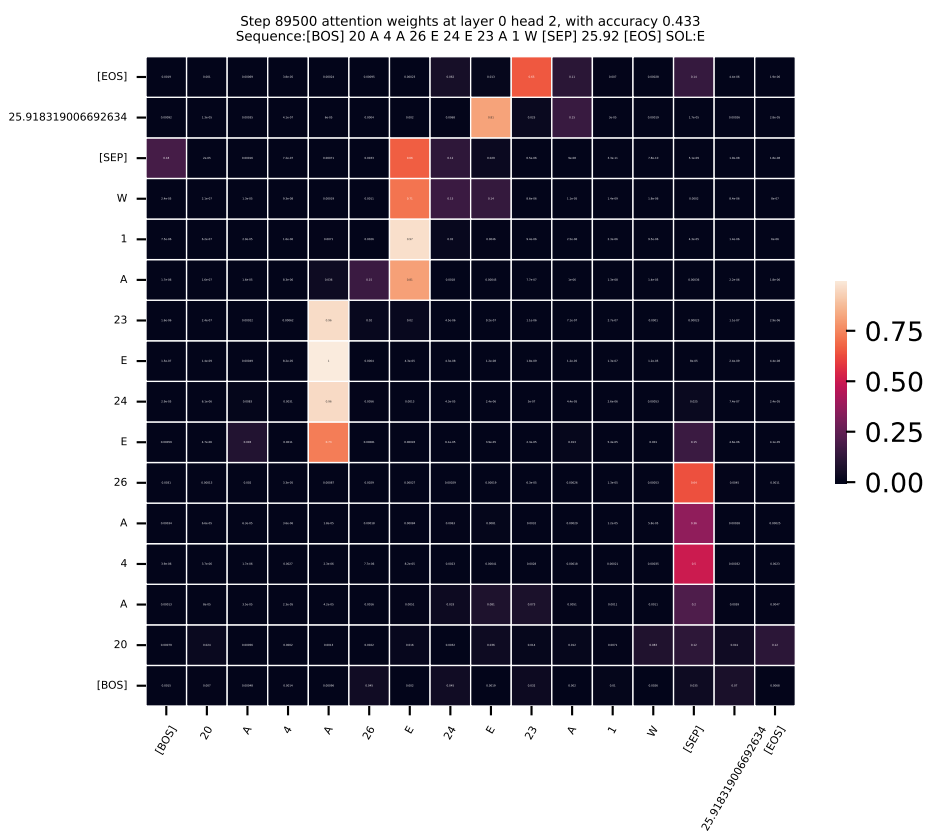

Step 80000 attention weights at layer 1 head 1, with accuracy 0.391
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.075
0.050
0.025

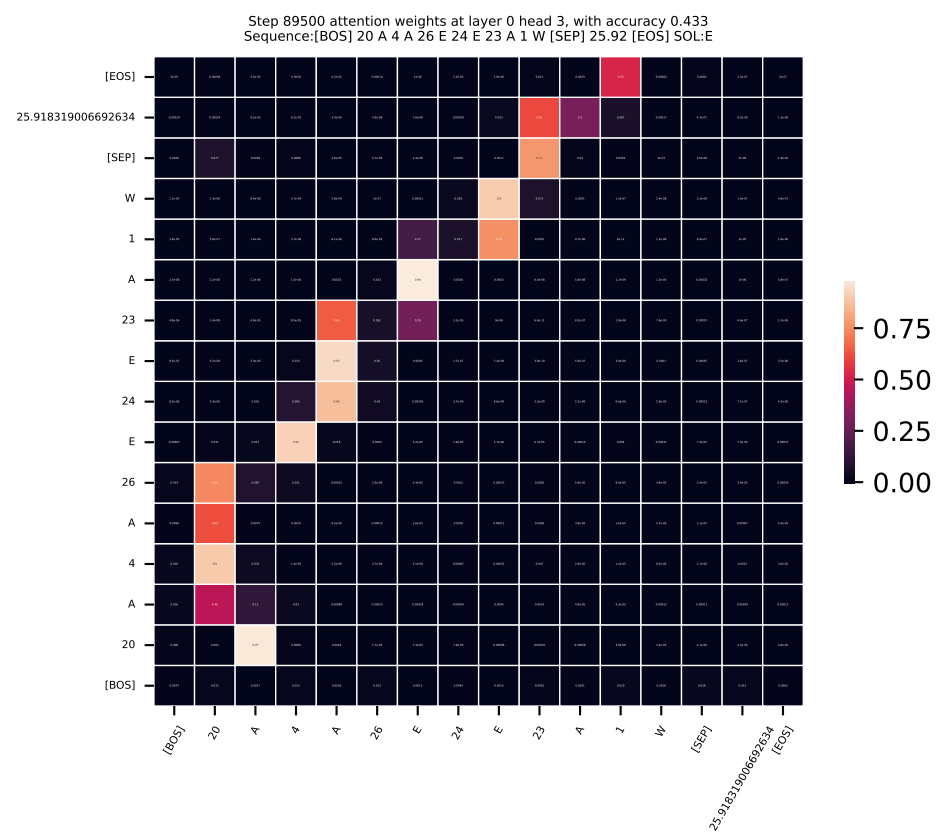

Step 80000 attention weights at layer 1 head 2, with accuracy 0.391
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.10
0.05

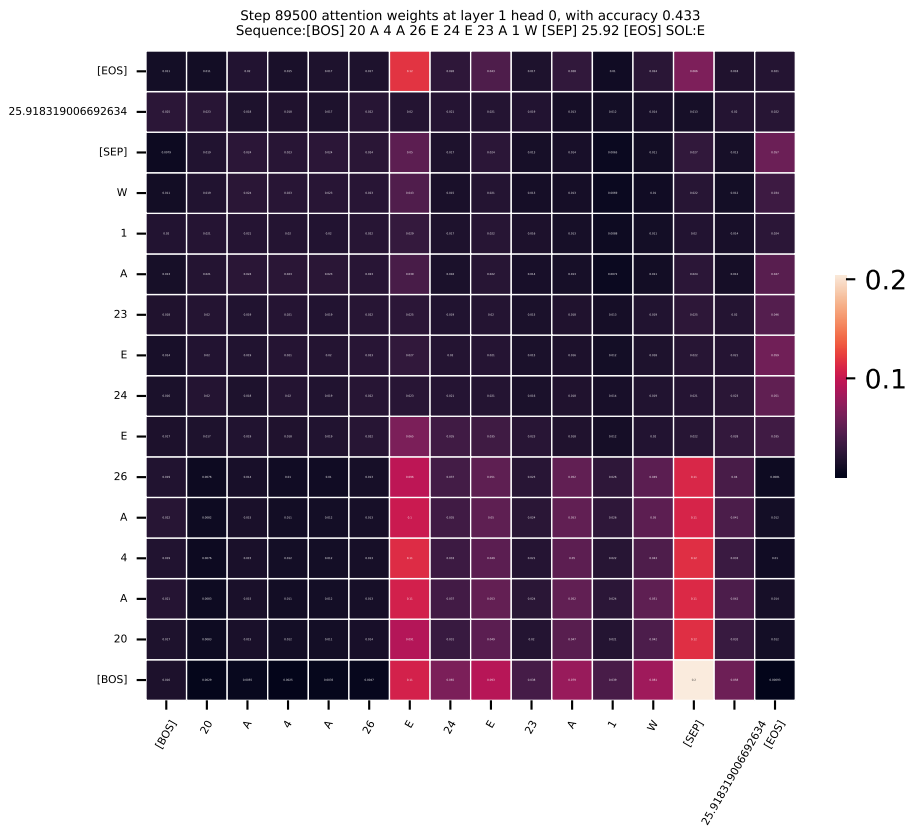

Step 80000 attention weights at layer 1 head 3, with accuracy 0.391
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

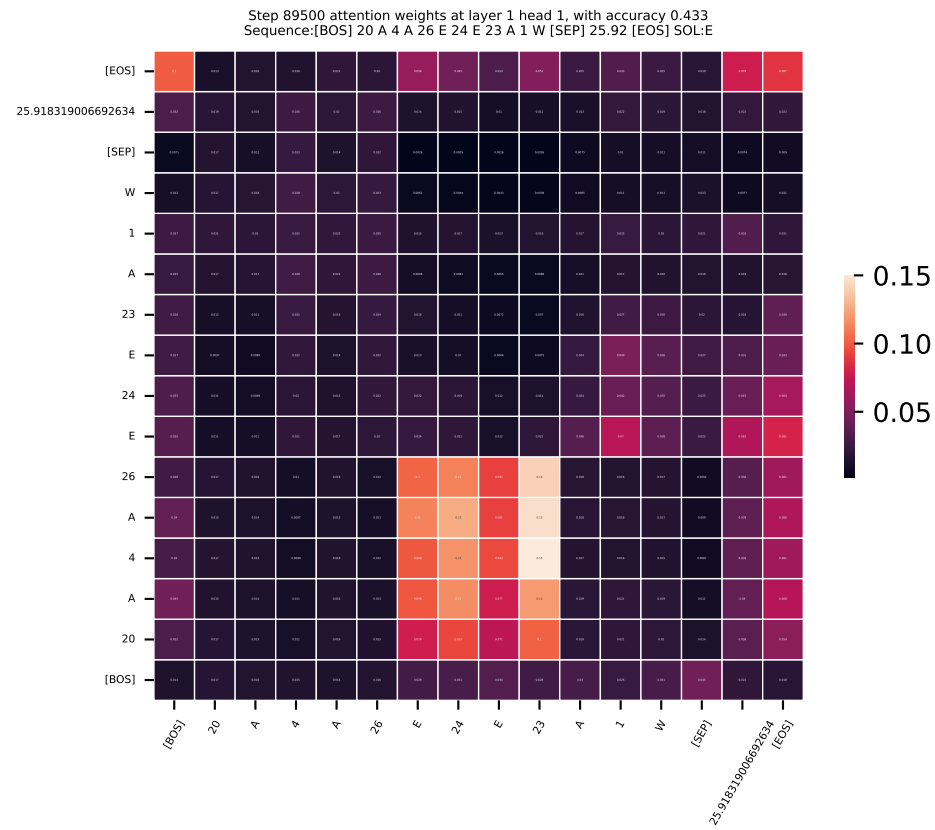

Step 89500 attention weights at layer 0 head 0, with accuracy 0.433
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

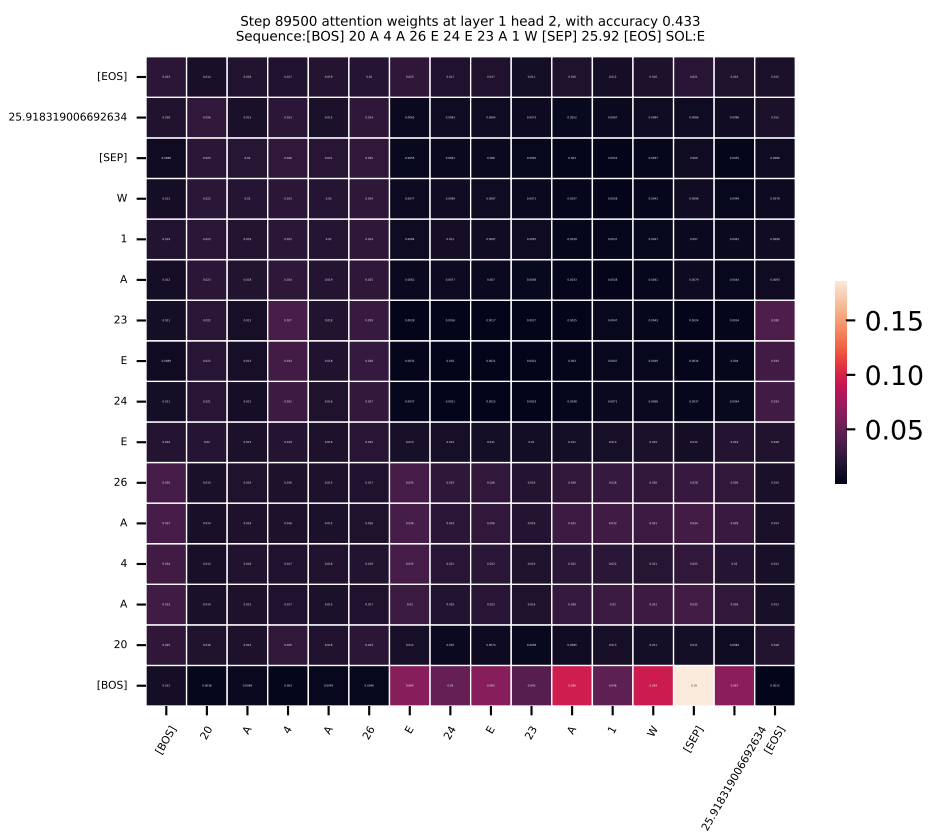

Step 89500 attention weights at layer 0 head 1, with accuracy 0.433
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

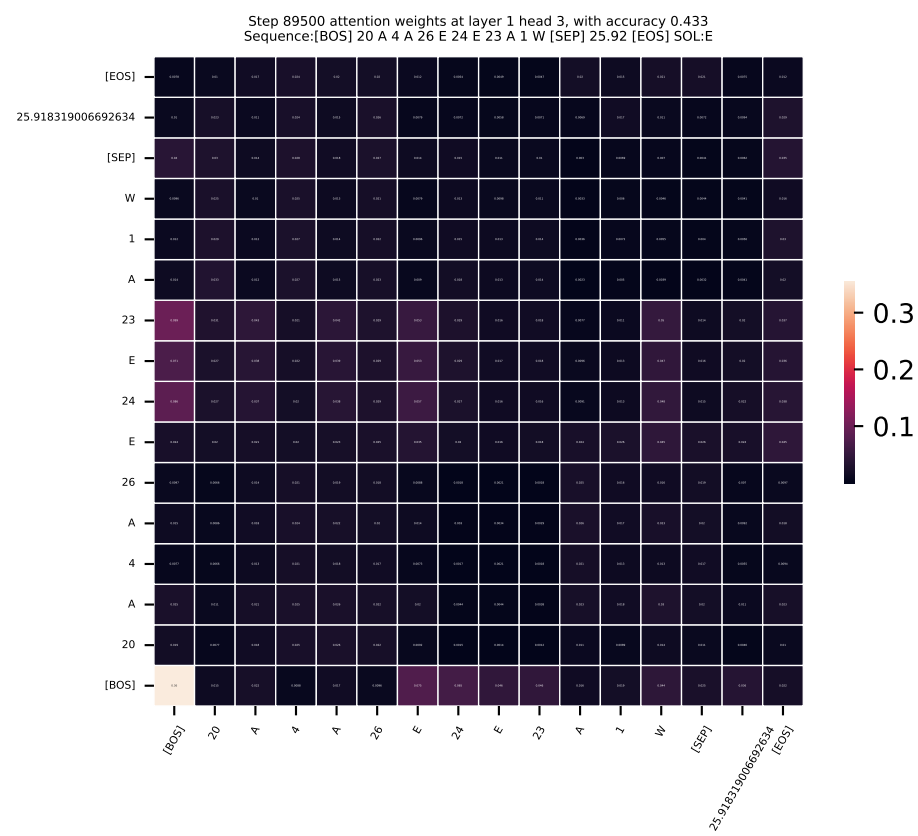

Step 89500 attention weights at layer 0 head 2, with accuracy 0.433
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

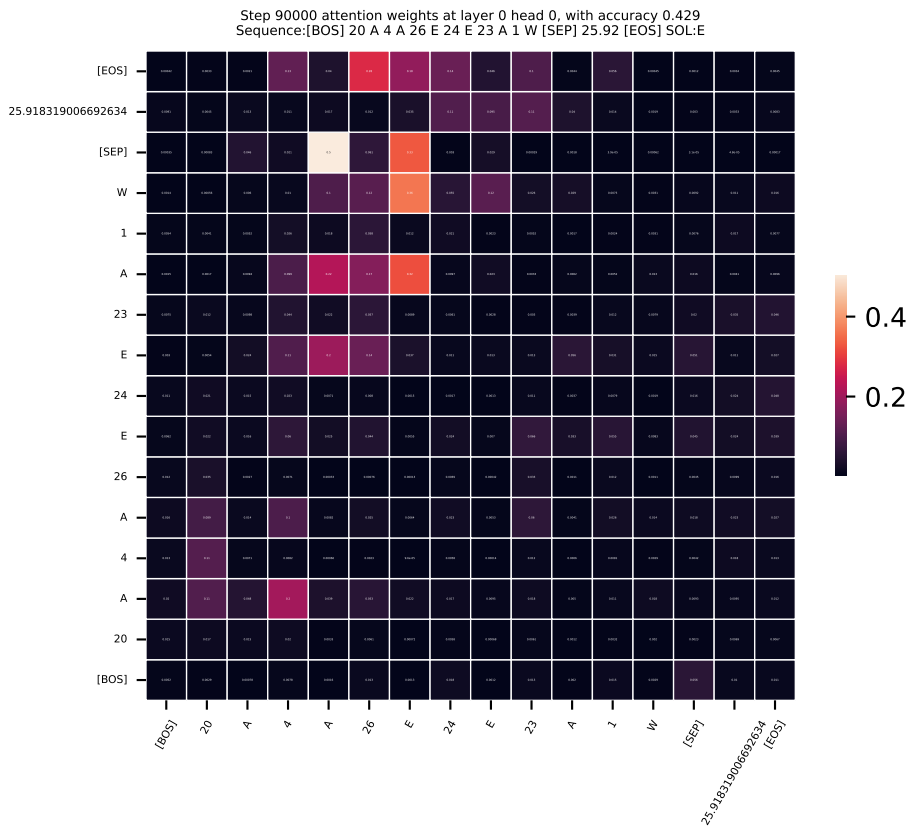

Step 89500 attention weights at layer 0 head 3, with accuracy 0.433
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

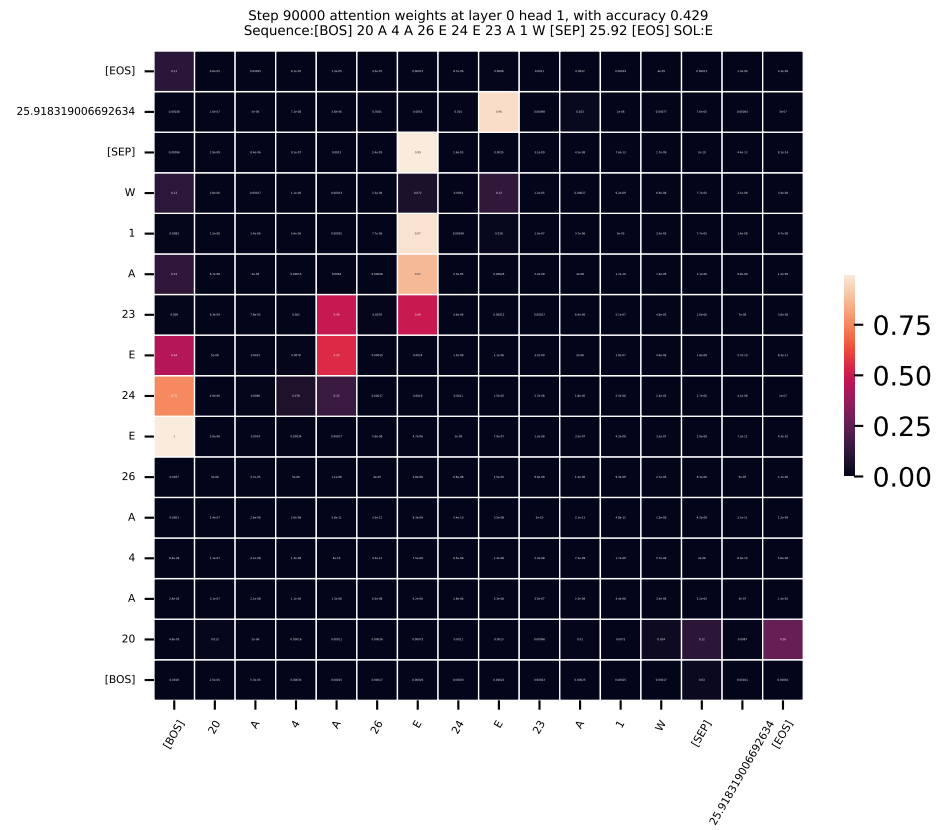

Step 89500 attention weights at layer 1 head 0, with accuracy 0.433
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.2
0.1

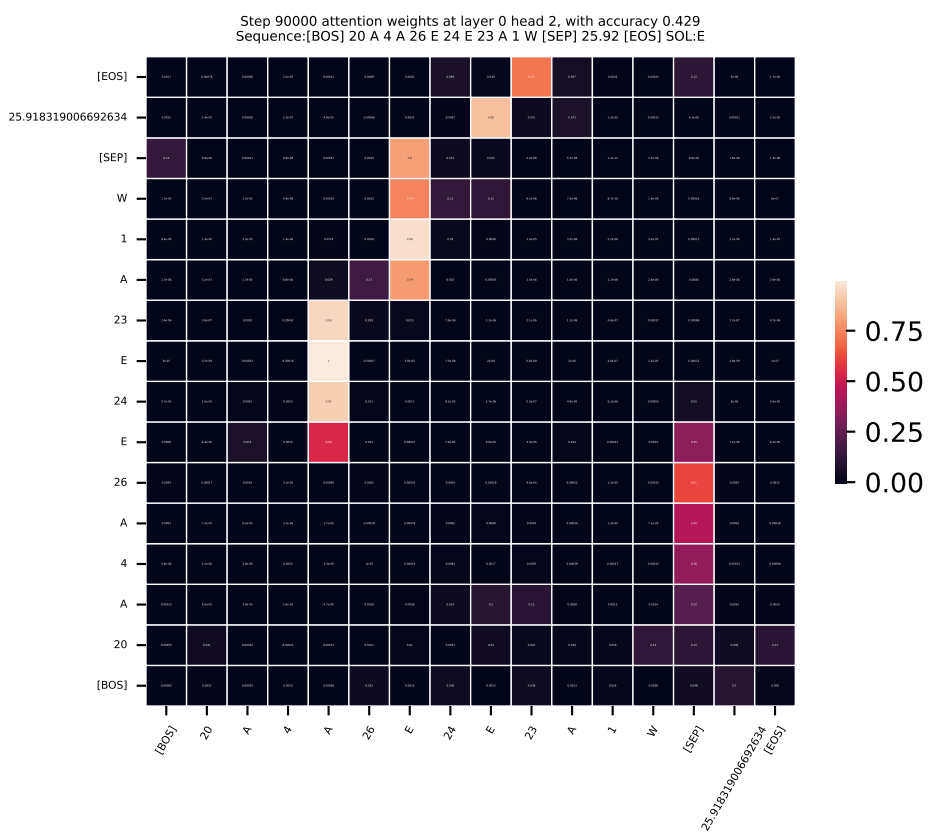

Step 89500 attention weights at layer 1 head 1, with accuracy 0.433
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.15
0.10
0.05

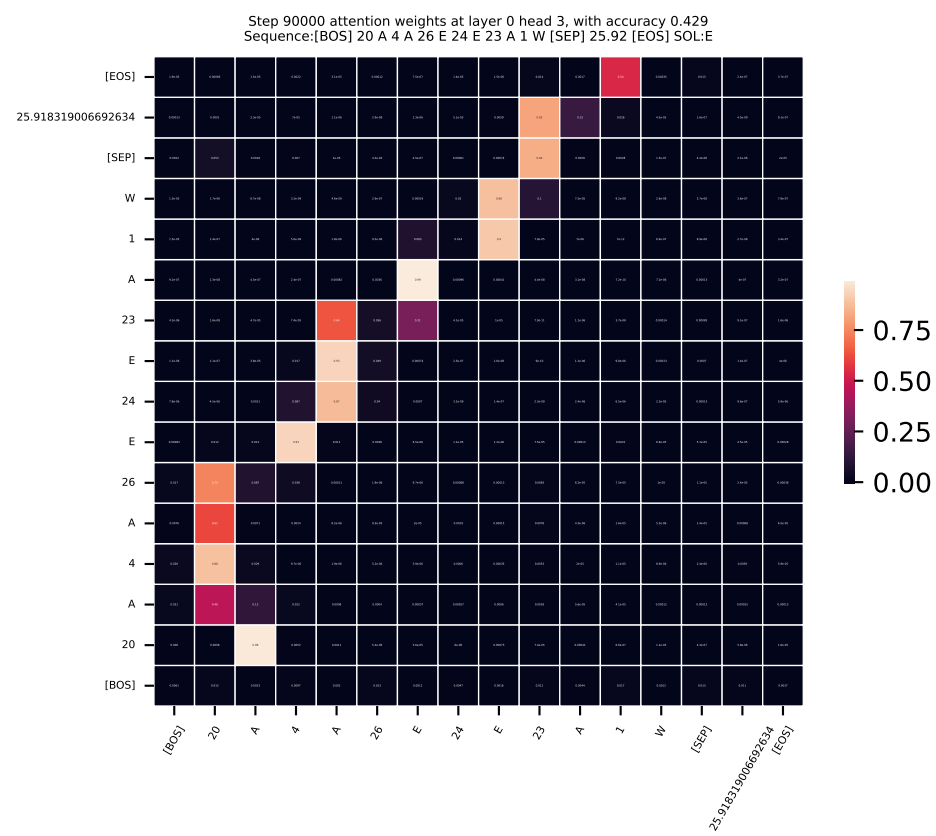

Step 89500 attention weights at layer 1 head 2, with accuracy 0.433
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.15
0.10
0.05

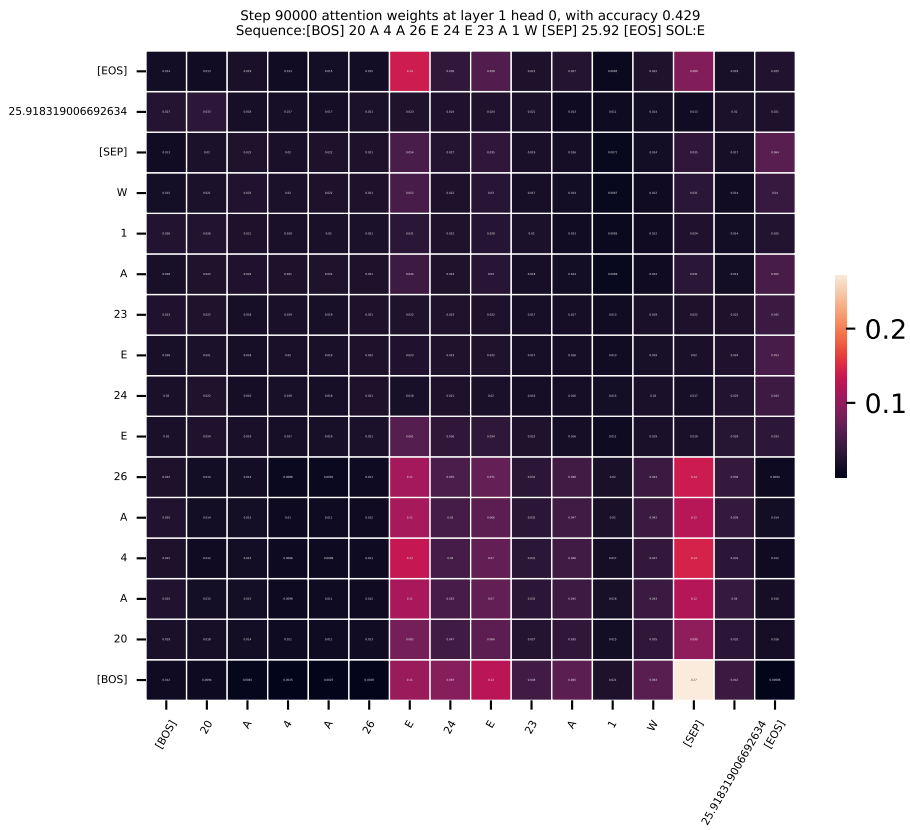

Step 89500 attention weights at layer 1 head 3, with accuracy 0.433
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

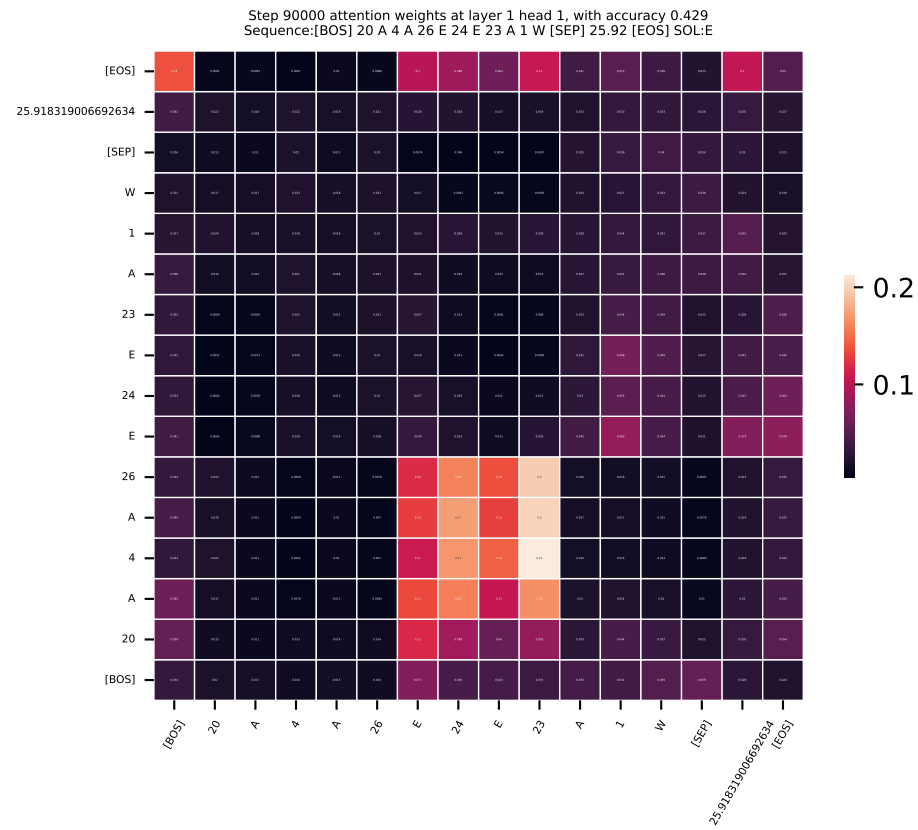

Step 90000 attention weights at layer 0 head 0, with accuracy 0.429
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

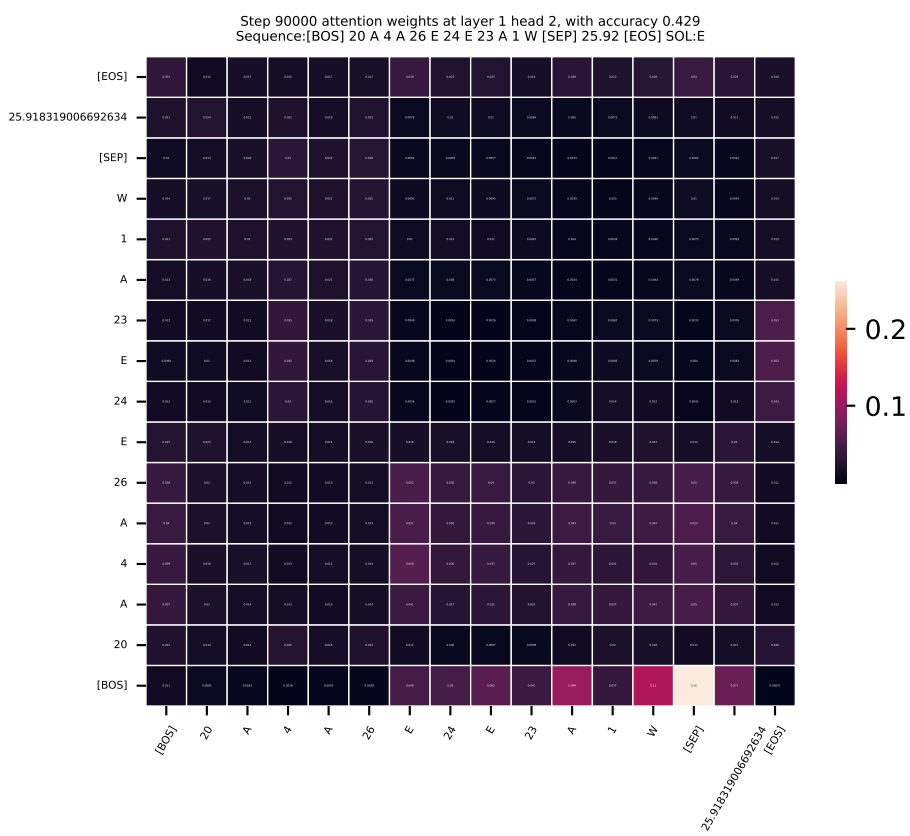

Step 90000 attention weights at layer 0 head 1, with accuracy 0.429
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

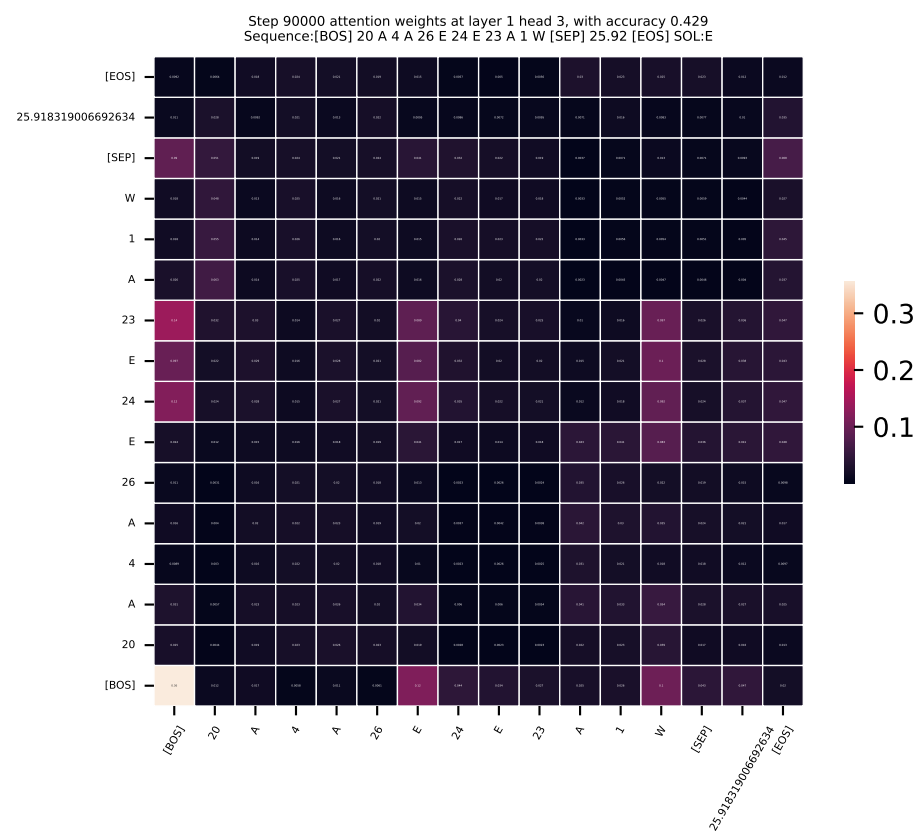

Step 90000 attention weights at layer 0 head 2, with accuracy 0.429
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

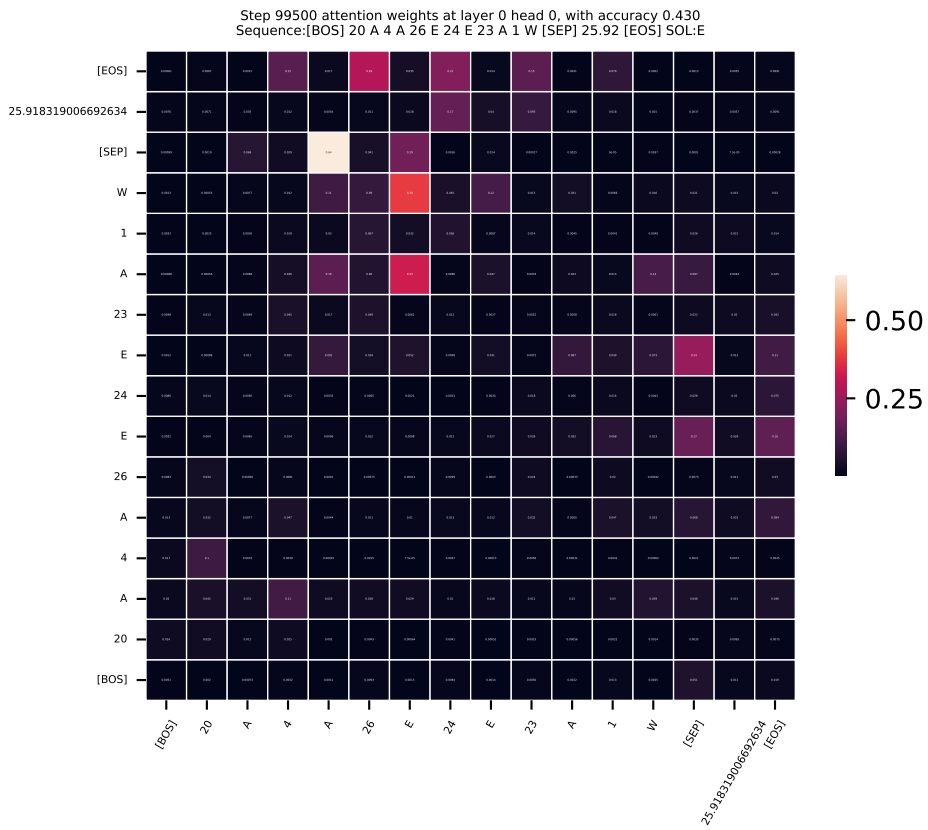

Step 90000 attention weights at layer 0 head 3, with accuracy 0.429
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

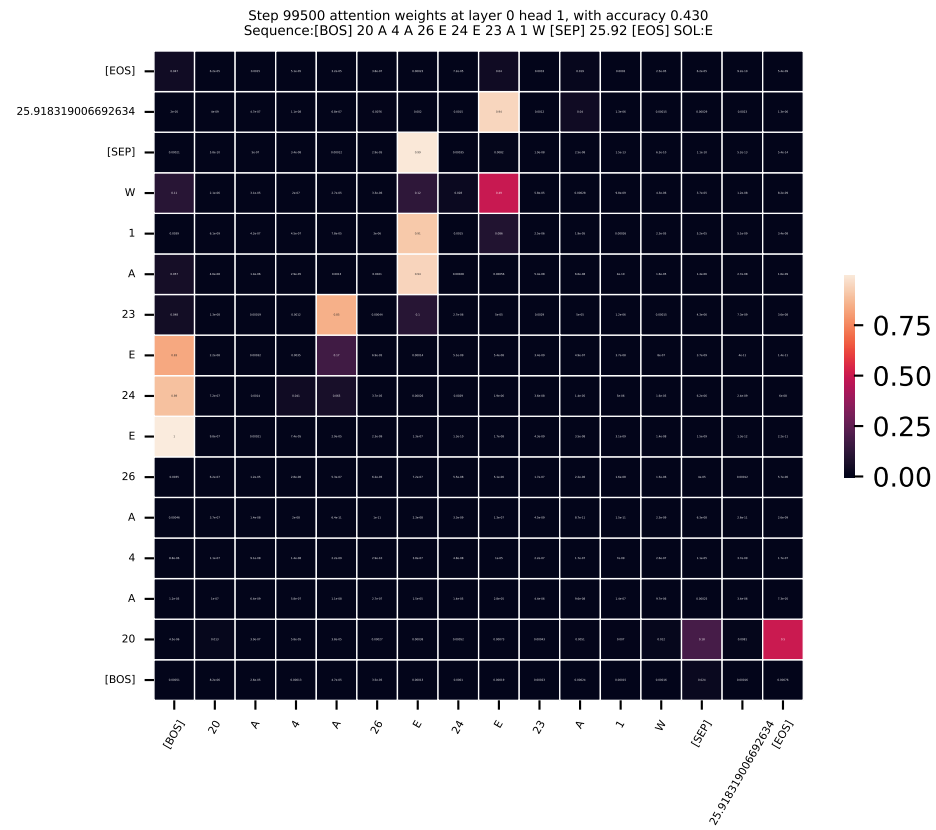

Step 90000 attention weights at layer 1 head 0, with accuracy 0.429
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.2
0.1

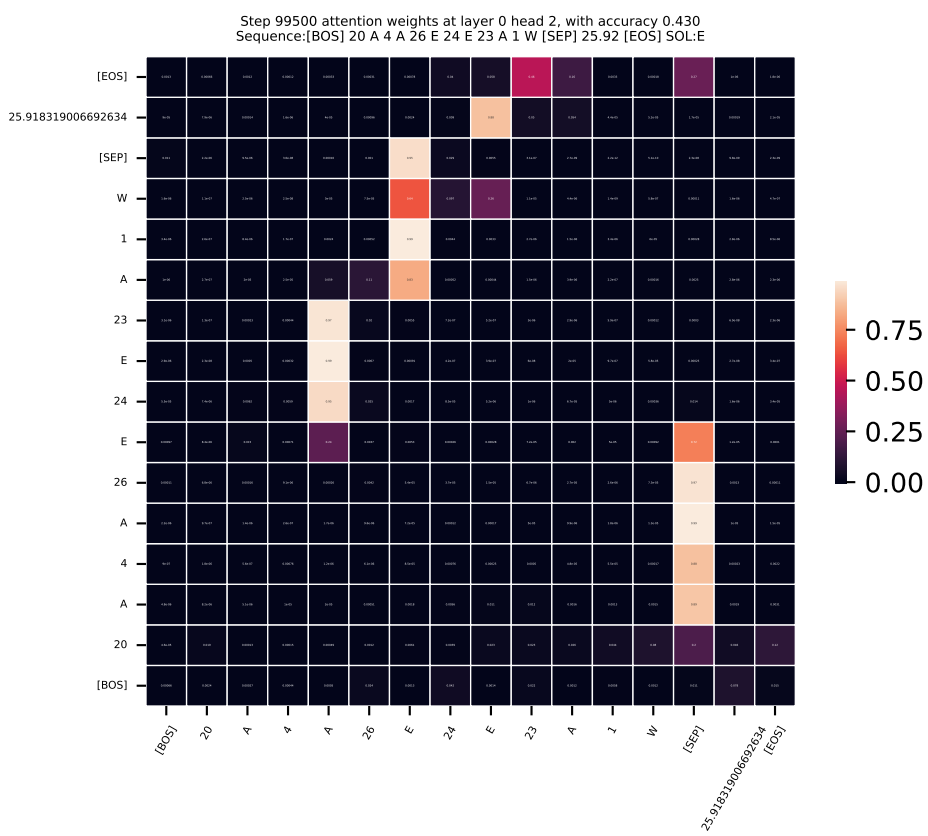

Step 90000 attention weights at layer 1 head 1, with accuracy 0.429
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.2
0.1

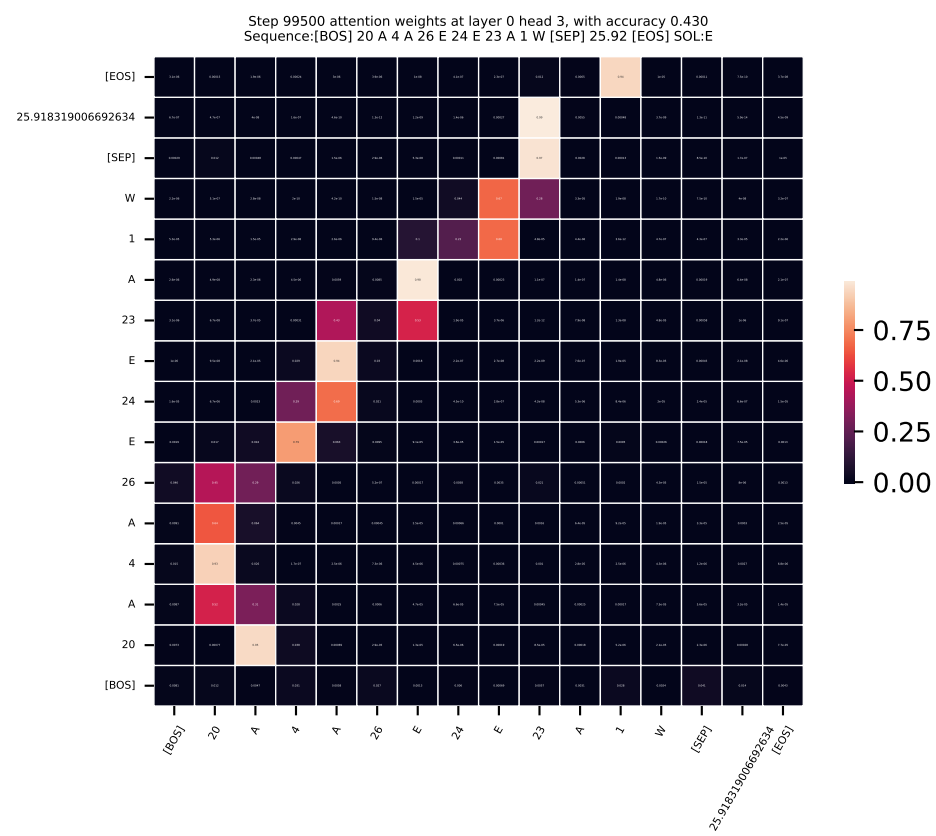

Step 90000 attention weights at layer 1 head 2, with accuracy 0.429
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.2
0.1

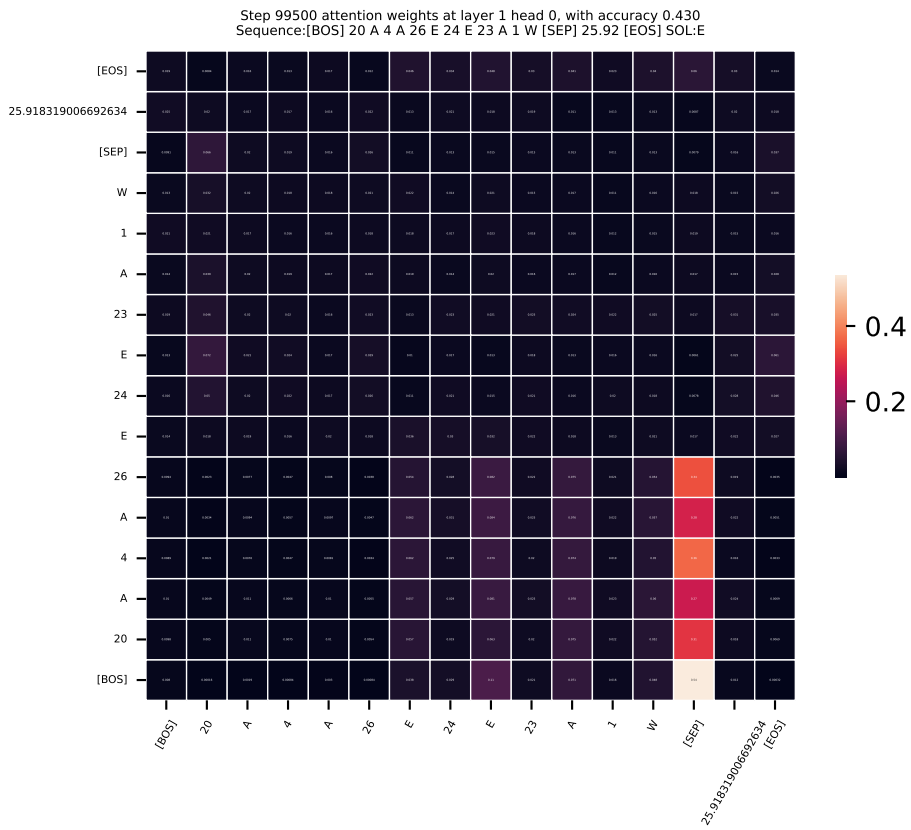

Step 90000 attention weights at layer 1 head 3, with accuracy 0.429
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

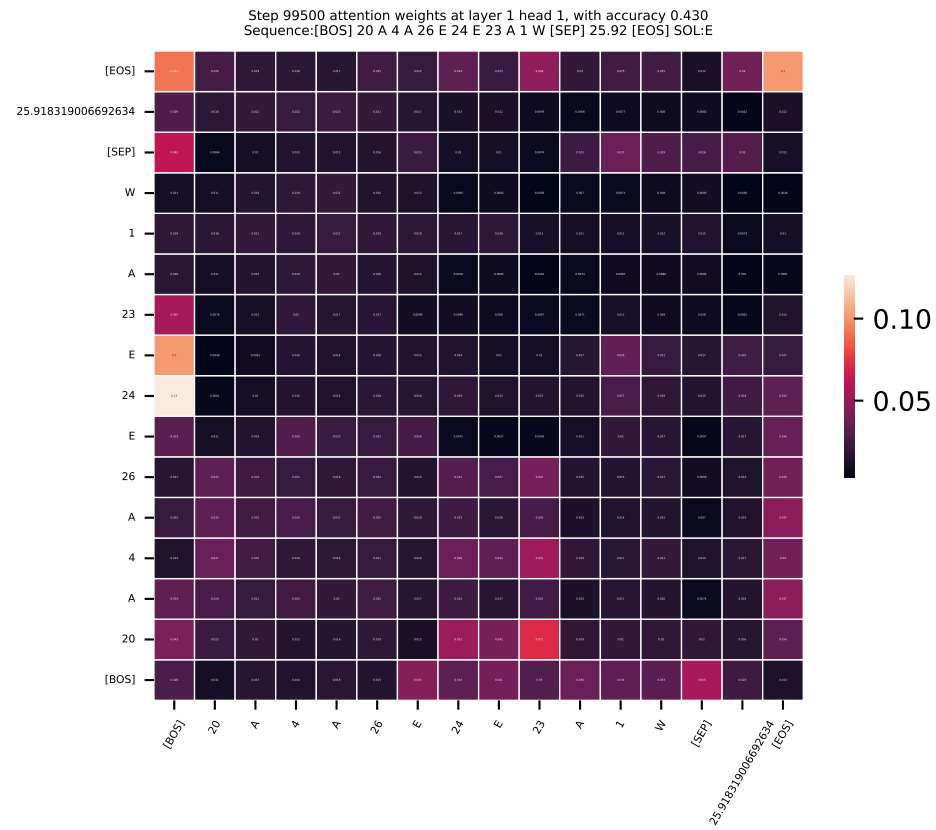

Step 99500 attention weights at layer 0 head 0, with accuracy 0.430
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.50
0.25

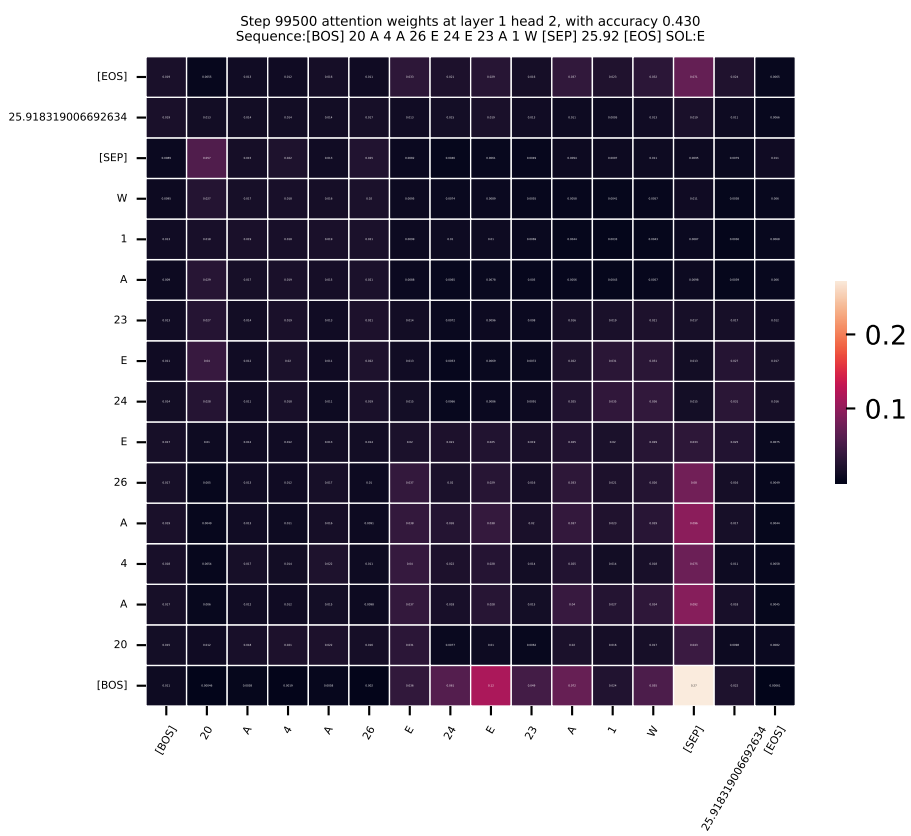

Step 99500 attention weights at layer 0 head 1, with accuracy 0.430
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

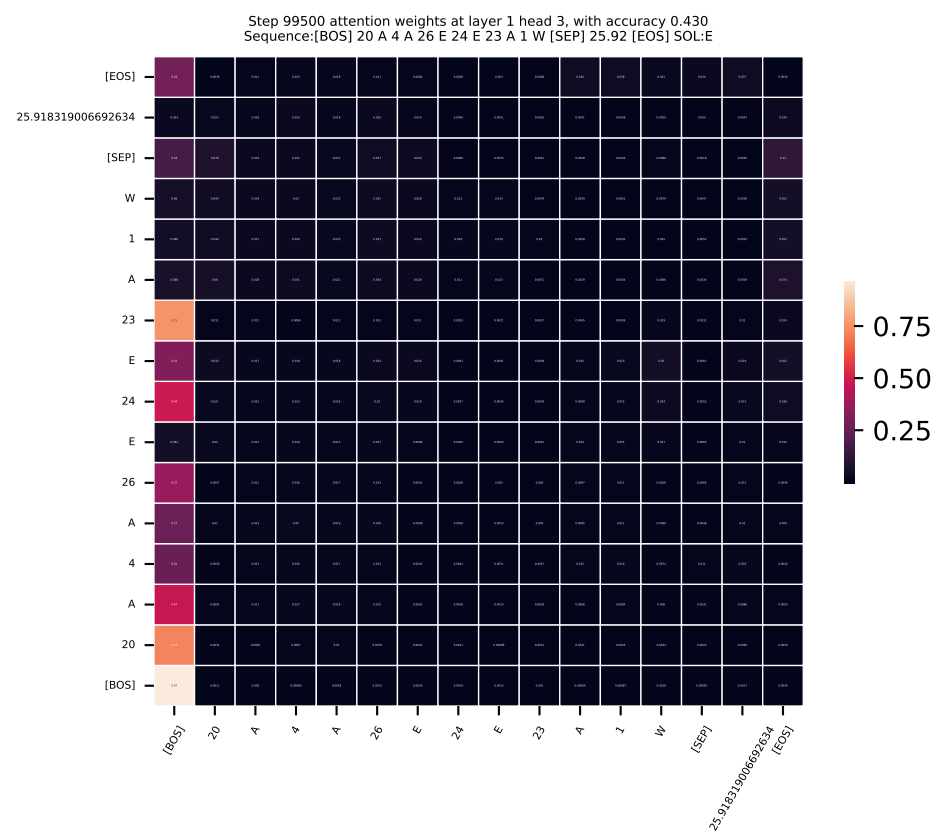

Step 99500 attention weights at layer 0 head 2, with accuracy 0.430
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

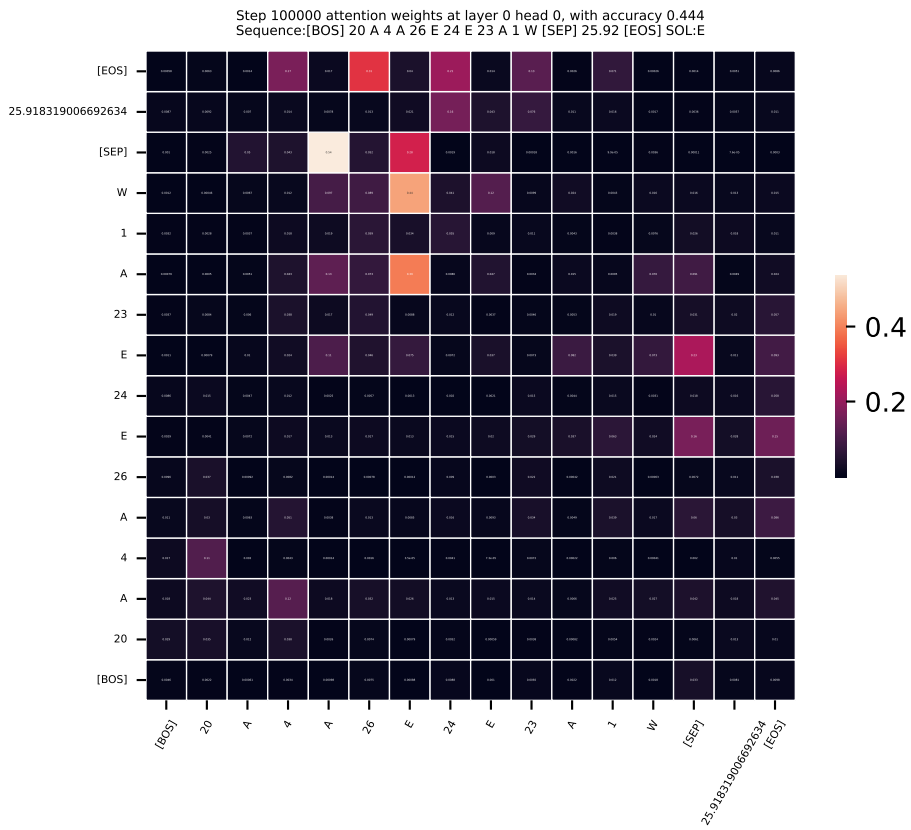

Step 99500 attention weights at layer 0 head 3, with accuracy 0.430
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

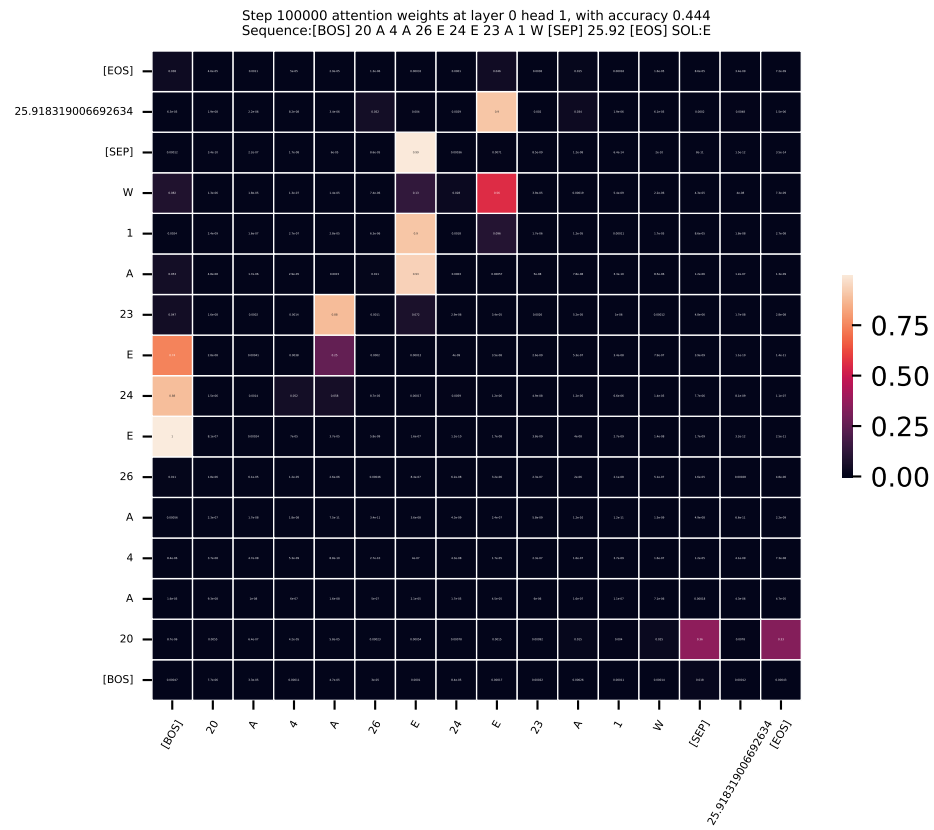

Step 99500 attention weights at layer 1 head 0, with accuracy 0.430
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

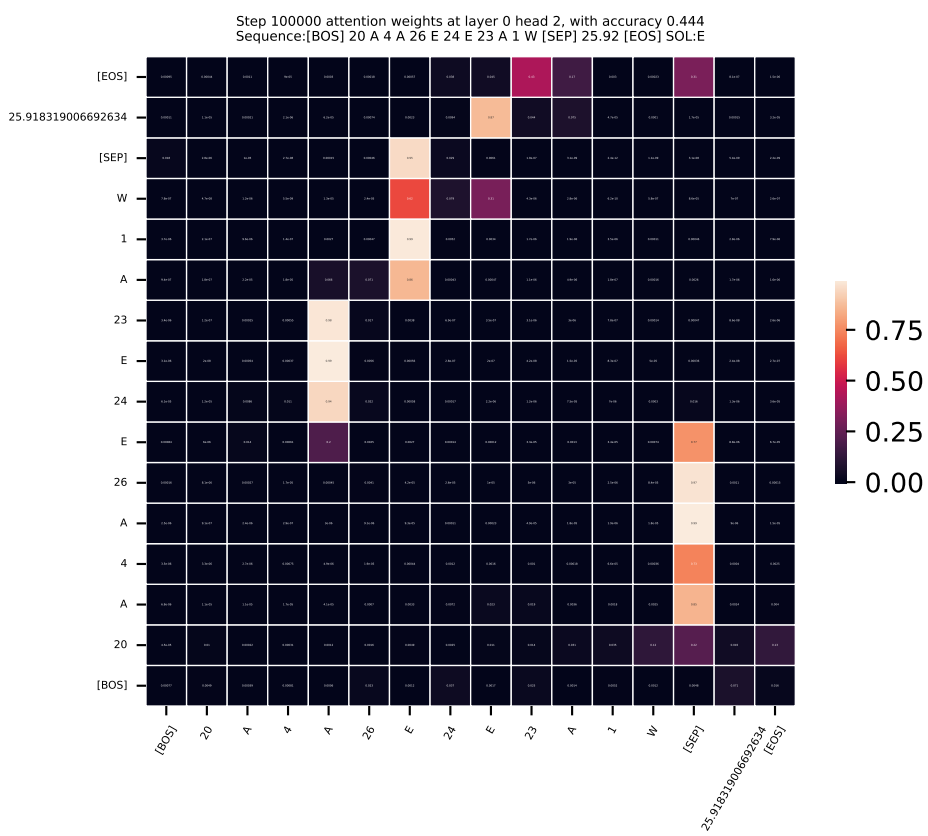

Step 99500 attention weights at layer 1 head 1, with accuracy 0.430
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.10
0.05

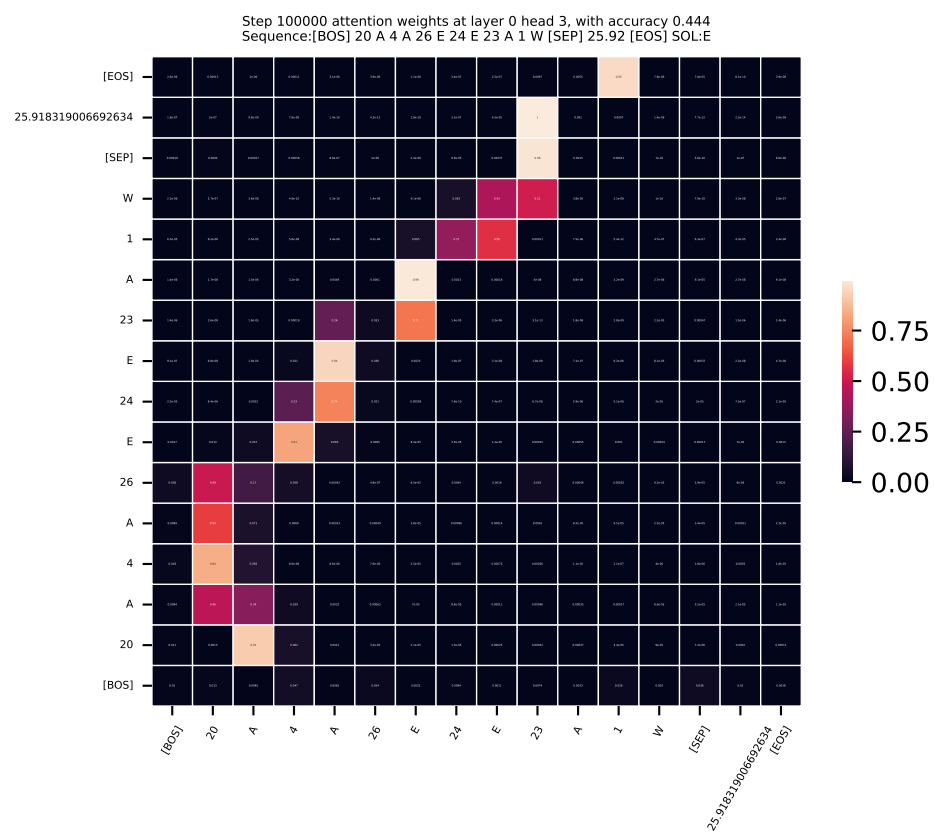

Step 99500 attention weights at layer 1 head 2, with accuracy 0.430
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.2
0.1

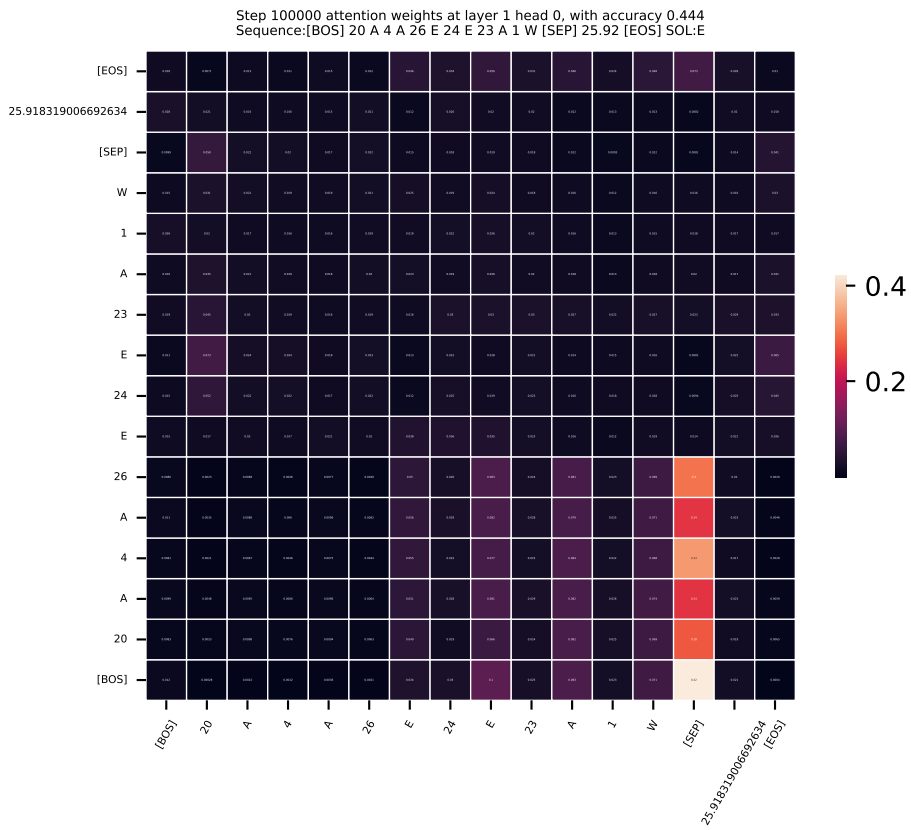

Step 99500 attention weights at layer 1 head 3, with accuracy 0.430
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

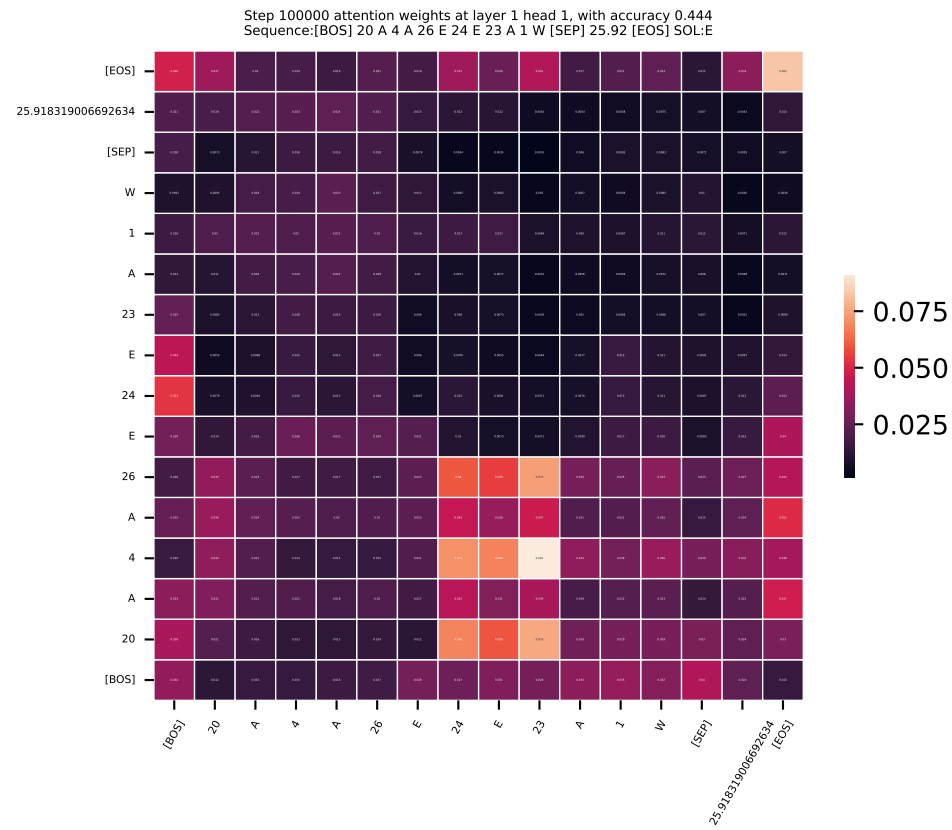

Step 100000 attention weights at layer 0 head 0, with accuracy 0.444
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

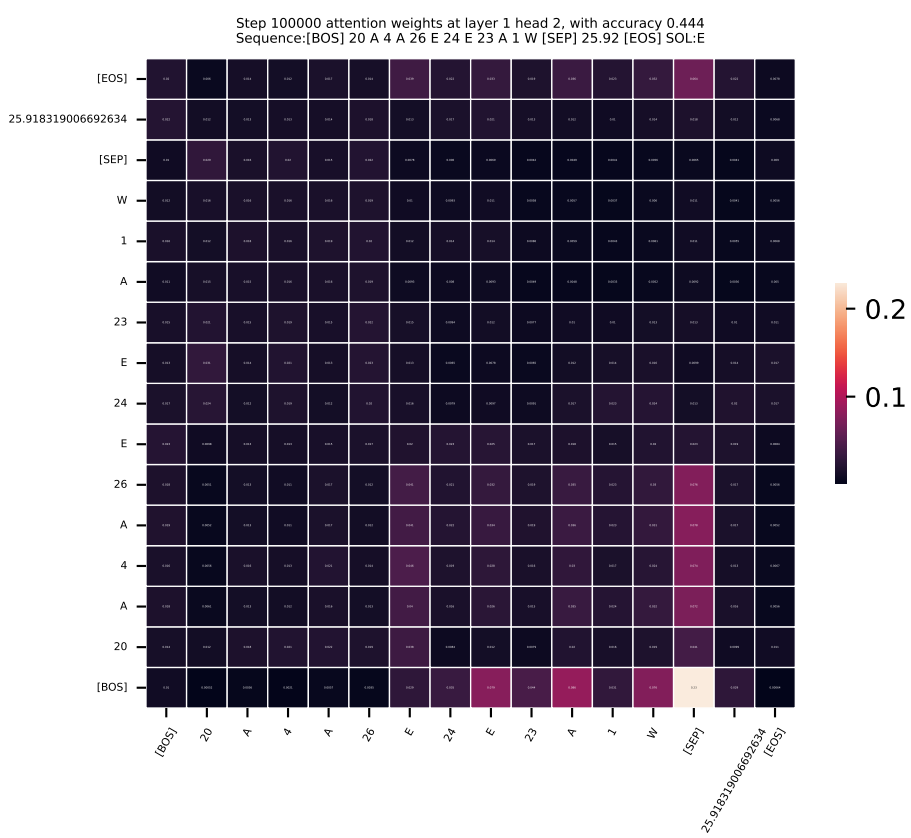

Step 100000 attention weights at layer 0 head 1, with accuracy 0.444
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

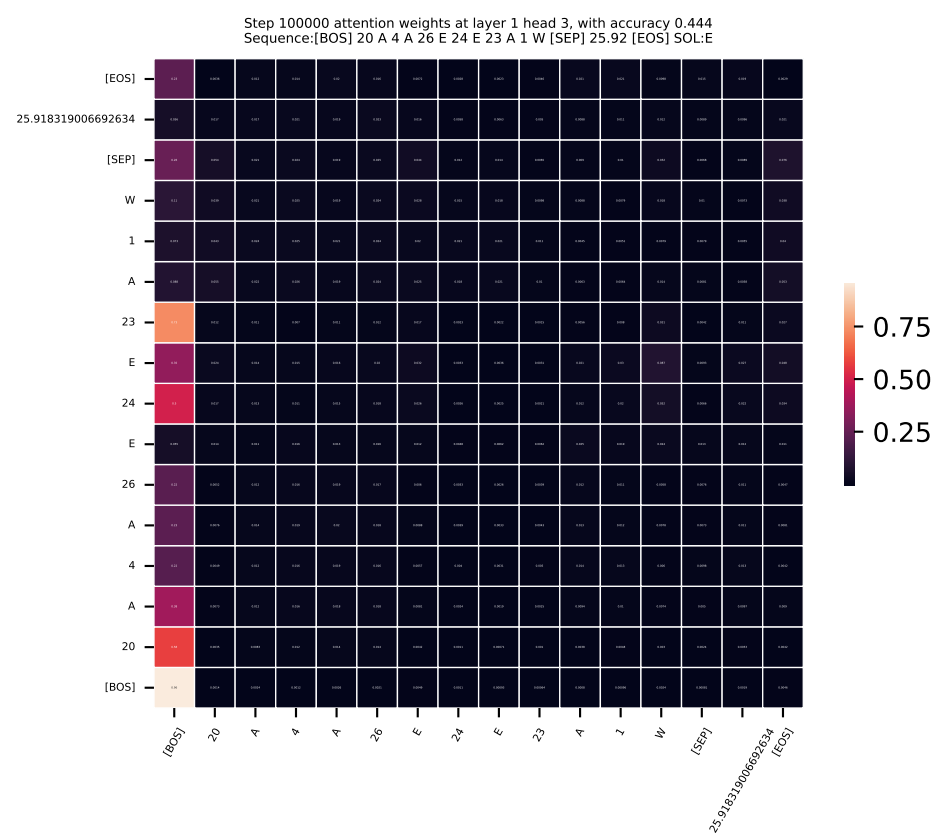

Step 100000 attention weights at layer 0 head 2, with accuracy 0.444
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

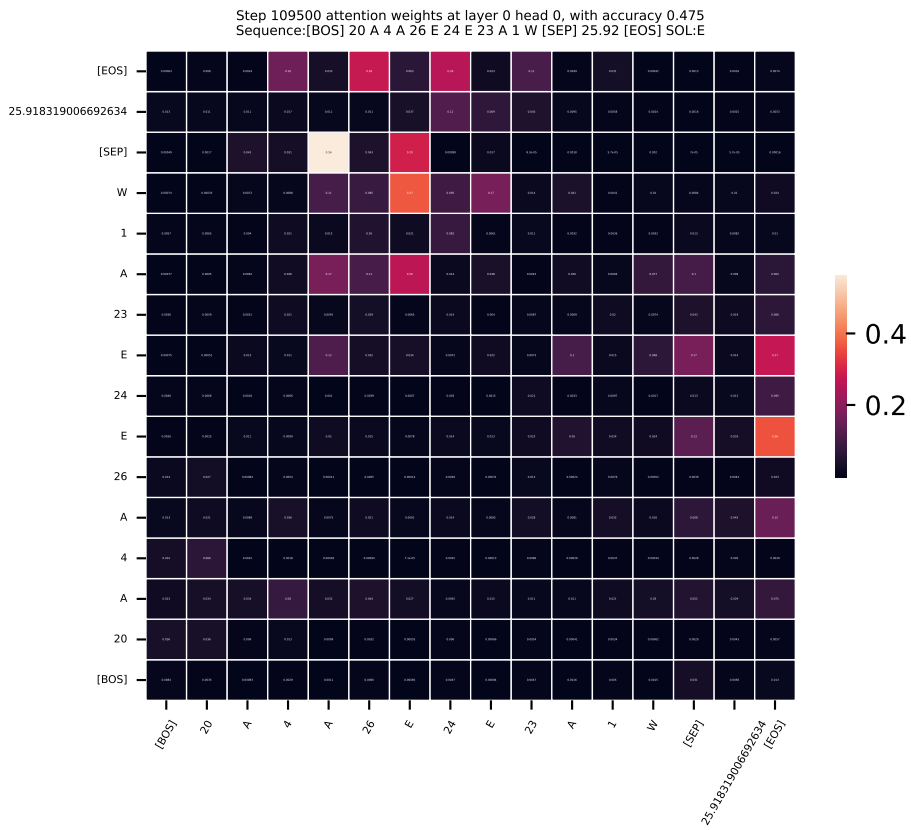

Step 100000 attention weights at layer 0 head 3, with accuracy 0.444
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

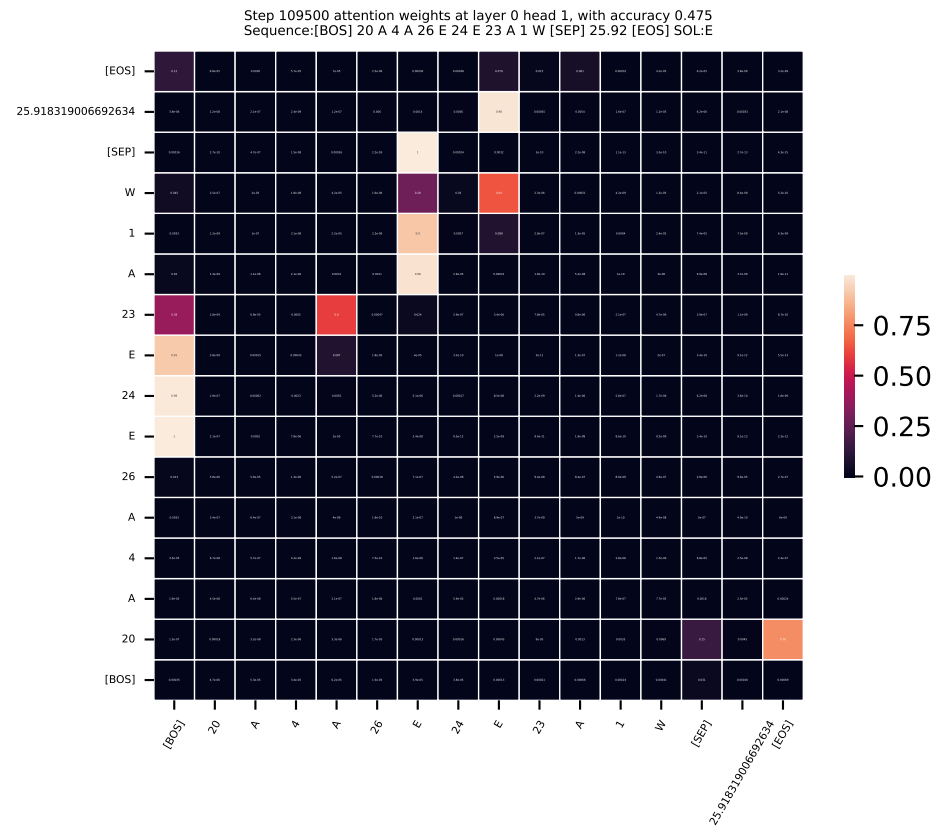

Step 100000 attention weights at layer 1 head 0, with accuracy 0.444
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

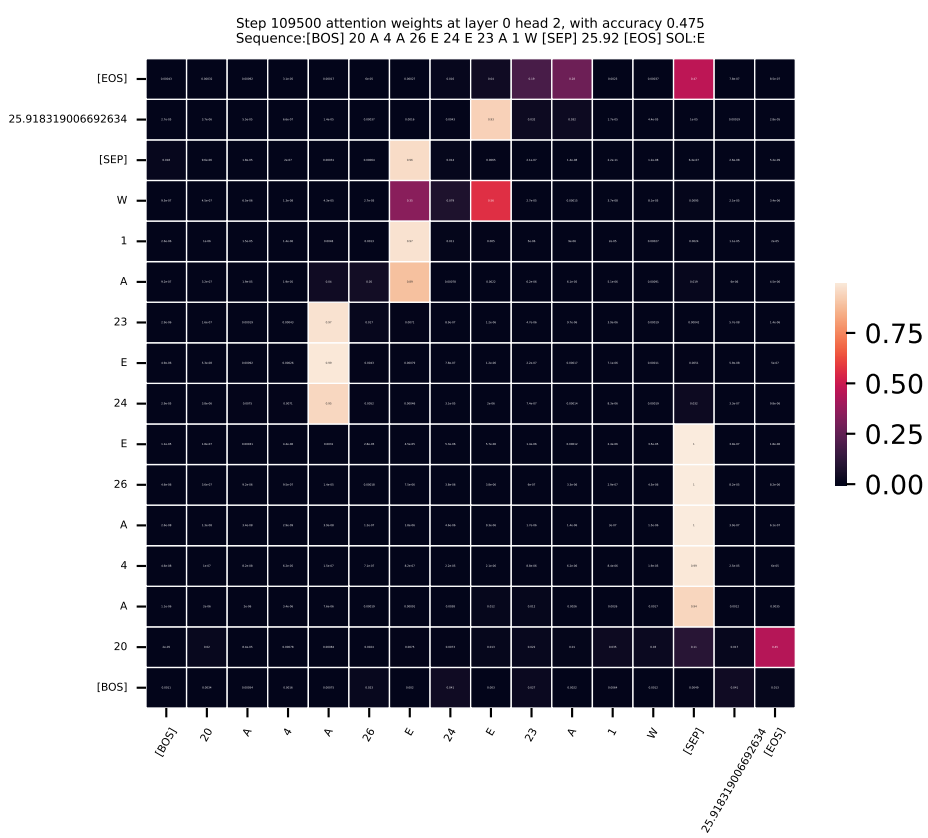

Step 100000 attention weights at layer 1 head 1, with accuracy 0.444
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.075
0.050
0.025

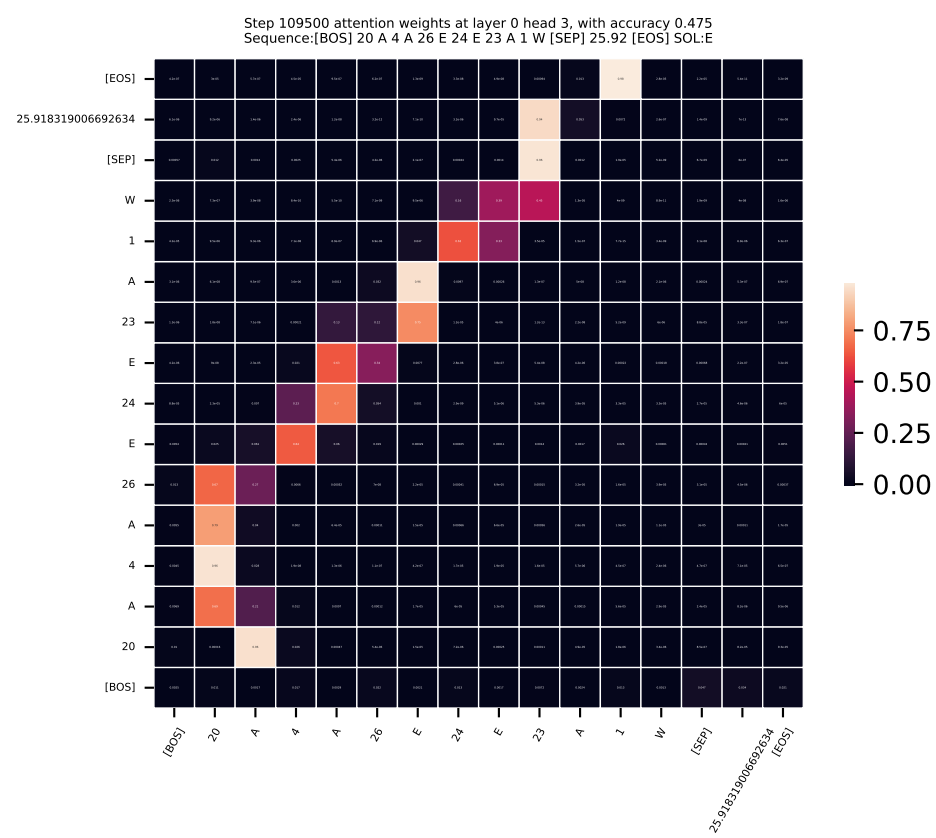

Step 100000 attention weights at layer 1 head 2, with accuracy 0.444
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.2
0.1

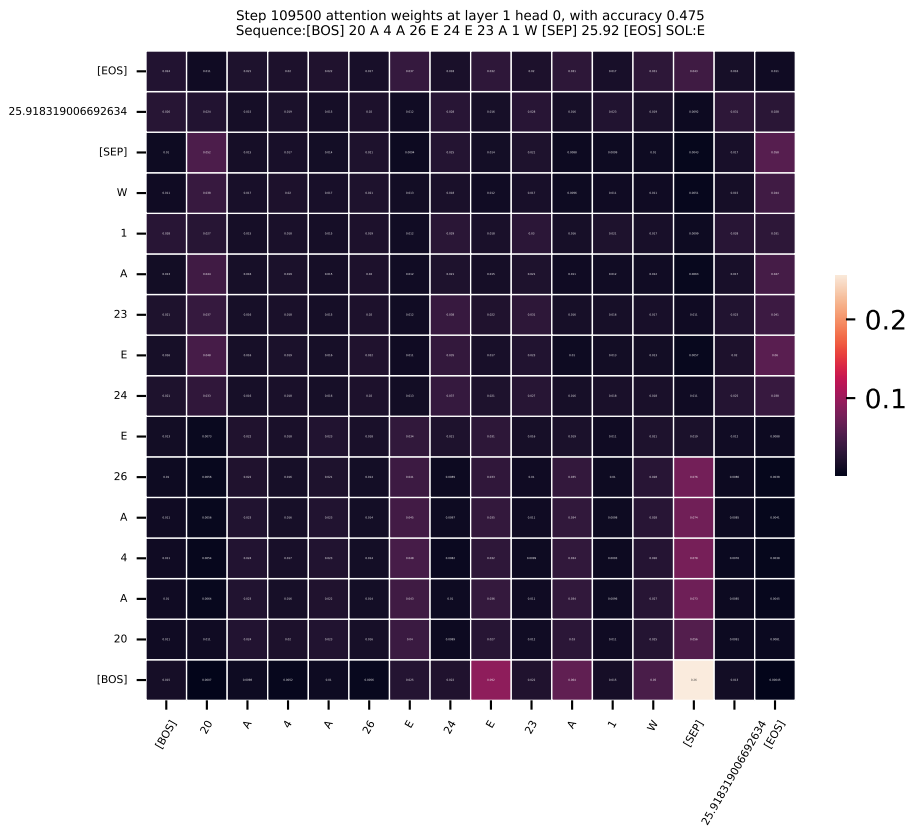

Step 100000 attention weights at layer 1 head 3, with accuracy 0.444
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

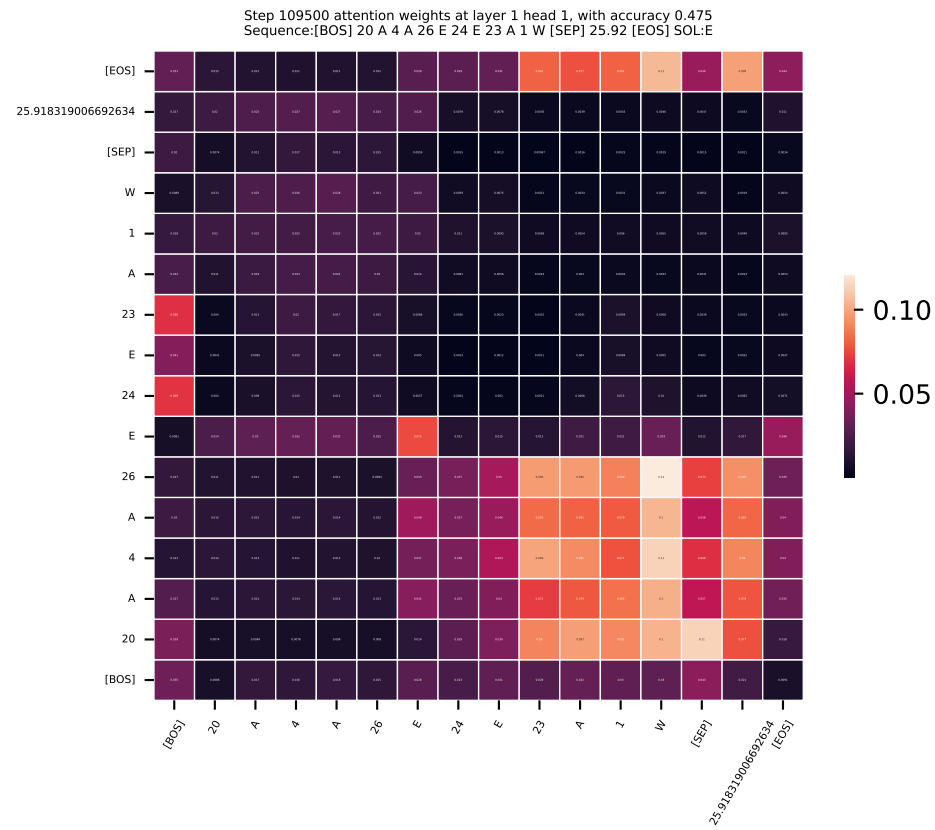

Step 109500 attention weights at layer 0 head 0, with accuracy 0.475
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

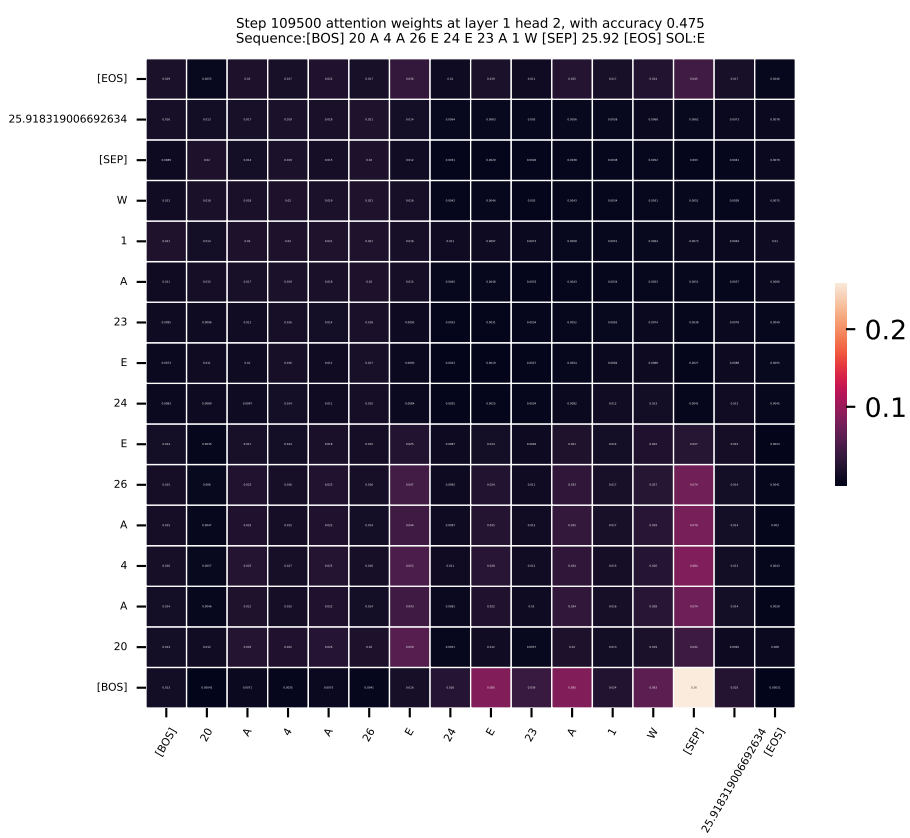

Step 109500 attention weights at layer 0 head 1, with accuracy 0.475
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

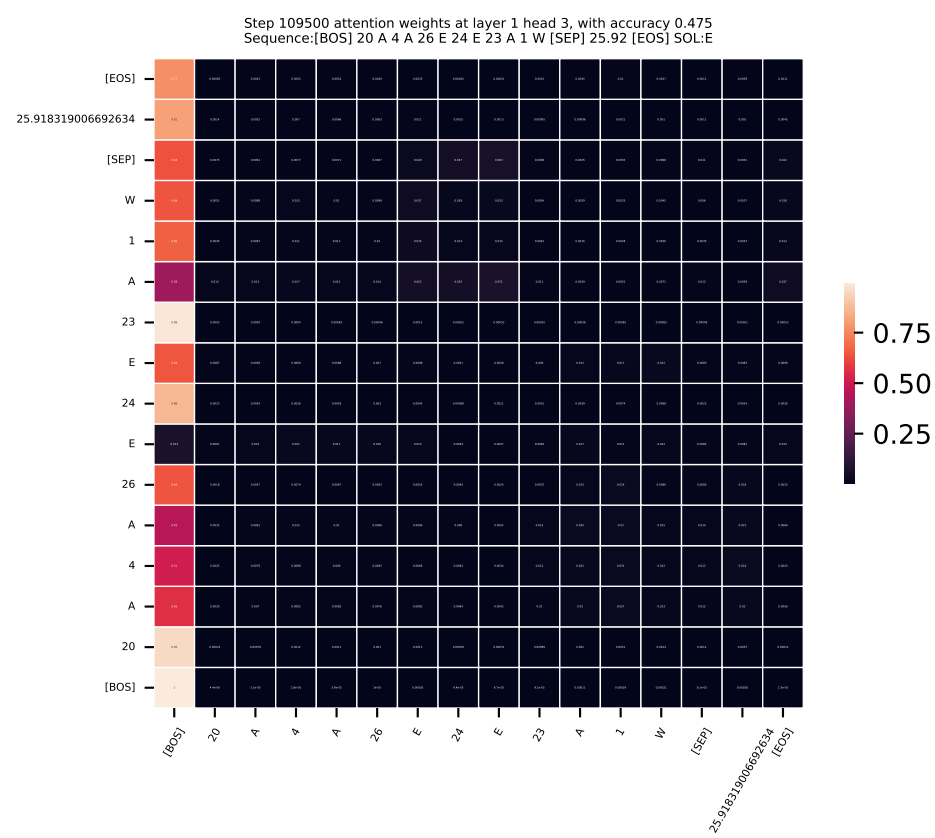

Step 109500 attention weights at layer 0 head 2, with accuracy 0.475
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

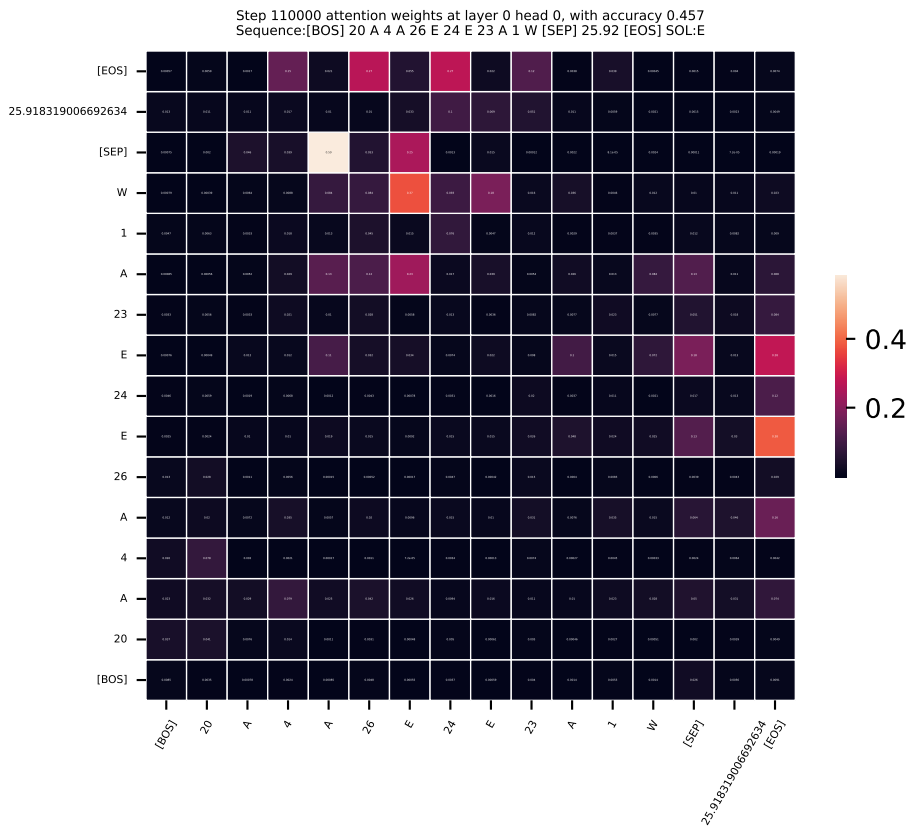

Step 109500 attention weights at layer 0 head 3, with accuracy 0.475
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

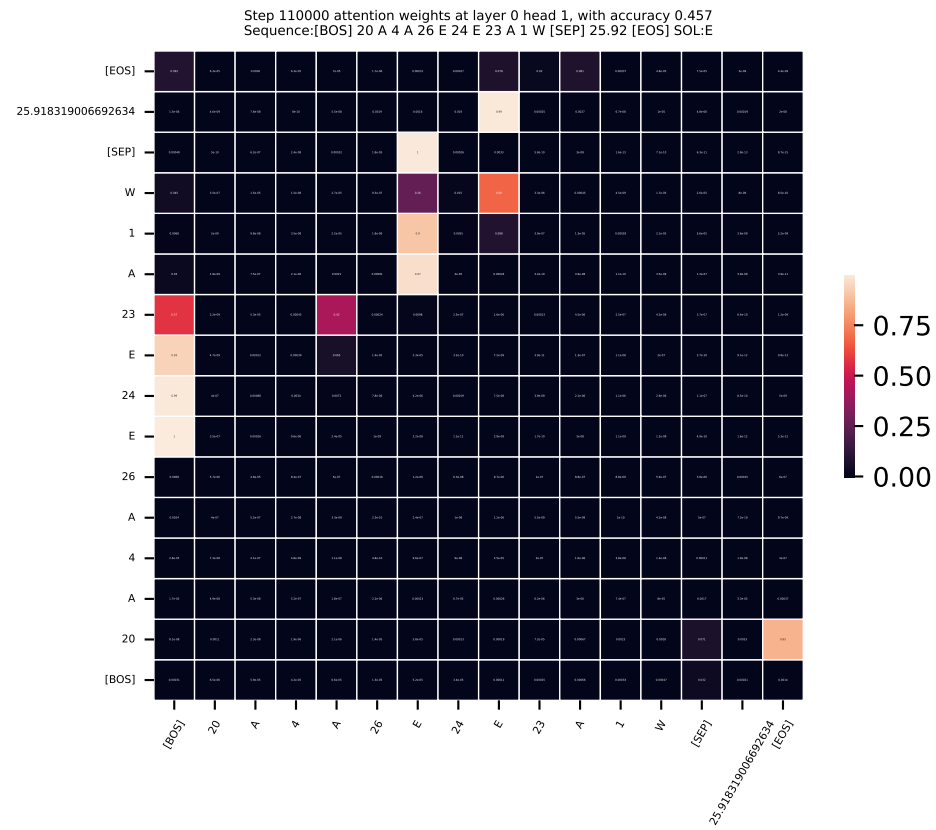

Step 109500 attention weights at layer 1 head 0, with accuracy 0.475
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.2
0.1

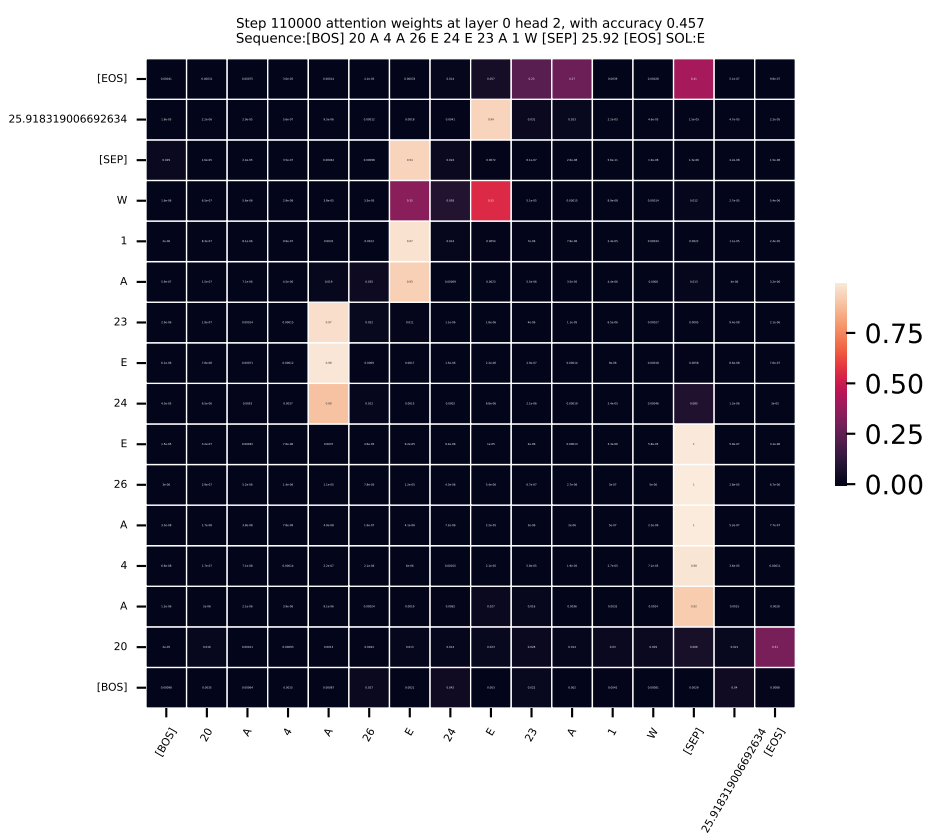

Step 109500 attention weights at layer 1 head 1, with accuracy 0.475
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.10
0.05

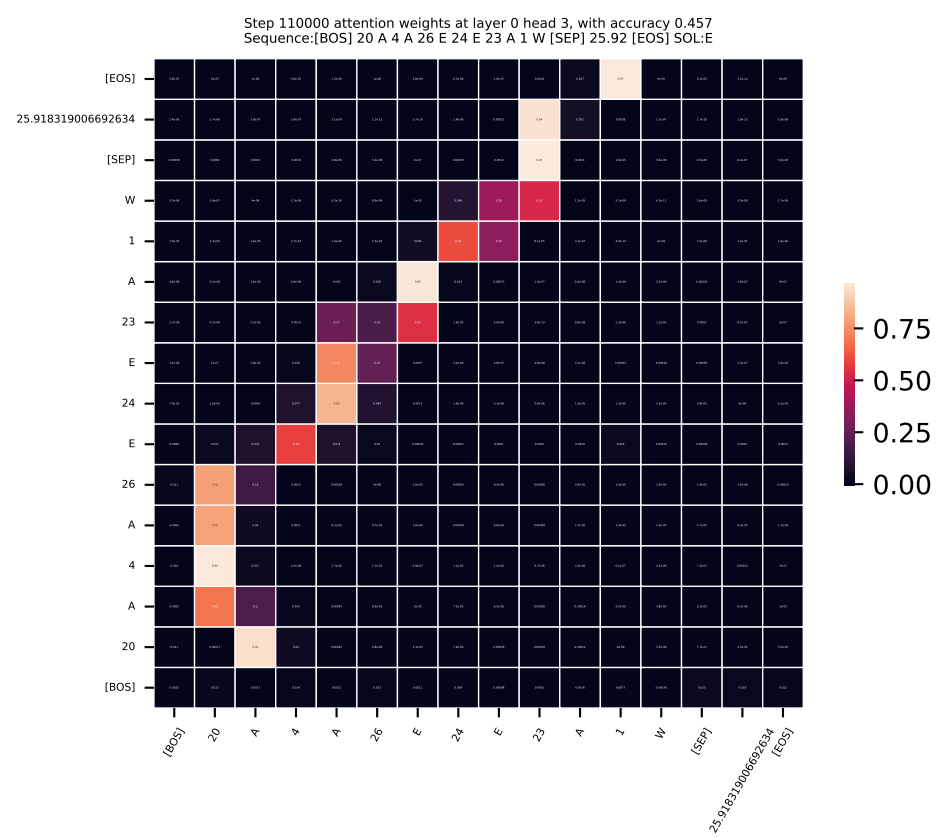

Step 109500 attention weights at layer 1 head 2, with accuracy 0.475
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.2
0.1

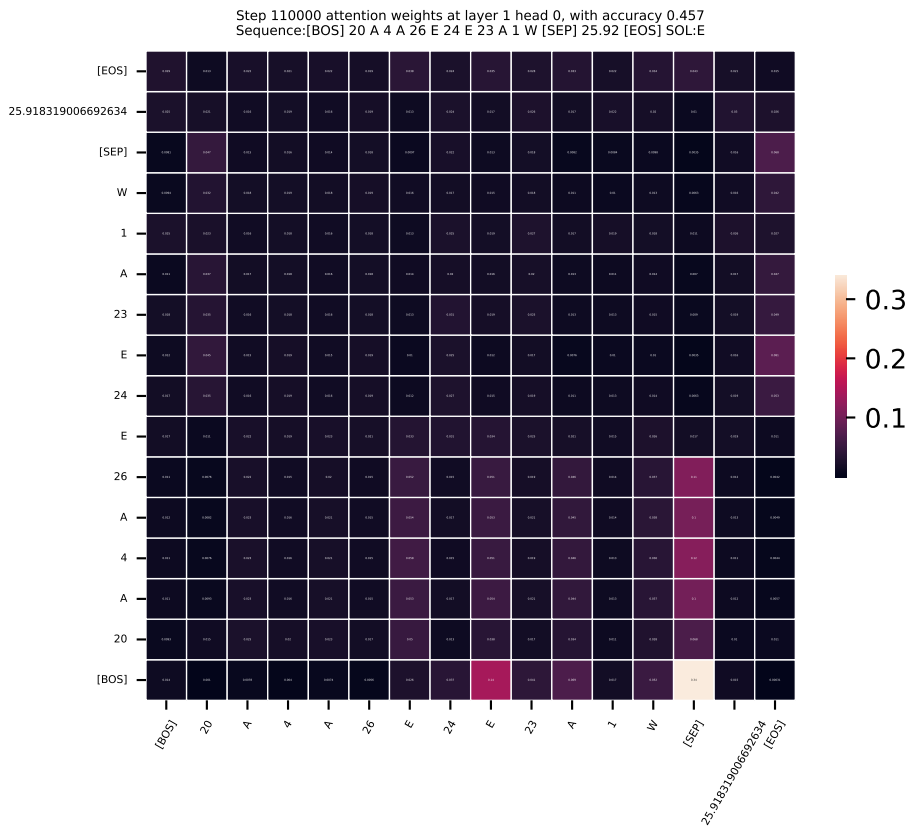

Step 109500 attention weights at layer 1 head 3, with accuracy 0.475
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

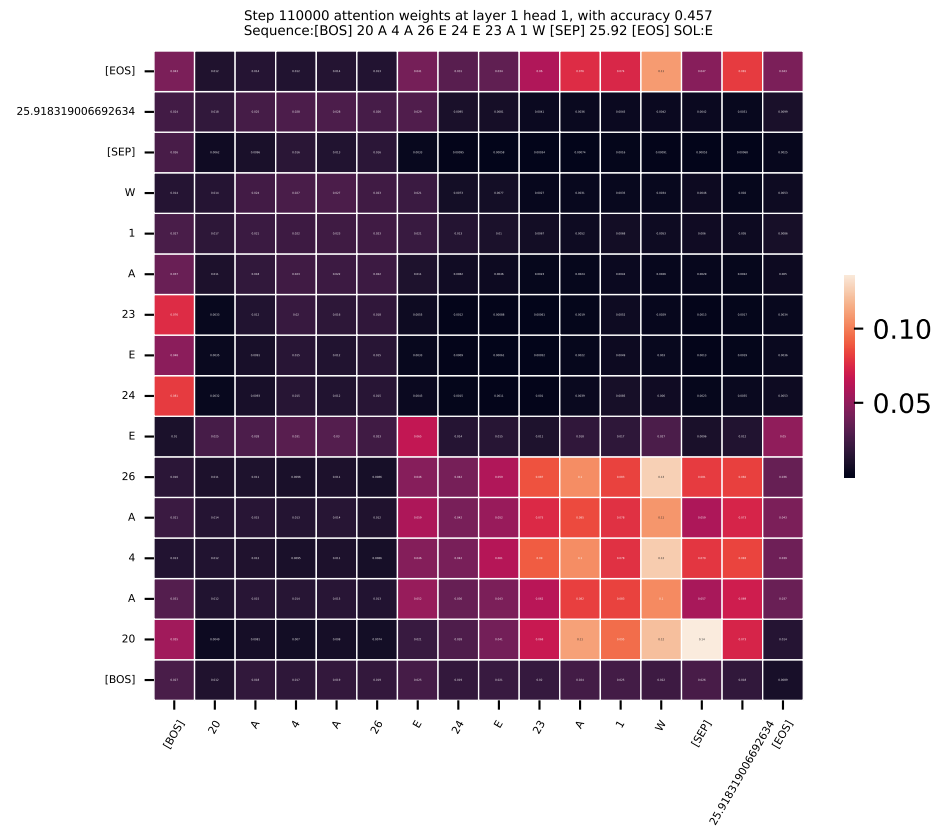

Step 110000 attention weights at layer 0 head 0, with accuracy 0.457
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

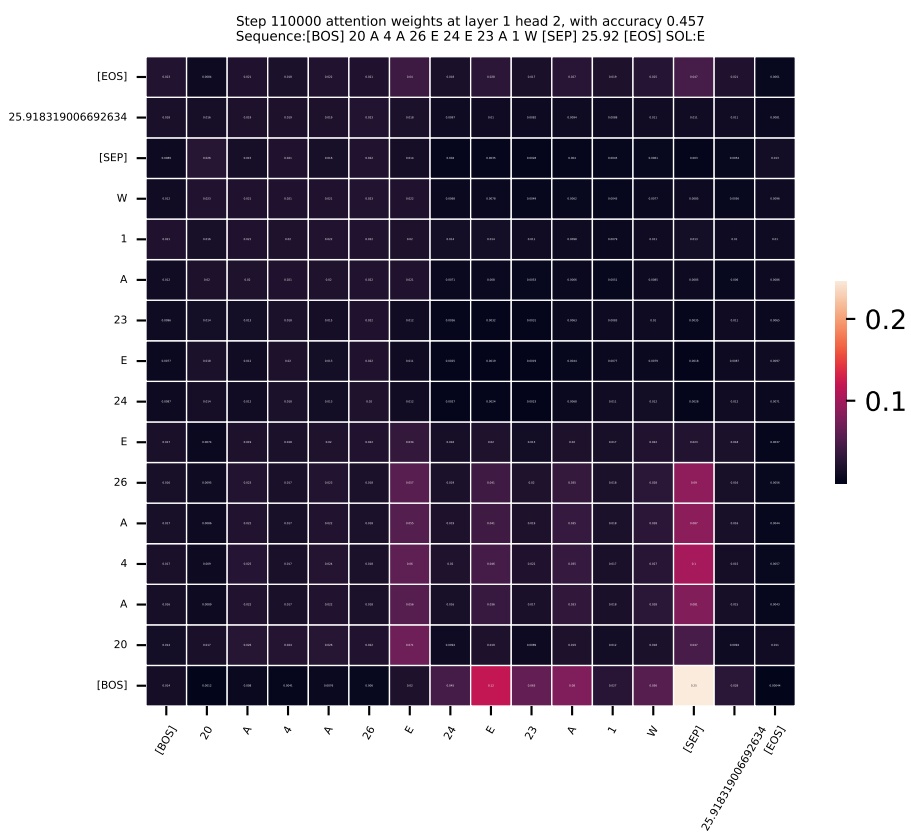

Step 110000 attention weights at layer 0 head 1, with accuracy 0.457
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

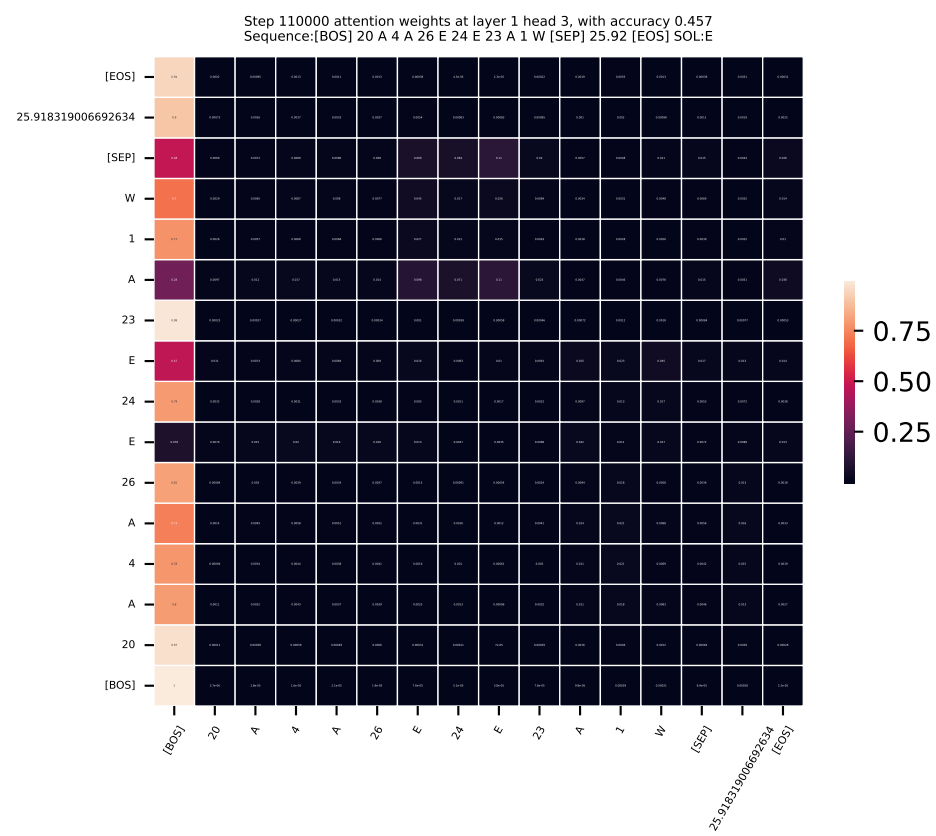

Step 110000 attention weights at layer 0 head 2, with accuracy 0.457
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

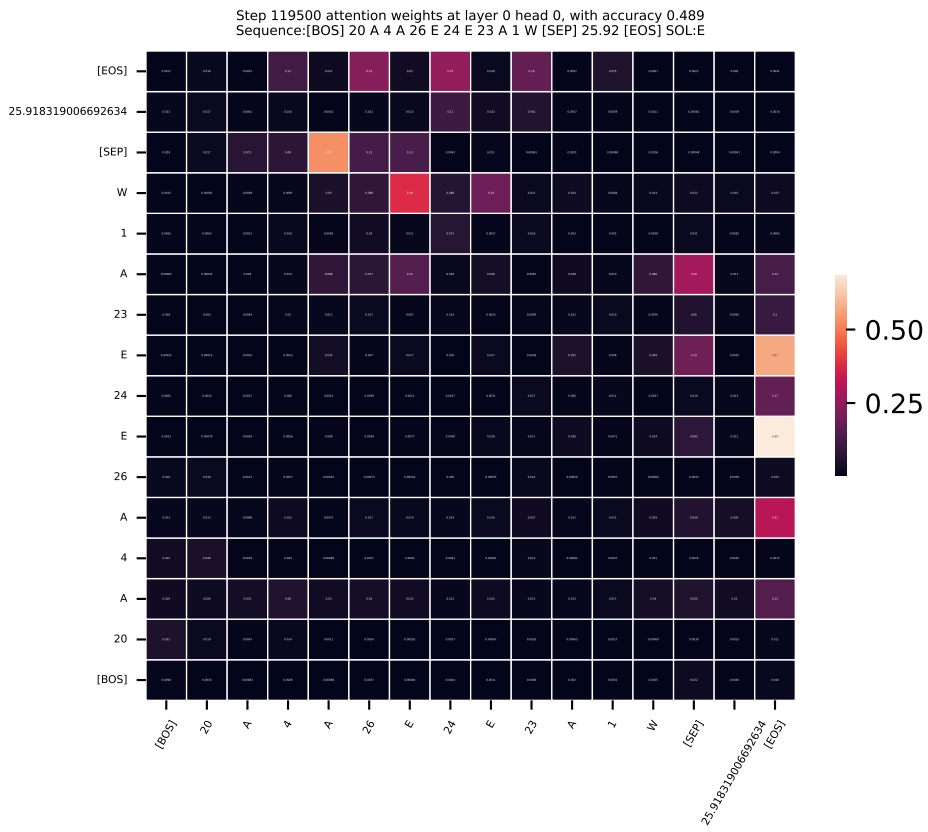

Step 110000 attention weights at layer 0 head 3, with accuracy 0.457
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

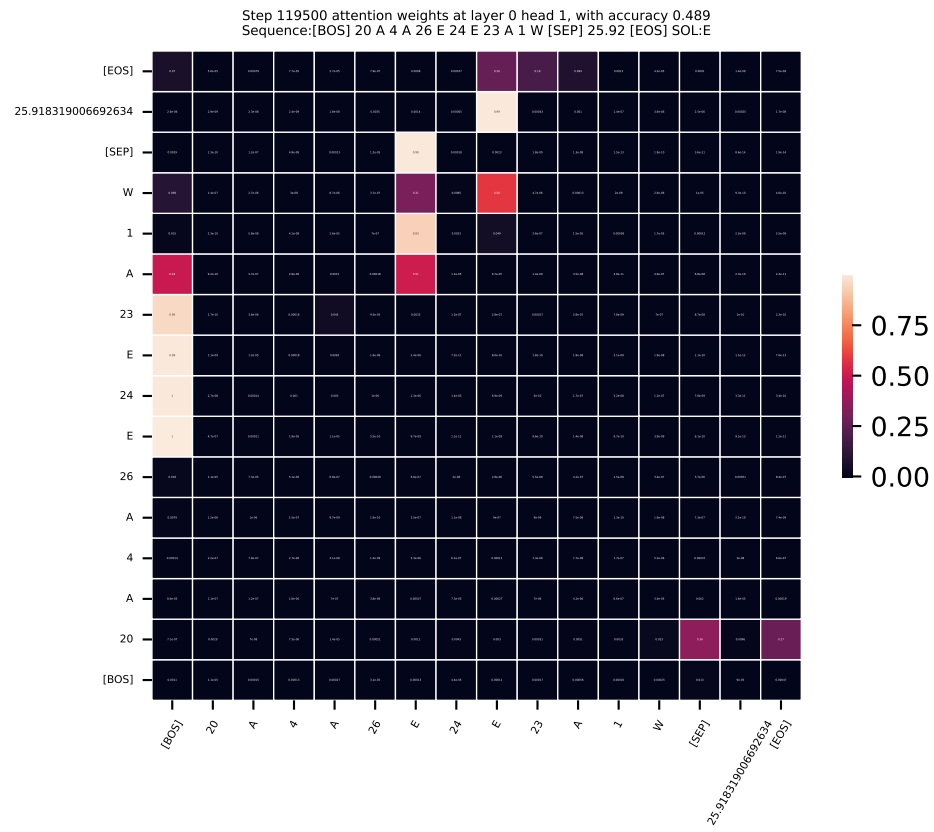

Step 110000 attention weights at layer 1 head 0, with accuracy 0.457
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.3
0.2
0.1

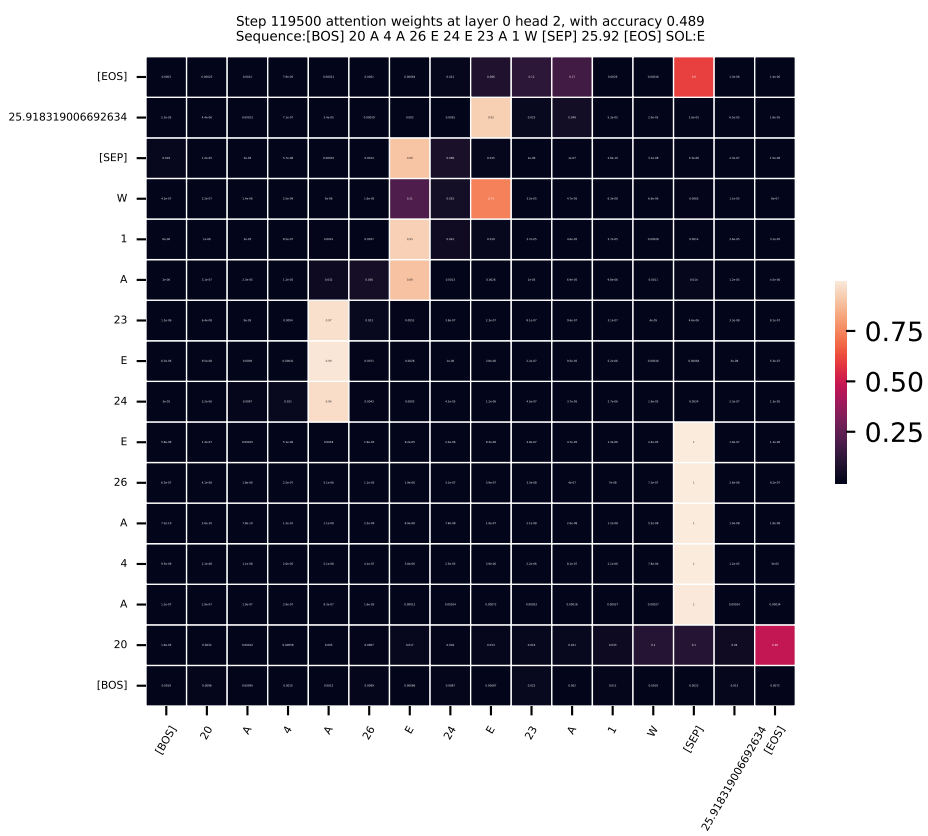

Step 110000 attention weights at layer 1 head 1, with accuracy 0.457
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.10
0.05

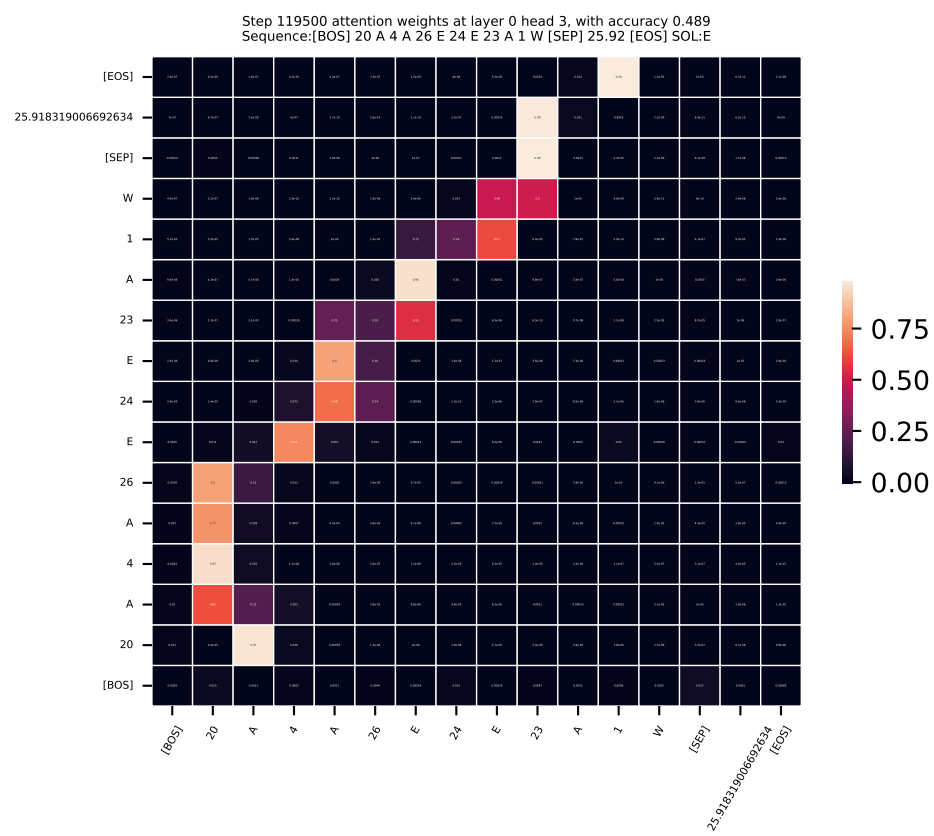

Step 110000 attention weights at layer 1 head 2, with accuracy 0.457
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.2
0.1

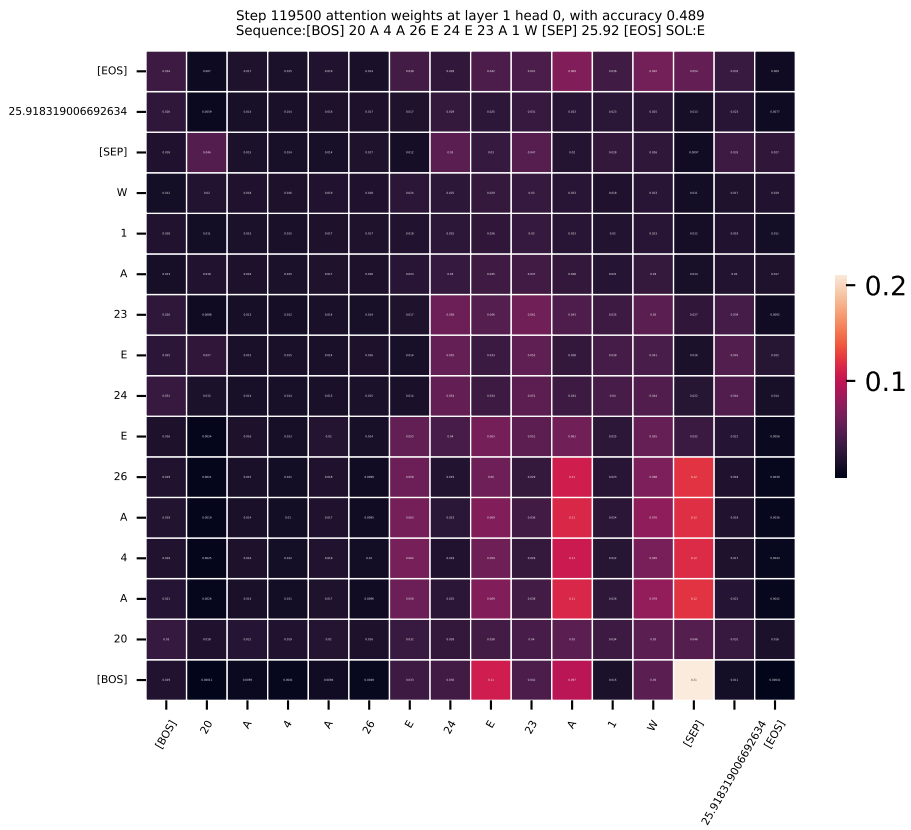

Step 110000 attention weights at layer 1 head 3, with accuracy 0.457
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

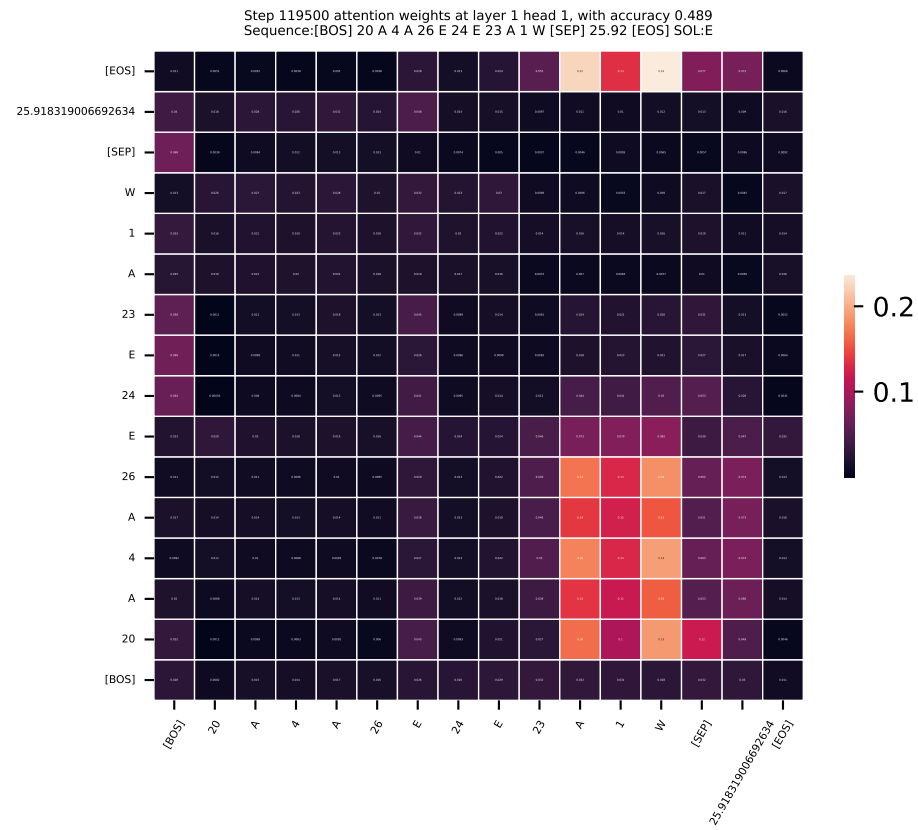

Step 119500 attention weights at layer 0 head 0, with accuracy 0.489
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

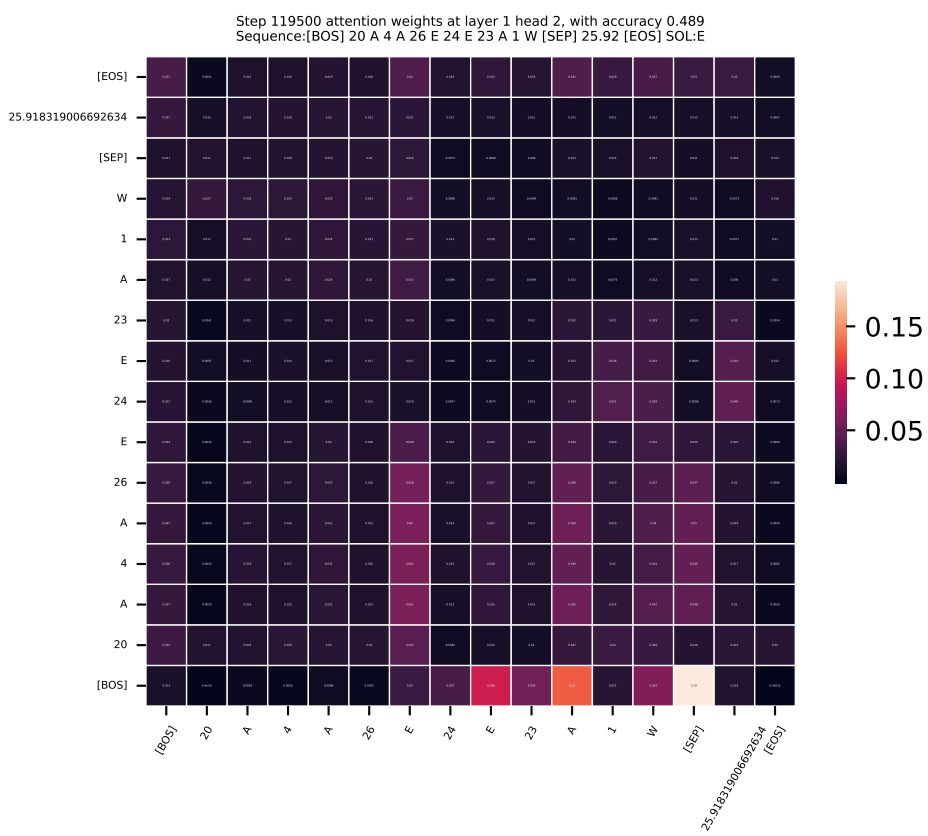

Step 119500 attention weights at layer 0 head 1, with accuracy 0.489
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

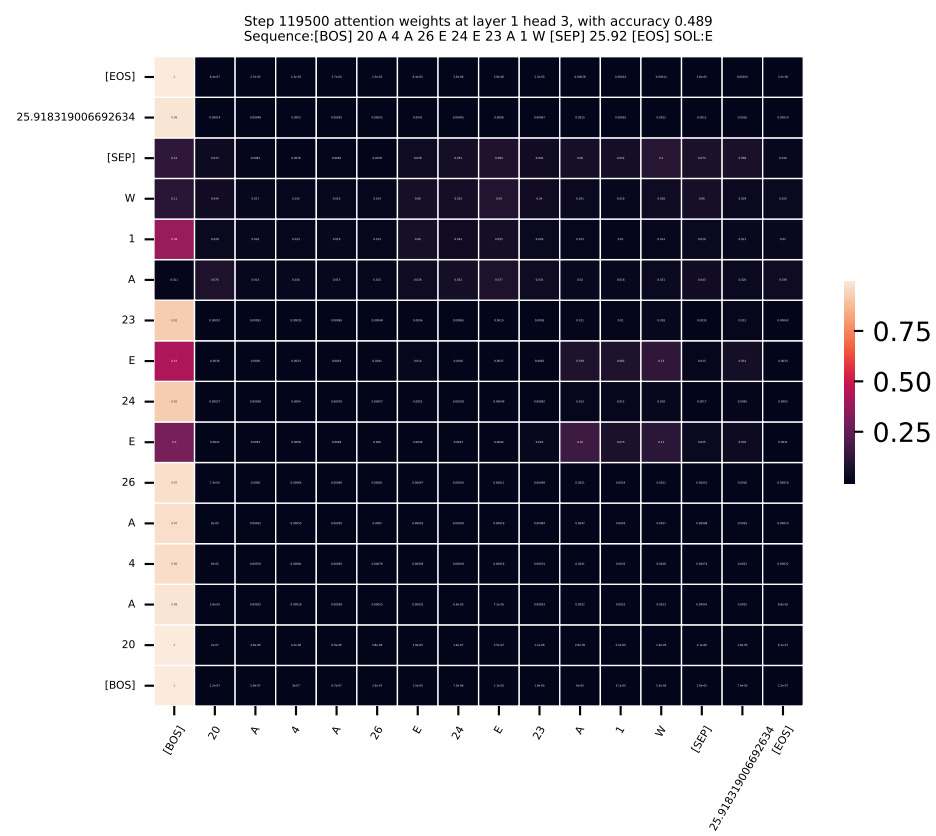

Step 119500 attention weights at layer 0 head 2, with accuracy 0.489
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

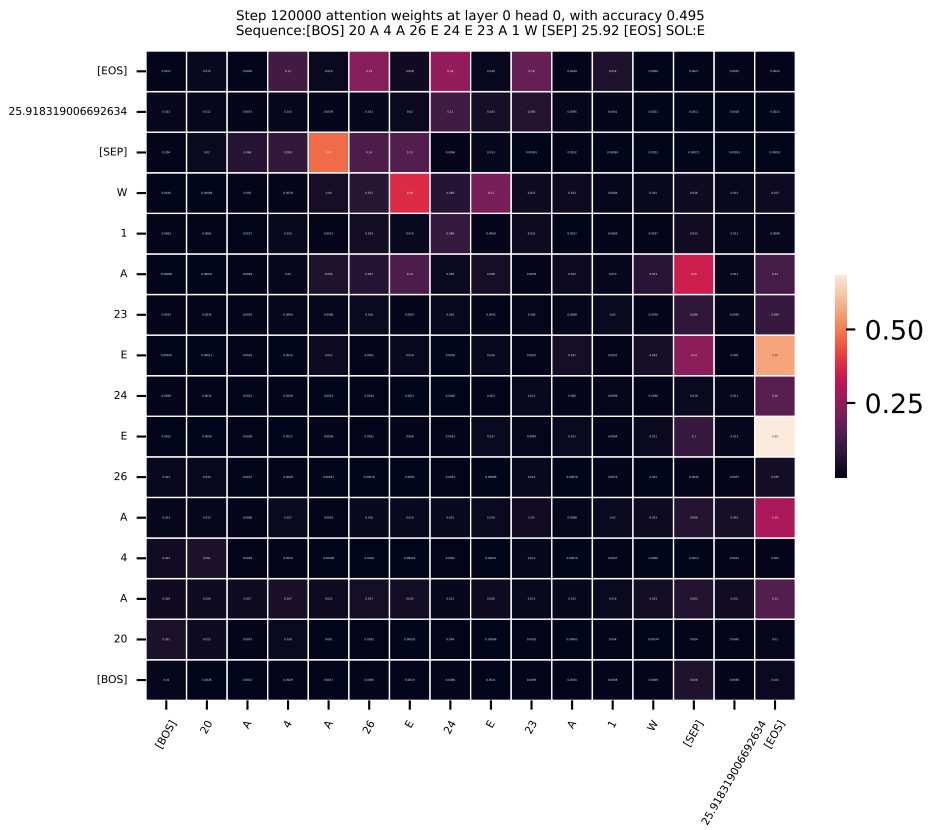

Step 119500 attention weights at layer 0 head 3, with accuracy 0.489
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

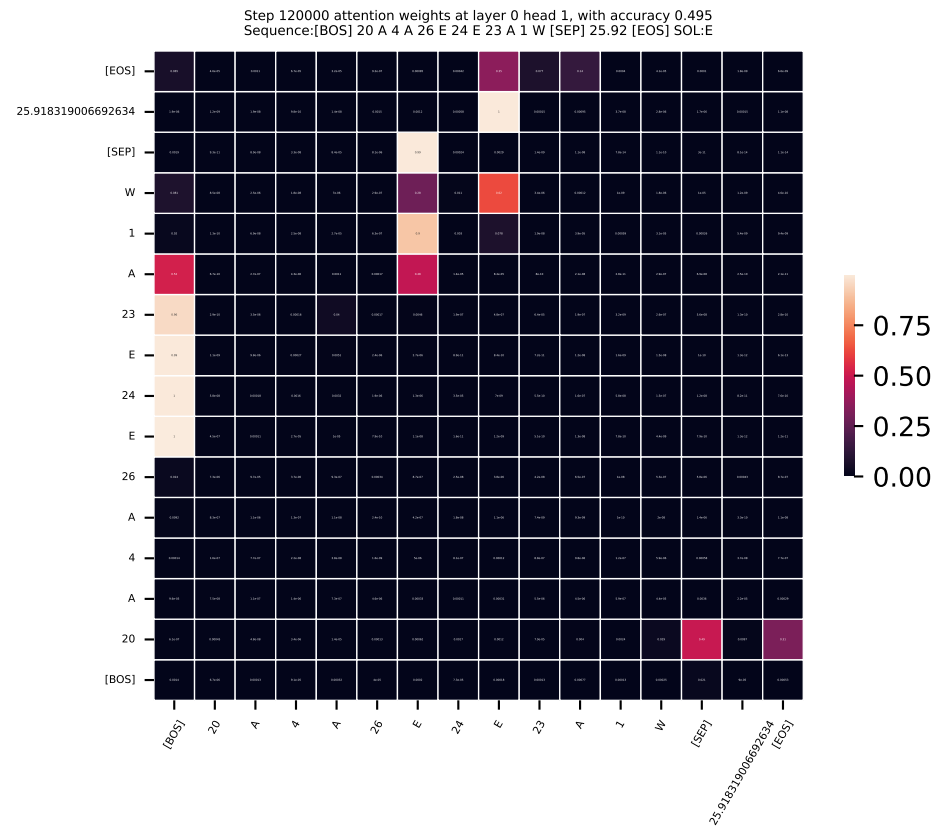

Step 119500 attention weights at layer 1 head 0, with accuracy 0.489
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.2
0.1

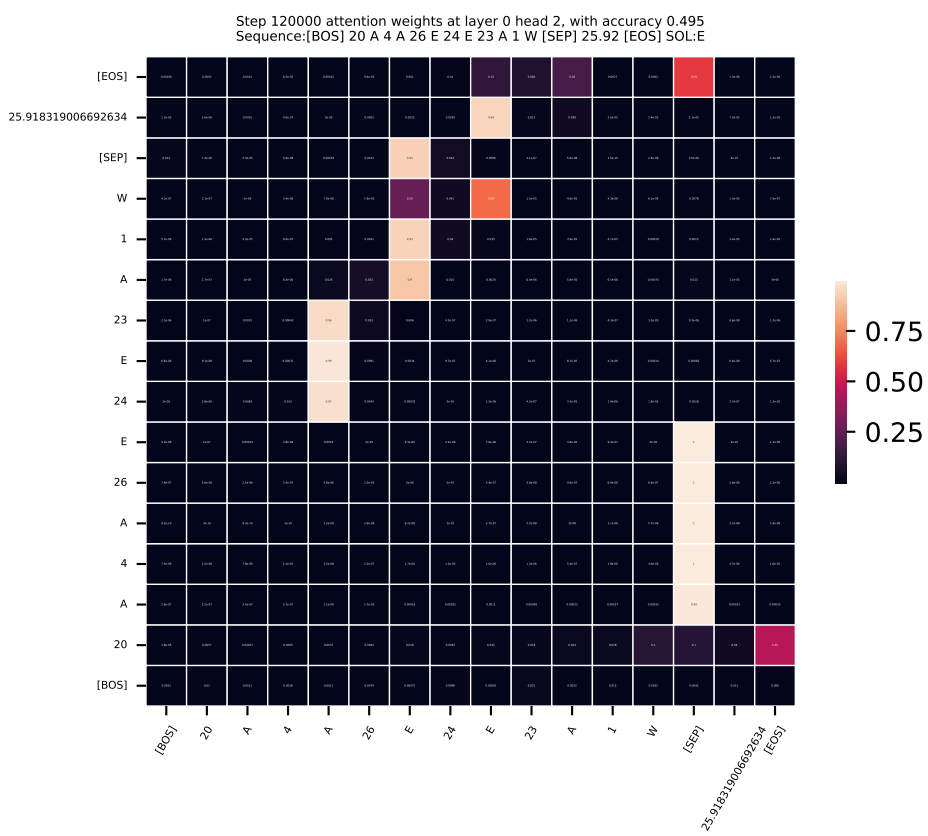

Step 119500 attention weights at layer 1 head 1, with accuracy 0.489
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.2
0.1

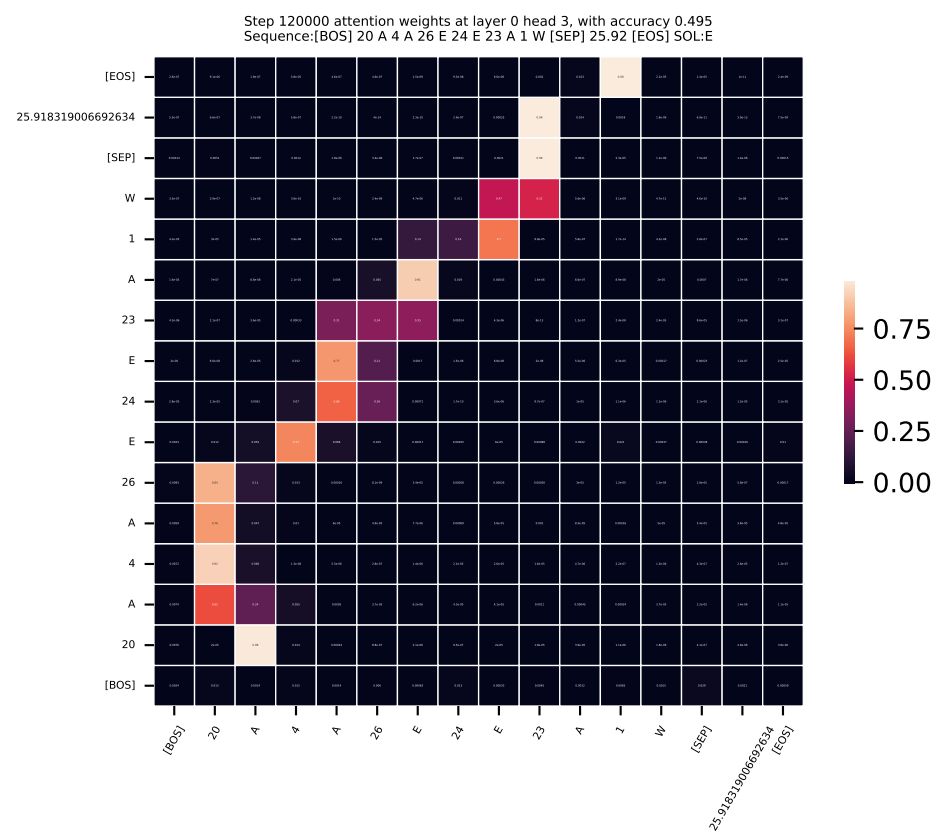

Step 119500 attention weights at layer 1 head 2, with accuracy 0.489
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.15
0.10
0.05

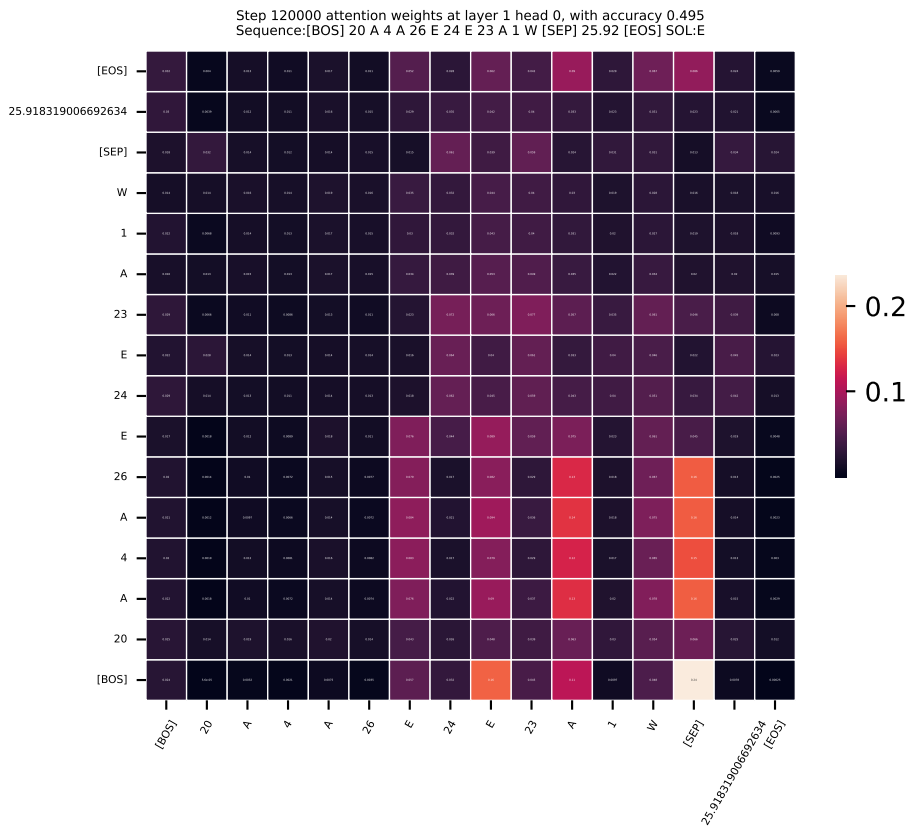

Step 119500 attention weights at layer 1 head 3, with accuracy 0.489
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

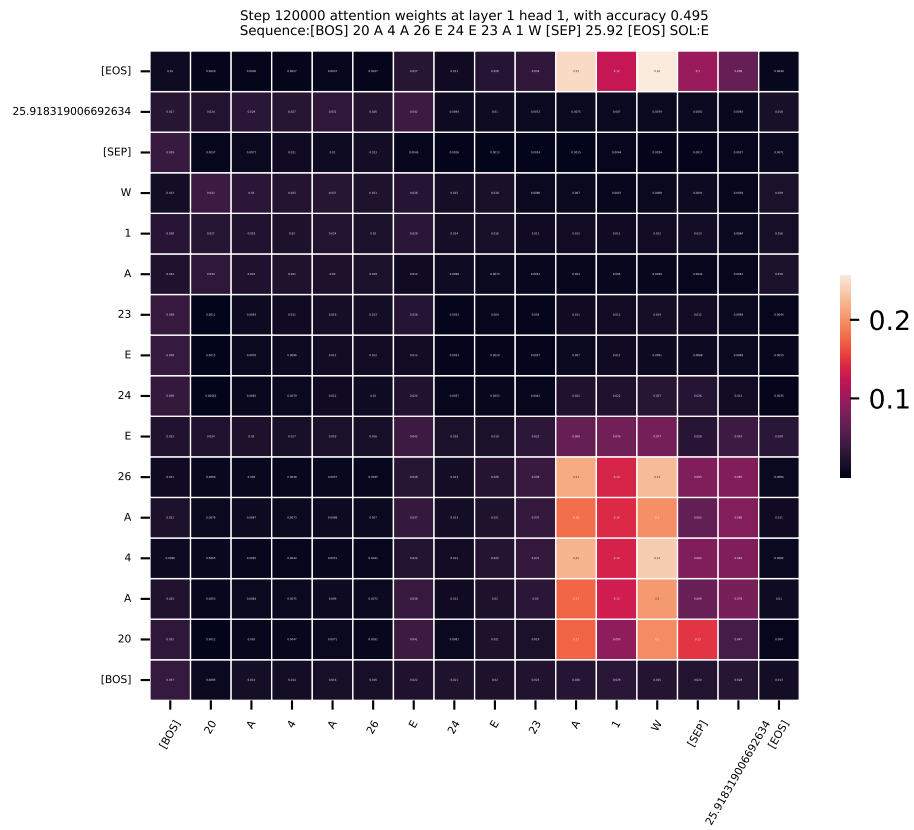

Step 120000 attention weights at layer 0 head 0, with accuracy 0.495
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

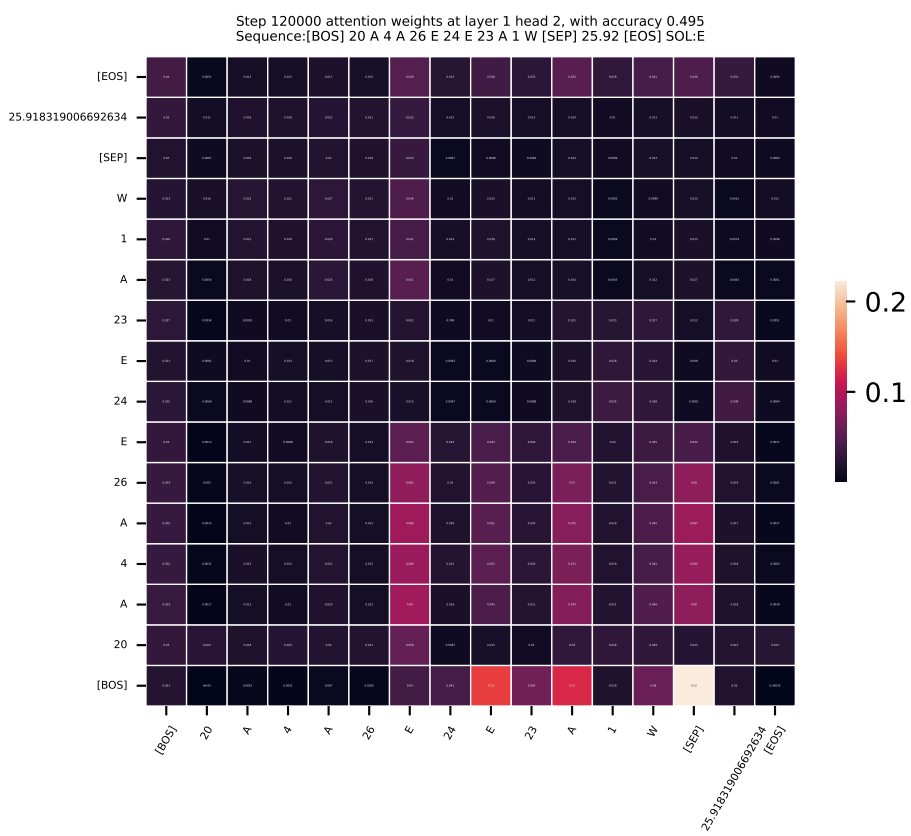

Step 120000 attention weights at layer 0 head 1, with accuracy 0.495
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

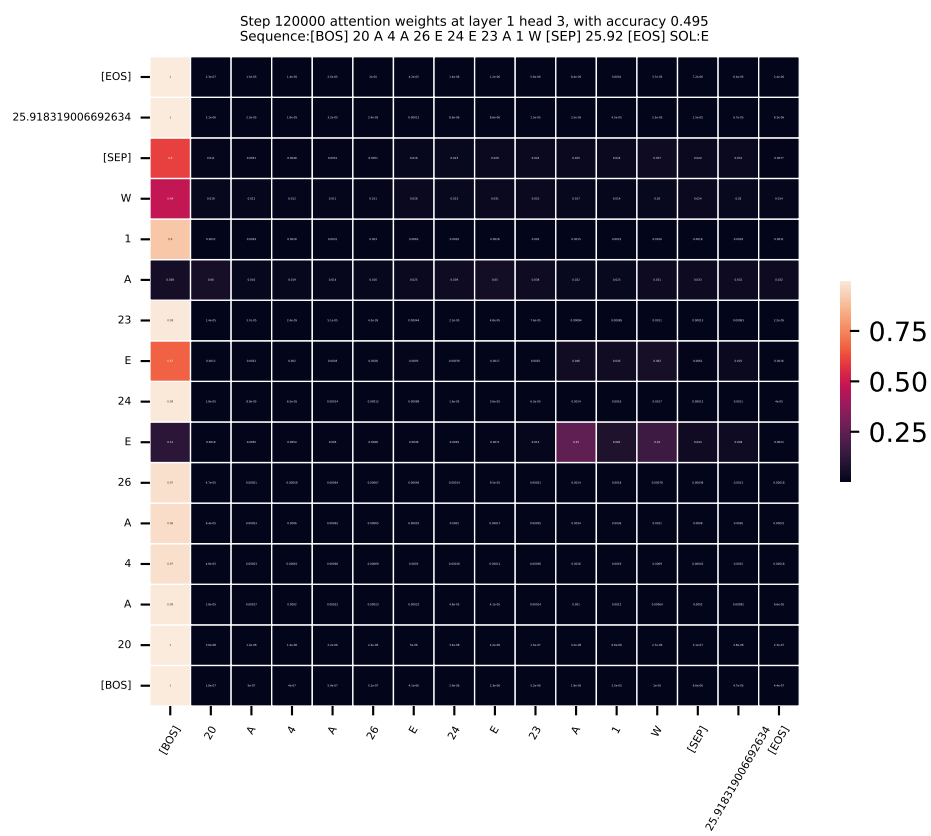

Step 120000 attention weights at layer 0 head 2, with accuracy 0.495
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

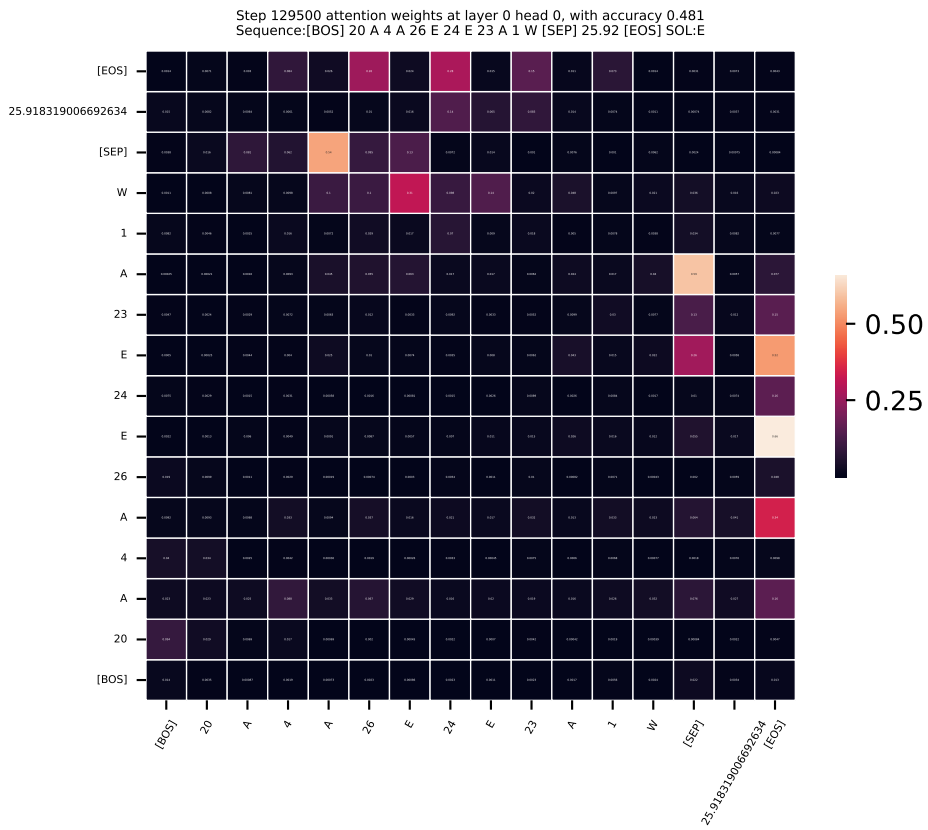

Step 120000 attention weights at layer 0 head 3, with accuracy 0.495
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

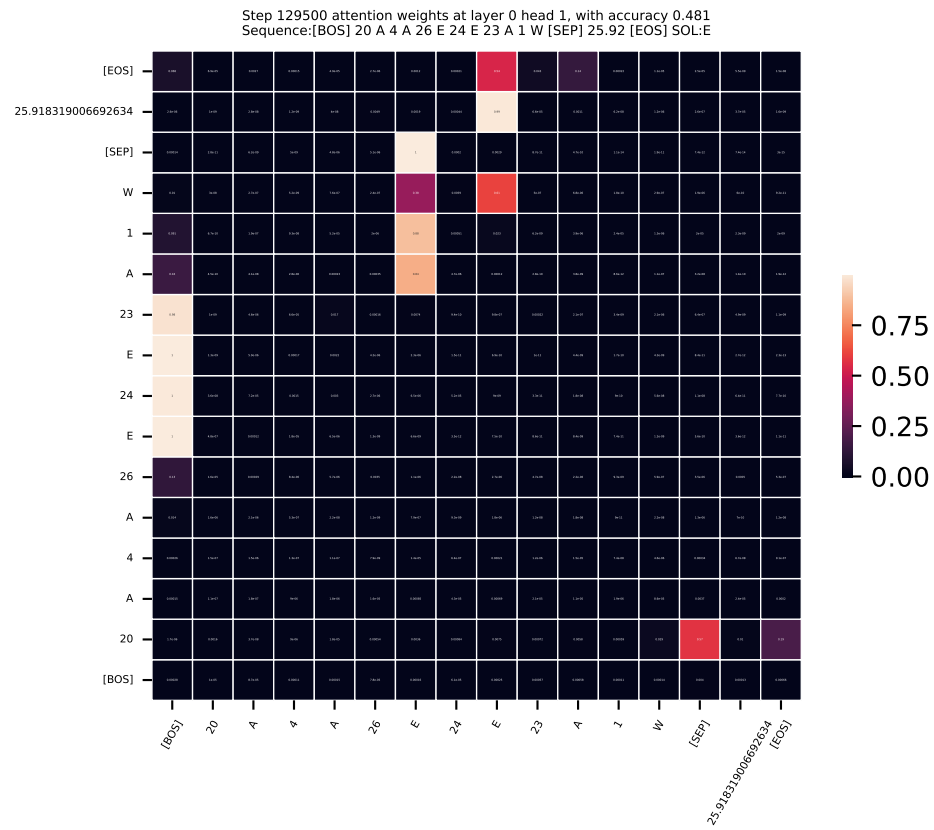

Step 120000 attention weights at layer 1 head 0, with accuracy 0.495
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.2
0.1

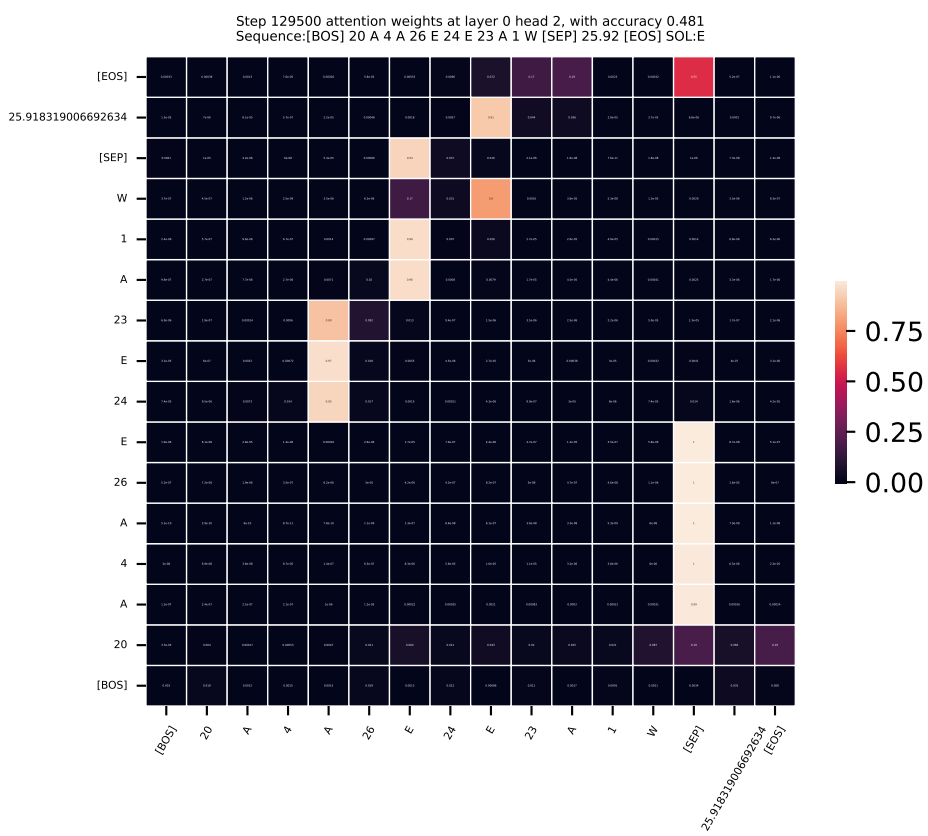

Step 120000 attention weights at layer 1 head 1, with accuracy 0.495
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.2
0.1

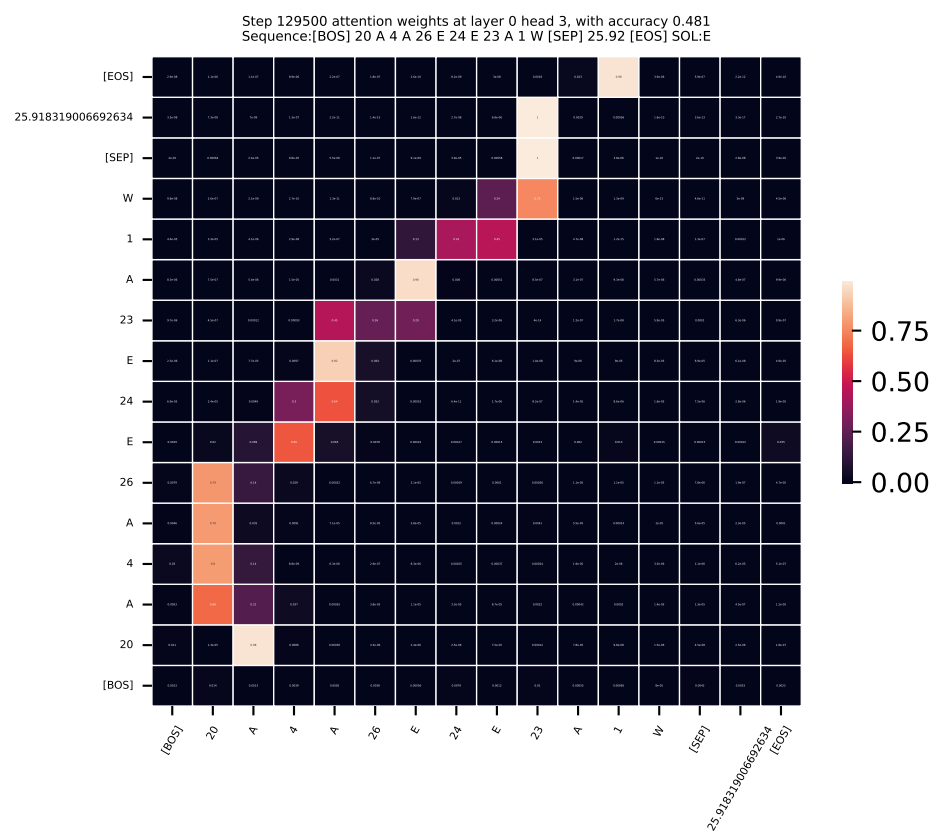

Step 120000 attention weights at layer 1 head 2, with accuracy 0.495
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.2
0.1

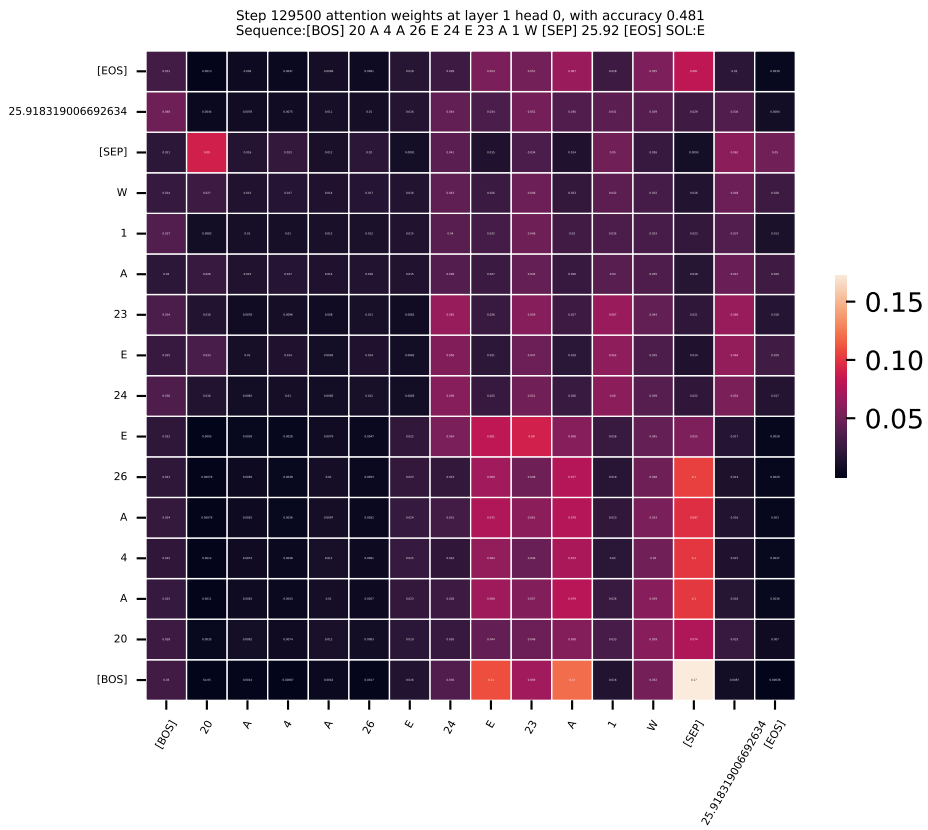

Step 120000 attention weights at layer 1 head 3, with accuracy 0.495
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

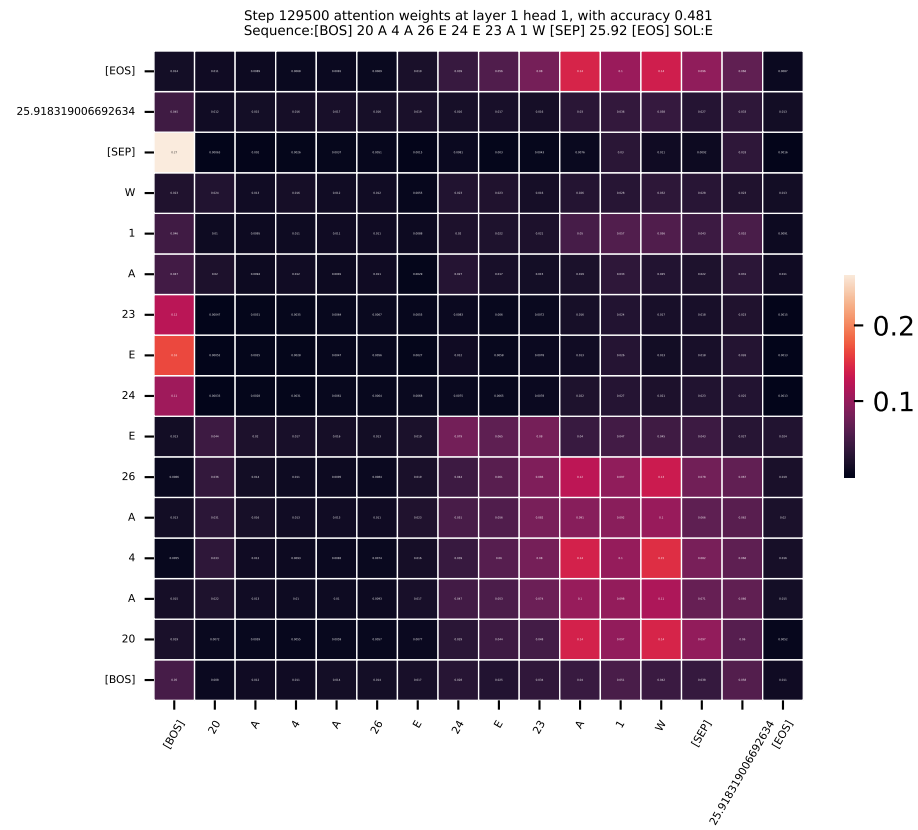

Step 129500 attention weights at layer 0 head 0, with accuracy 0.481
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

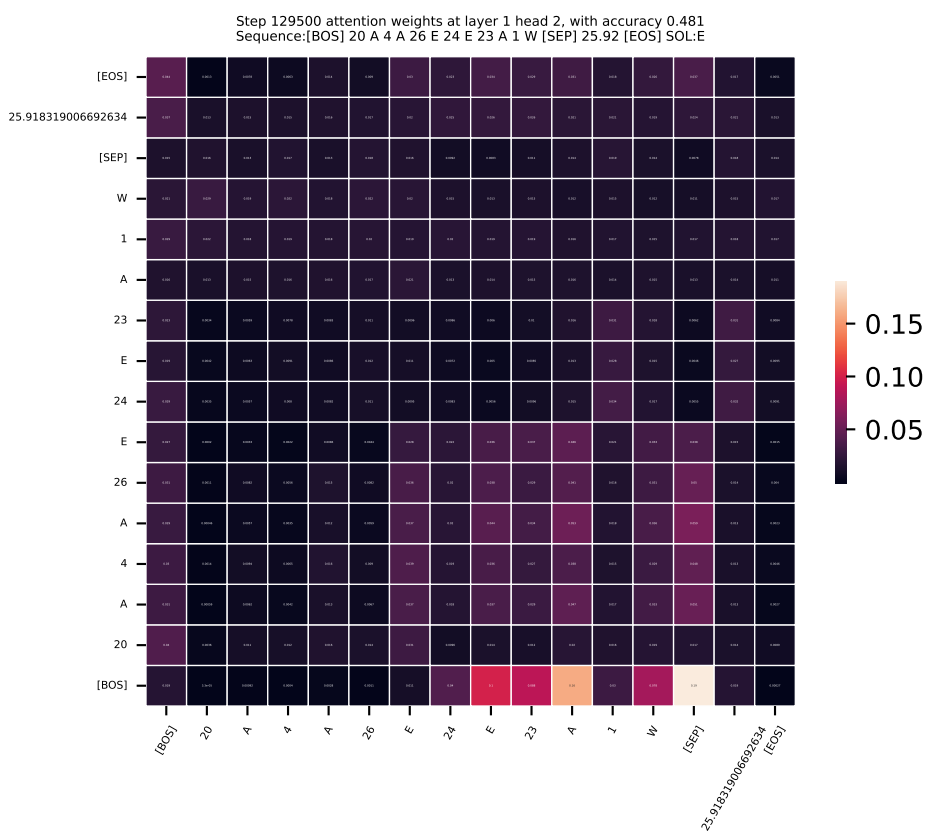

Step 129500 attention weights at layer 0 head 1, with accuracy 0.481
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

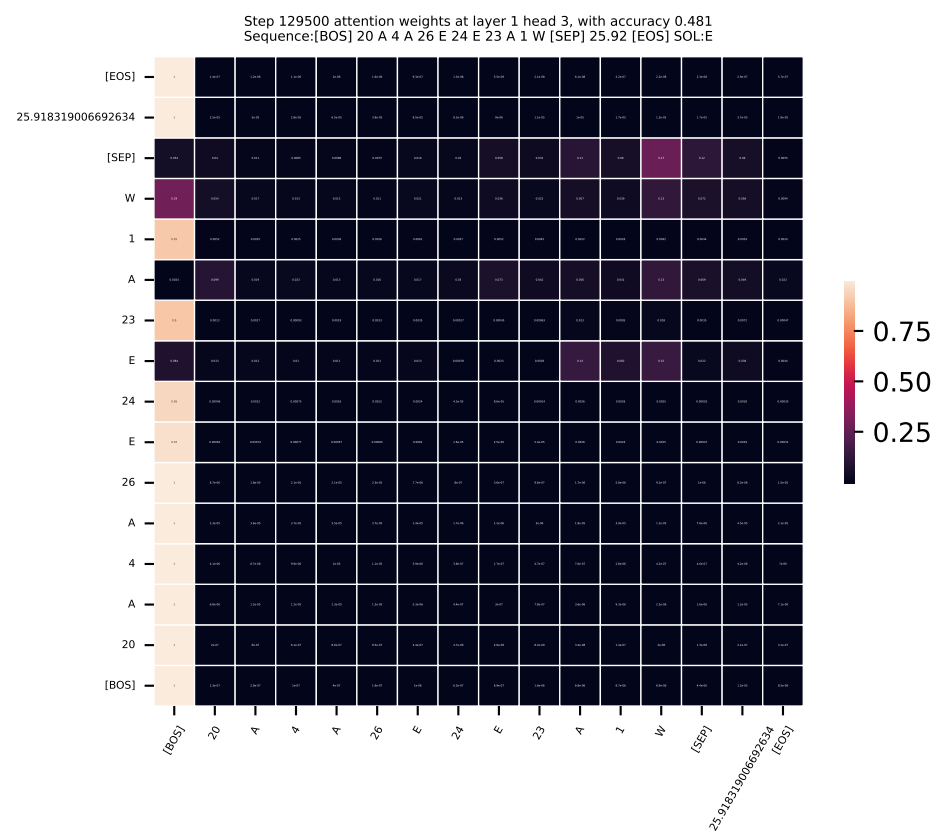

Step 129500 attention weights at layer 0 head 2, with accuracy 0.481
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

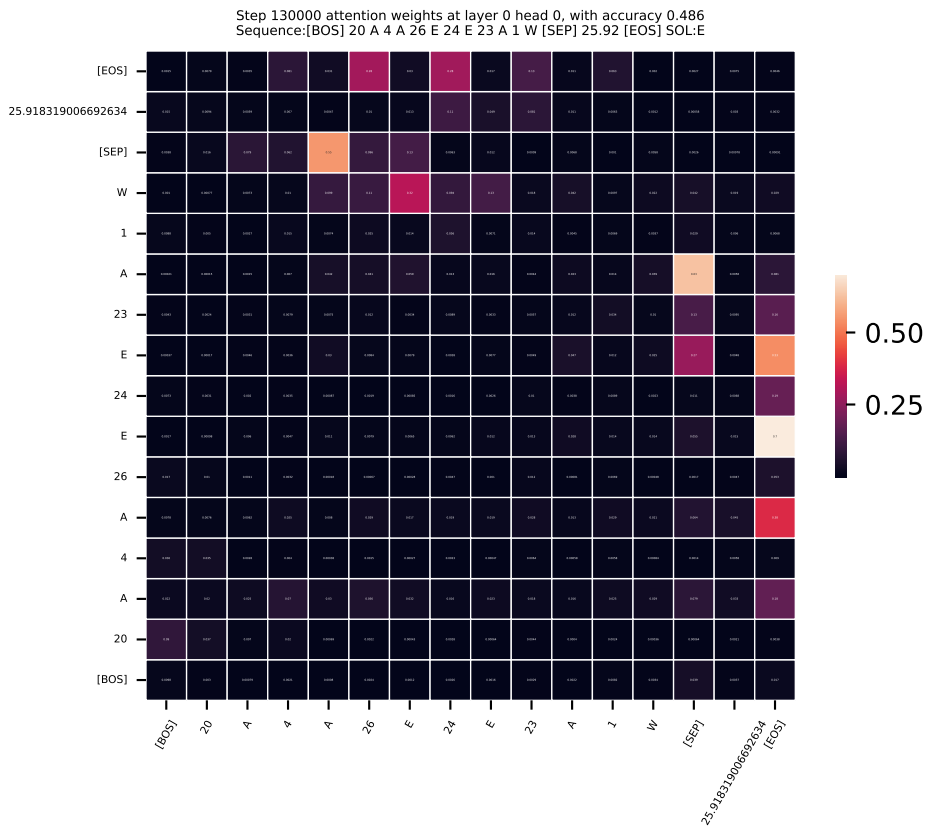

Step 129500 attention weights at layer 0 head 3, with accuracy 0.481
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

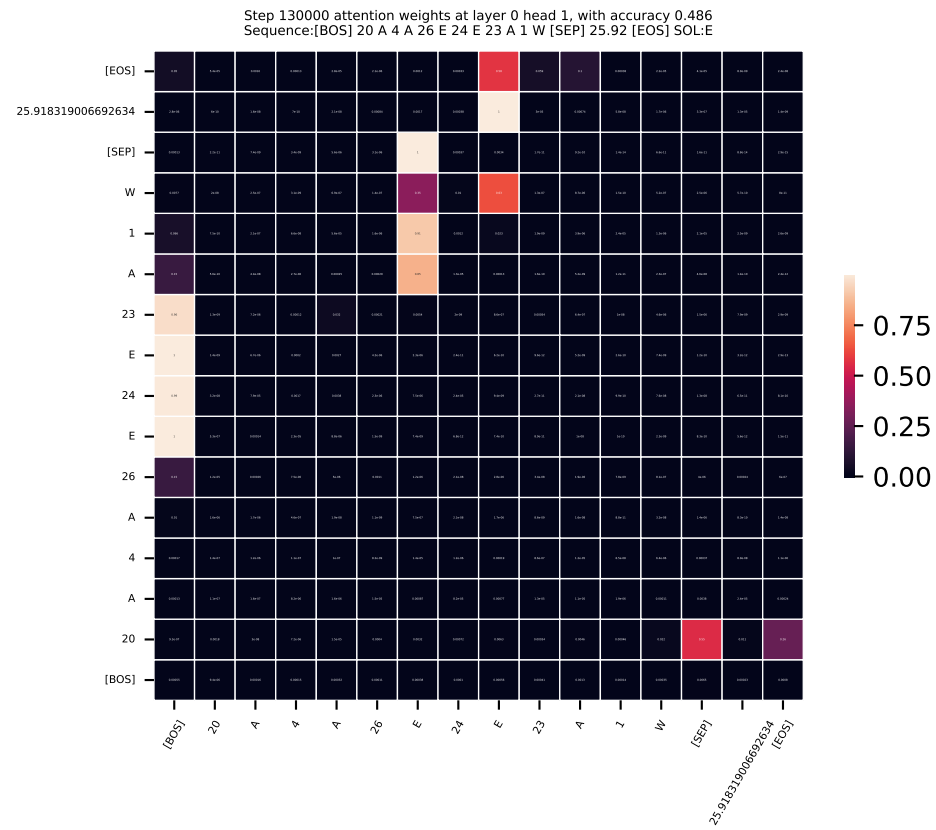

Step 129500 attention weights at layer 1 head 0, with accuracy 0.481
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.15
0.10
0.05

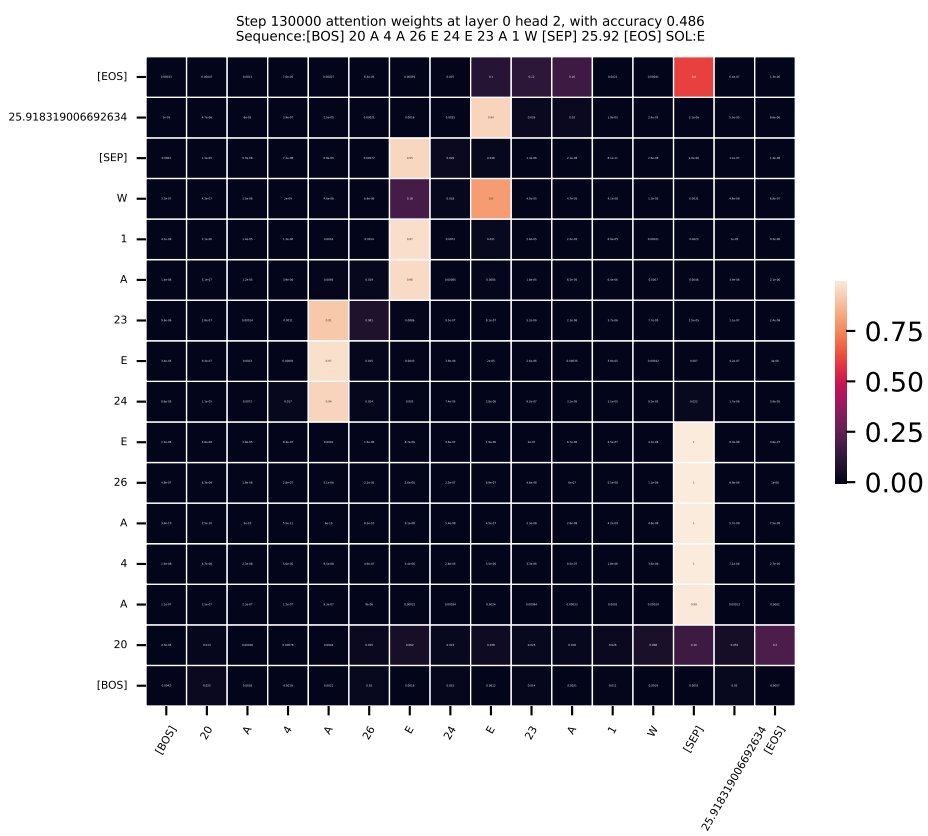

Step 129500 attention weights at layer 1 head 1, with accuracy 0.481
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.2
0.1

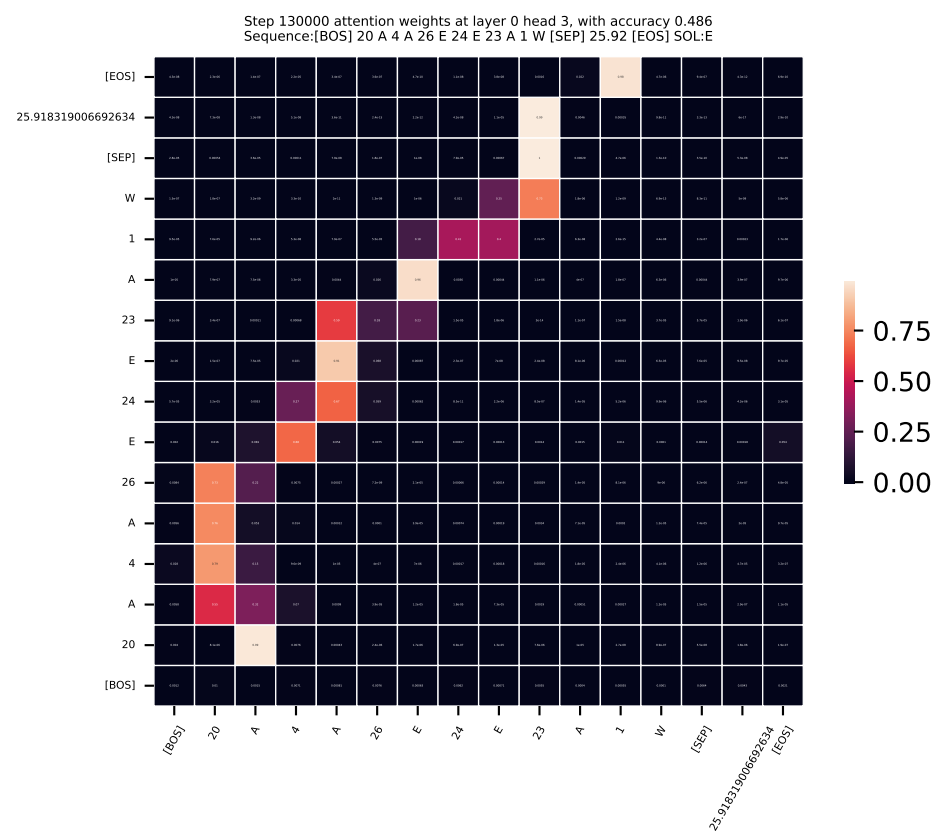

Step 129500 attention weights at layer 1 head 2, with accuracy 0.481
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.15
0.10
0.05

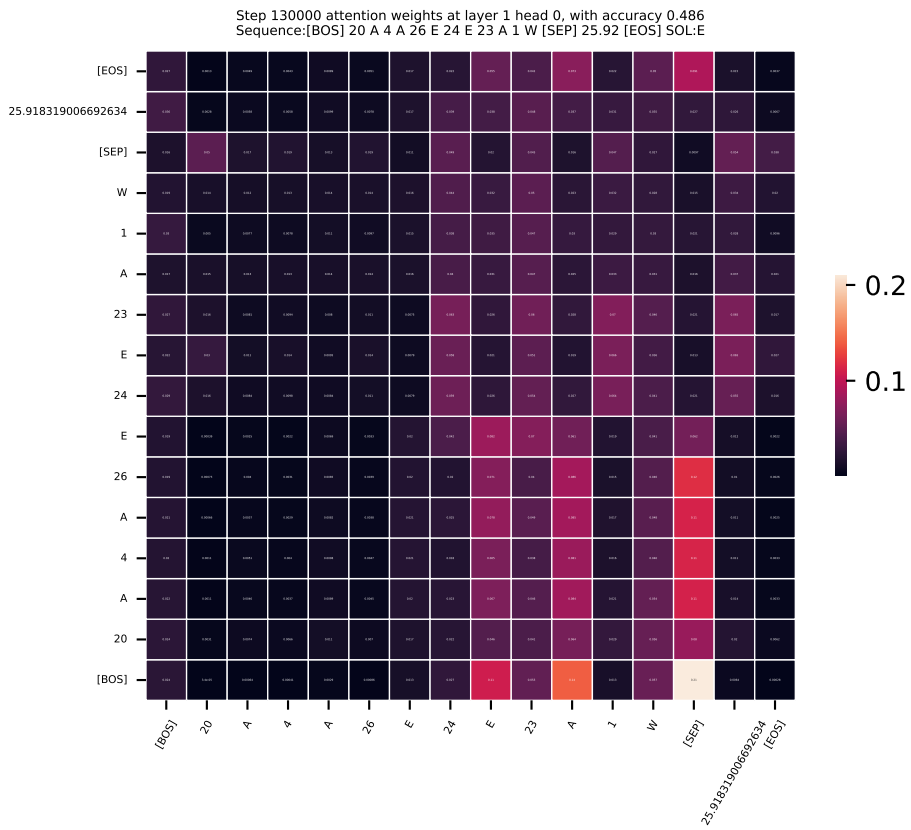

Step 129500 attention weights at layer 1 head 3, with accuracy 0.481
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

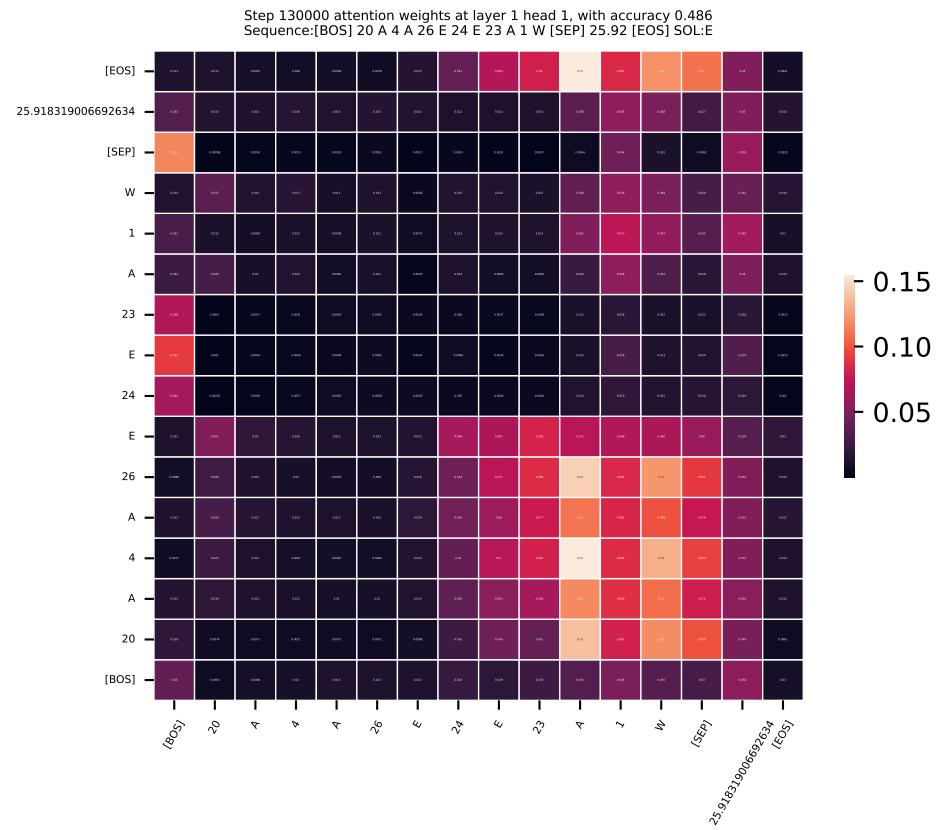

Step 130000 attention weights at layer 0 head 0, with accuracy 0.486
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

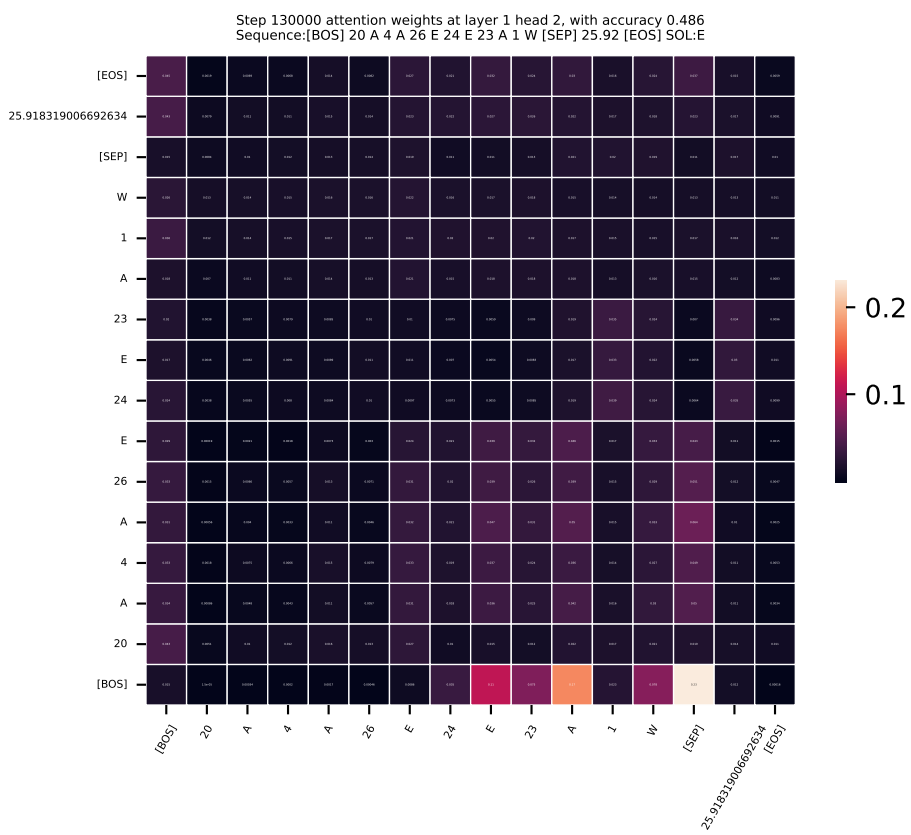

Step 130000 attention weights at layer 0 head 1, with accuracy 0.486
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

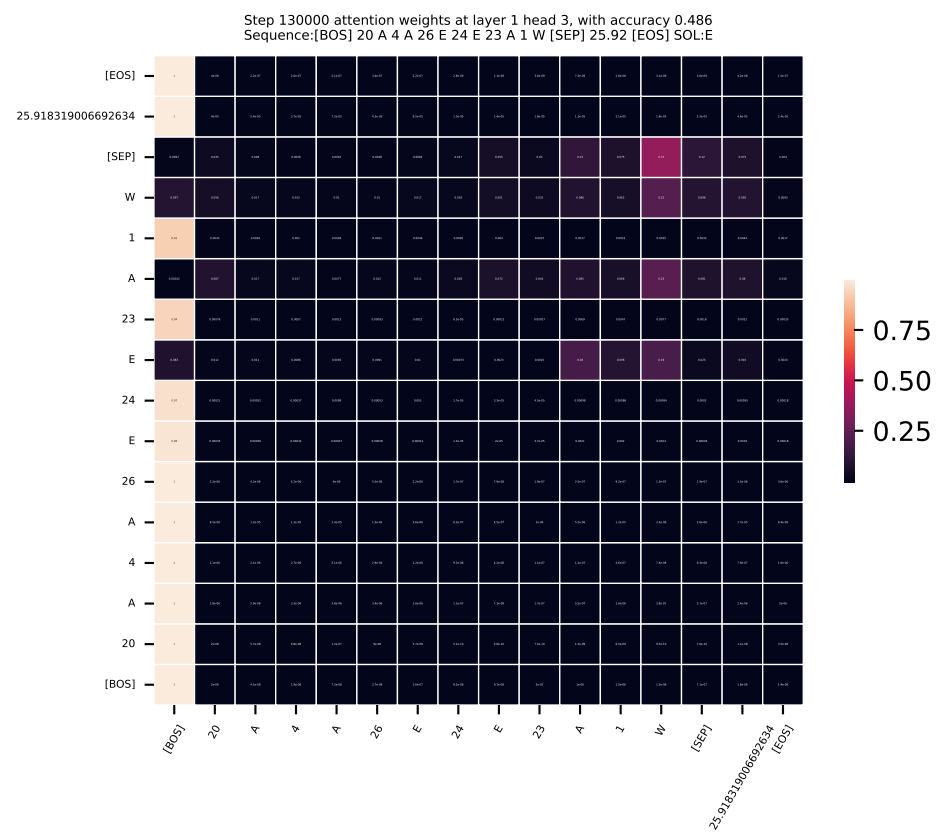

Step 130000 attention weights at layer 0 head 2, with accuracy 0.486
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

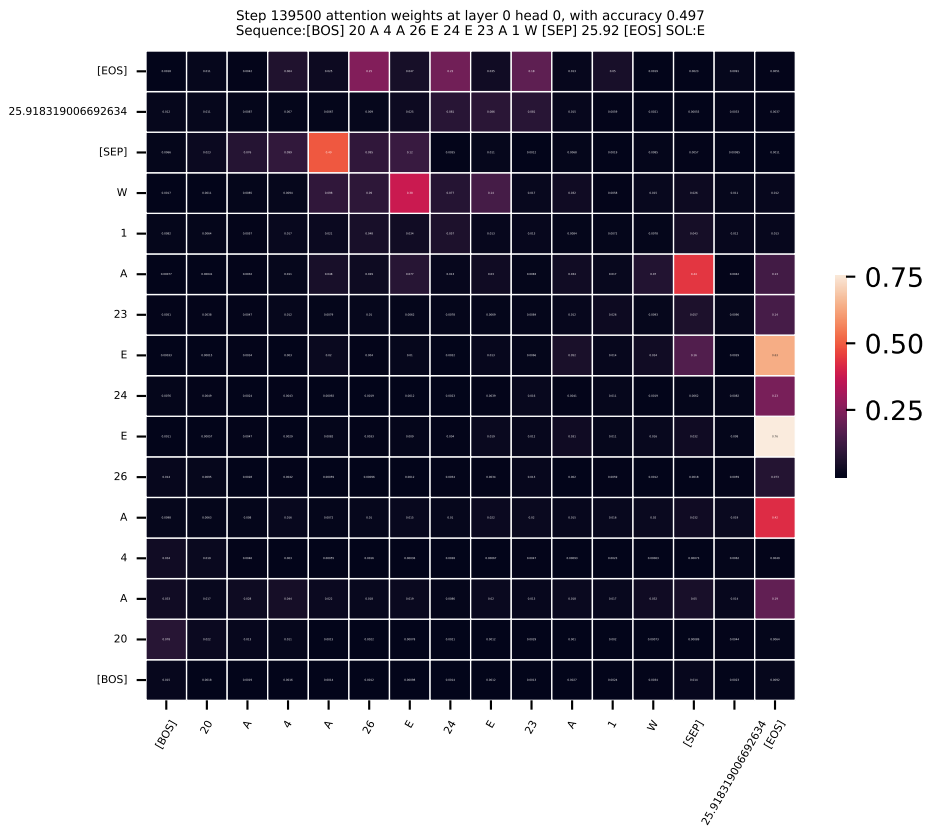

Step 130000 attention weights at layer 0 head 3, with accuracy 0.486
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

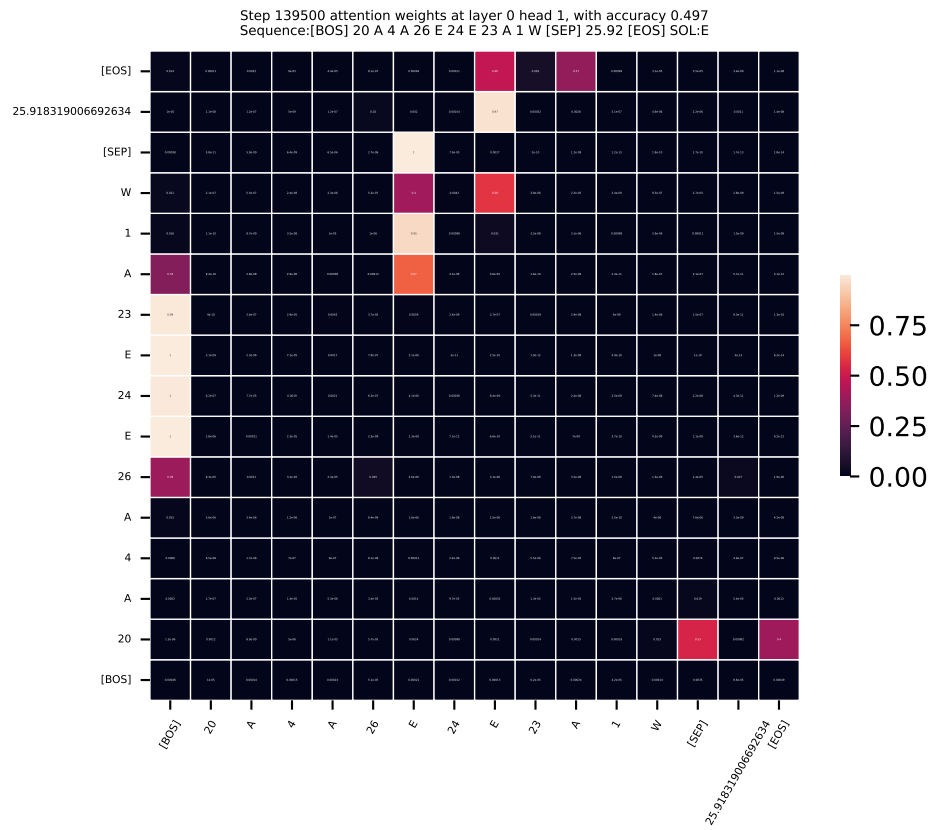

Step 130000 attention weights at layer 1 head 0, with accuracy 0.486
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.2
0.1

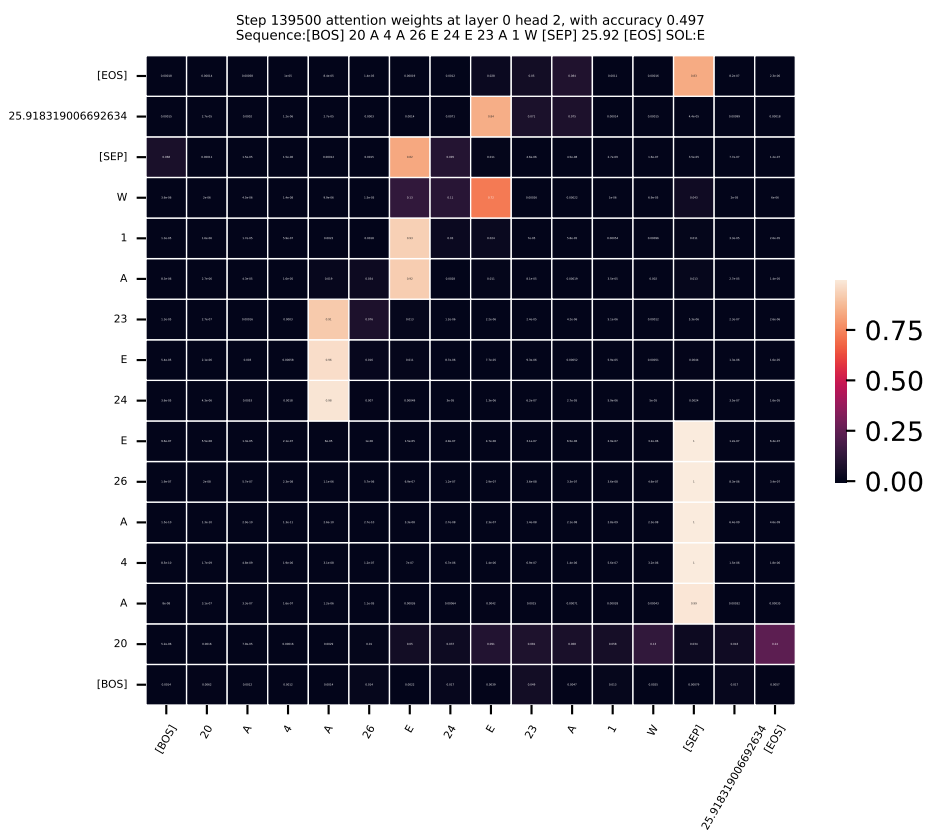

Step 130000 attention weights at layer 1 head 1, with accuracy 0.486
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.15
0.10
0.05

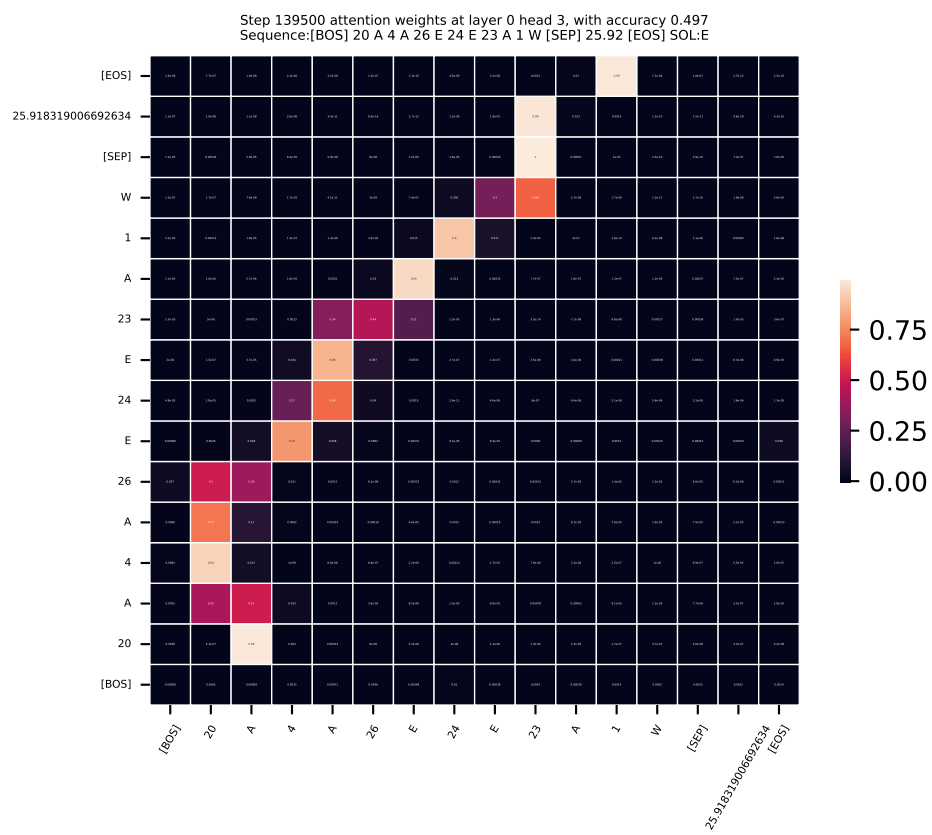

Step 130000 attention weights at layer 1 head 2, with accuracy 0.486
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.2
0.1

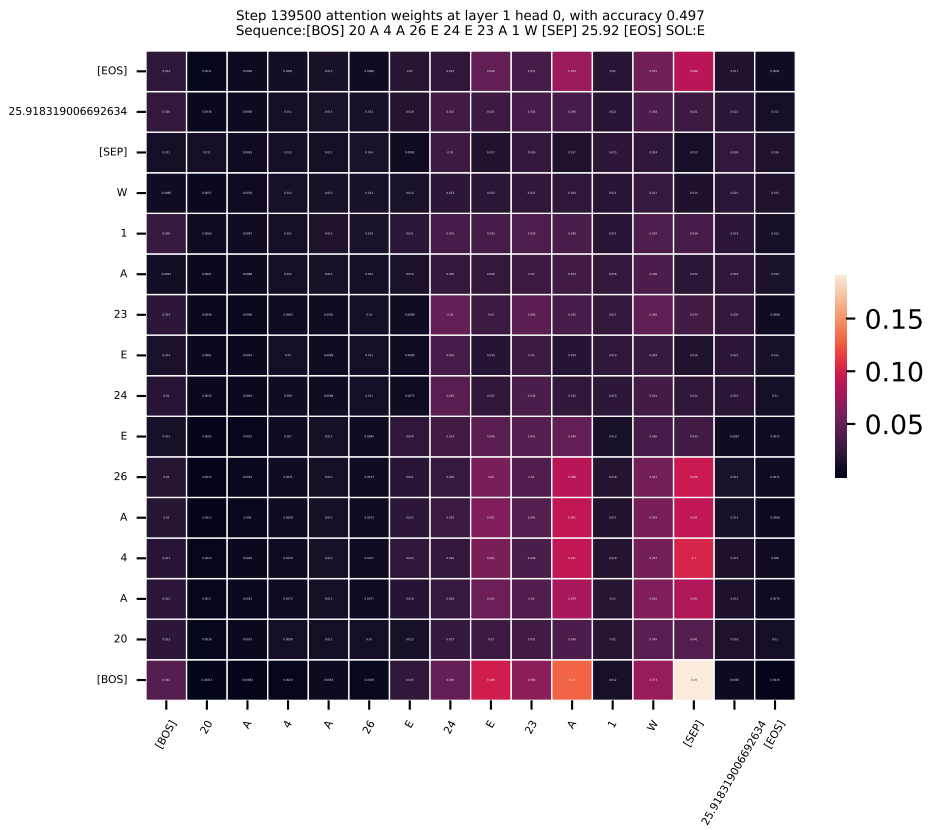

Step 130000 attention weights at layer 1 head 3, with accuracy 0.486
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

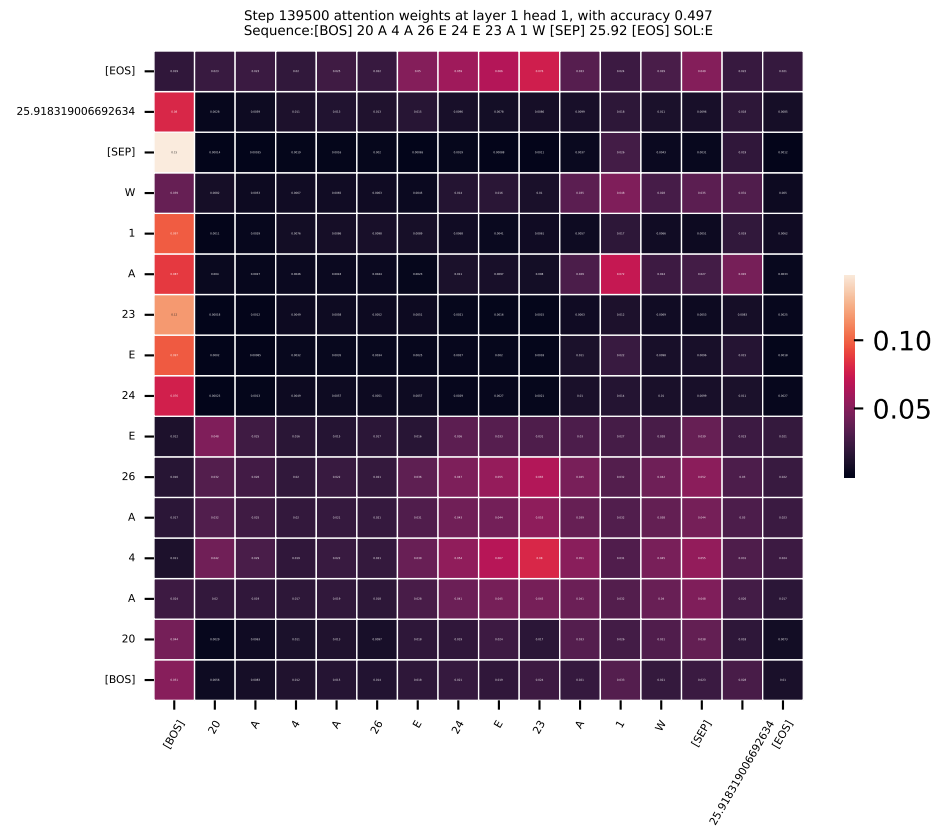

Step 139500 attention weights at layer 0 head 0, with accuracy 0.497
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

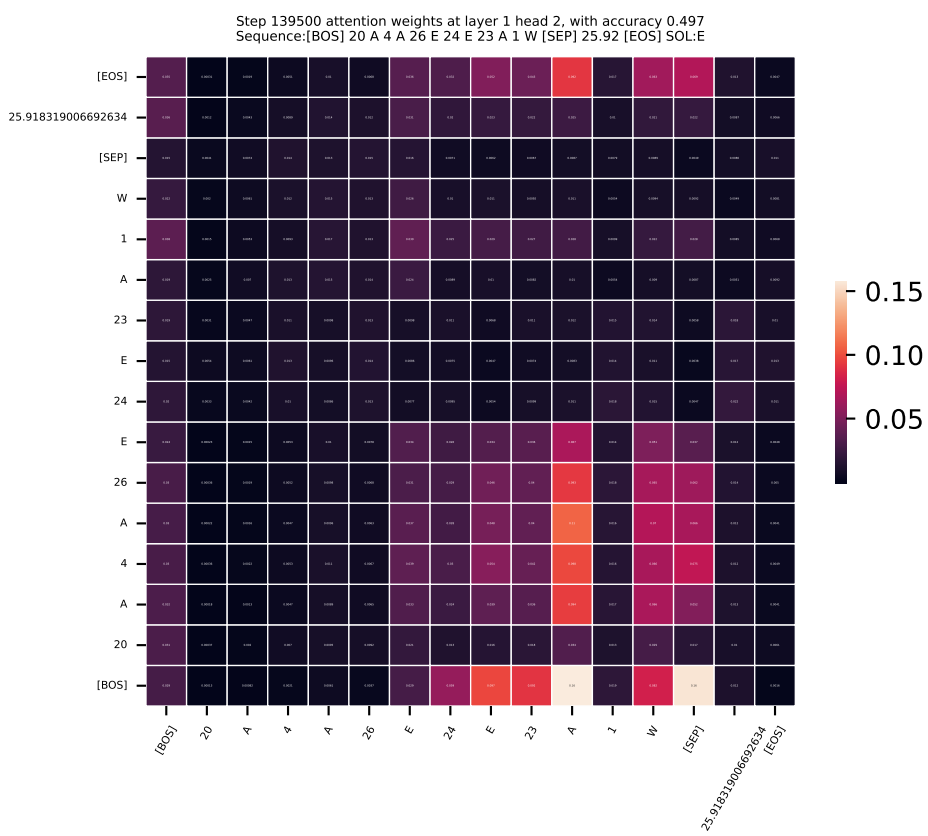

Step 139500 attention weights at layer 0 head 1, with accuracy 0.497
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

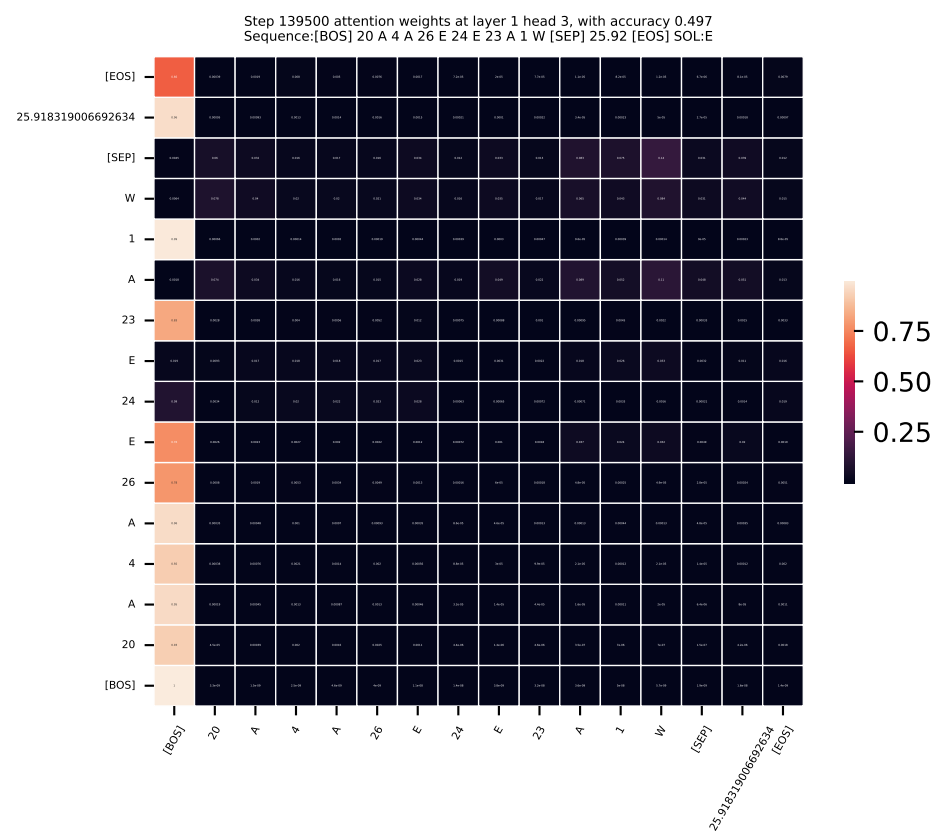

Step 139500 attention weights at layer 0 head 2, with accuracy 0.497
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

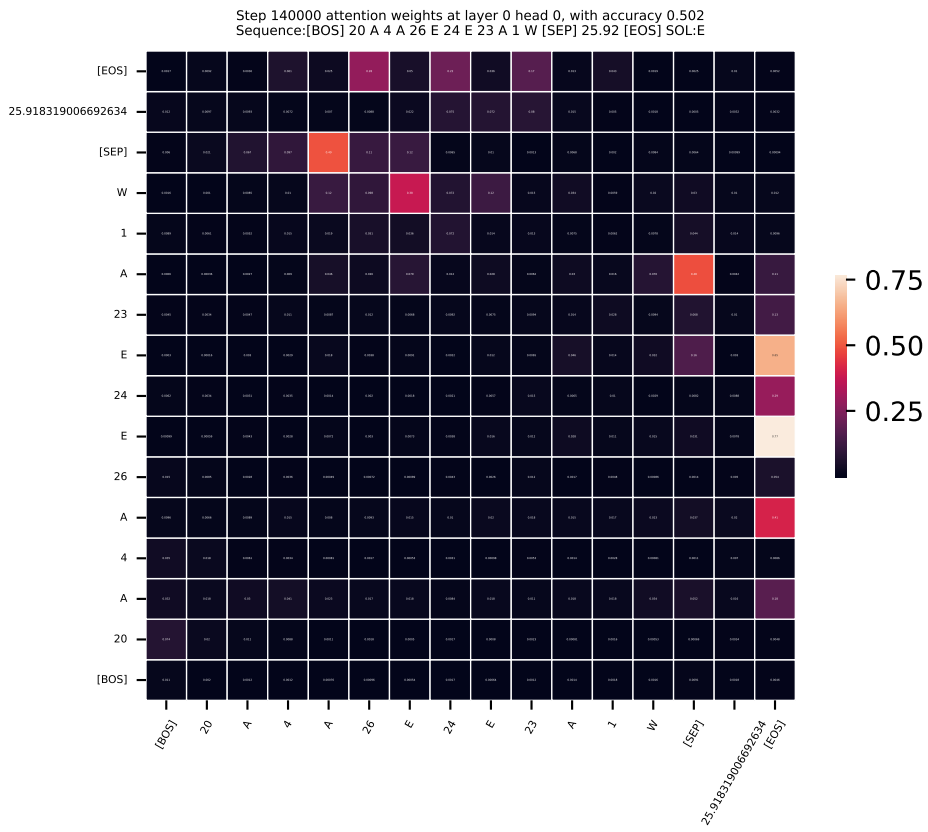

Step 139500 attention weights at layer 0 head 3, with accuracy 0.497
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

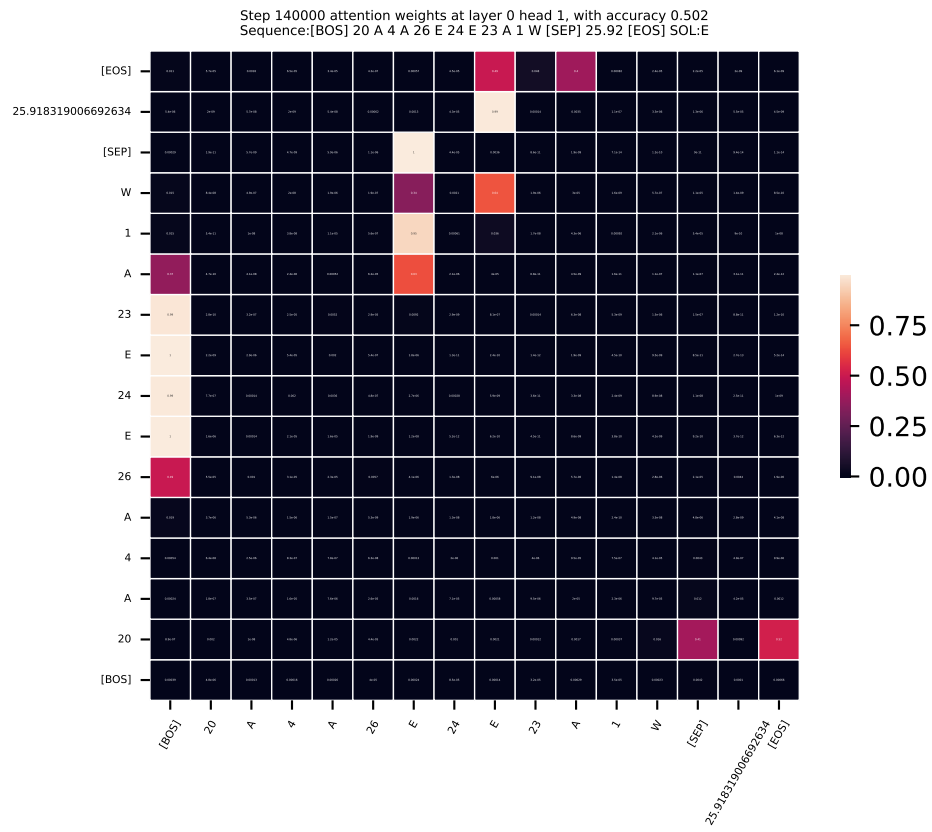

Step 139500 attention weights at layer 1 head 0, with accuracy 0.497
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.15
0.10
0.05

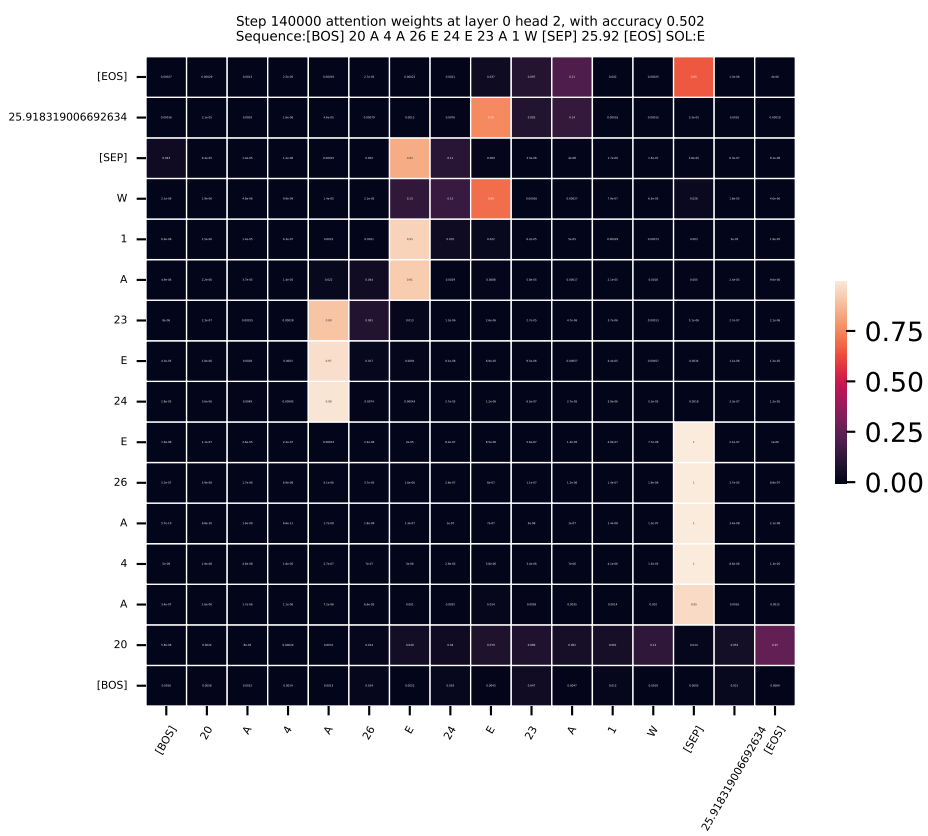

Step 139500 attention weights at layer 1 head 1, with accuracy 0.497
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.10
0.05

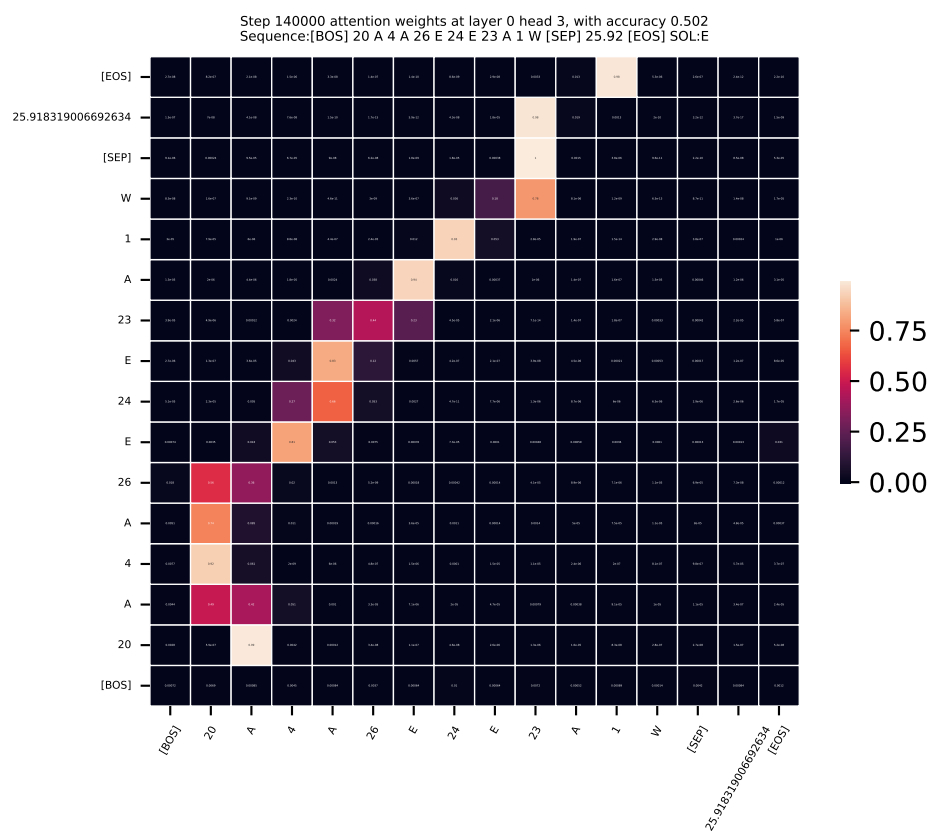

Step 139500 attention weights at layer 1 head 2, with accuracy 0.497
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.15
0.10
0.05

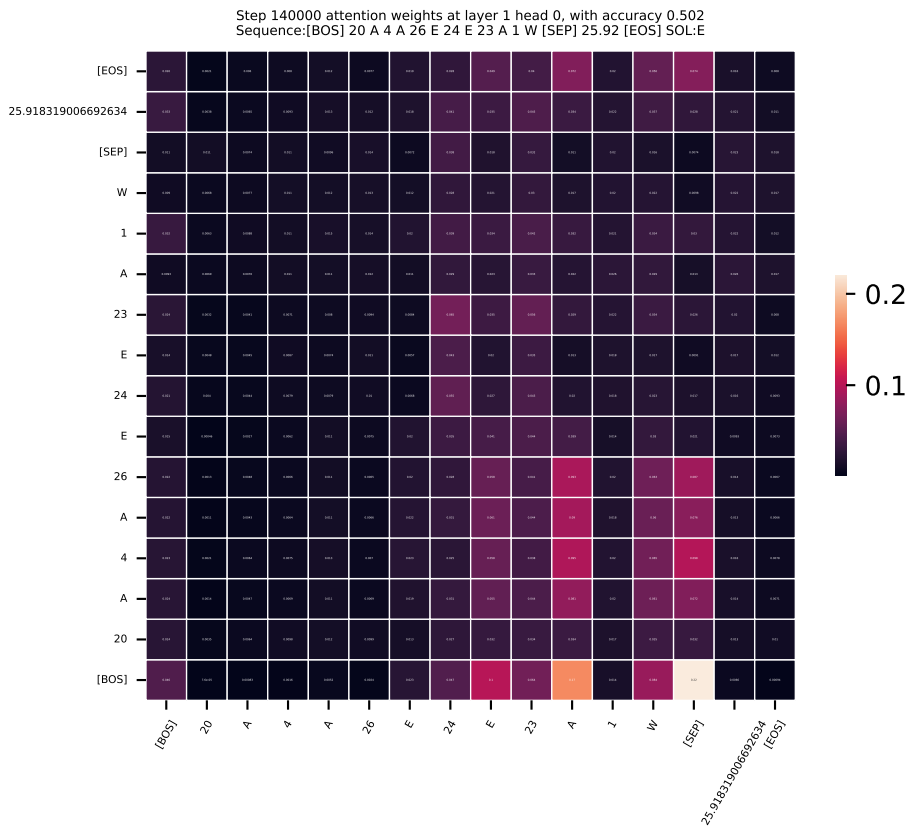

Step 139500 attention weights at layer 1 head 3, with accuracy 0.497
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

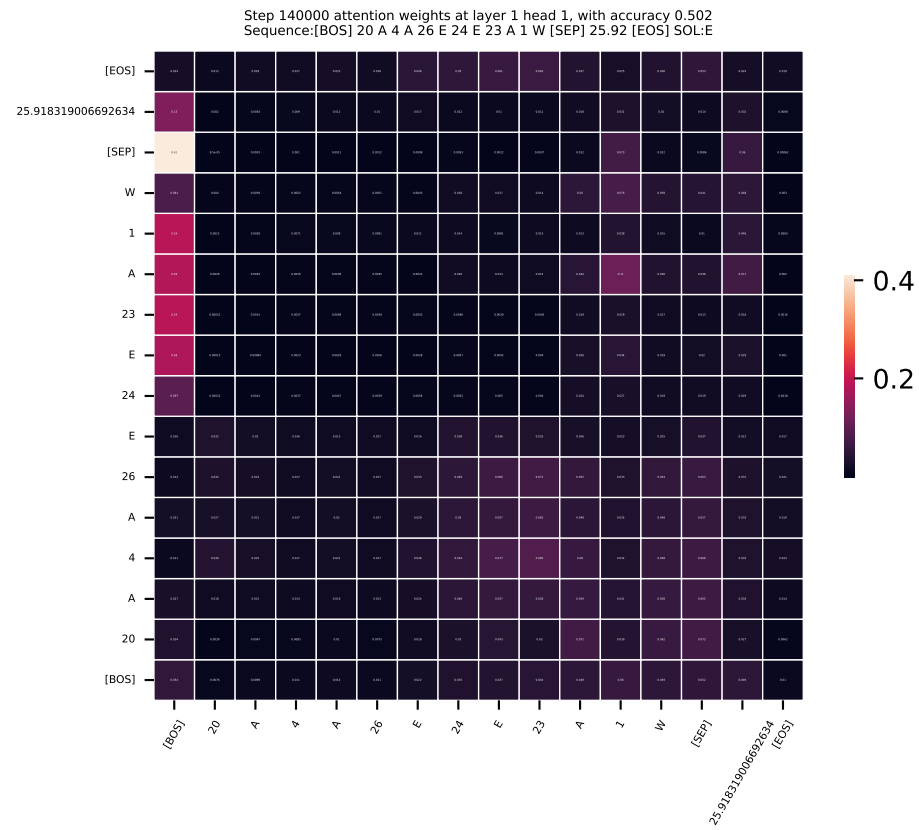

Step 140000 attention weights at layer 0 head 0, with accuracy 0.502
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

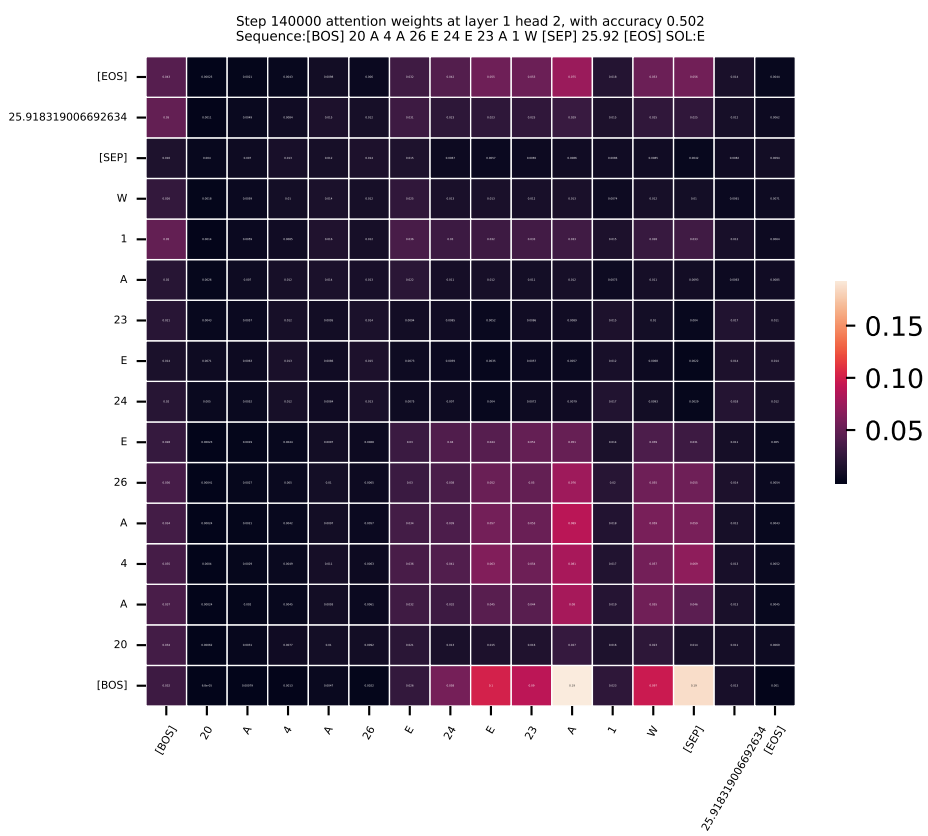

Step 140000 attention weights at layer 0 head 1, with accuracy 0.502
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

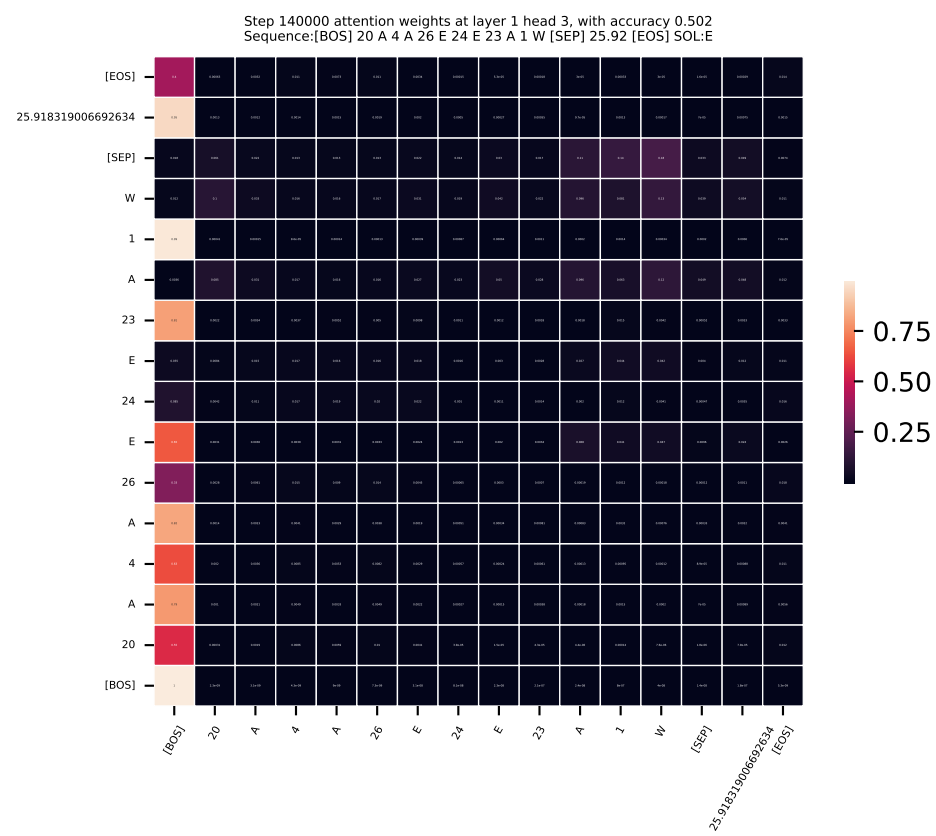

Step 140000 attention weights at layer 0 head 2, with accuracy 0.502
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

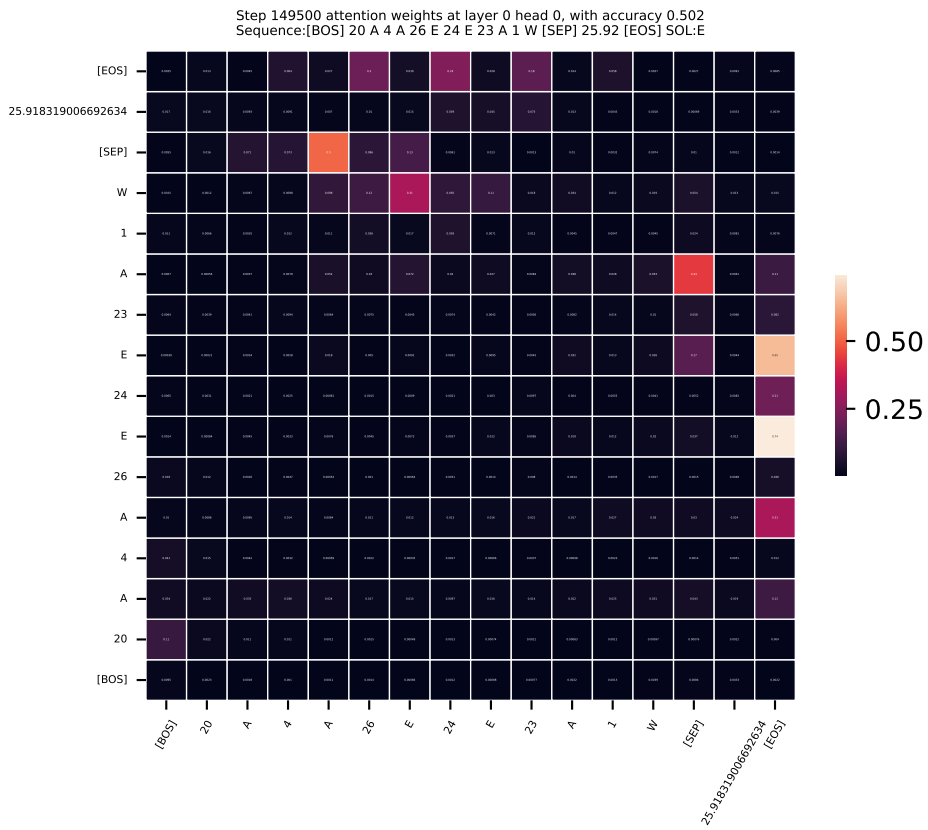

Step 140000 attention weights at layer 0 head 3, with accuracy 0.502
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

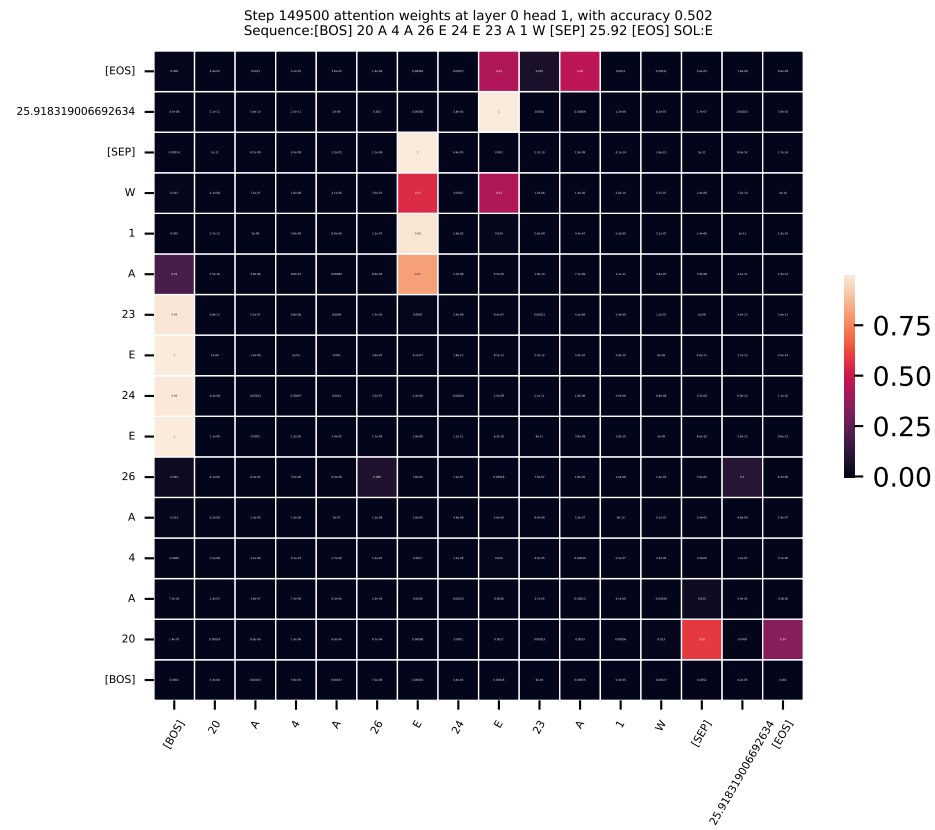

Step 140000 attention weights at layer 1 head 0, with accuracy 0.502
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.2
0.1

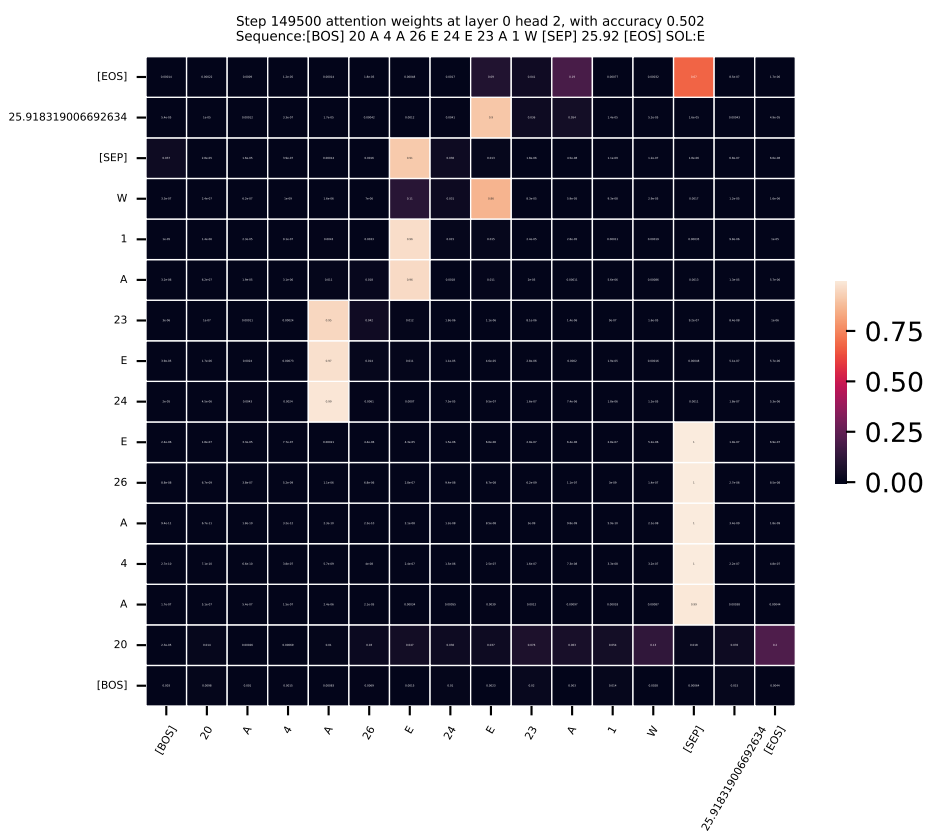

Step 140000 attention weights at layer 1 head 1, with accuracy 0.502
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

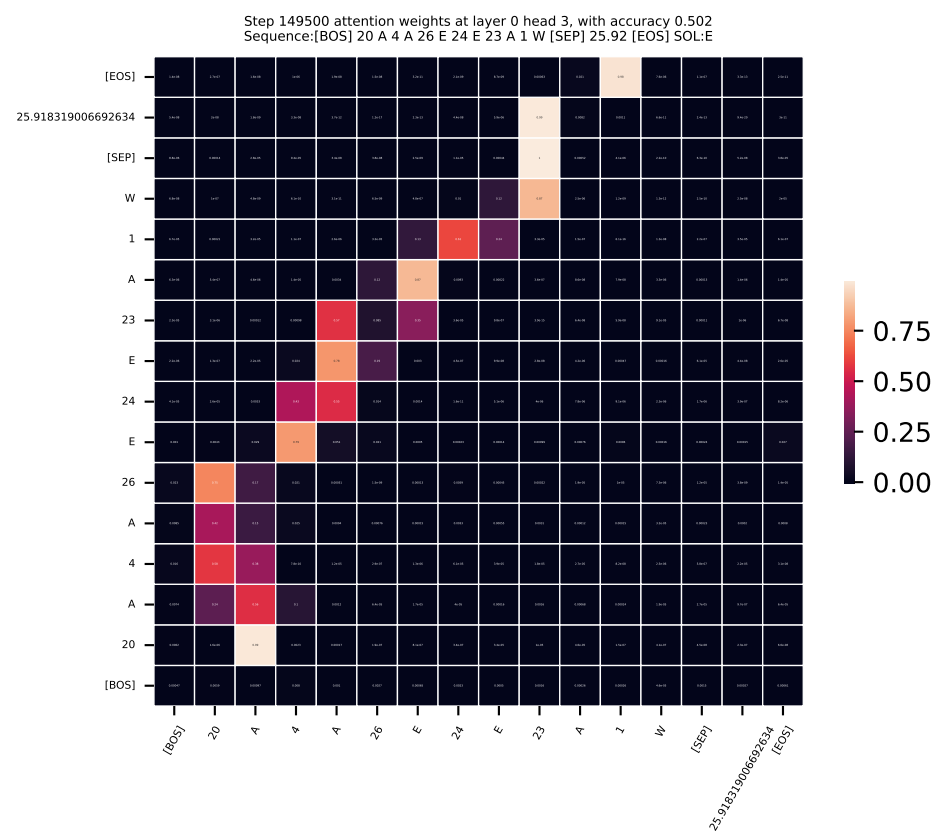

Step 140000 attention weights at layer 1 head 2, with accuracy 0.502
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.15
0.10
0.05

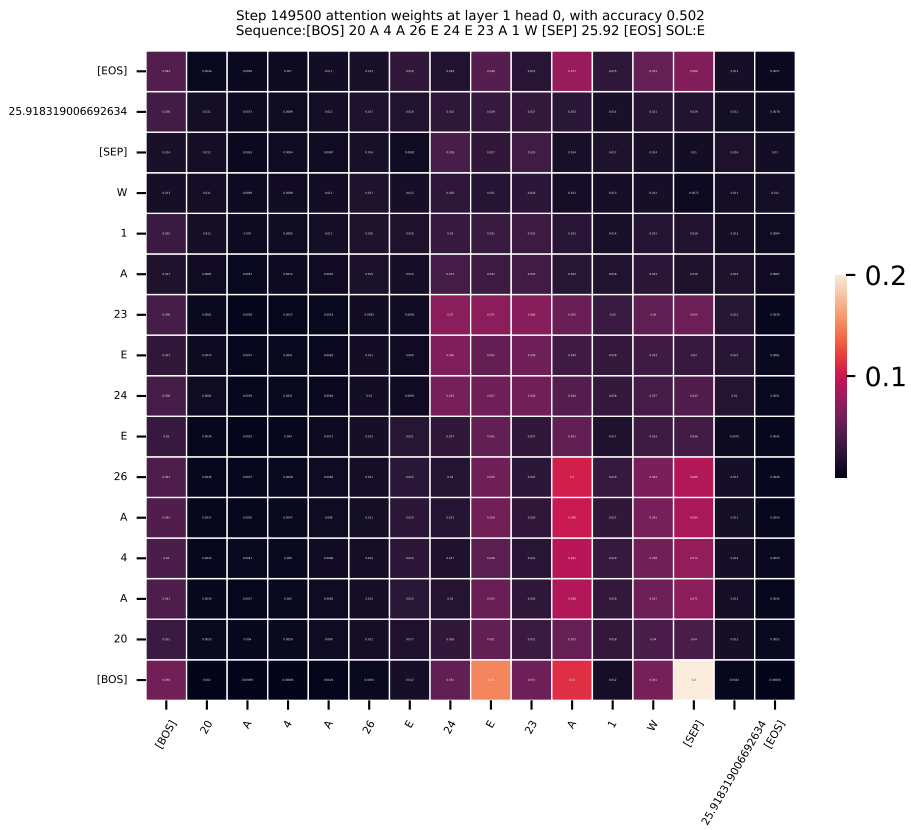

Step 140000 attention weights at layer 1 head 3, with accuracy 0.502
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

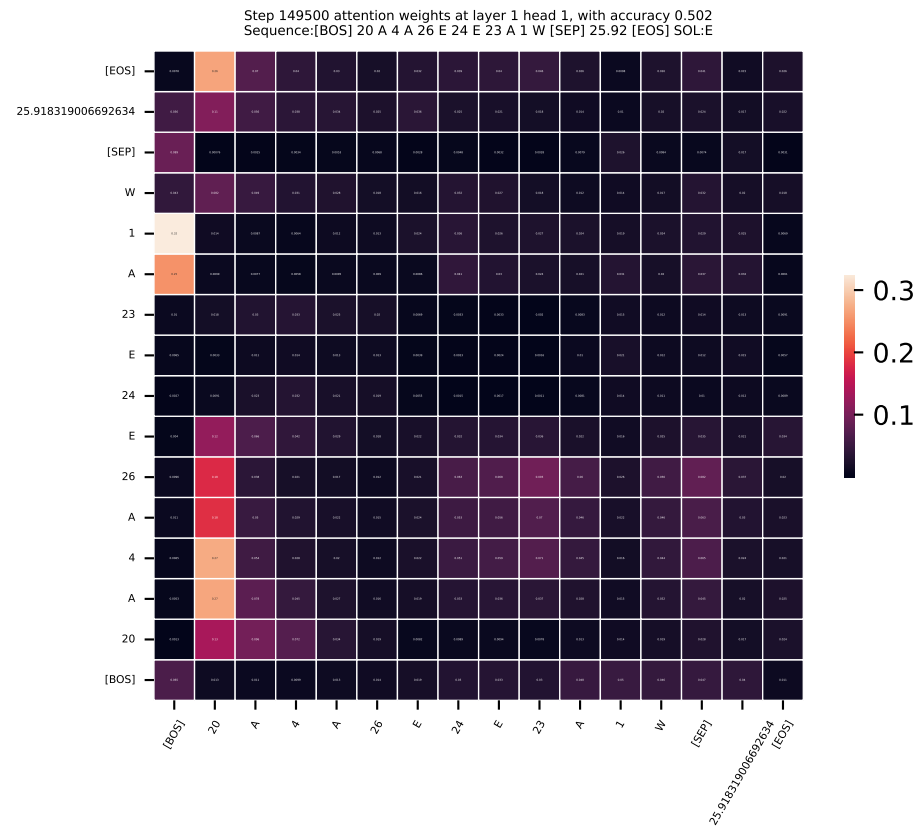

Step 149500 attention weights at layer 0 head 0, with accuracy 0.502
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

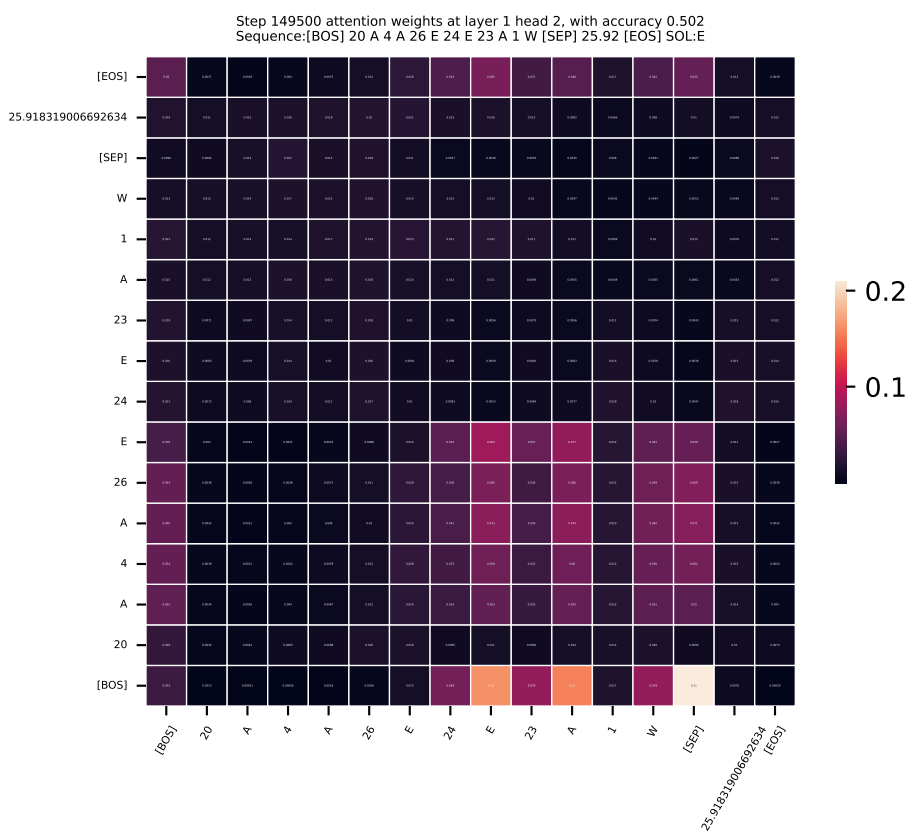

Step 149500 attention weights at layer 0 head 1, with accuracy 0.502
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

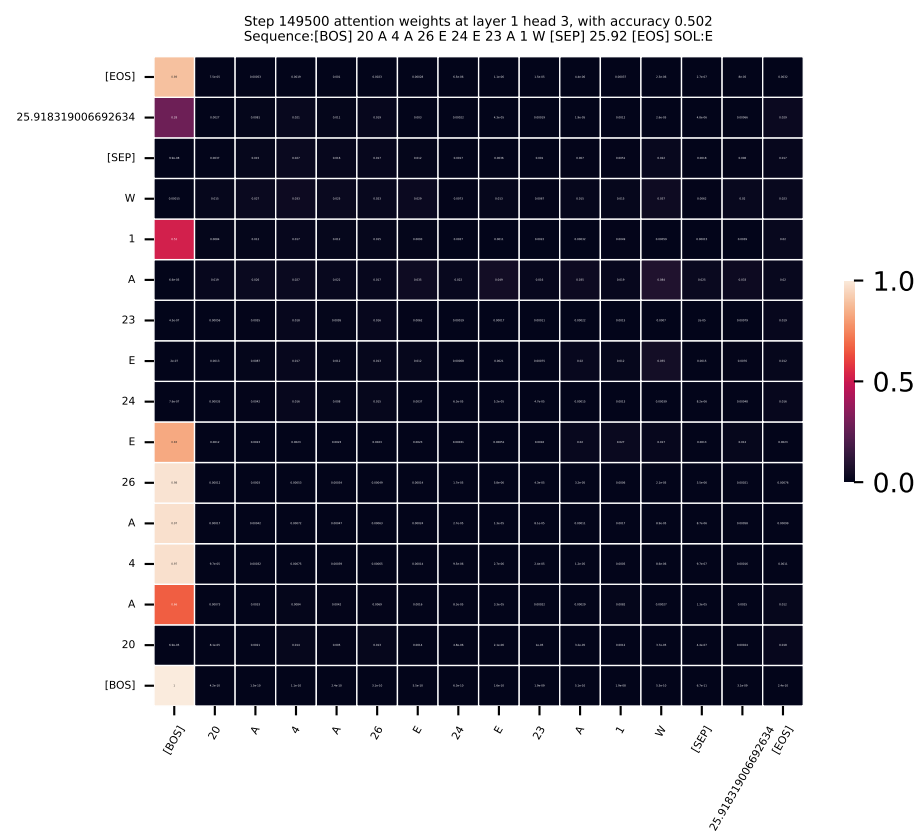

Step 149500 attention weights at layer 0 head 2, with accuracy 0.502
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

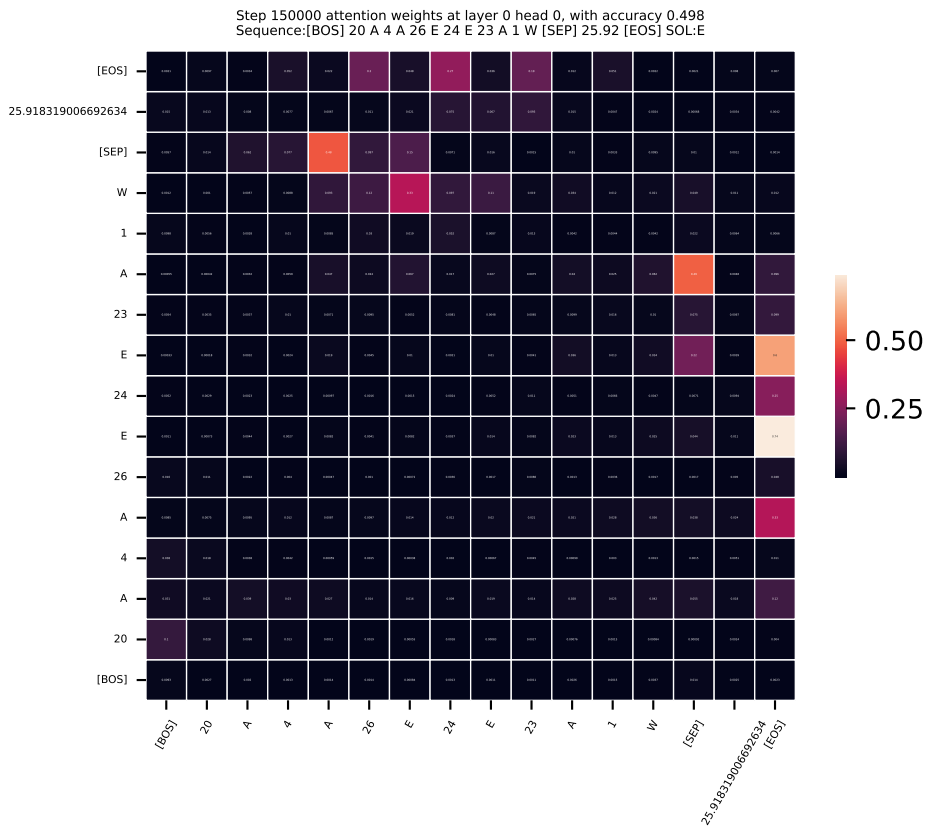

Step 149500 attention weights at layer 0 head 3, with accuracy 0.502
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

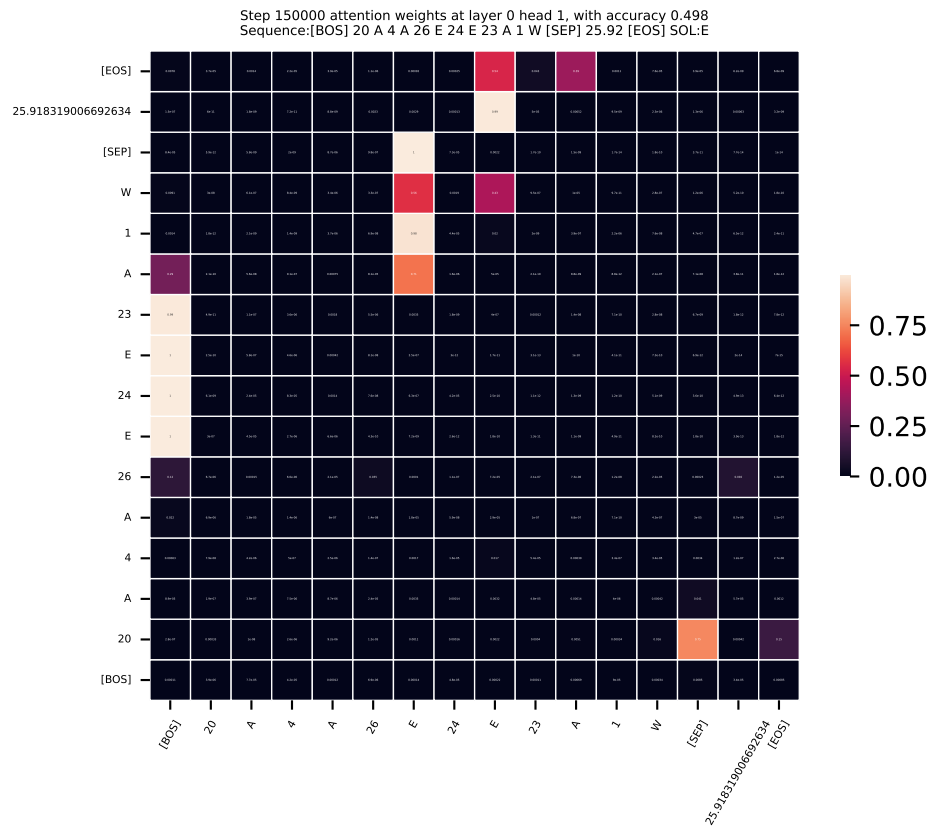

Step 149500 attention weights at layer 1 head 0, with accuracy 0.502
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.2
0.1

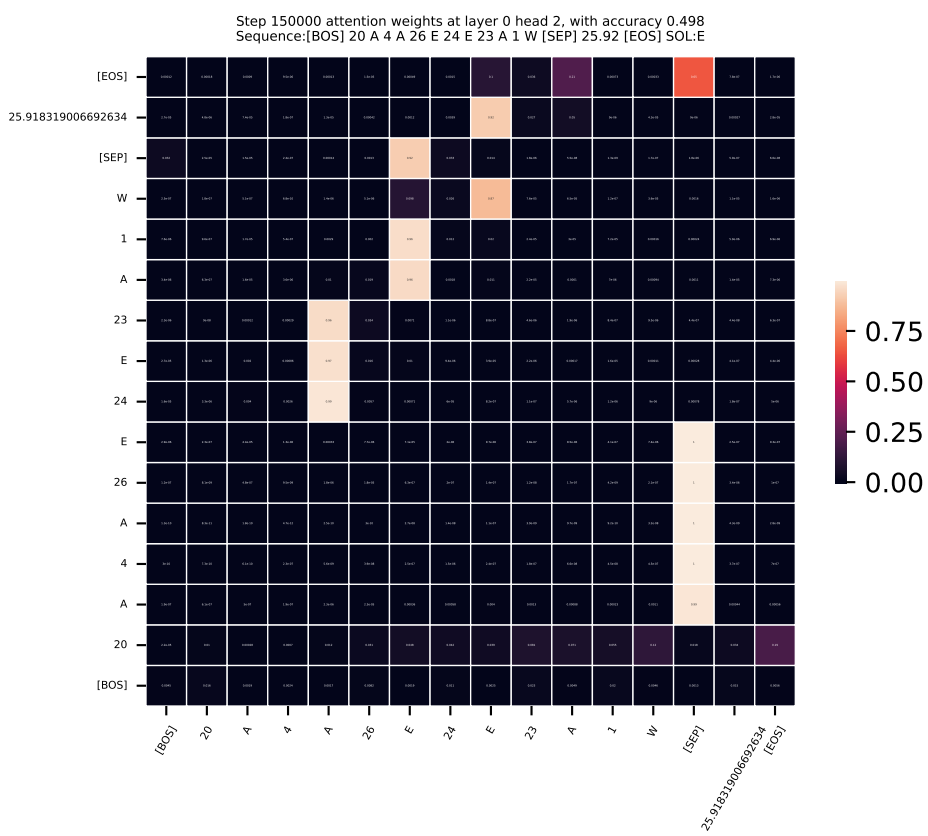

Step 149500 attention weights at layer 1 head 1, with accuracy 0.502
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.3
0.2
0.1

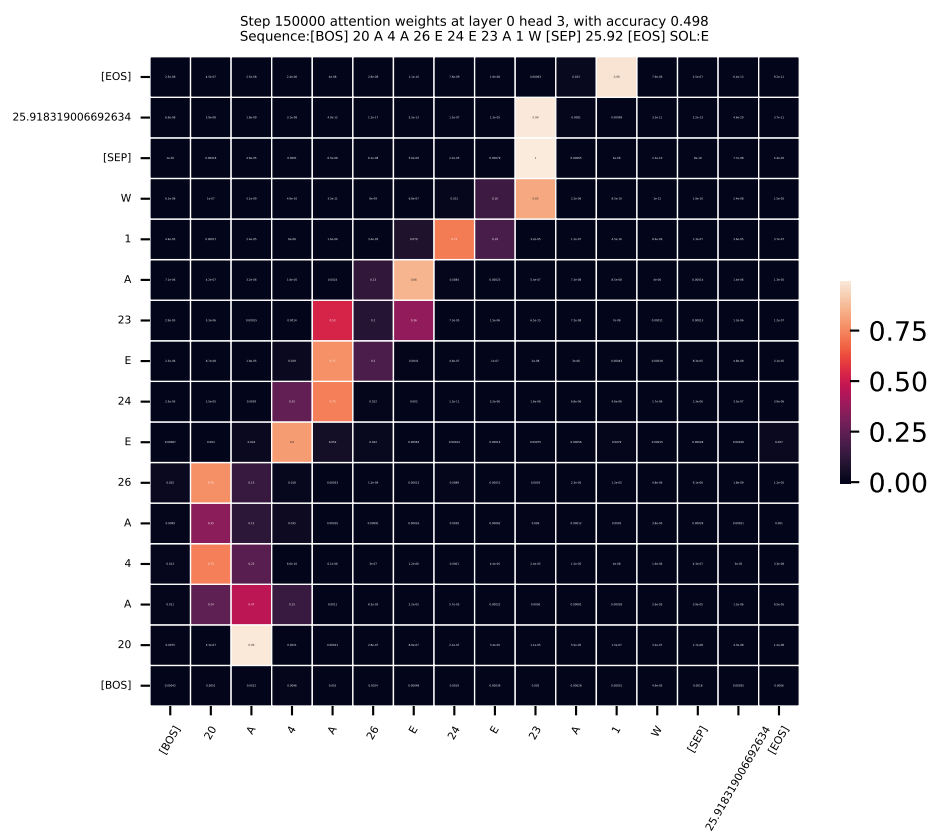

Step 149500 attention weights at layer 1 head 2, with accuracy 0.502
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.2
0.1

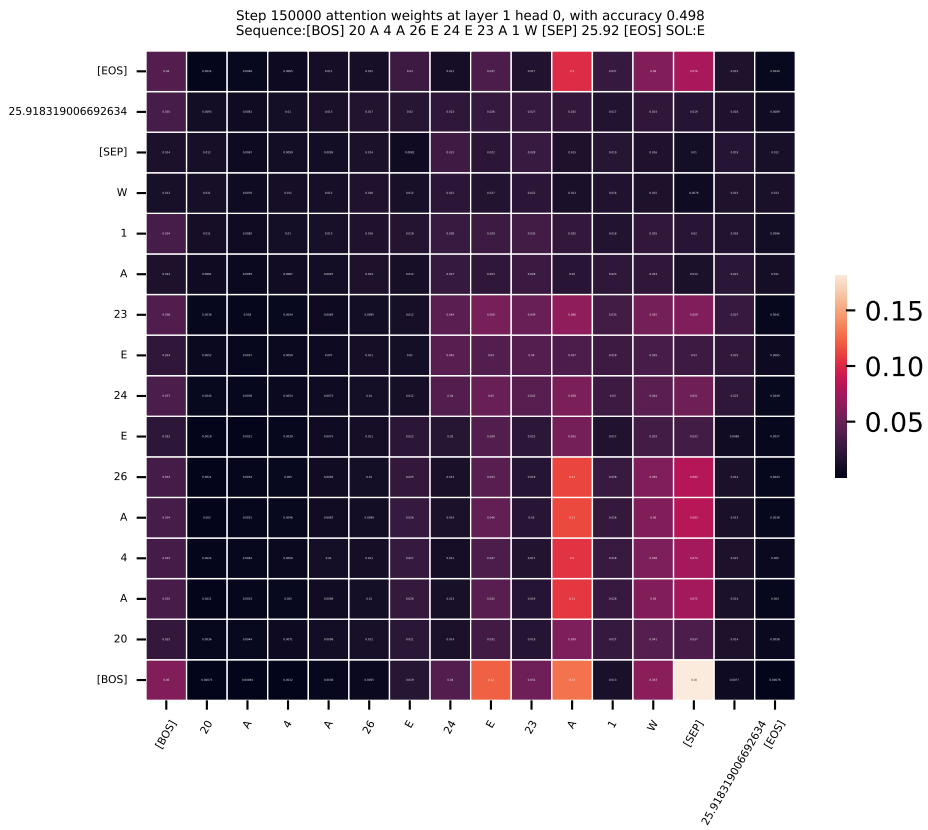

Step 149500 attention weights at layer 1 head 3, with accuracy 0.502
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
1.0
0.5
0.0

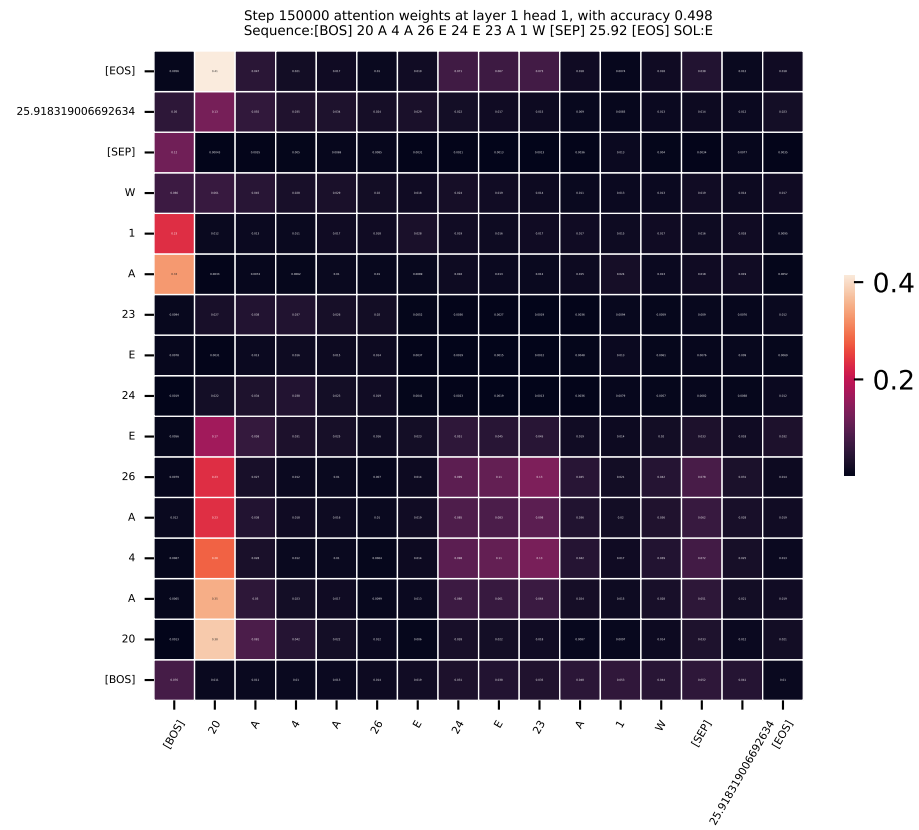

Step 150000 attention weights at layer 0 head 0, with accuracy 0.498
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

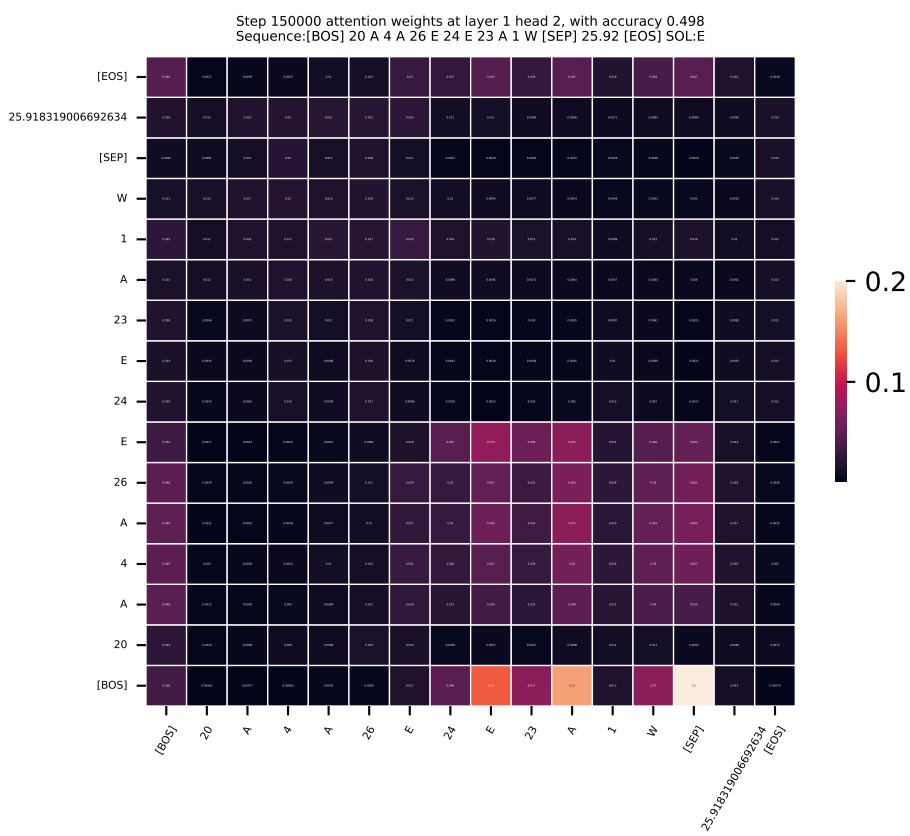

Step 150000 attention weights at layer 0 head 1, with accuracy 0.498
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

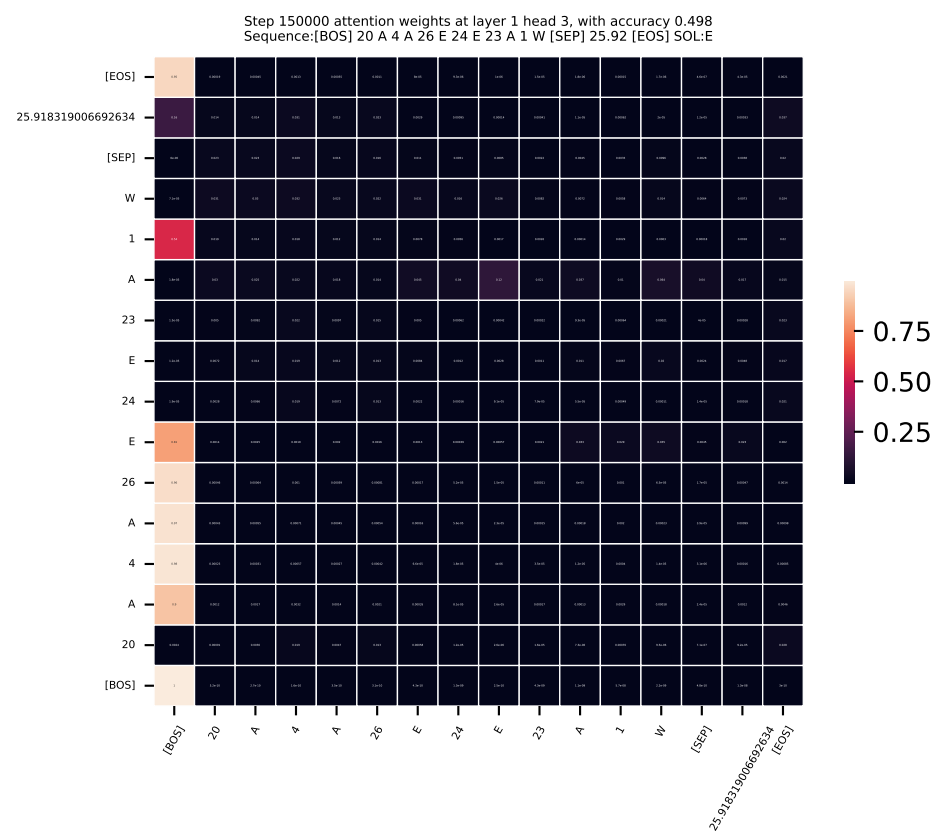

Step 150000 attention weights at layer 0 head 2, with accuracy 0.498
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

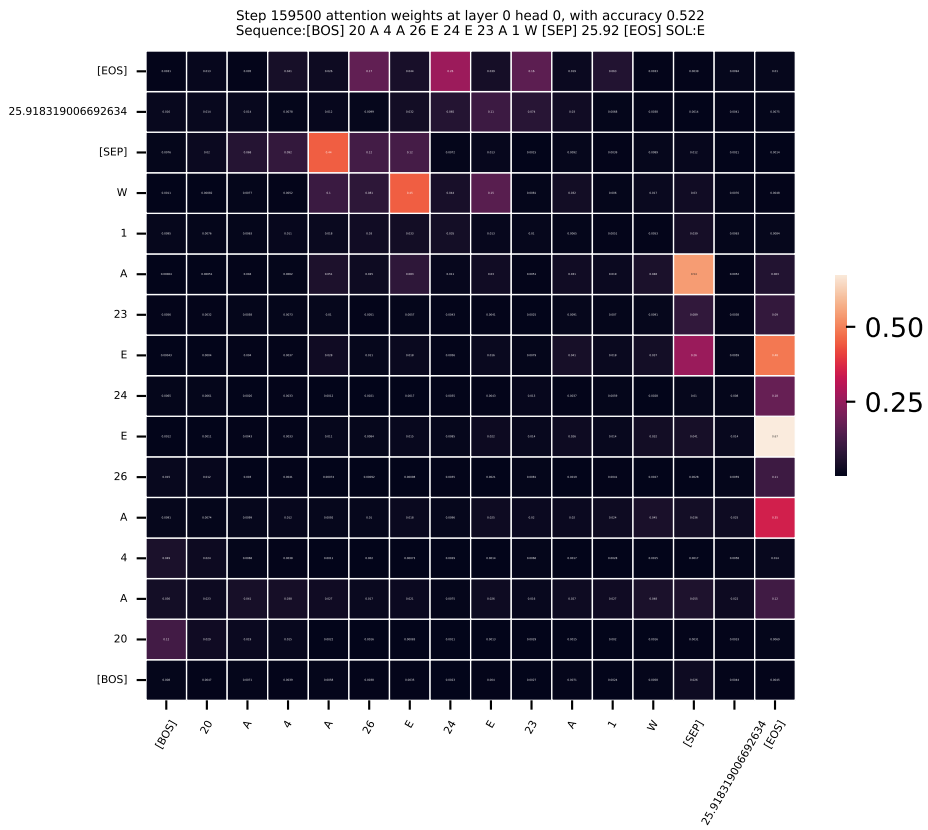

Step 150000 attention weights at layer 0 head 3, with accuracy 0.498
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

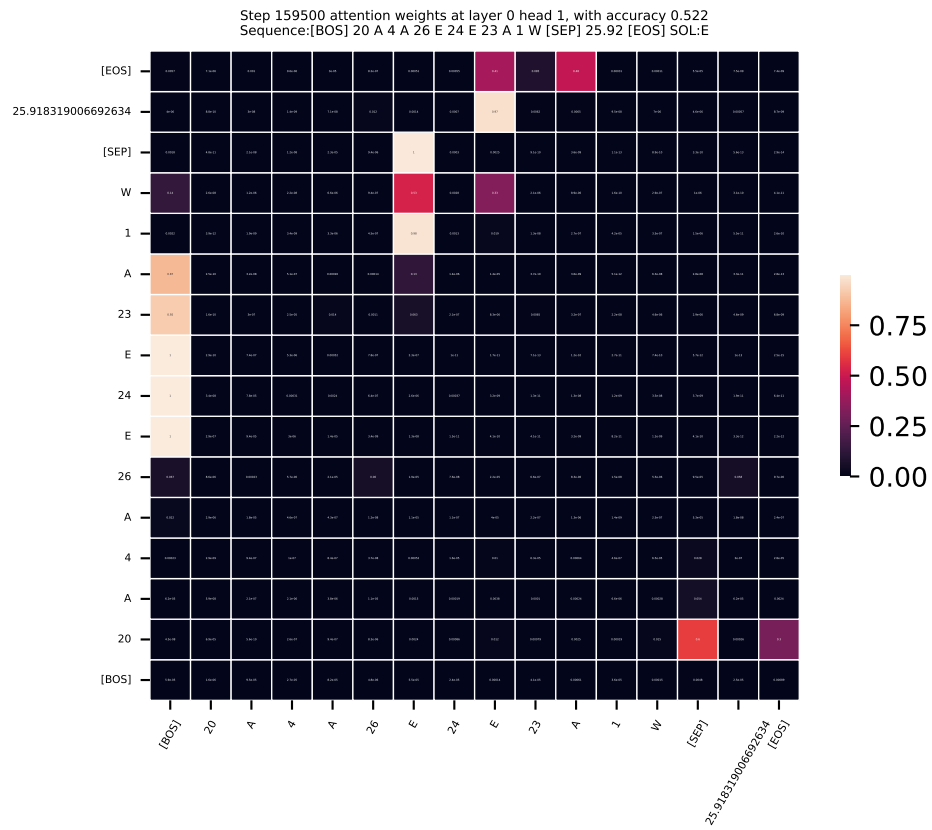

Step 150000 attention weights at layer 1 head 0, with accuracy 0.498
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.15
0.10
0.05

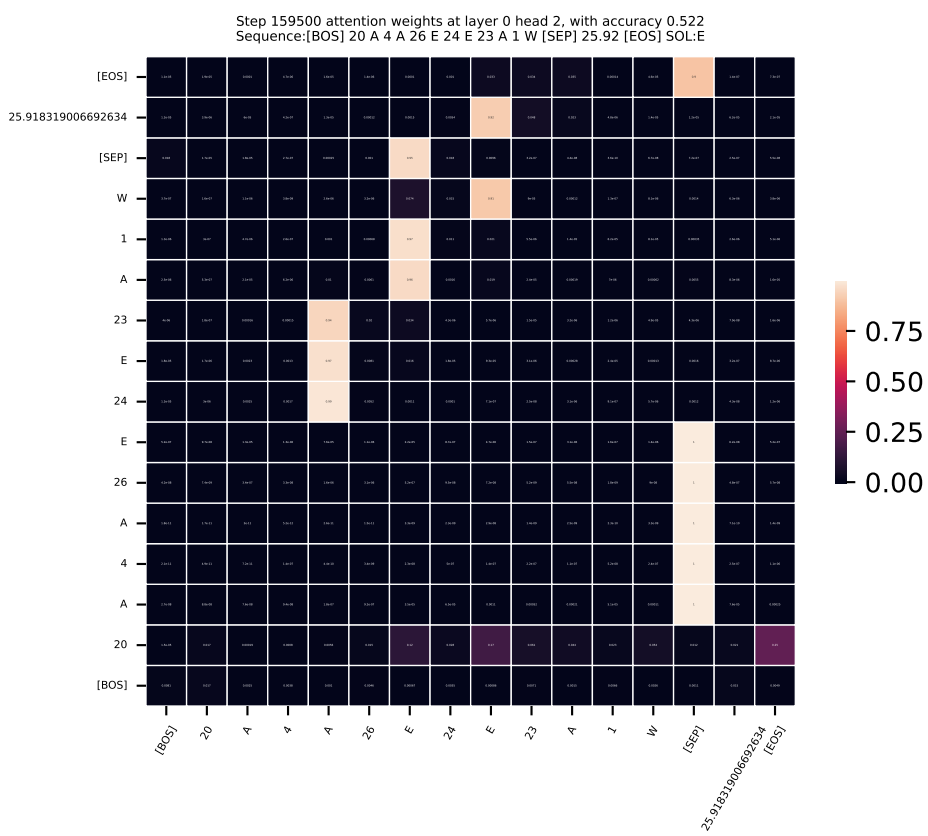

Step 150000 attention weights at layer 1 head 1, with accuracy 0.498
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

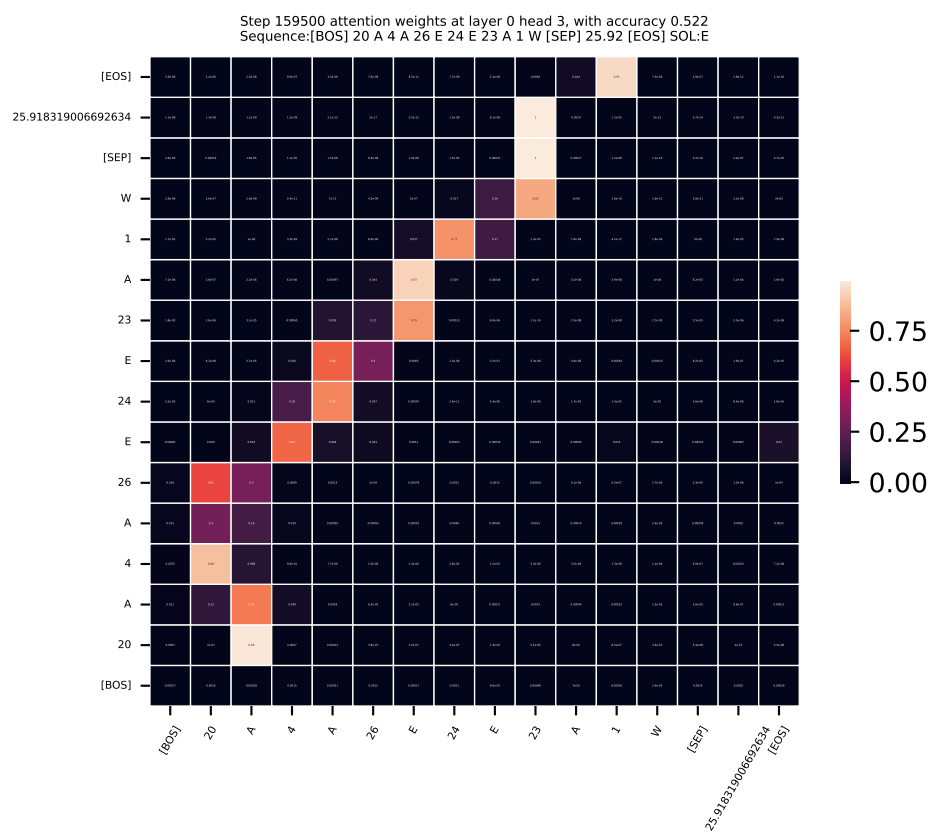

Step 150000 attention weights at layer 1 head 2, with accuracy 0.498
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.2
0.1

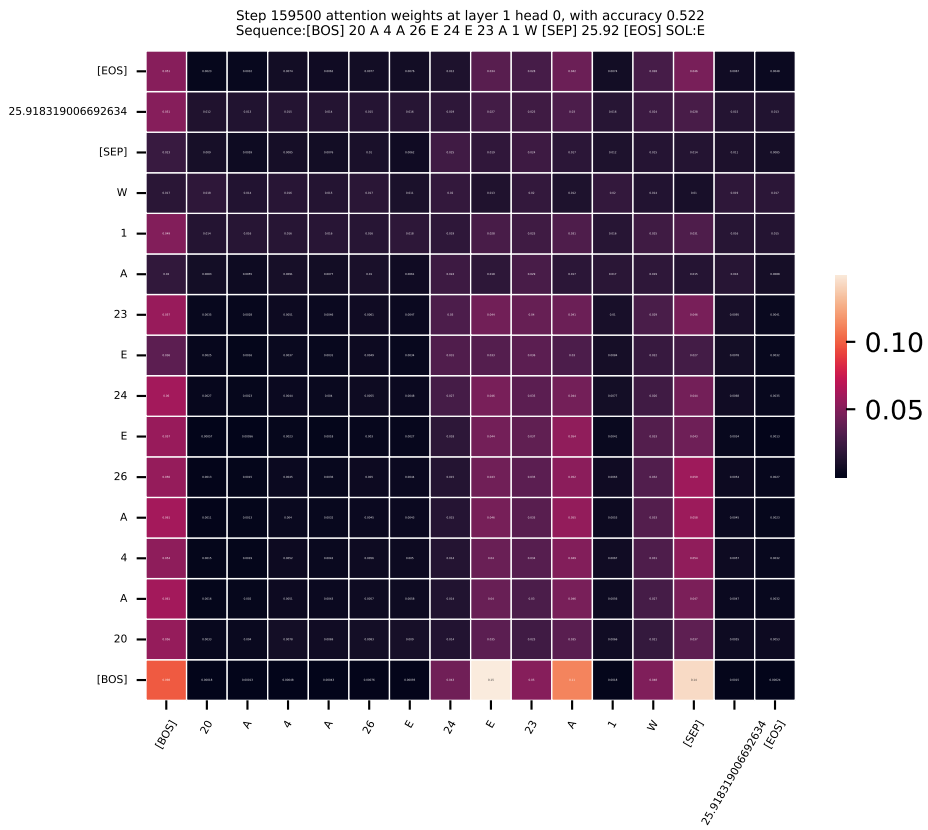

Step 150000 attention weights at layer 1 head 3, with accuracy 0.498
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

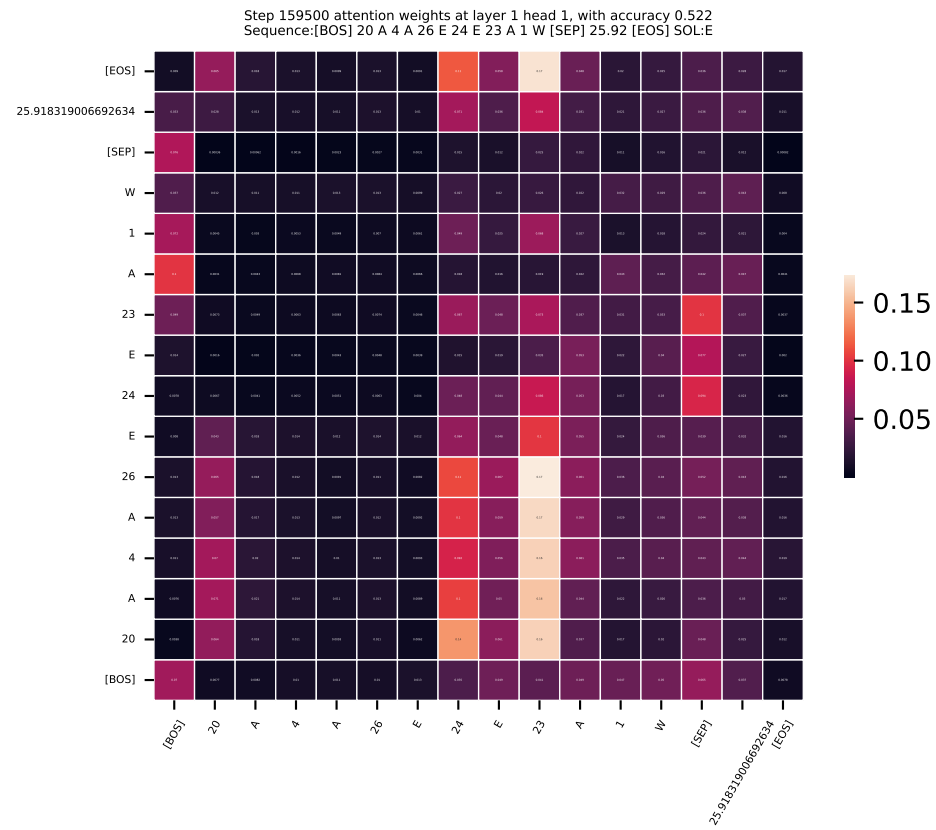

Step 159500 attention weights at layer 0 head 0, with accuracy 0.522
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.50
0.25

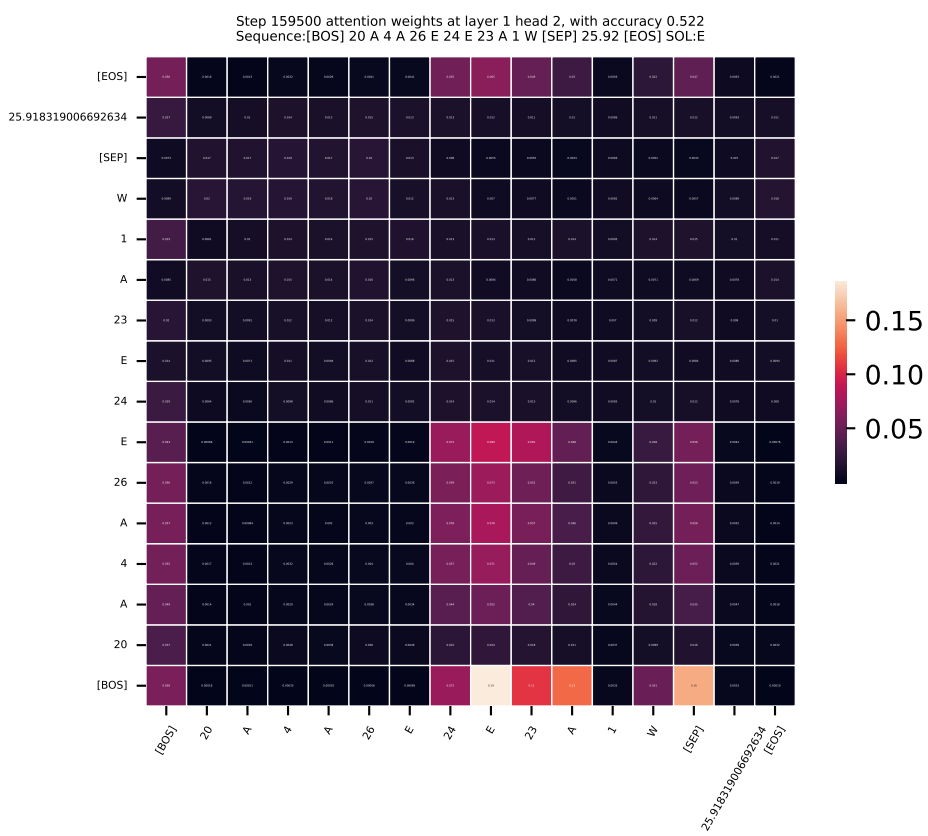

Step 159500 attention weights at layer 0 head 1, with accuracy 0.522
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25
0.00

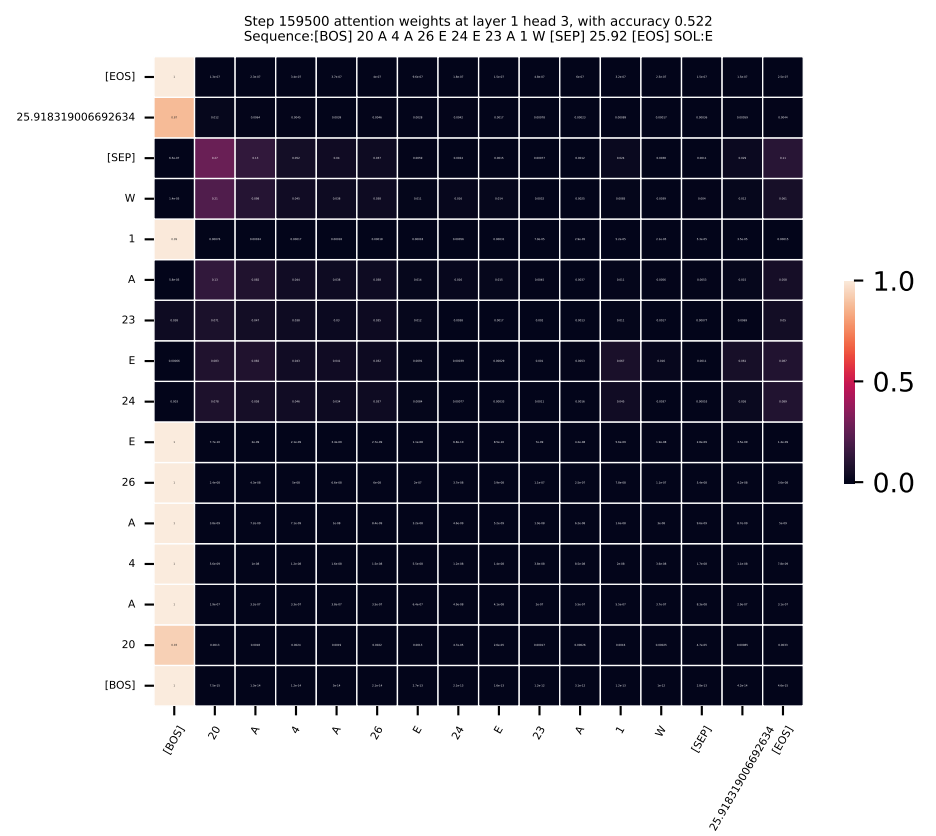

Step 159500 attention weights at layer 0 head 2, with accuracy 0.522
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25
0.00

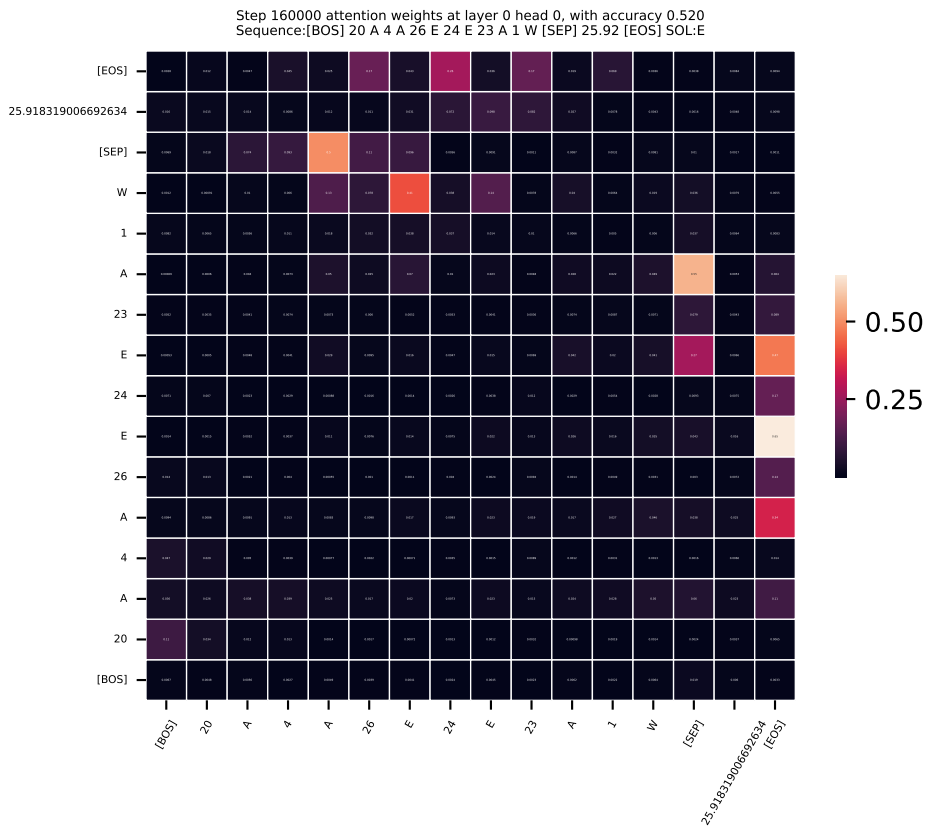

Step 159500 attention weights at layer 0 head 3, with accuracy 0.522
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25
0.00

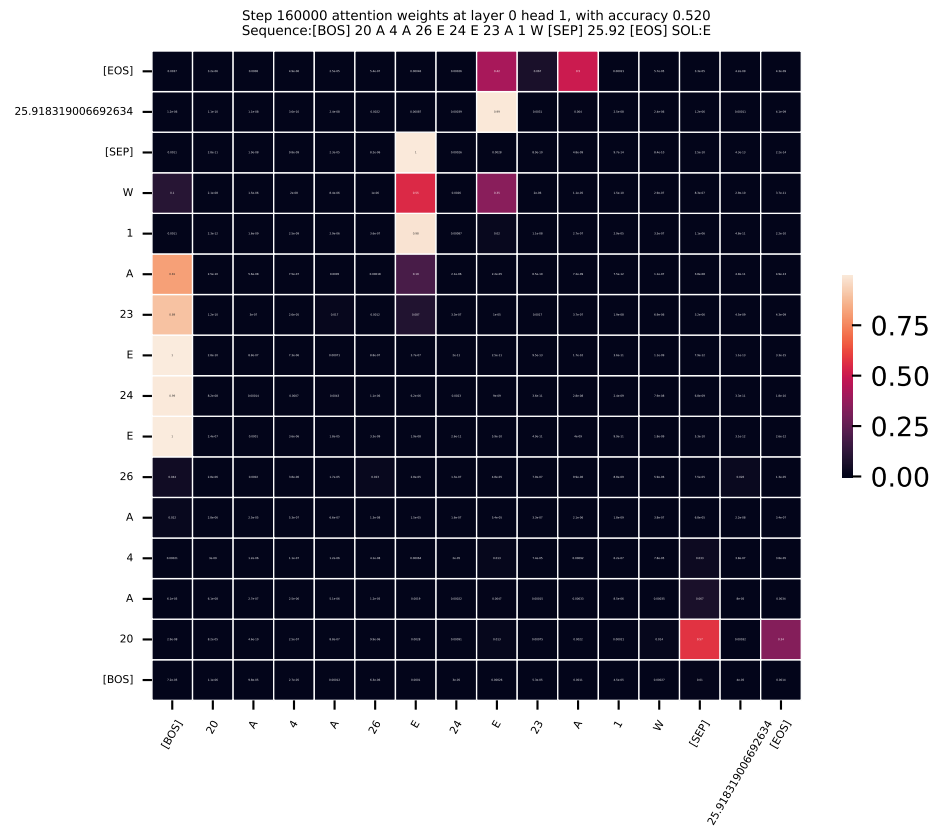

Step 159500 attention weights at layer 1 head 0, with accuracy 0.522
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.10
0.05

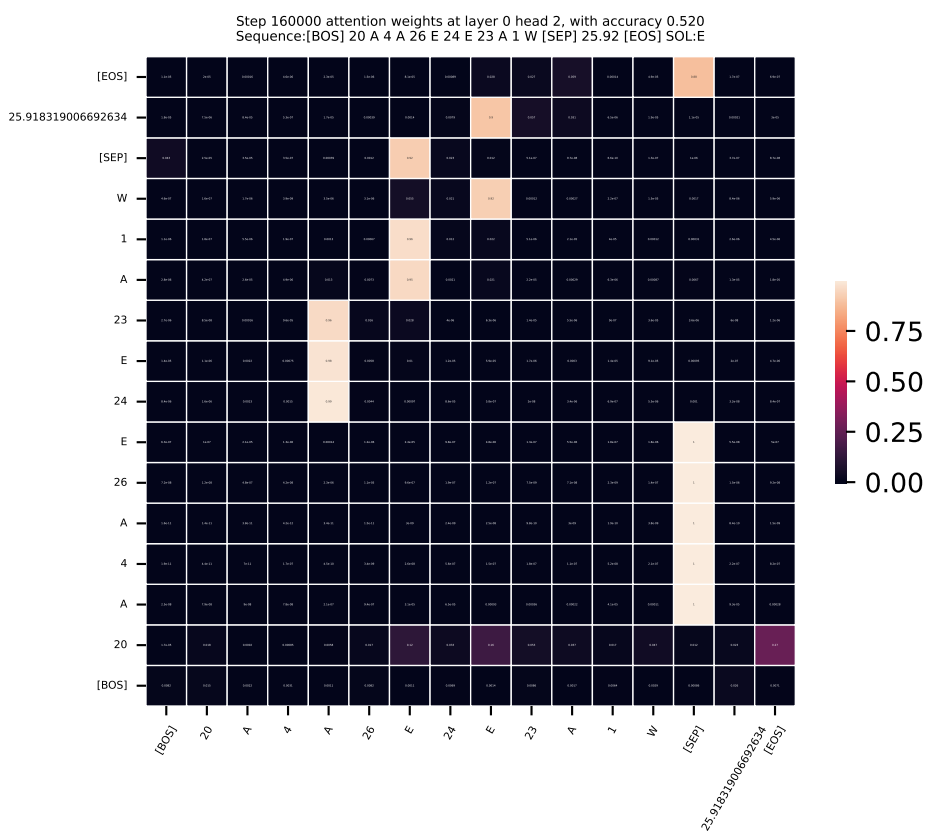

Step 159500 attention weights at layer 1 head 1, with accuracy 0.522
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.15
0.10
0.05

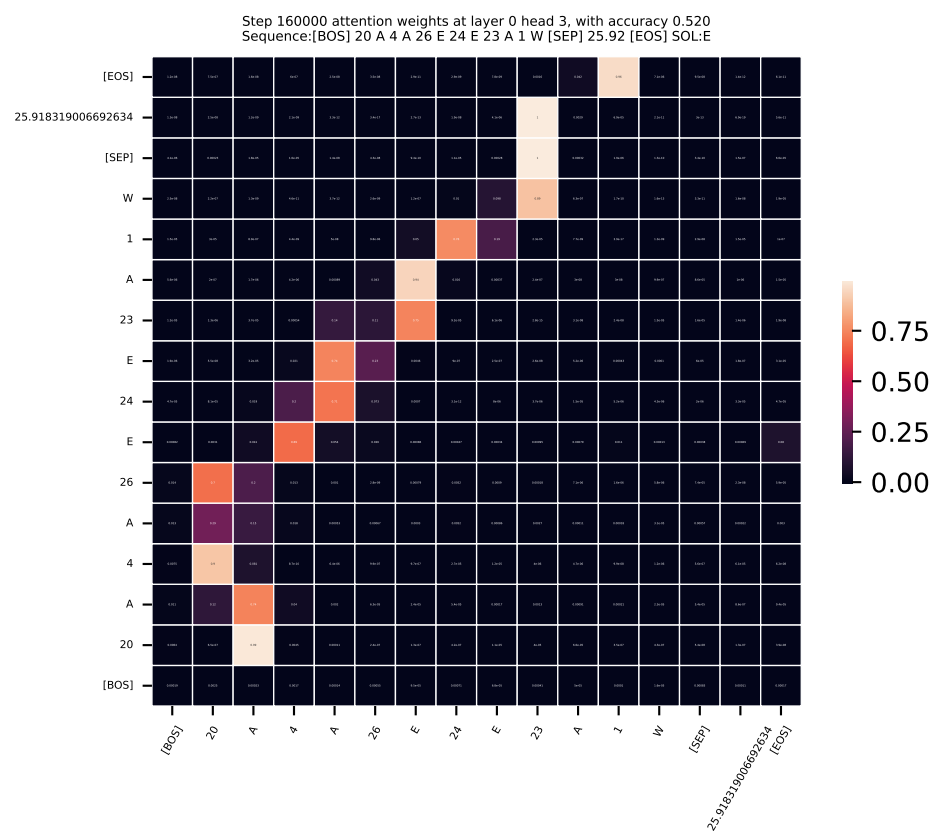

Step 159500 attention weights at layer 1 head 2, with accuracy 0.522
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.15
0.10
0.05

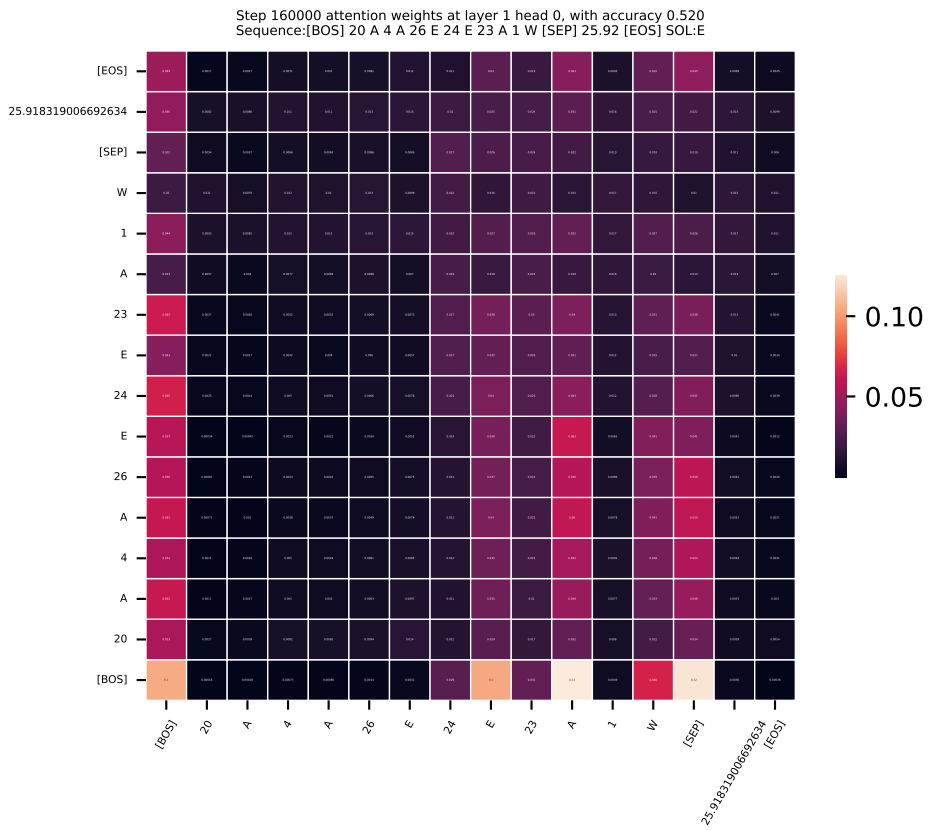

Step 159500 attention weights at layer 1 head 3, with accuracy 0.522
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
1.0
0.5
0.0

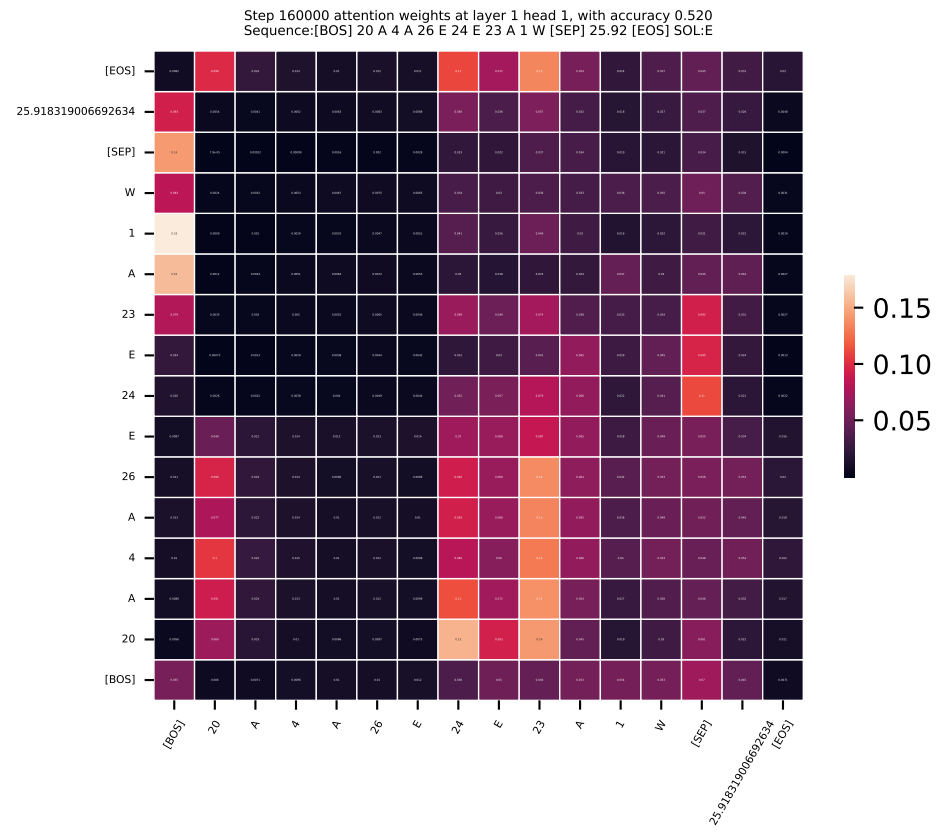

Step 160000 attention weights at layer 0 head 0, with accuracy 0.520
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

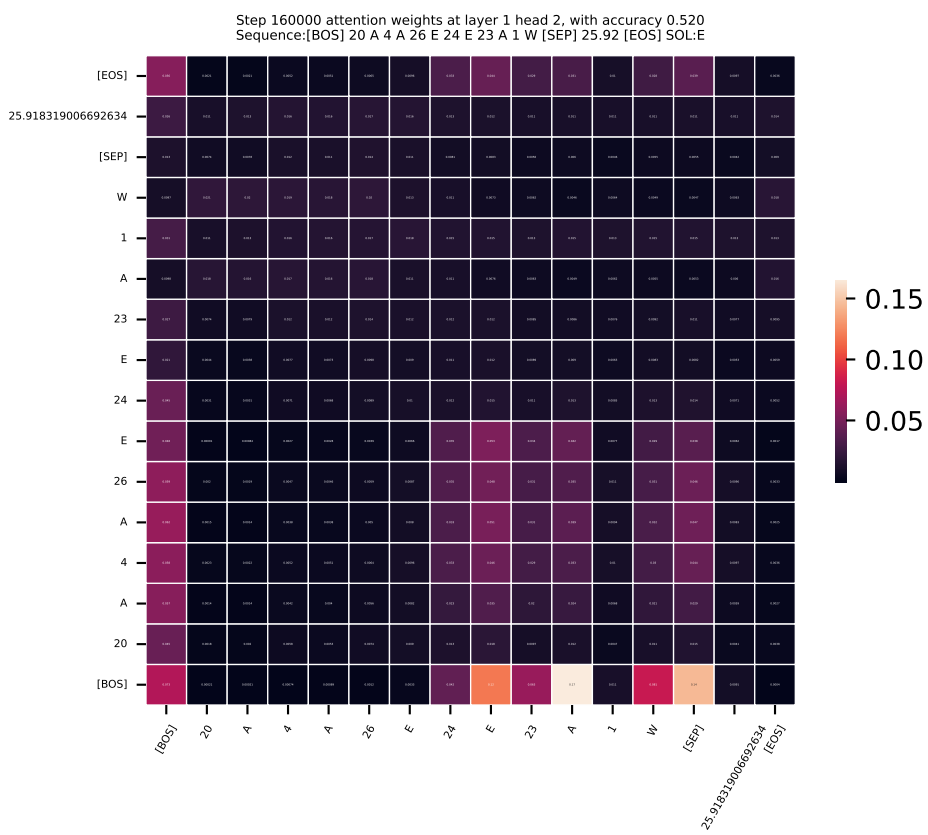

Step 160000 attention weights at layer 0 head 1, with accuracy 0.520
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

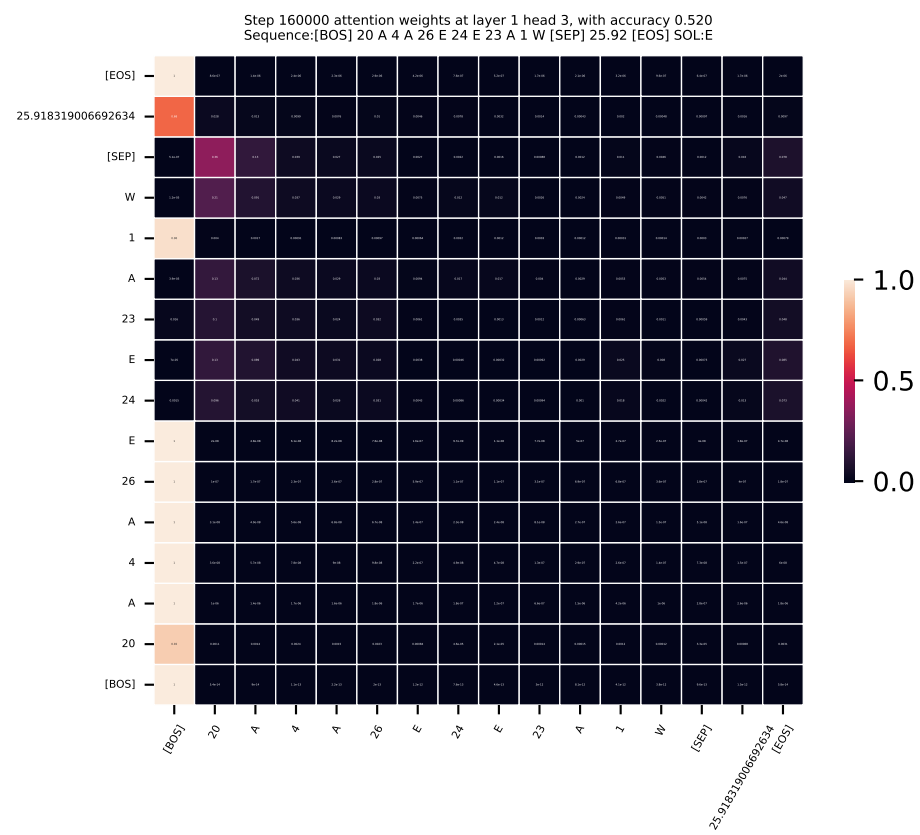

Step 160000 attention weights at layer 0 head 2, with accuracy 0.520
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

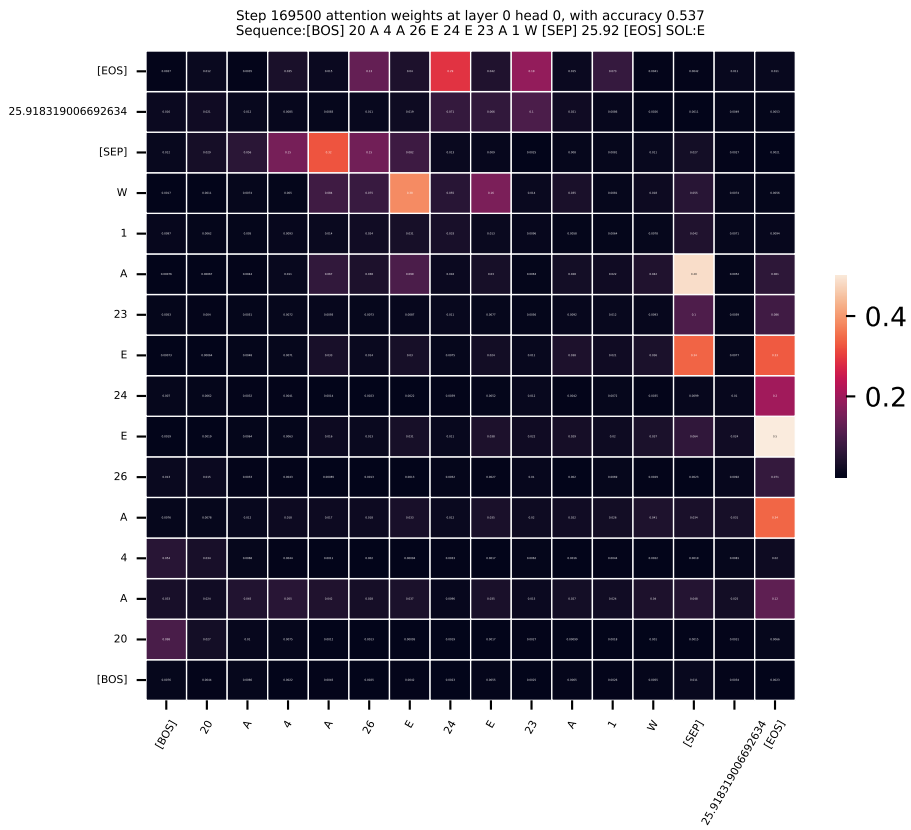

Step 160000 attention weights at layer 0 head 3, with accuracy 0.520
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

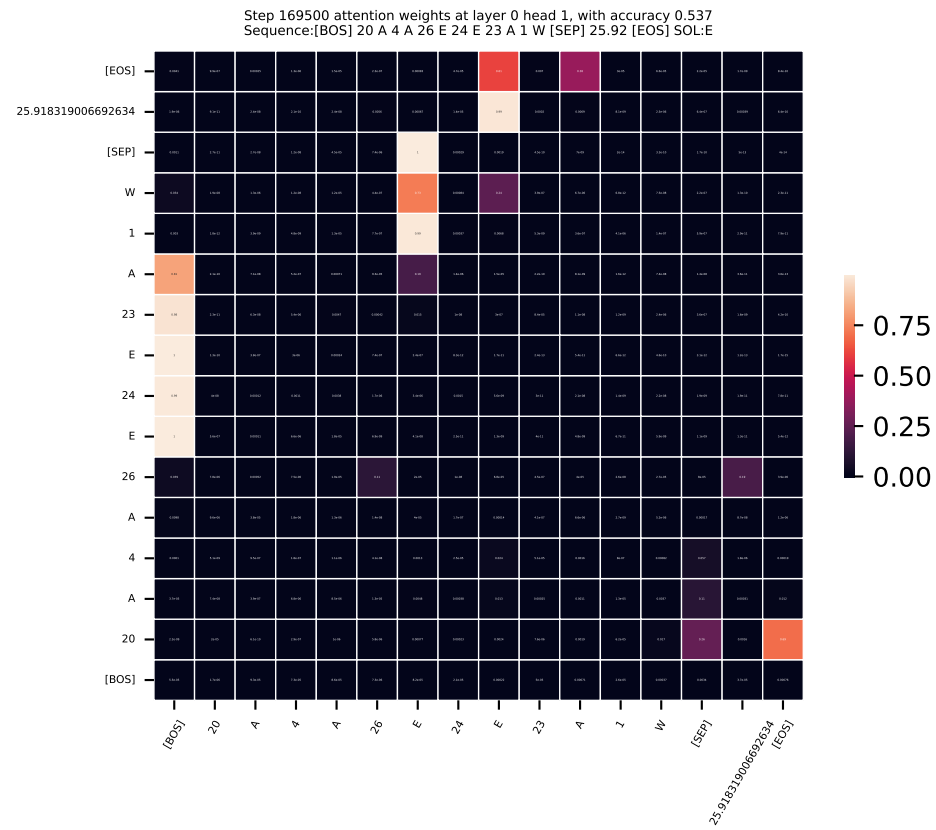

Step 160000 attention weights at layer 1 head 0, with accuracy 0.520
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.10
0.05

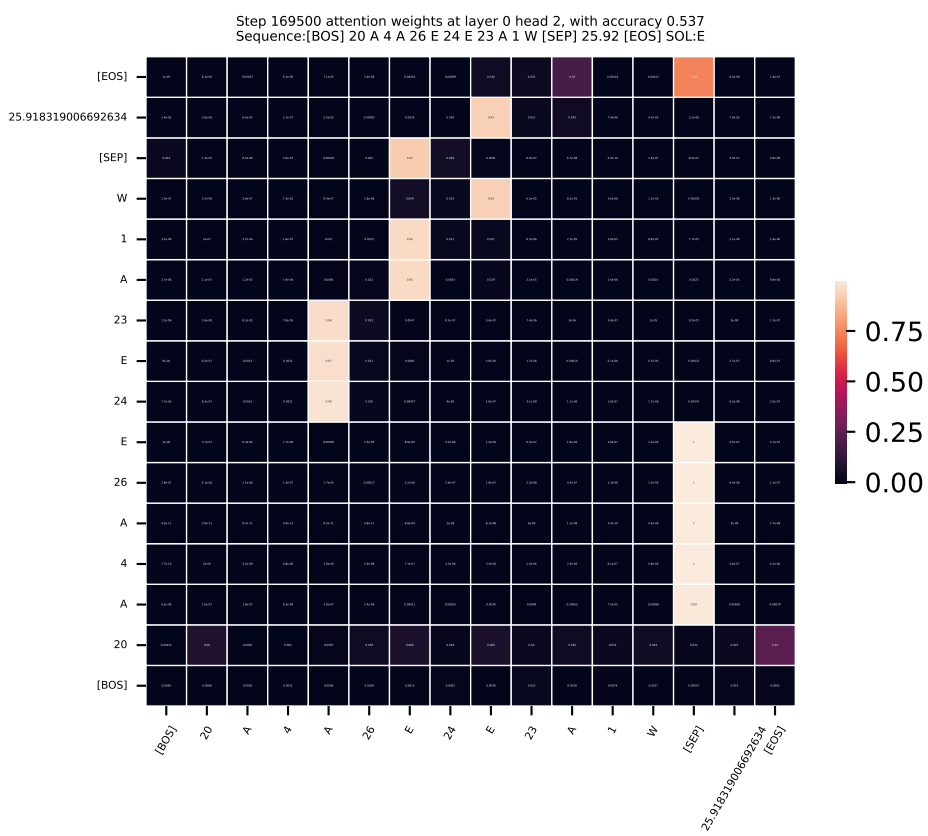

Step 160000 attention weights at layer 1 head 1, with accuracy 0.520
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.15
0.10
0.05

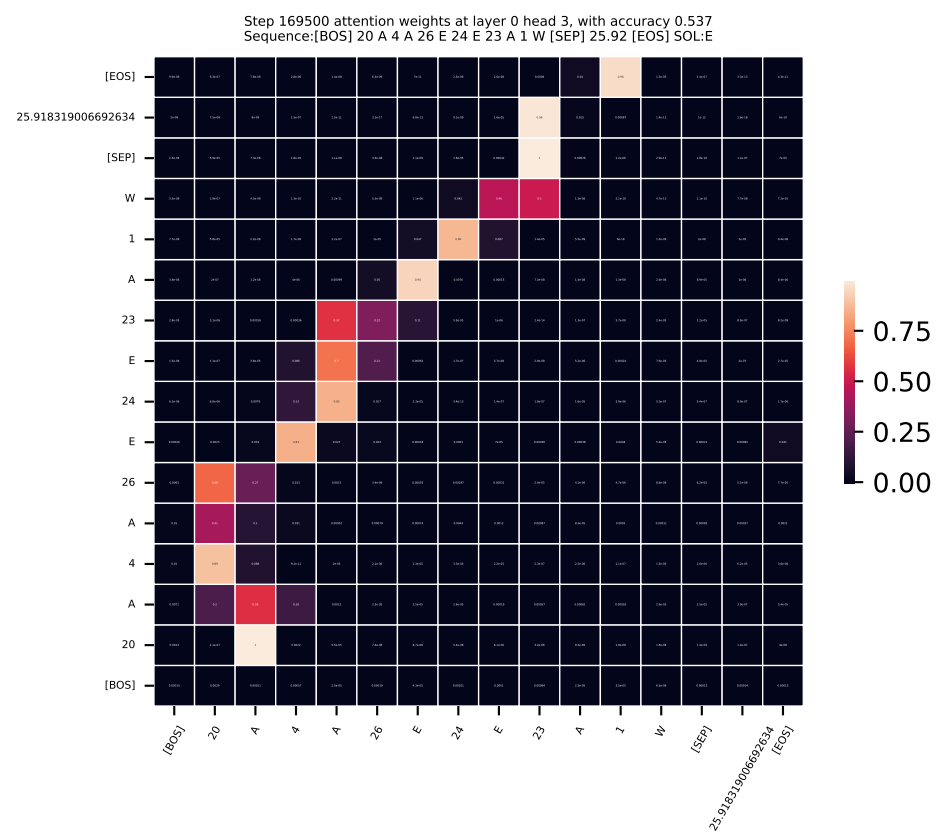

Step 160000 attention weights at layer 1 head 2, with accuracy 0.520
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.15
0.10
0.05

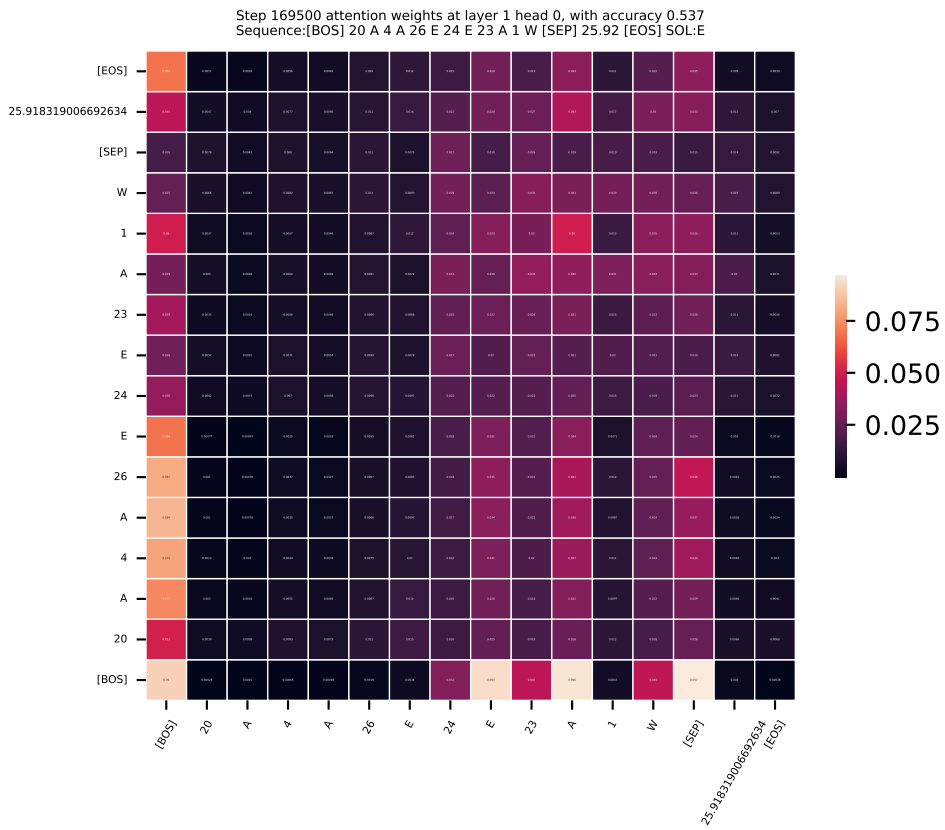

Step 160000 attention weights at layer 1 head 3, with accuracy 0.520
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
1.0
0.5
0.0

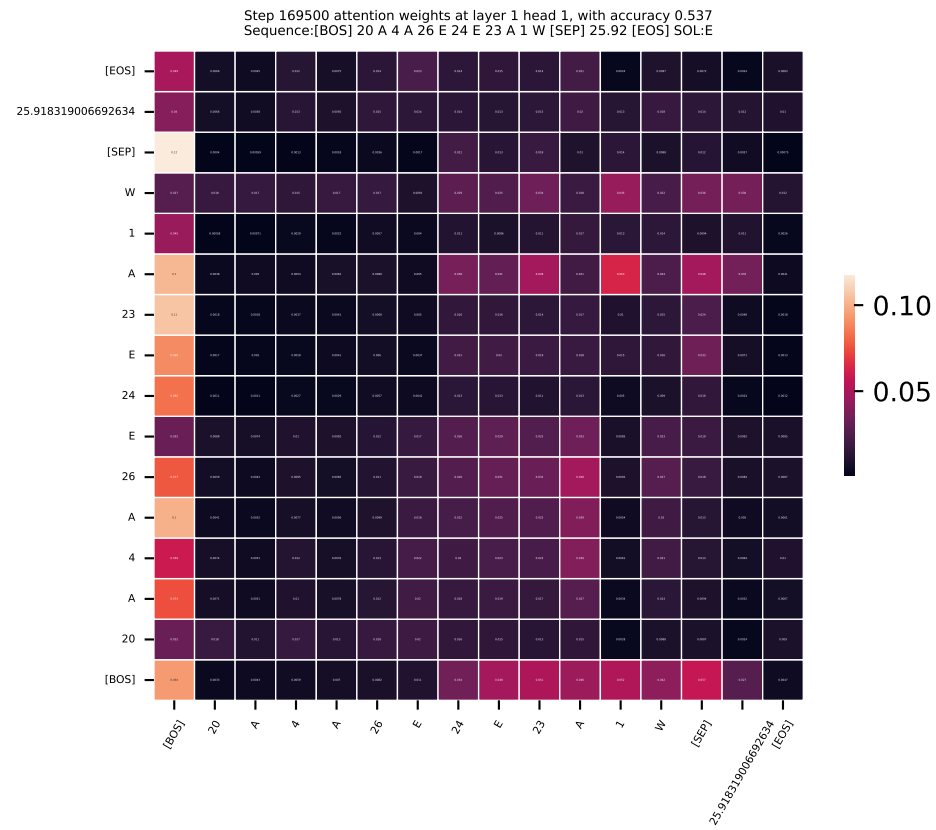

Step 169500 attention weights at layer 0 head 0, with accuracy 0.537
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.4
0.2

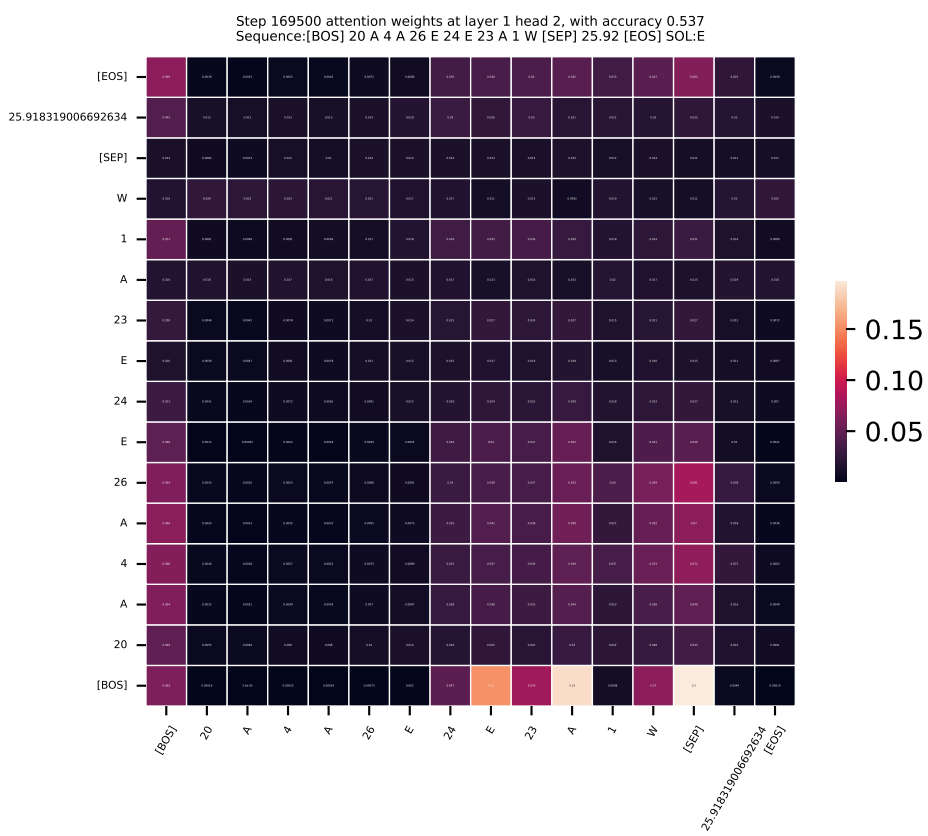

Step 169500 attention weights at layer 0 head 1, with accuracy 0.537
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

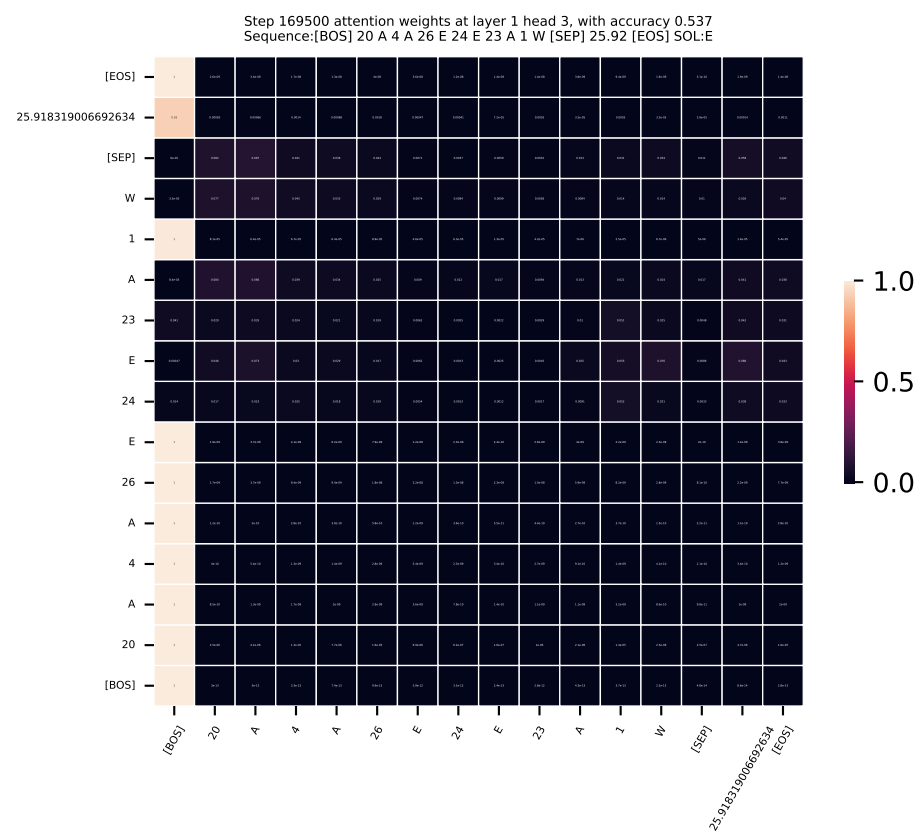

Step 169500 attention weights at layer 0 head 2, with accuracy 0.537
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

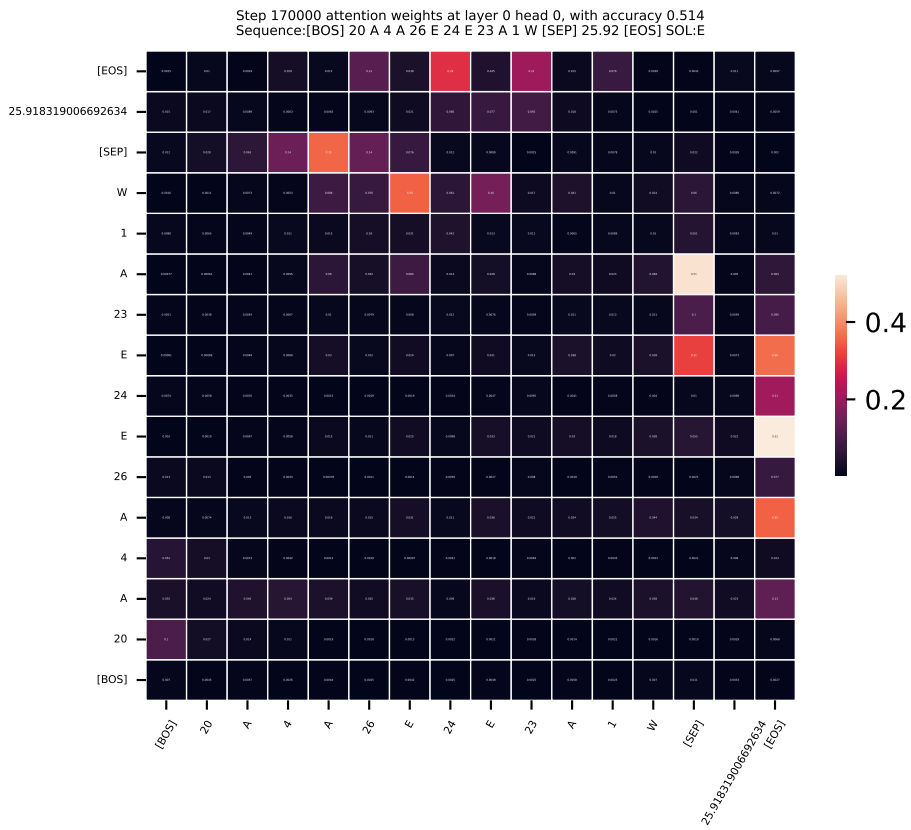

Step 169500 attention weights at layer 0 head 3, with accuracy 0.537
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

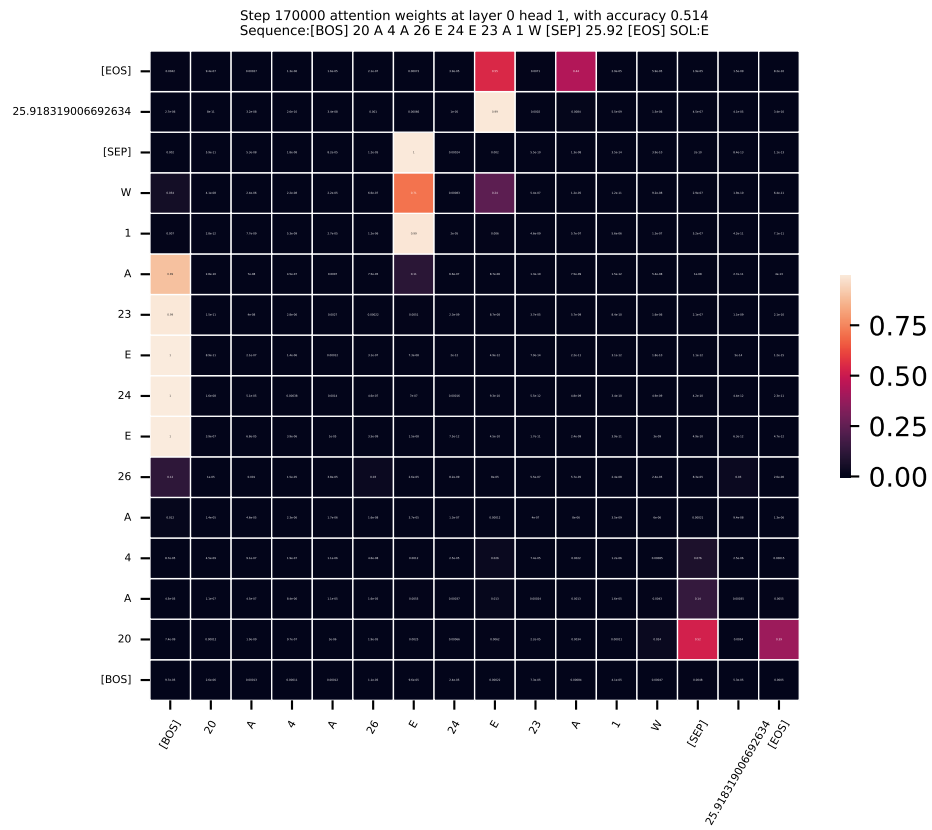

Step 169500 attention weights at layer 1 head 0, with accuracy 0.537
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.075
0.050
0.025

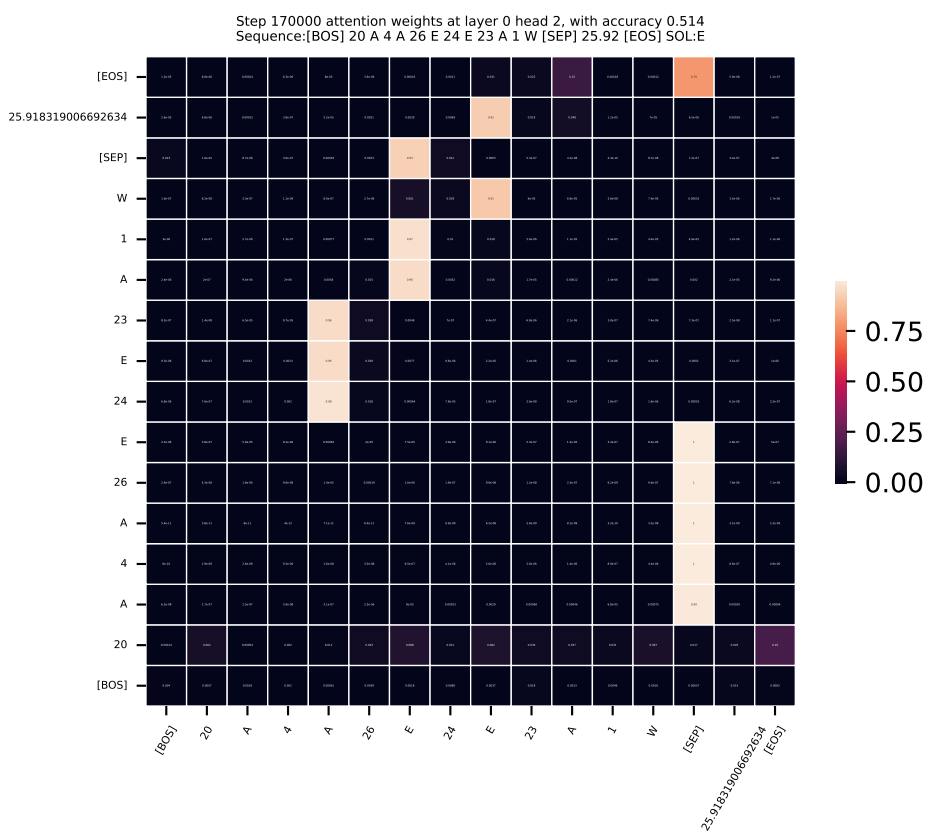

Step 169500 attention weights at layer 1 head 1, with accuracy 0.537
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.10
0.05

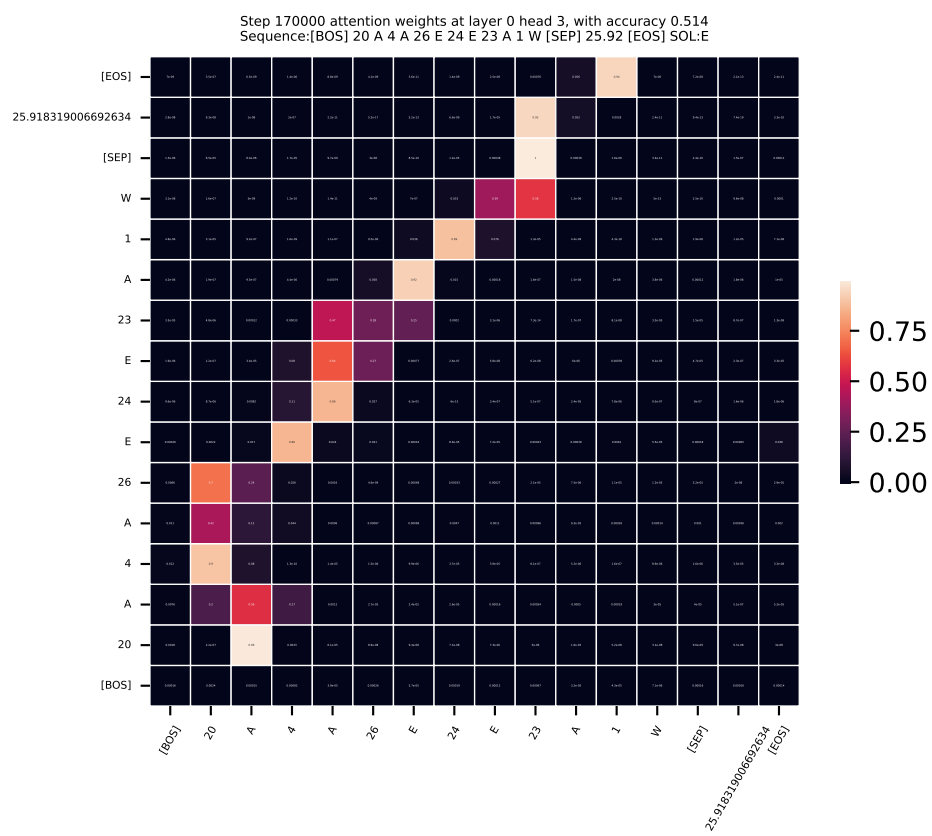

Step 169500 attention weights at layer 1 head 2, with accuracy 0.537
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.15
0.10
0.05

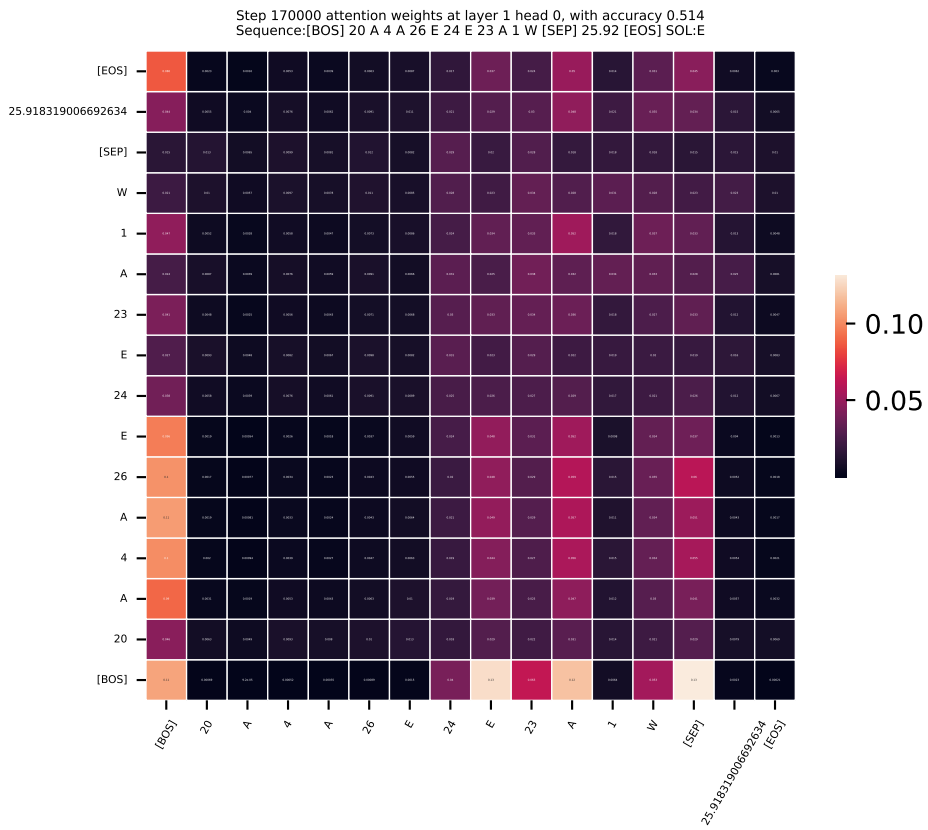

Step 169500 attention weights at layer 1 head 3, with accuracy 0.537
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
1.0
0.5
0.0

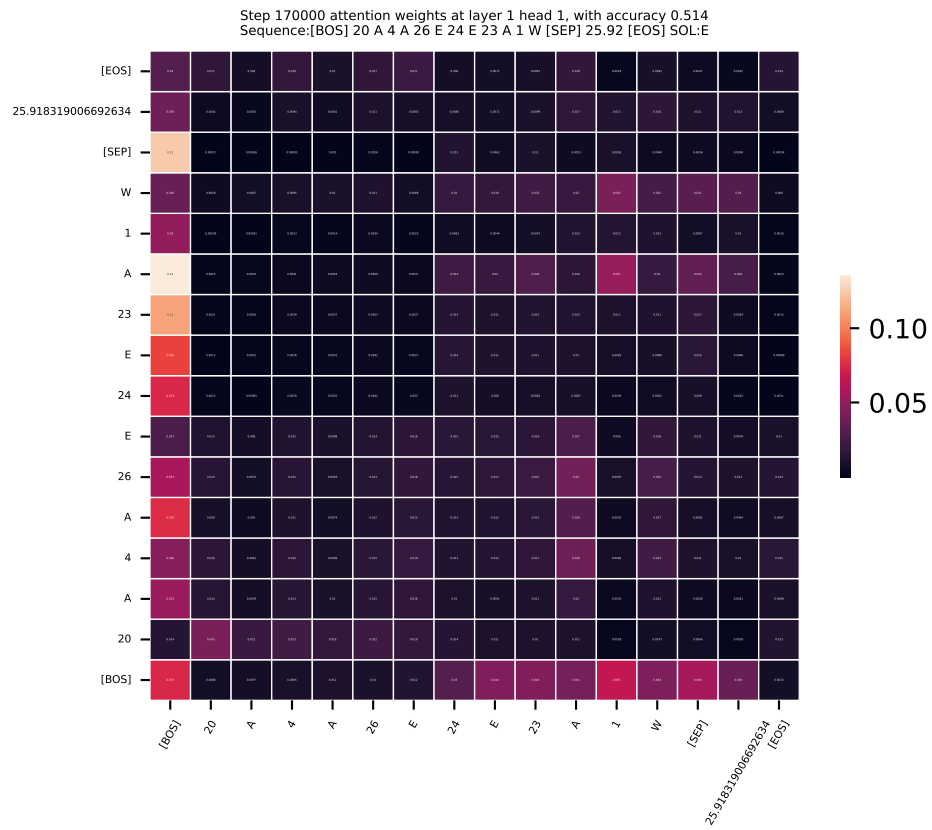

Step 170000 attention weights at layer 0 head 0, with accuracy 0.514
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

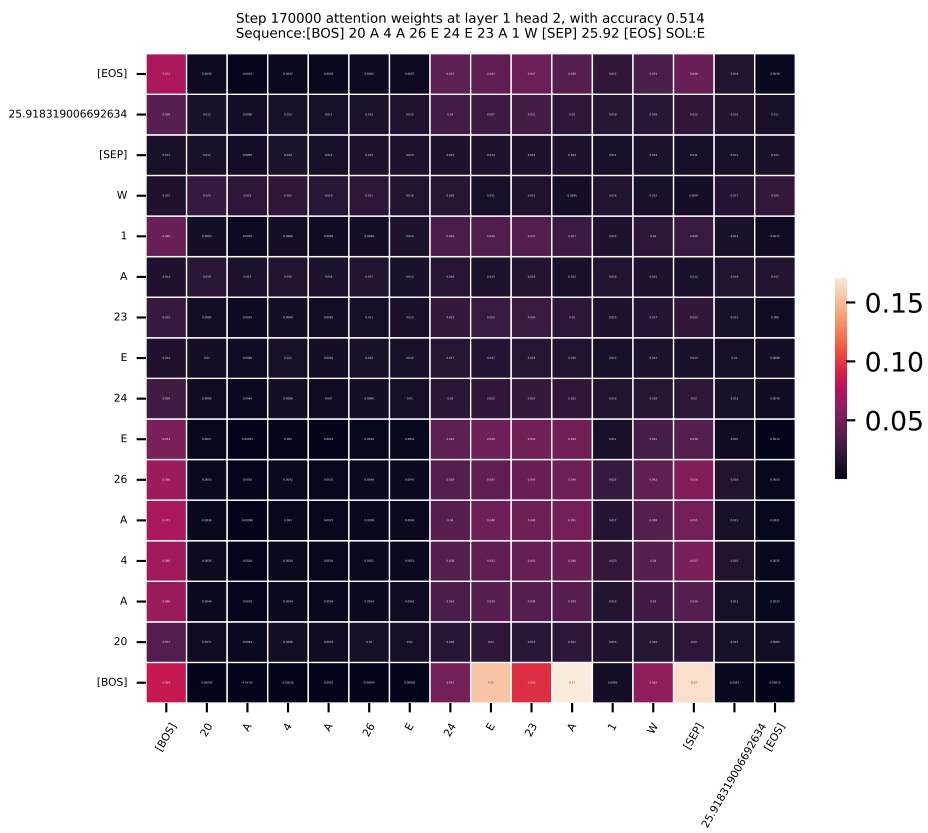

Step 170000 attention weights at layer 0 head 1, with accuracy 0.514
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

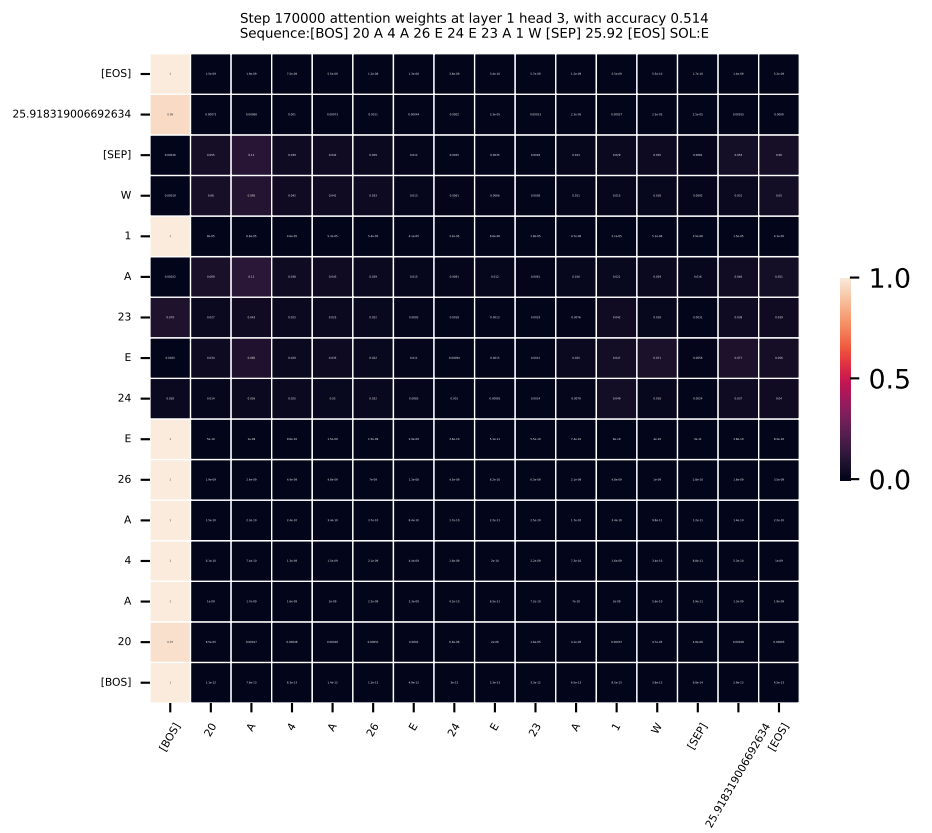

Step 170000 attention weights at layer 0 head 2, with accuracy 0.514
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

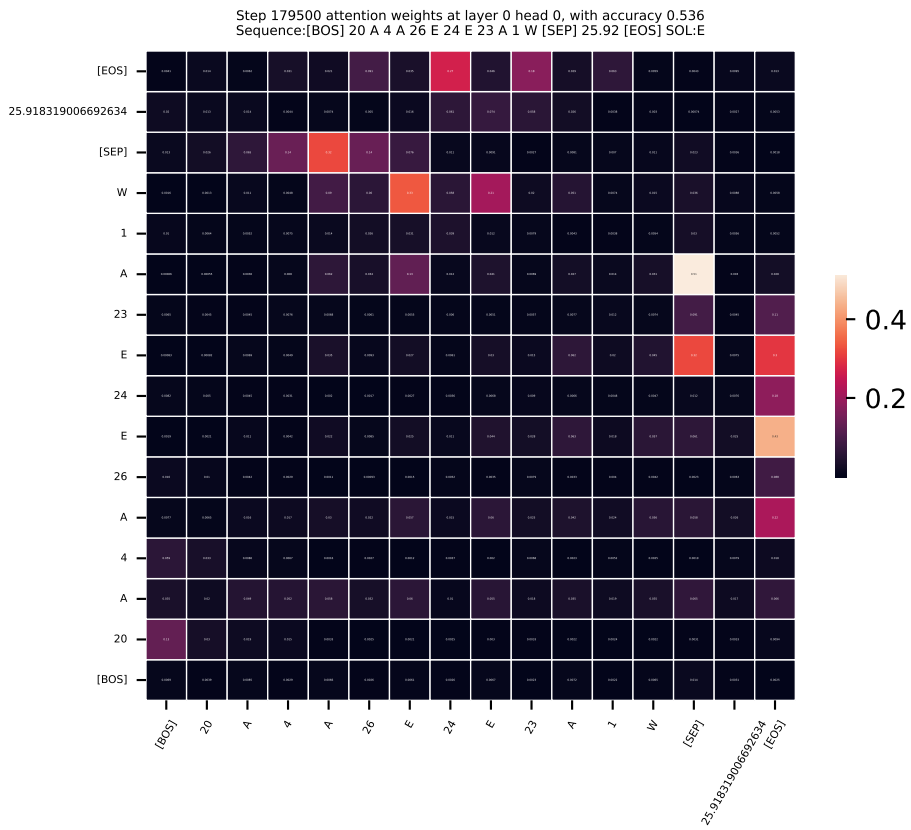

Step 170000 attention weights at layer 0 head 3, with accuracy 0.514
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

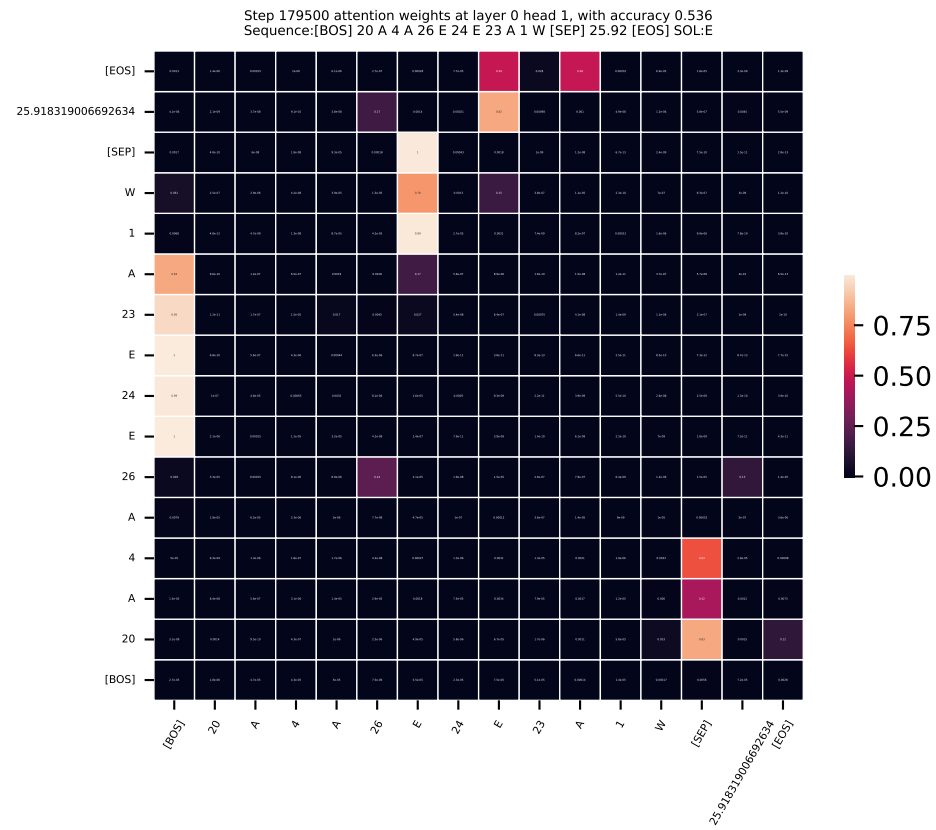

Step 170000 attention weights at layer 1 head 0, with accuracy 0.514
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.10
0.05

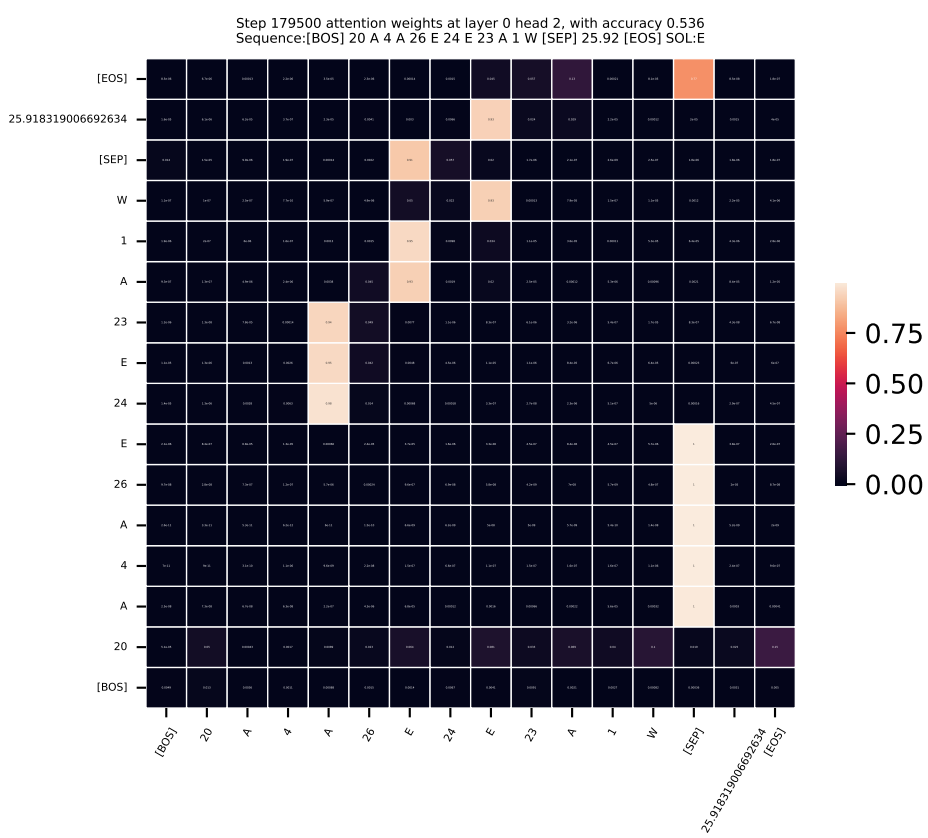

Step 170000 attention weights at layer 1 head 1, with accuracy 0.514
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.10
0.05

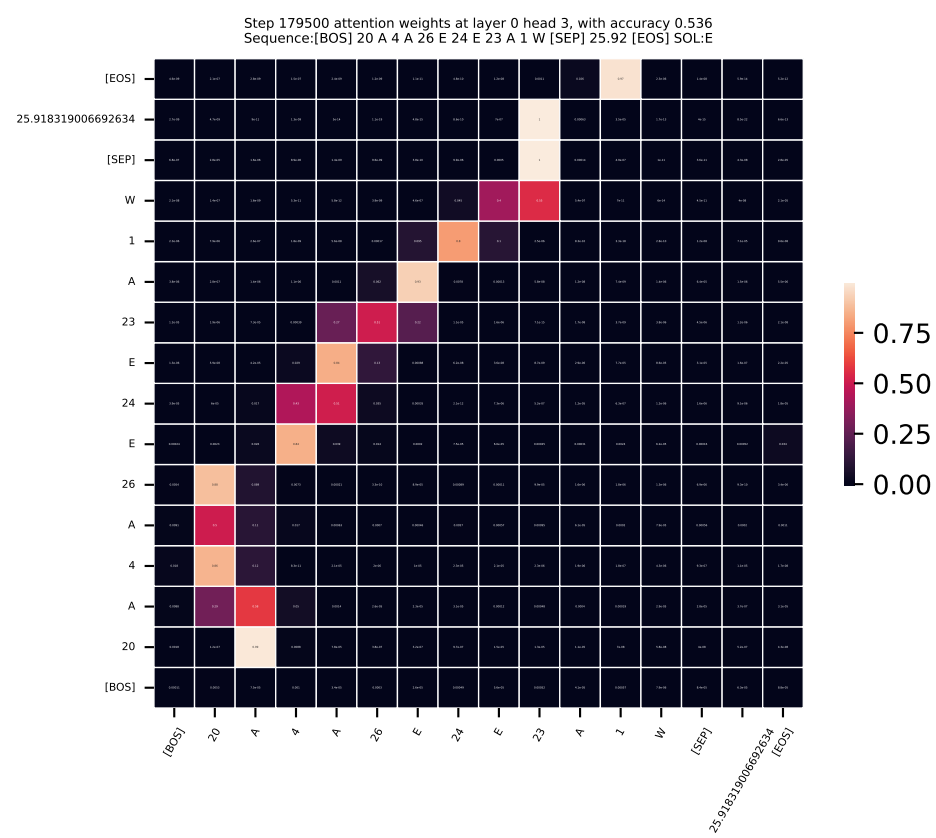

Step 170000 attention weights at layer 1 head 2, with accuracy 0.514
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.15
0.10
0.05

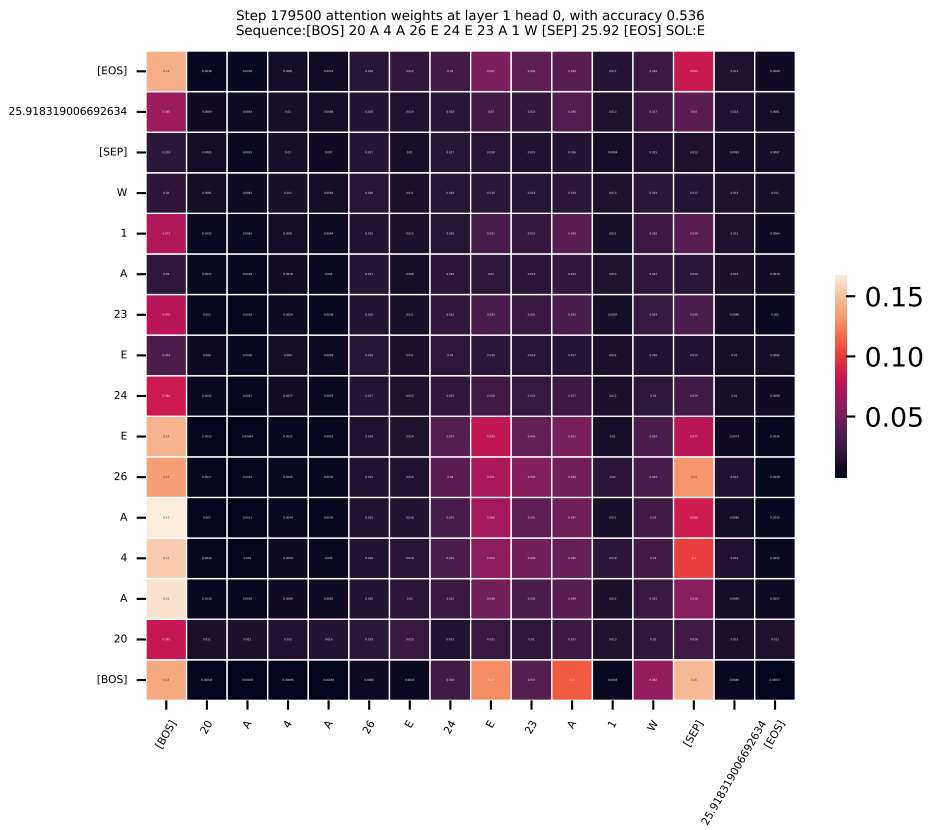

Step 170000 attention weights at layer 1 head 3, with accuracy 0.514
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
1.0
0.5
0.0

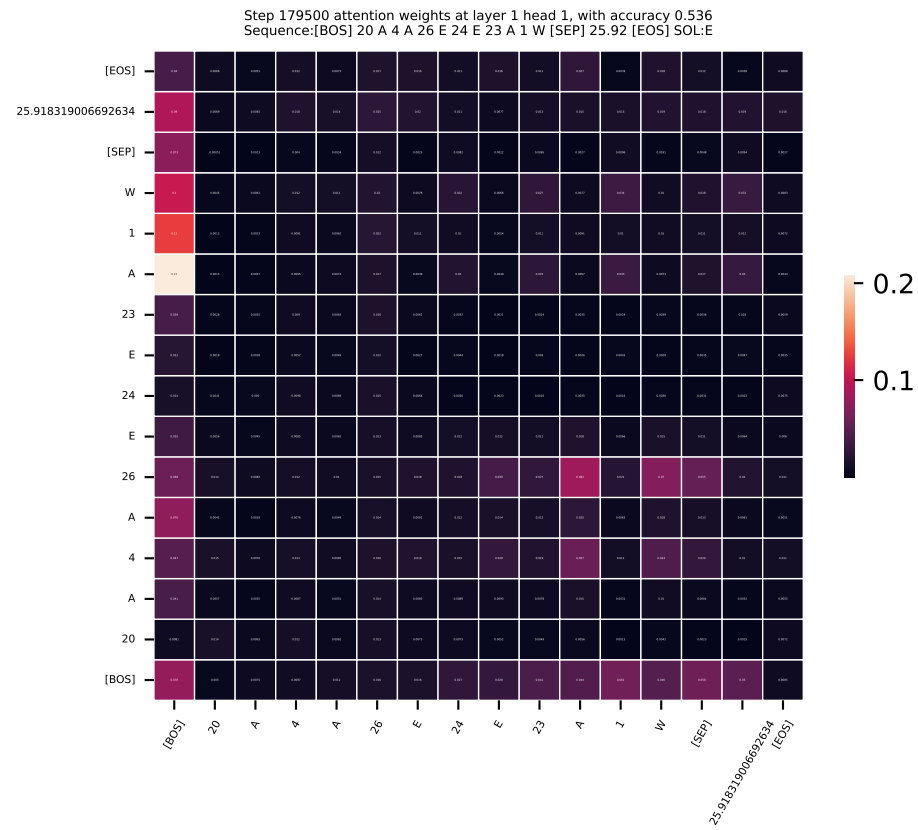

Step 179500 attention weights at layer 0 head 0, with accuracy 0.536
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

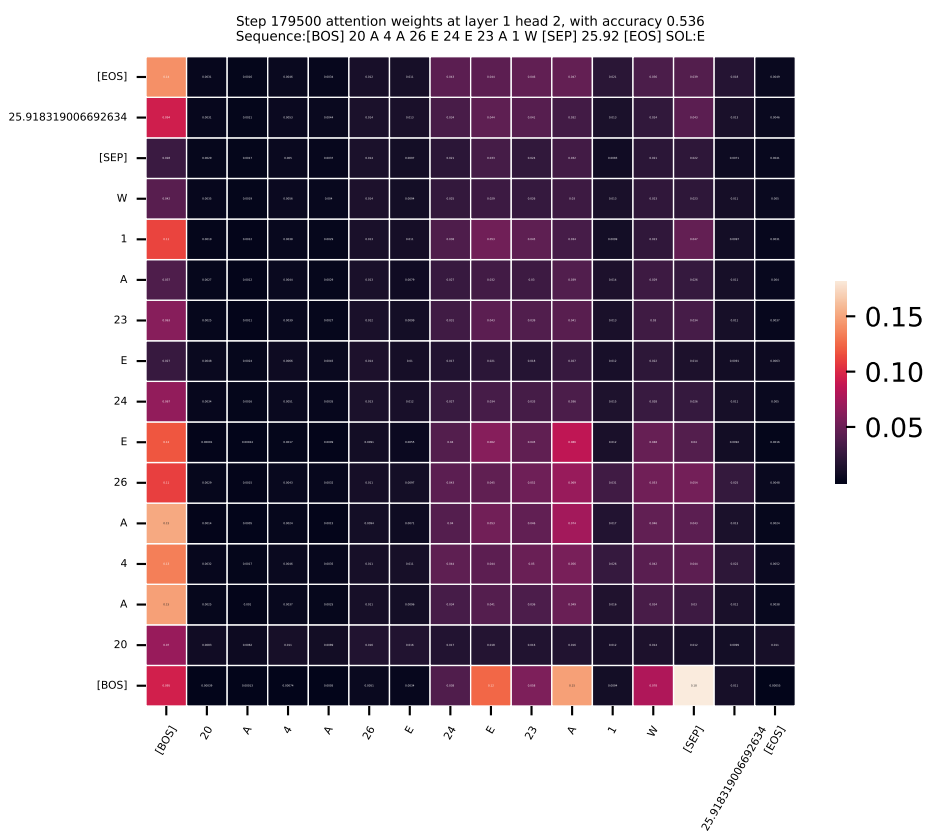

Step 179500 attention weights at layer 0 head 1, with accuracy 0.536
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

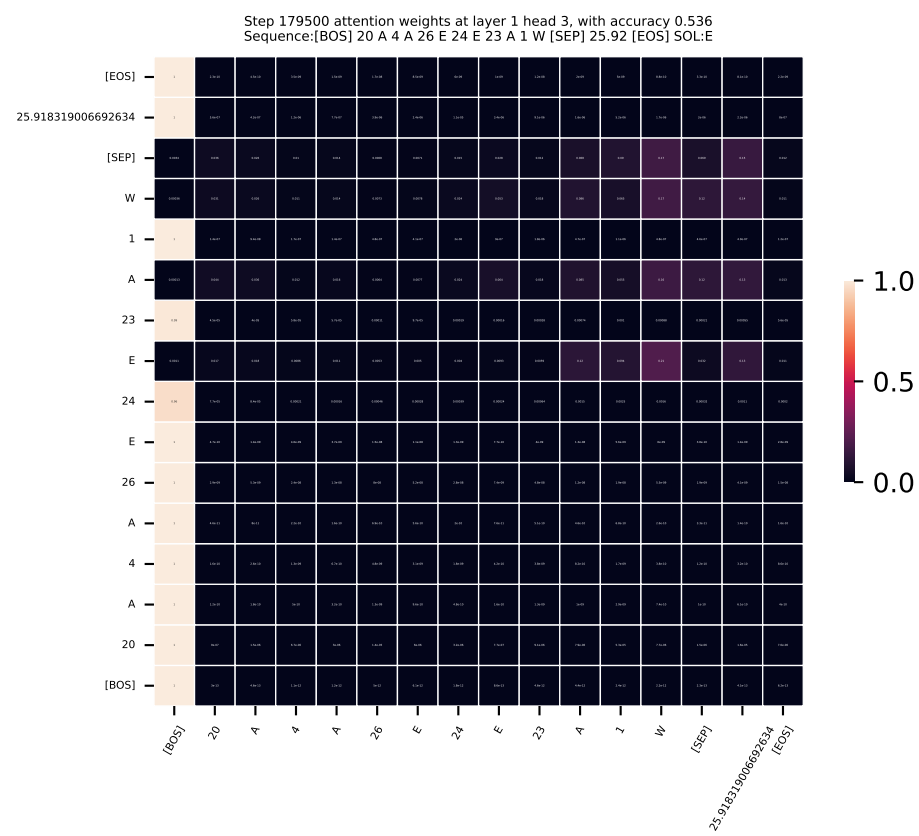

Step 179500 attention weights at layer 0 head 2, with accuracy 0.536
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

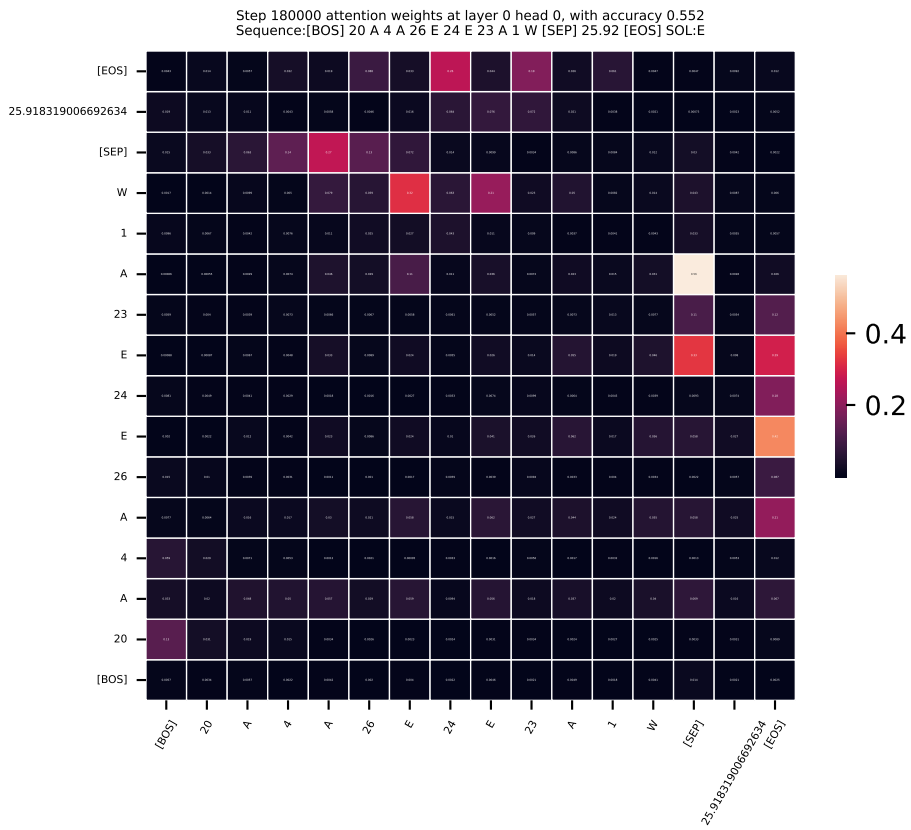

Step 179500 attention weights at layer 0 head 3, with accuracy 0.536
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

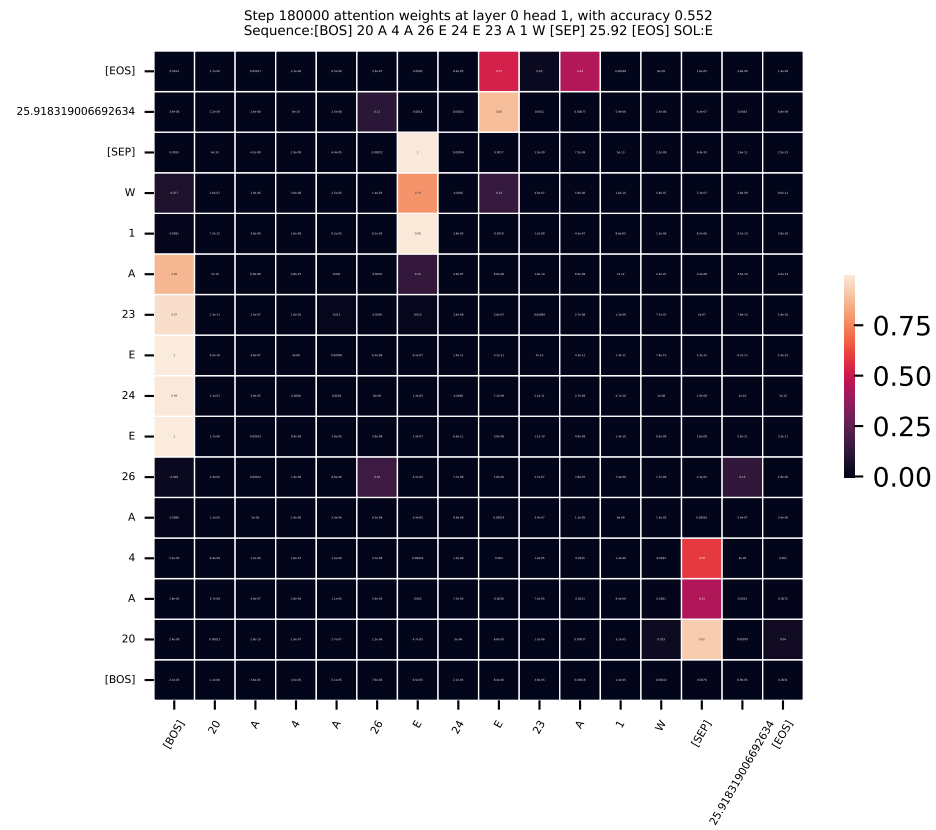

Step 179500 attention weights at layer 1 head 0, with accuracy 0.536
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.15
0.10
0.05

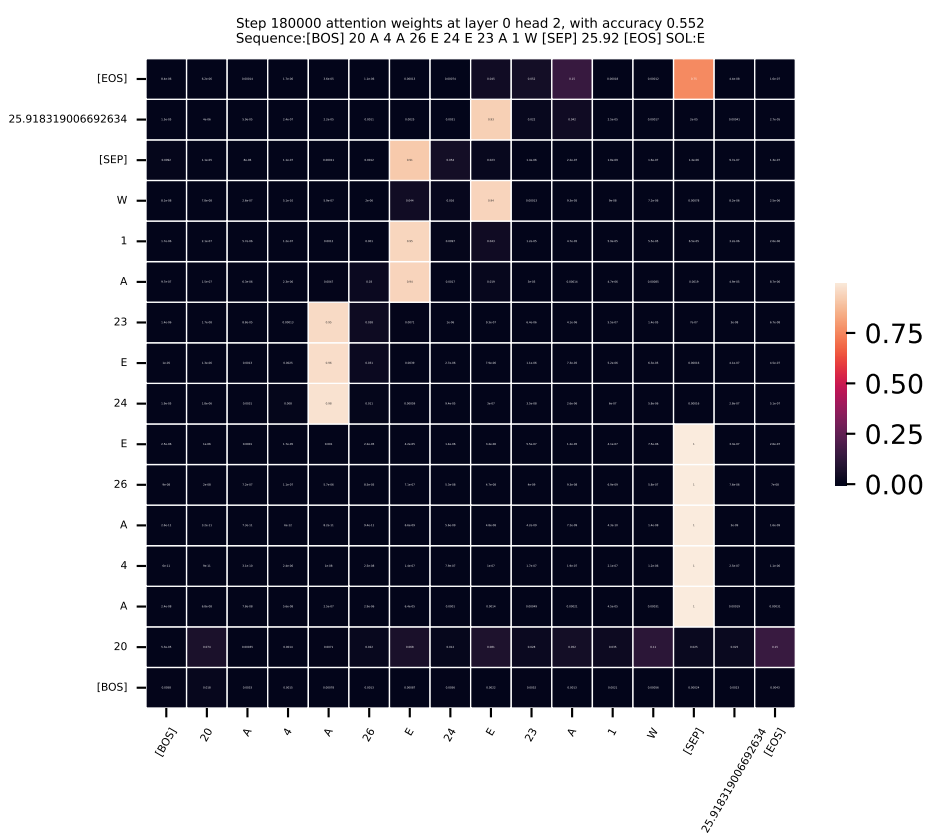

Step 179500 attention weights at layer 1 head 1, with accuracy 0.536
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.2
0.1

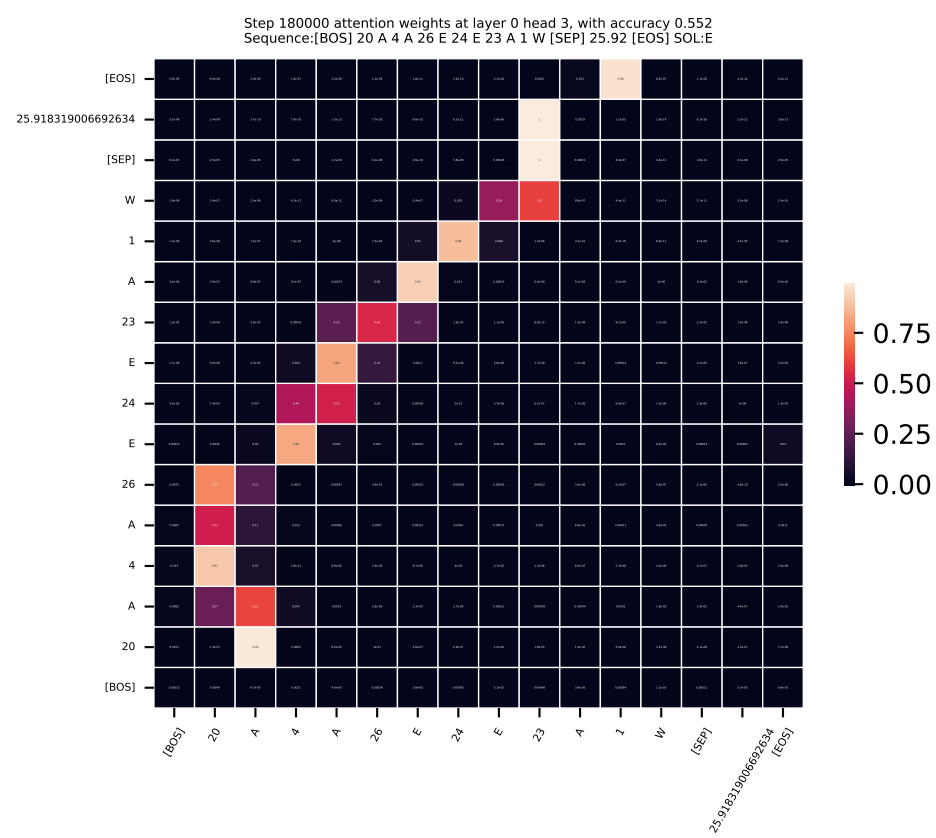

Step 179500 attention weights at layer 1 head 2, with accuracy 0.536
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.15
0.10
0.05

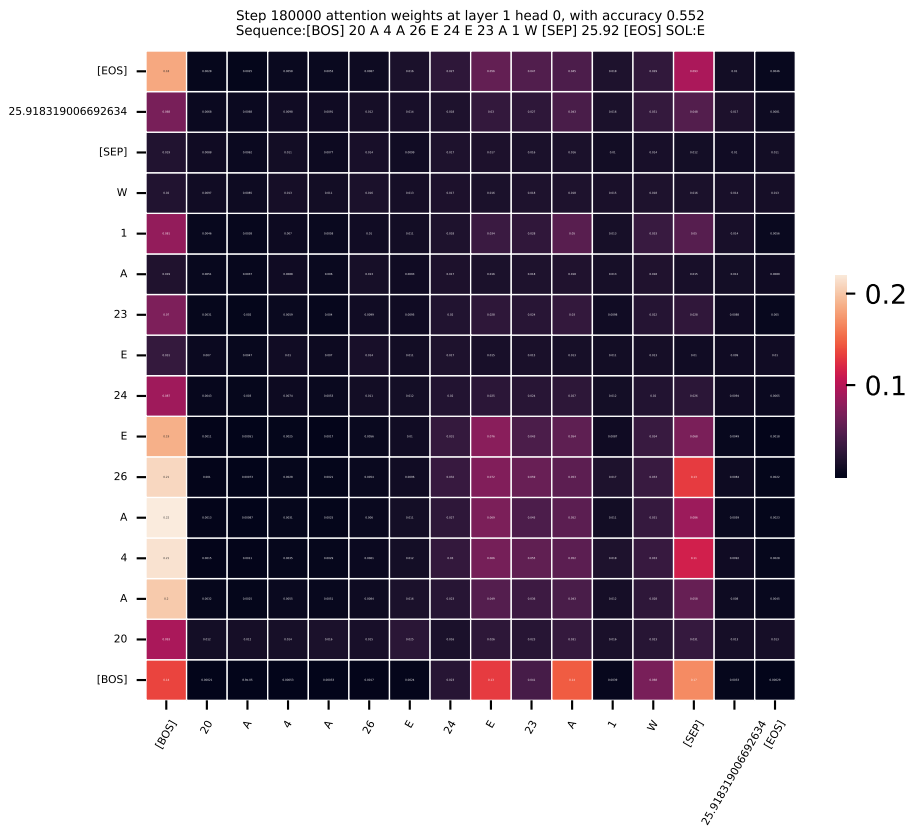

Step 179500 attention weights at layer 1 head 3, with accuracy 0.536
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
1.0
0.5
0.0

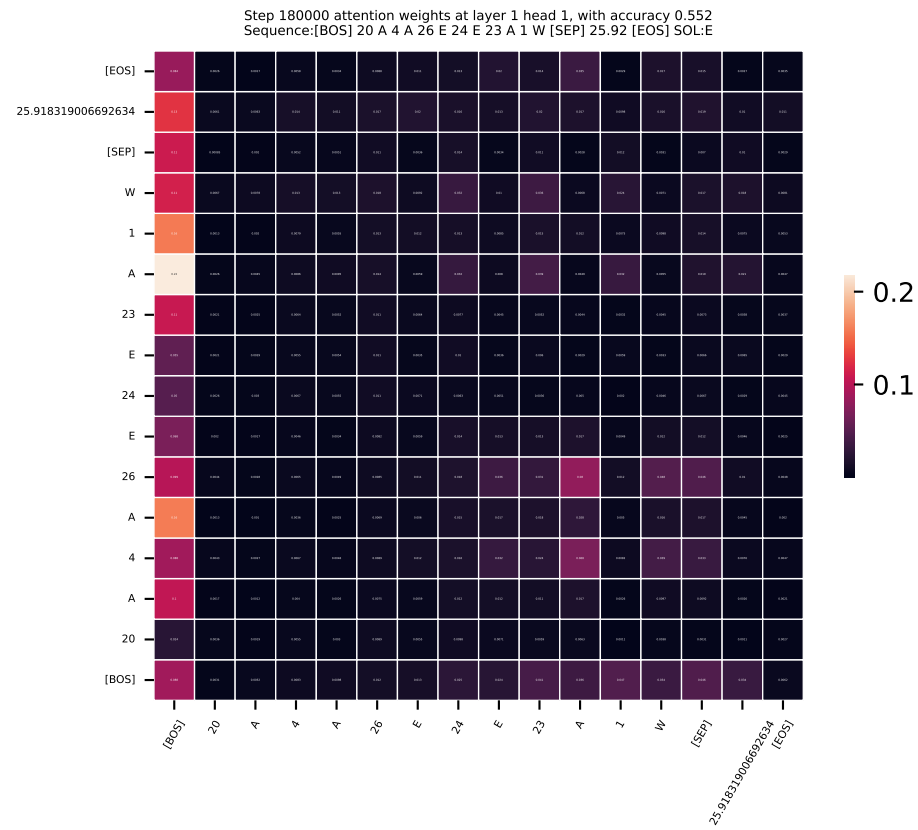

Step 180000 attention weights at layer 0 head 0, with accuracy 0.552
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

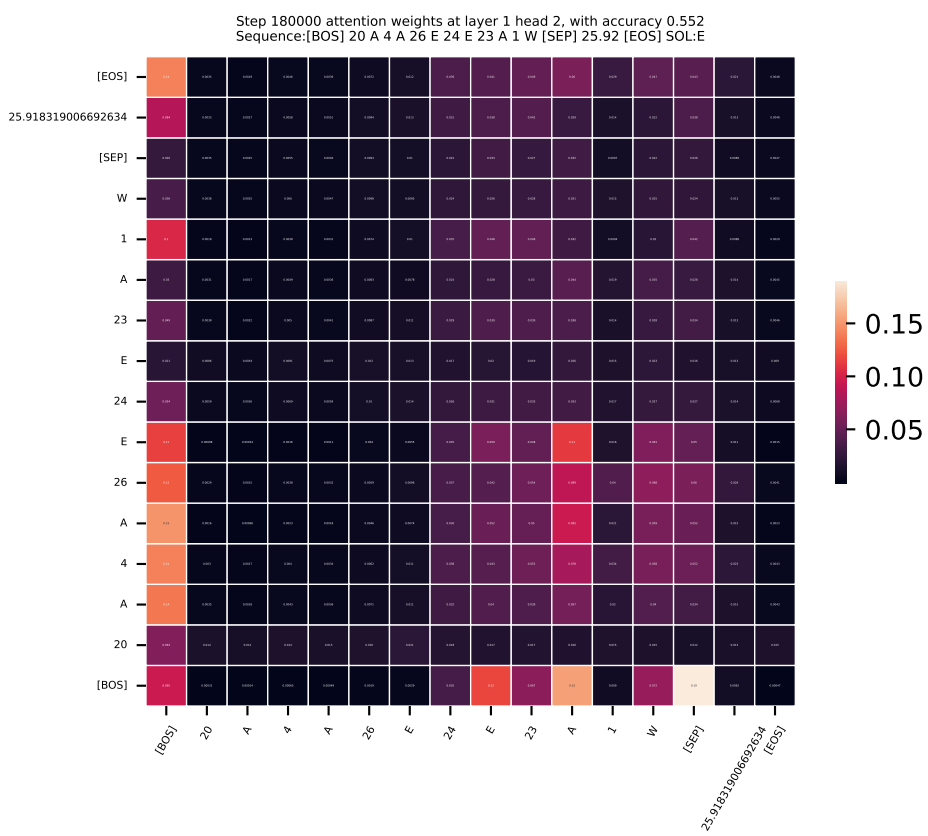

Step 180000 attention weights at layer 0 head 1, with accuracy 0.552
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

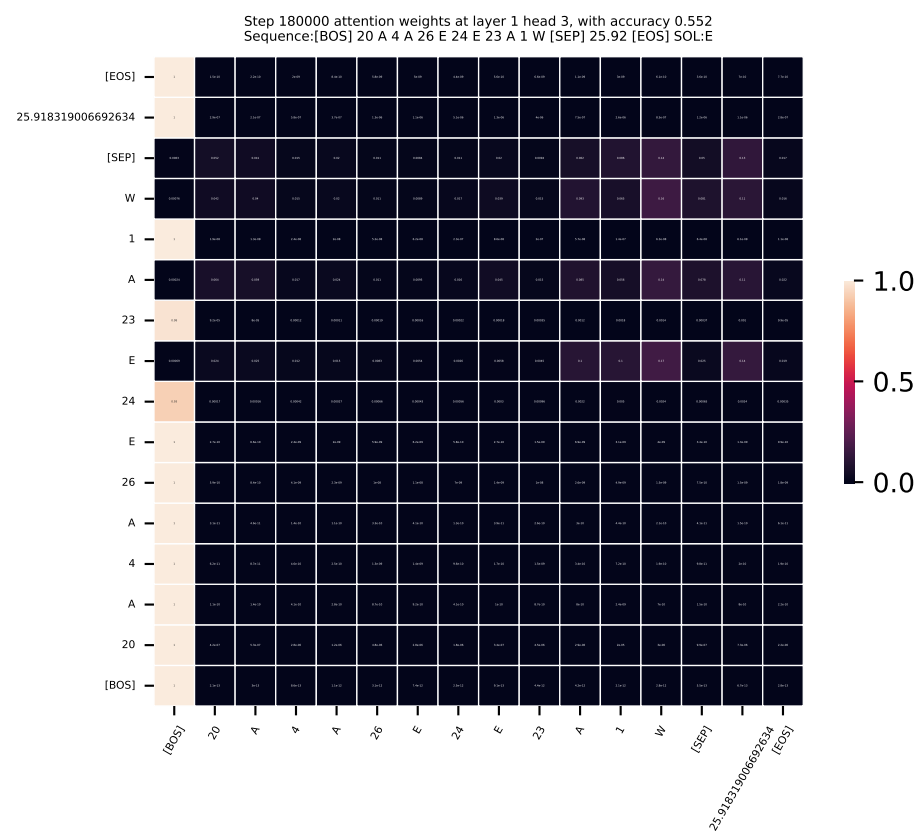

Step 180000 attention weights at layer 0 head 2, with accuracy 0.552
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

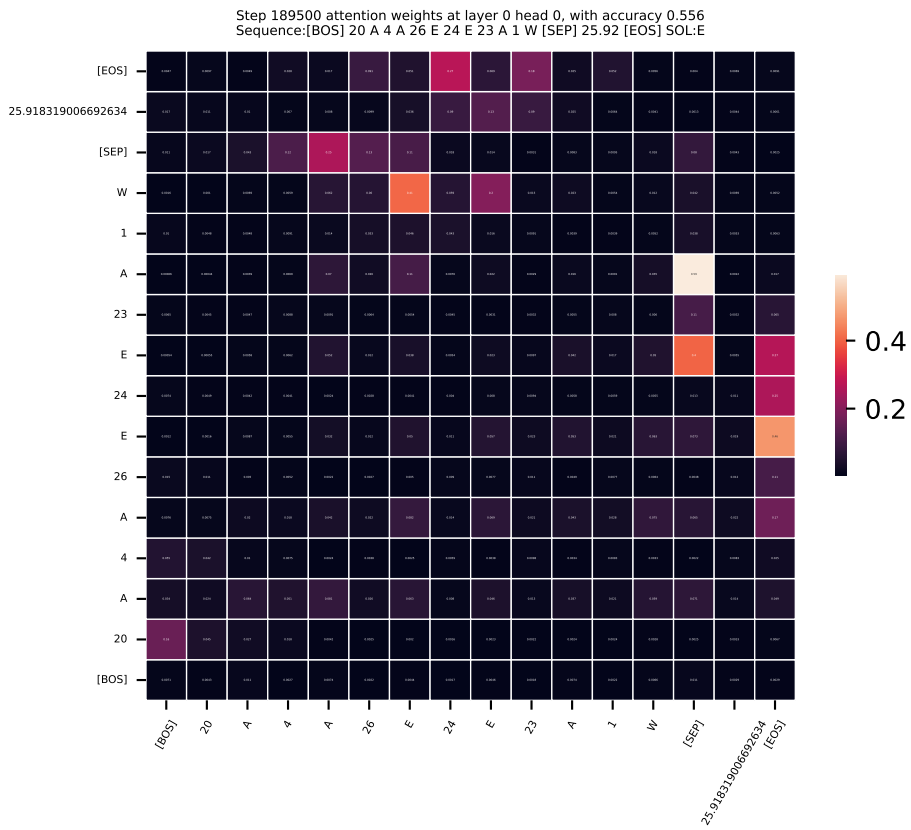

Step 180000 attention weights at layer 0 head 3, with accuracy 0.552
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

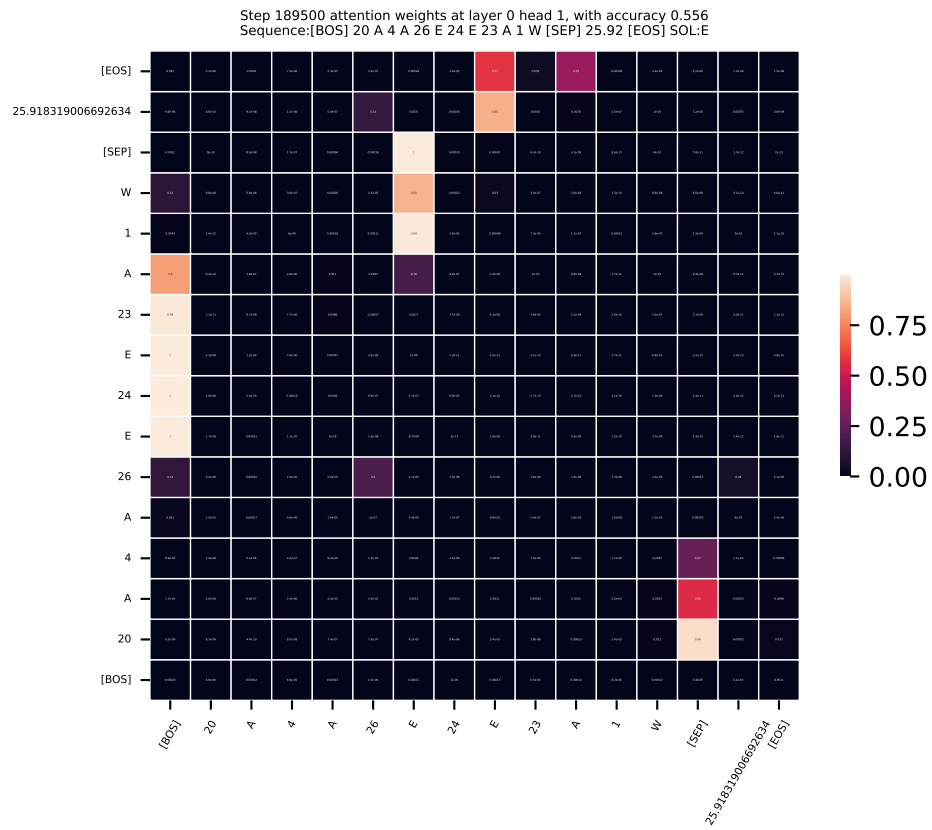

Step 180000 attention weights at layer 1 head 0, with accuracy 0.552
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.2
0.1

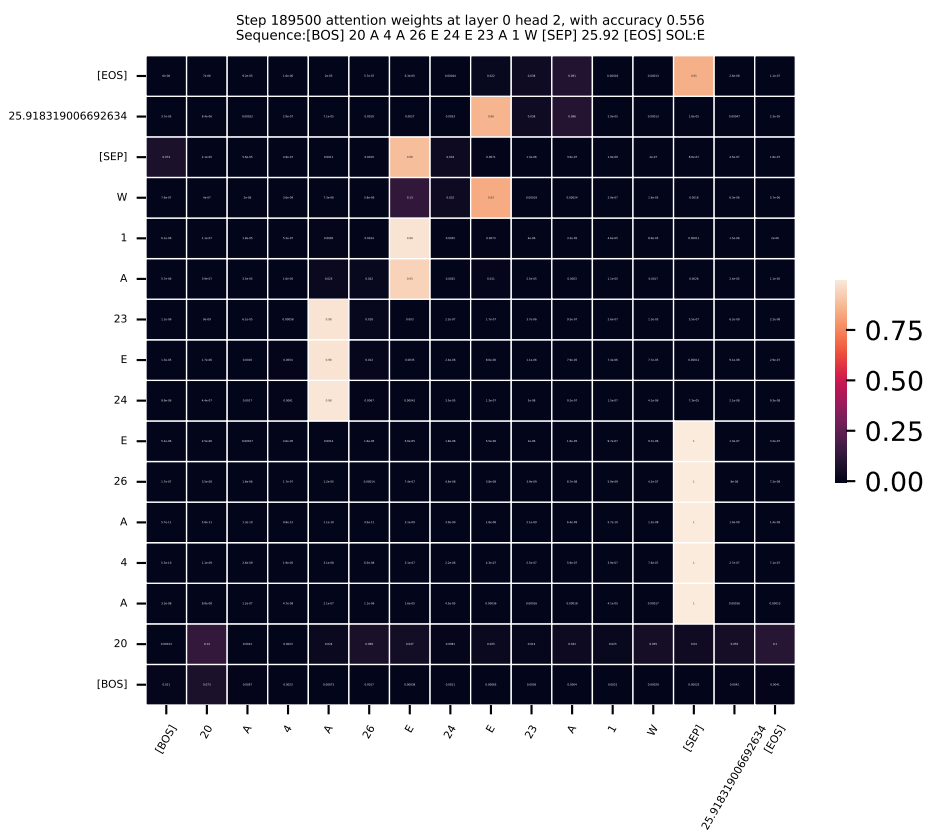

Step 180000 attention weights at layer 1 head 1, with accuracy 0.552
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.2
0.1

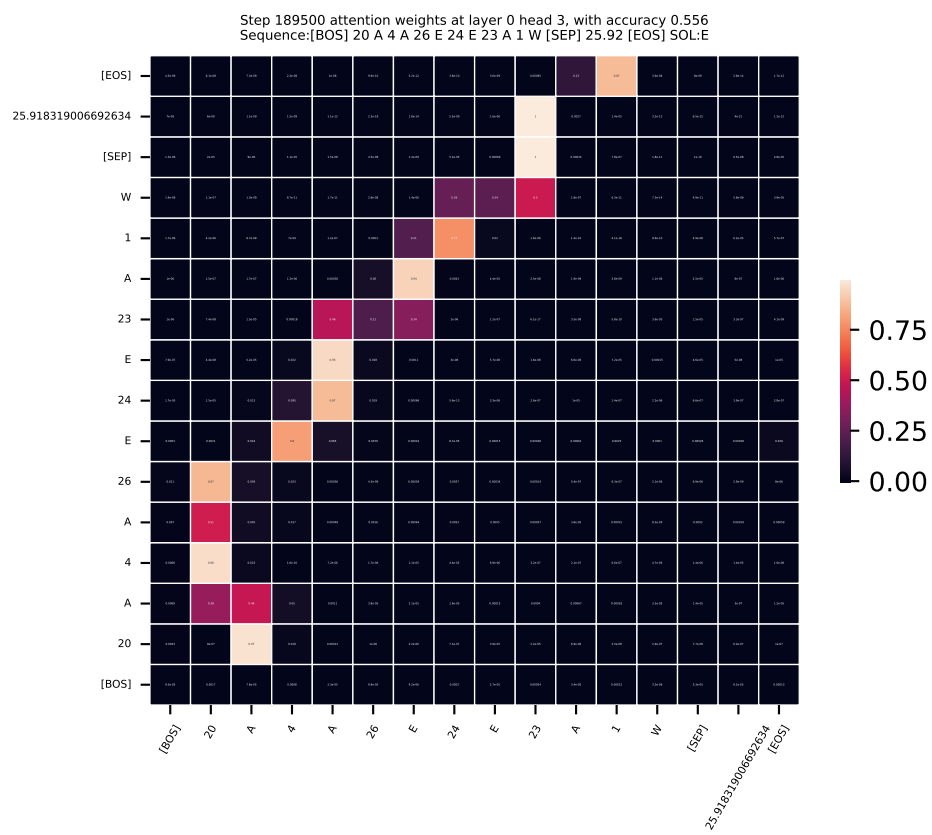

Step 180000 attention weights at layer 1 head 2, with accuracy 0.552
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.15
0.10
0.05

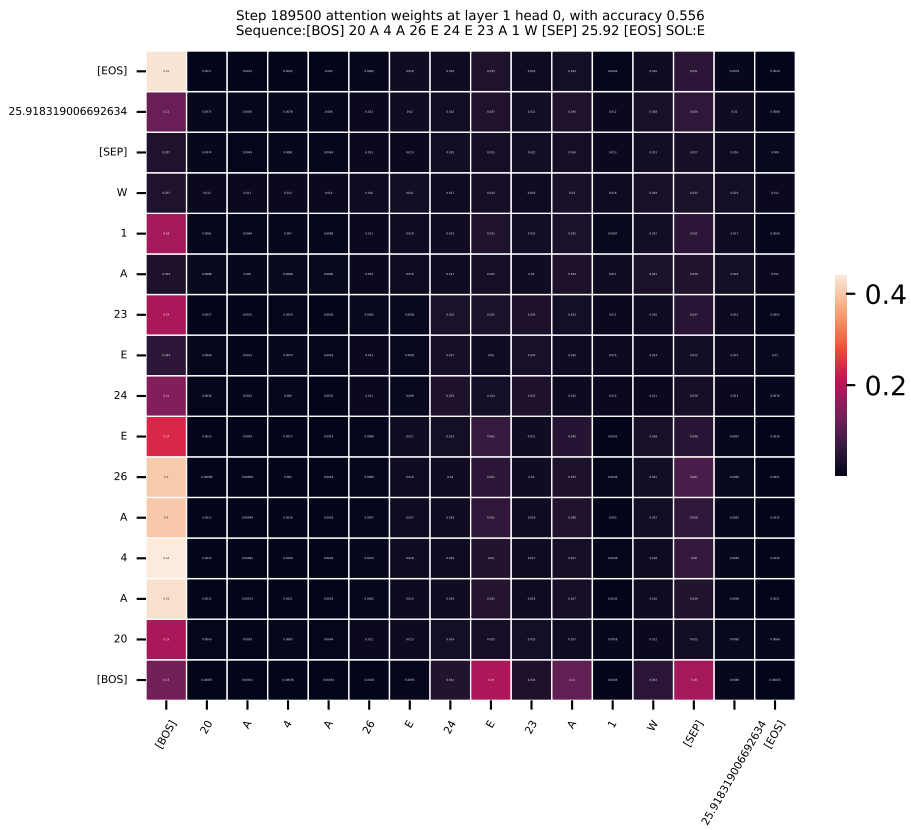

Step 180000 attention weights at layer 1 head 3, with accuracy 0.552
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
1.0
0.5
0.0

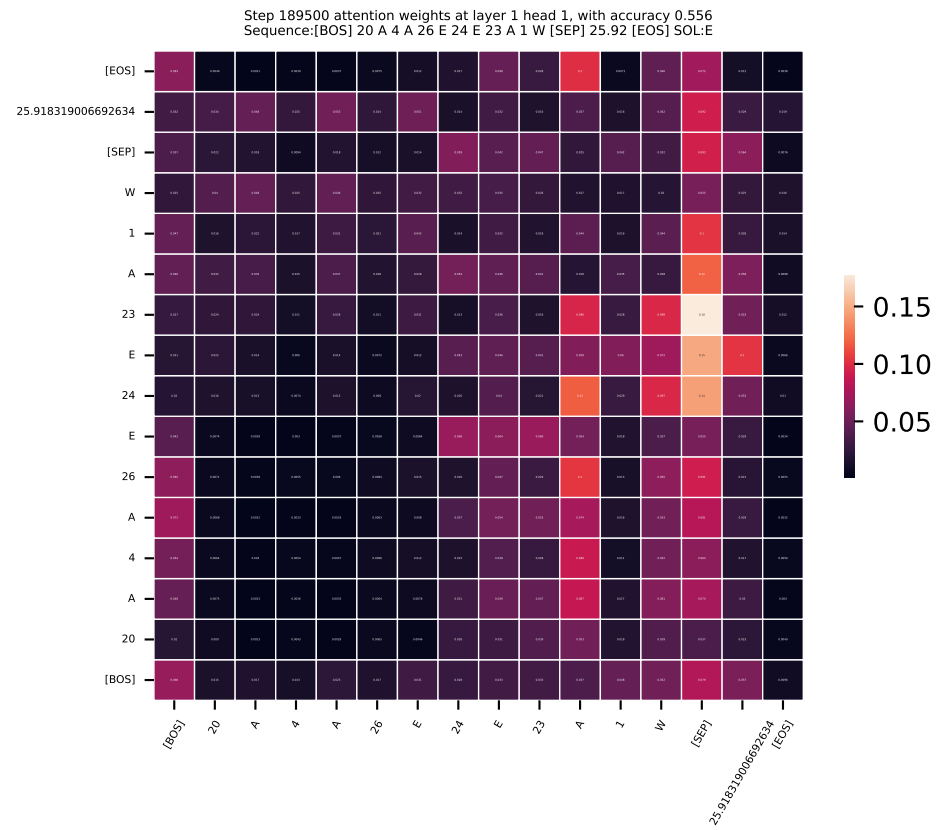

Step 189500 attention weights at layer 0 head 0, with accuracy 0.556
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

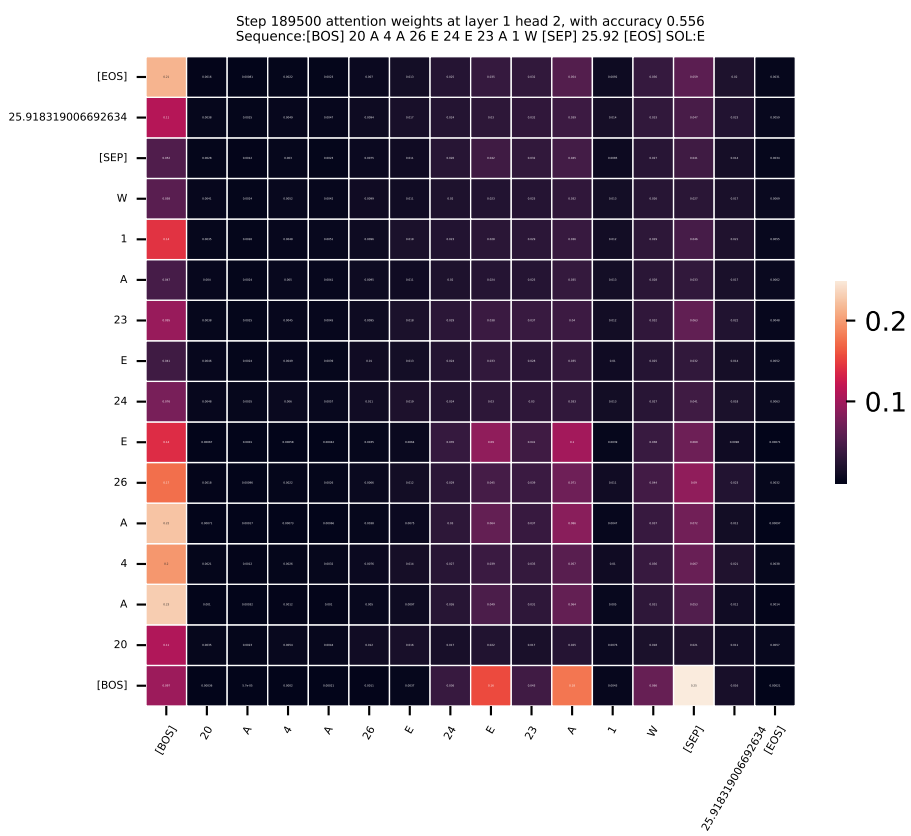

Step 189500 attention weights at layer 0 head 1, with accuracy 0.556
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

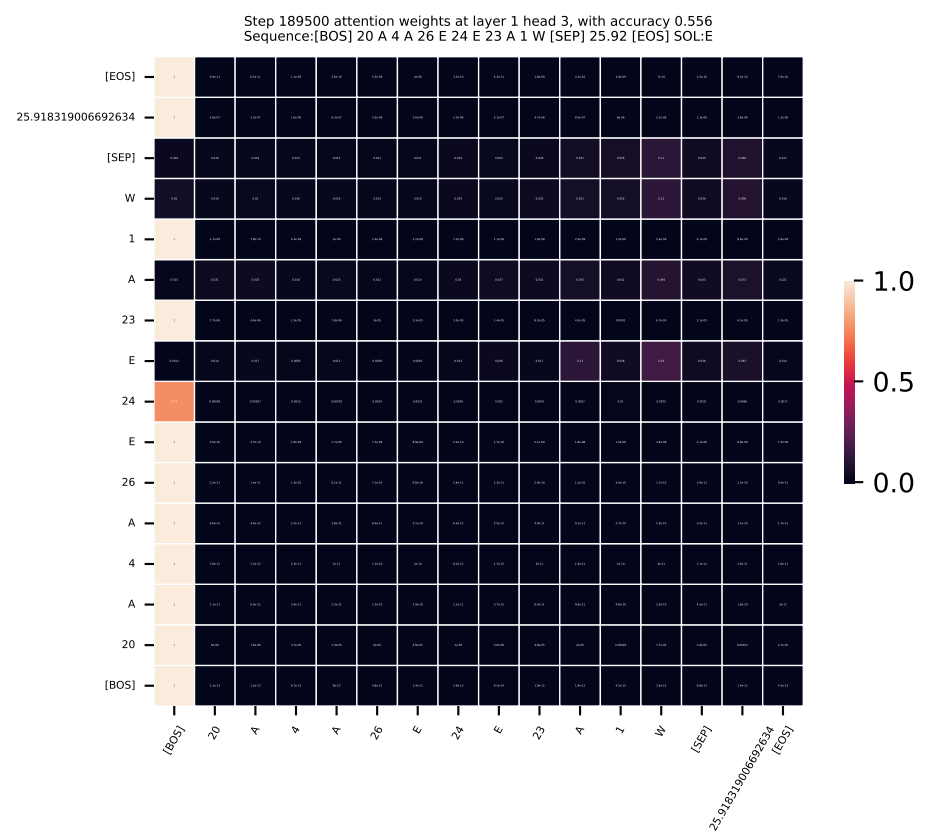

Step 189500 attention weights at layer 0 head 2, with accuracy 0.556
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

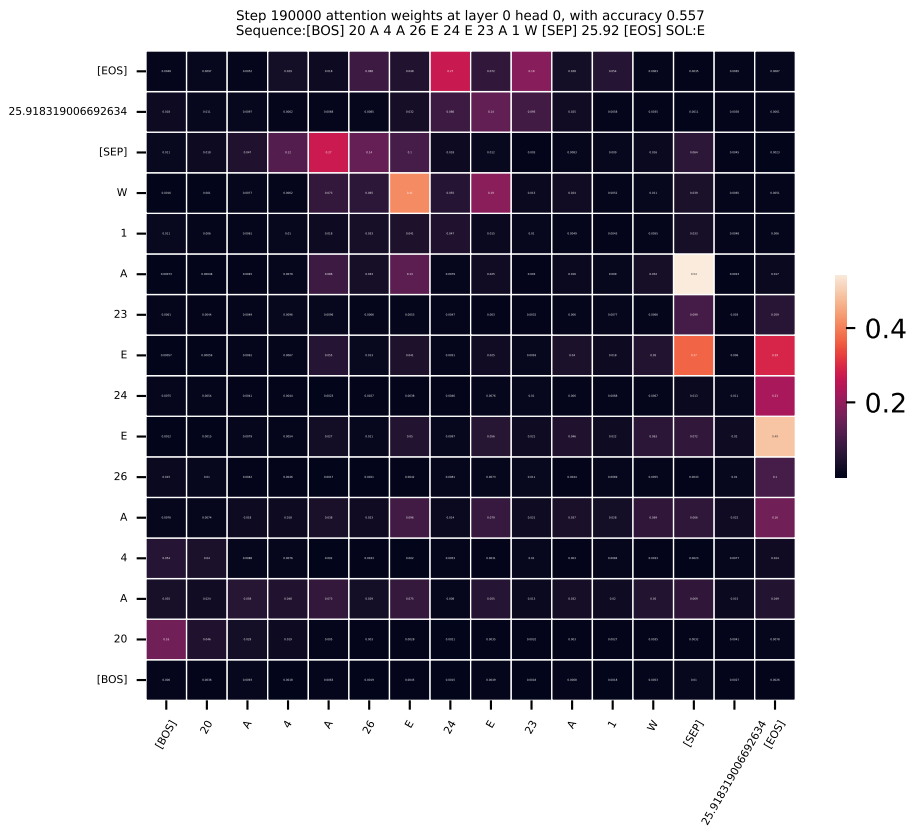

Step 189500 attention weights at layer 0 head 3, with accuracy 0.556
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

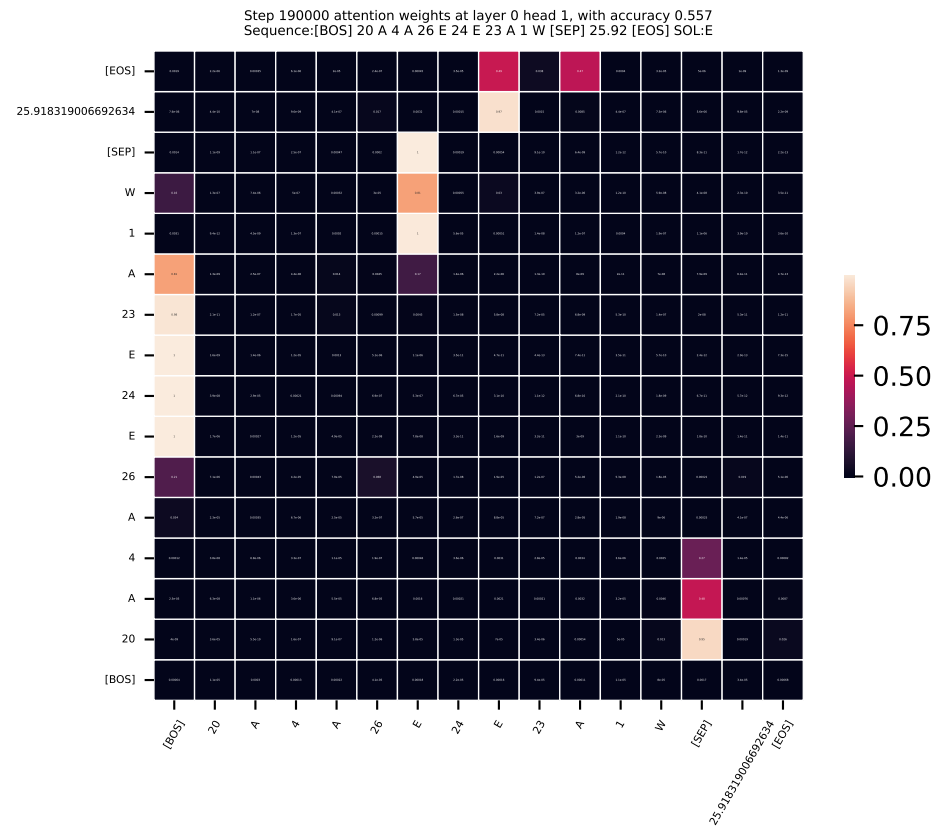

Step 189500 attention weights at layer 1 head 0, with accuracy 0.556
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

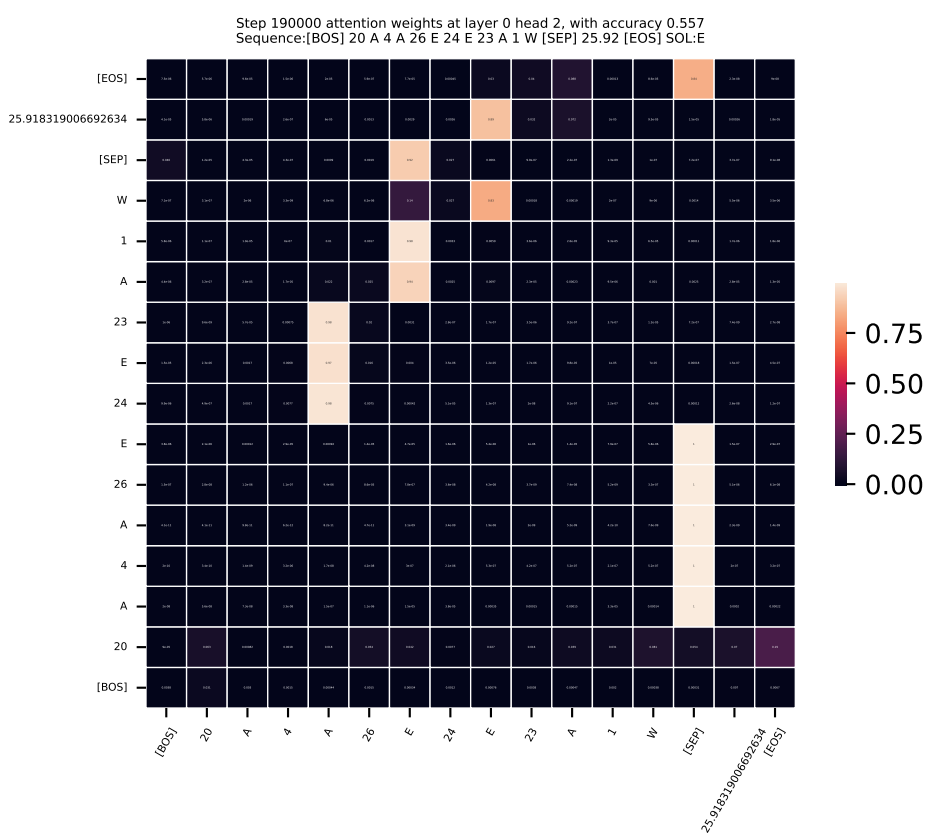

Step 189500 attention weights at layer 1 head 1, with accuracy 0.556
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.15
0.10
0.05

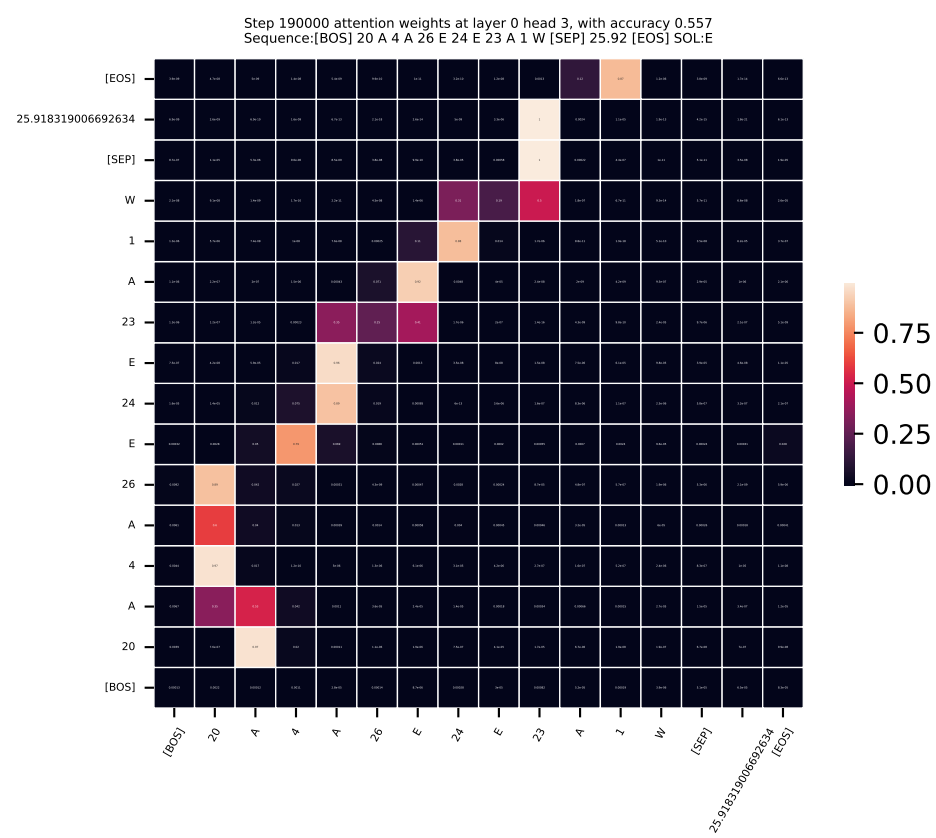

Step 189500 attention weights at layer 1 head 2, with accuracy 0.556
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.2
0.1

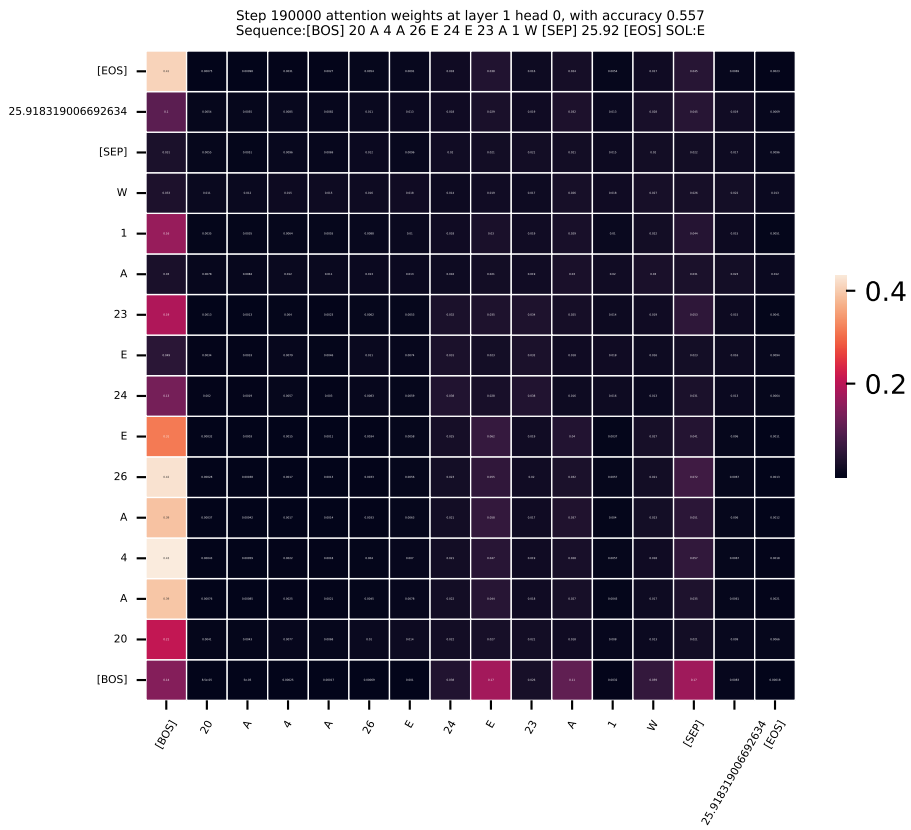

Step 189500 attention weights at layer 1 head 3, with accuracy 0.556
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
1.0
0.5
0.0

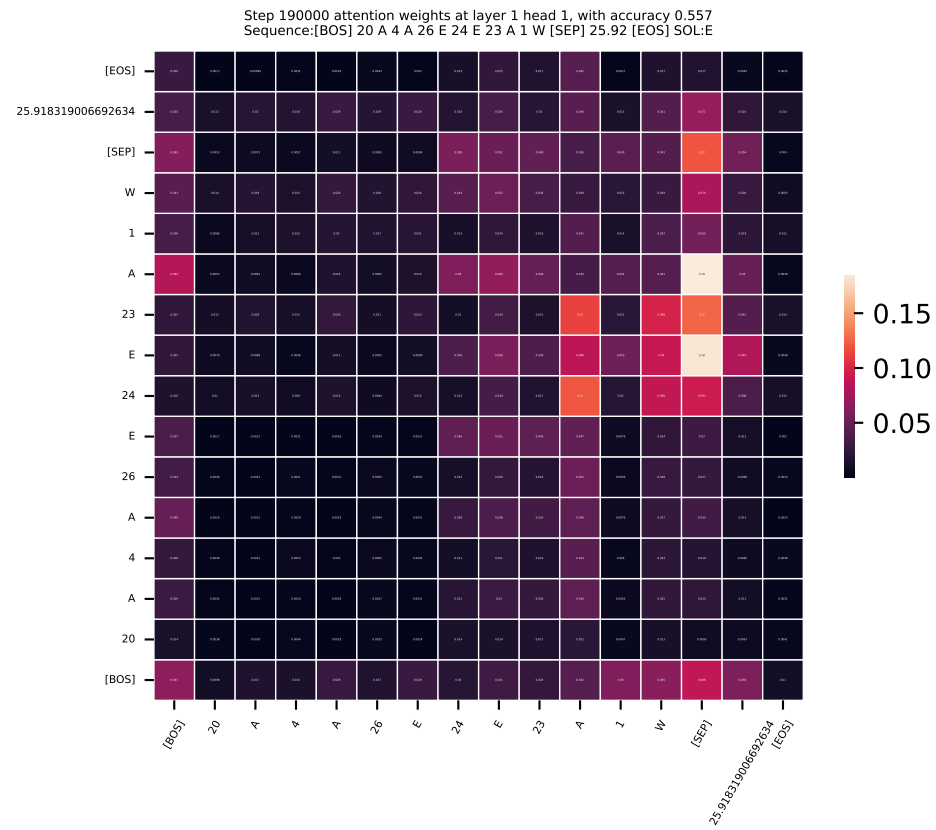

Step 190000 attention weights at layer 0 head 0, with accuracy 0.557
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

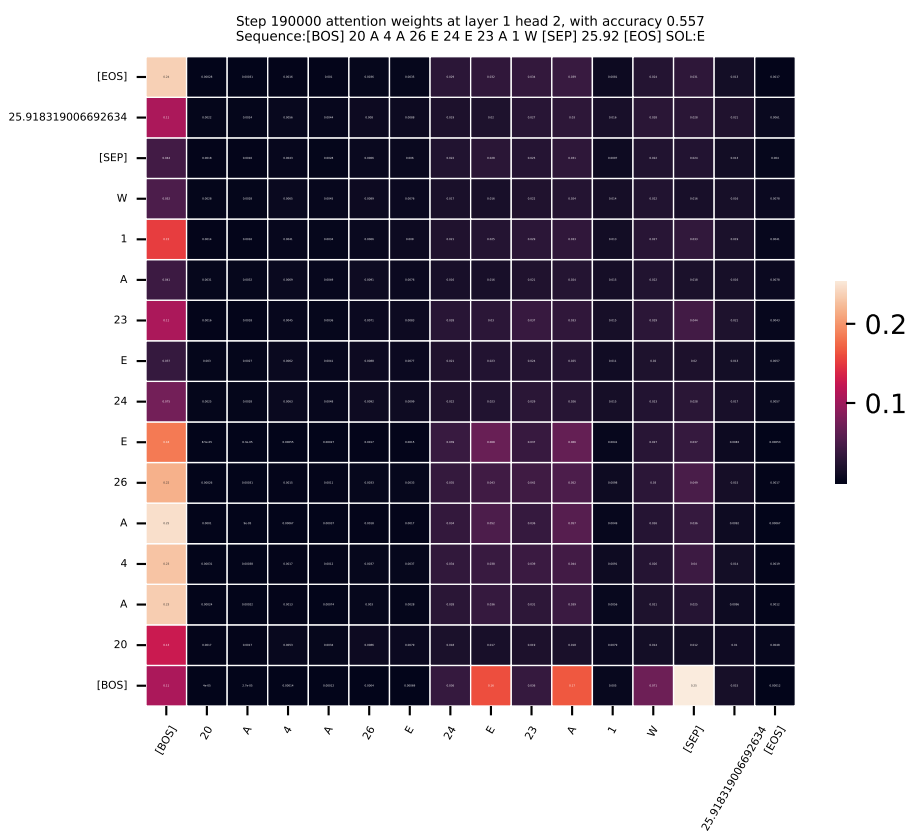

Step 190000 attention weights at layer 0 head 1, with accuracy 0.557
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

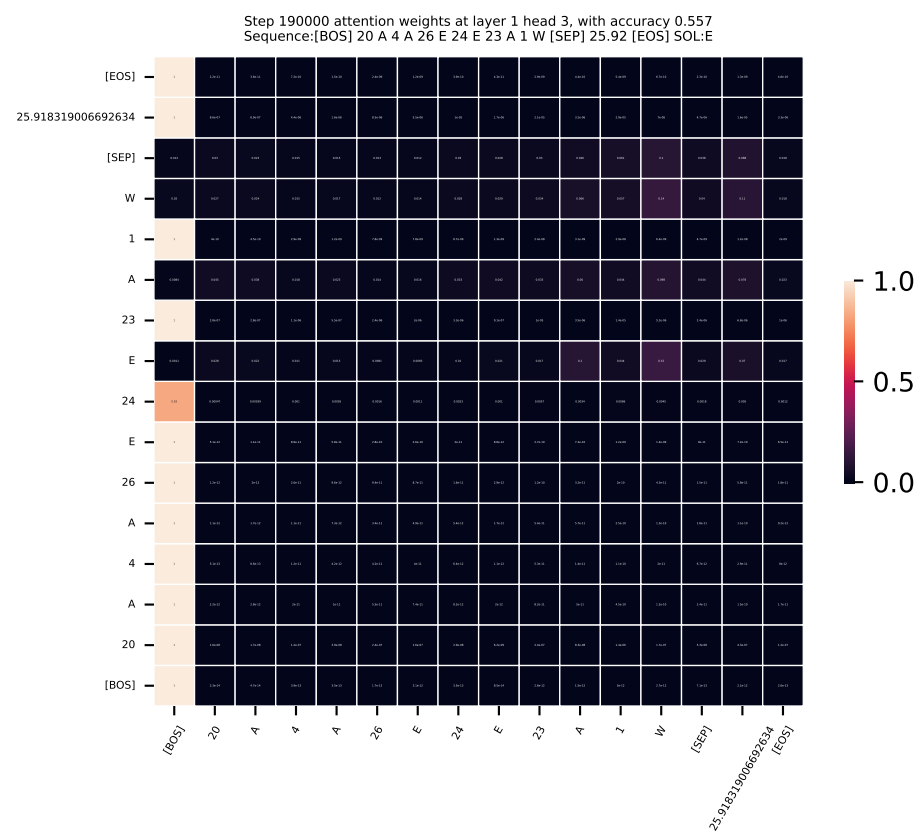

Step 190000 attention weights at layer 0 head 2, with accuracy 0.557
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

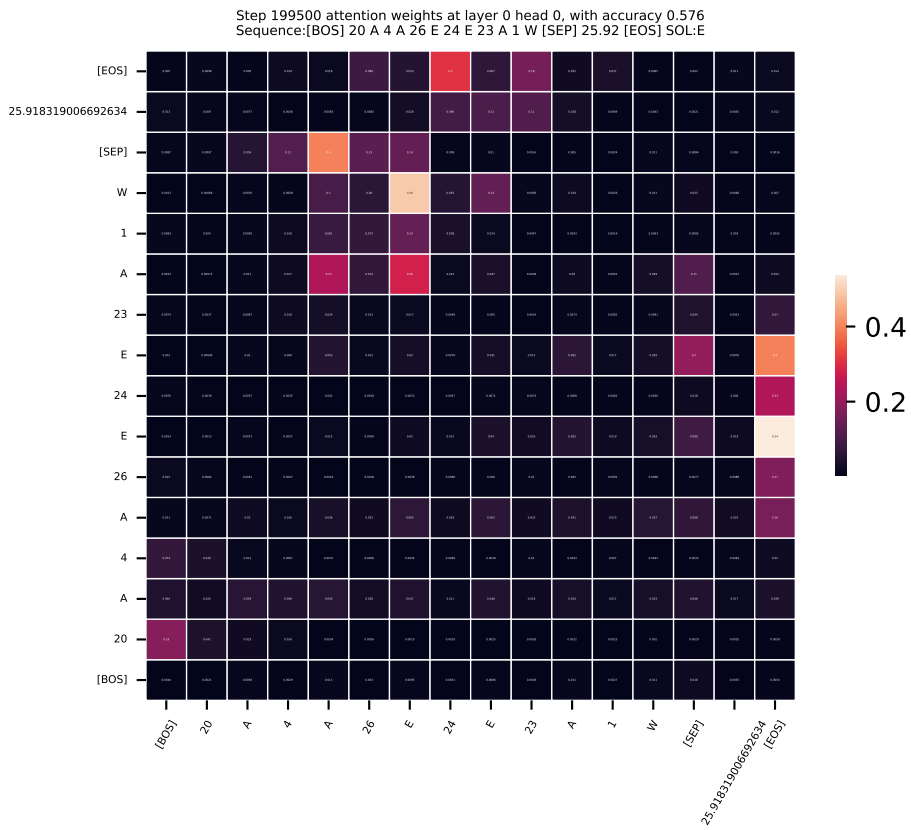

Step 190000 attention weights at layer 0 head 3, with accuracy 0.557
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

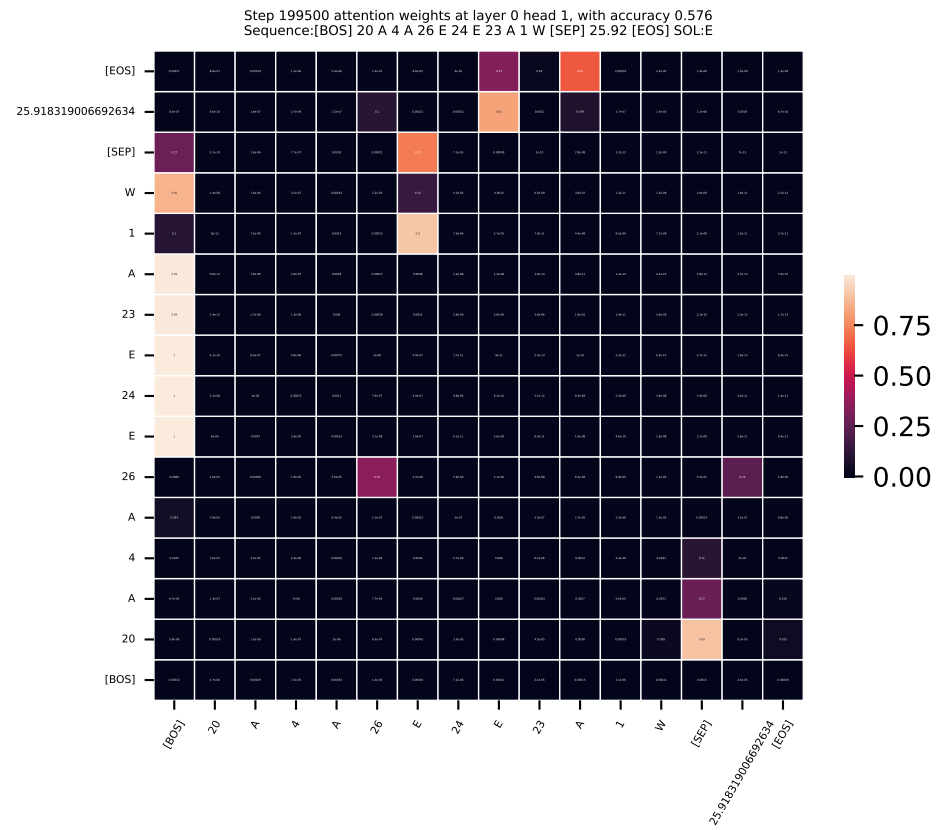

Step 190000 attention weights at layer 1 head 0, with accuracy 0.557
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

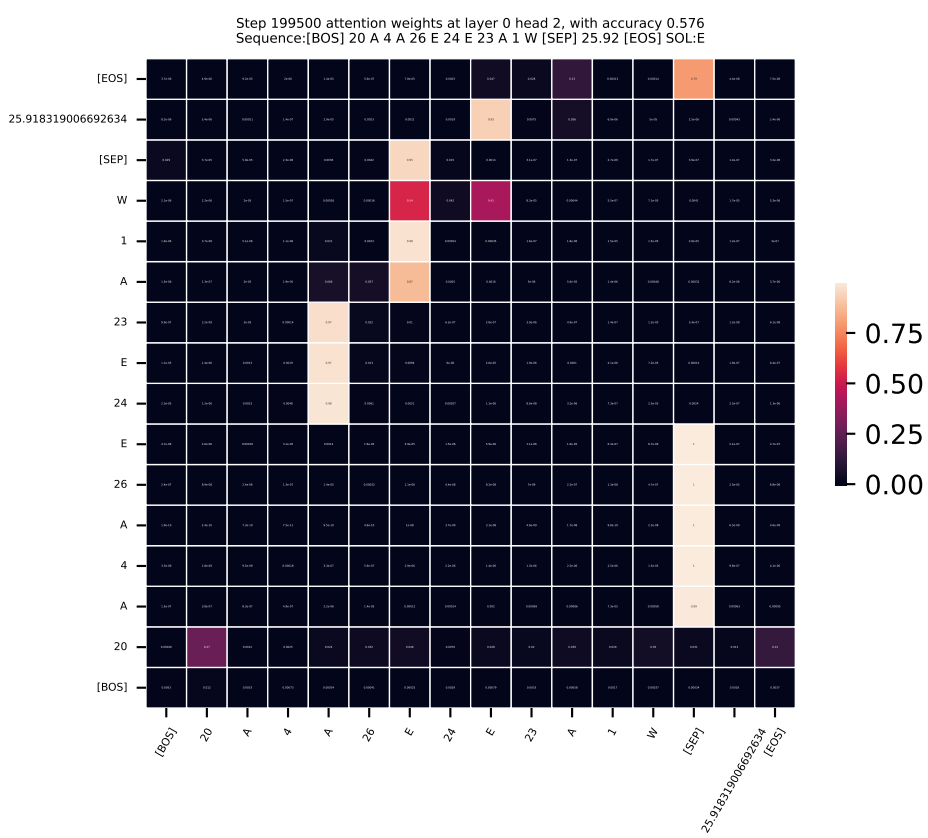

Step 190000 attention weights at layer 1 head 1, with accuracy 0.557
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.15
0.10
0.05

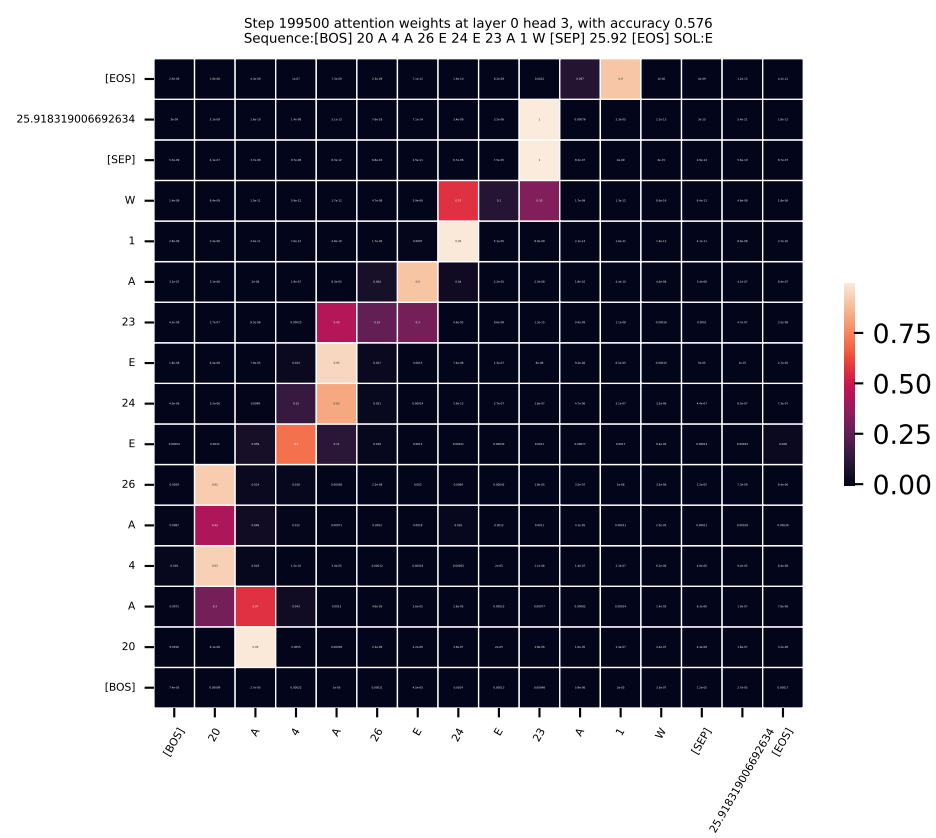

Step 190000 attention weights at layer 1 head 2, with accuracy 0.557
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.2
0.1

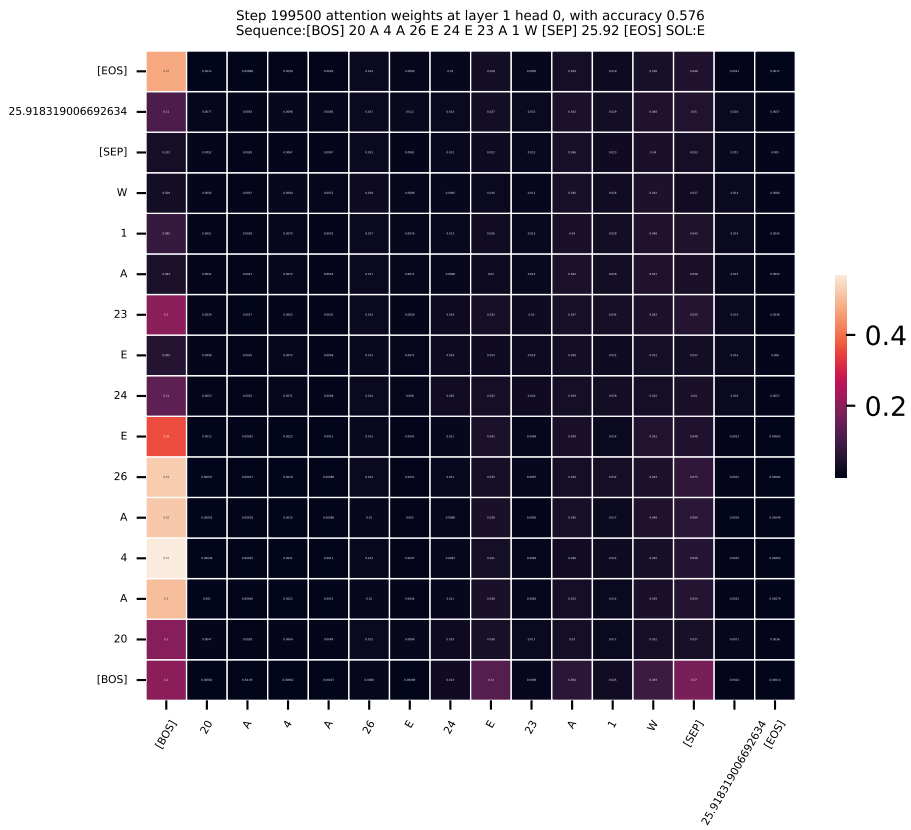

Step 190000 attention weights at layer 1 head 3, with accuracy 0.557
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
1.0
0.5
0.0

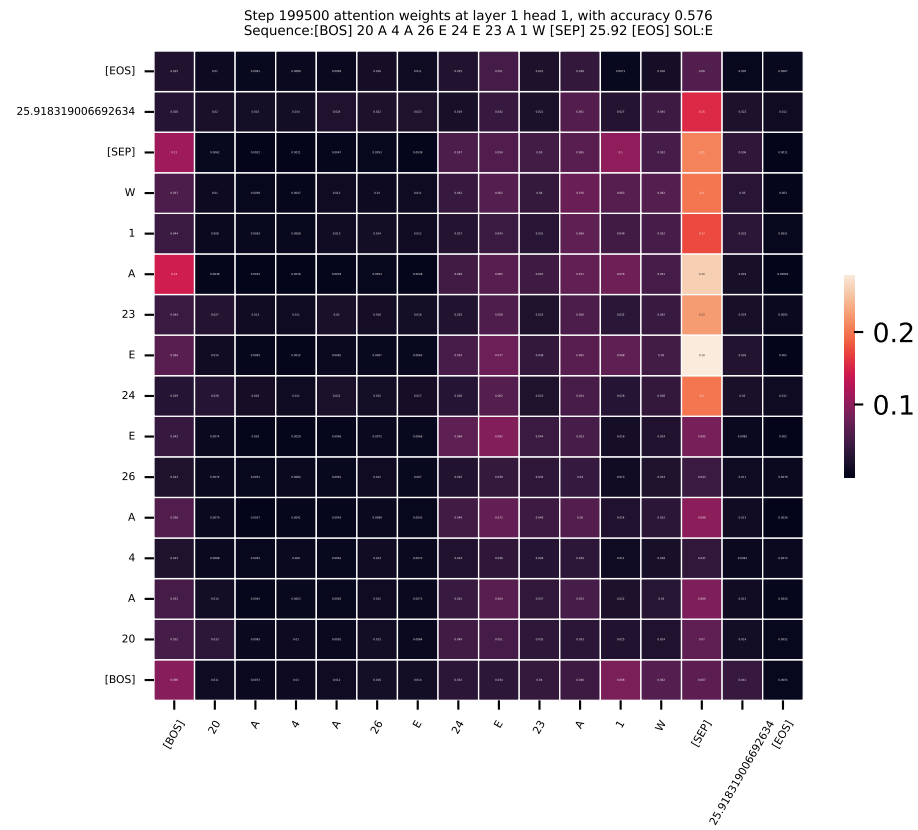

Step 199500 attention weights at layer 0 head 0, with accuracy 0.576
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.4
0.2

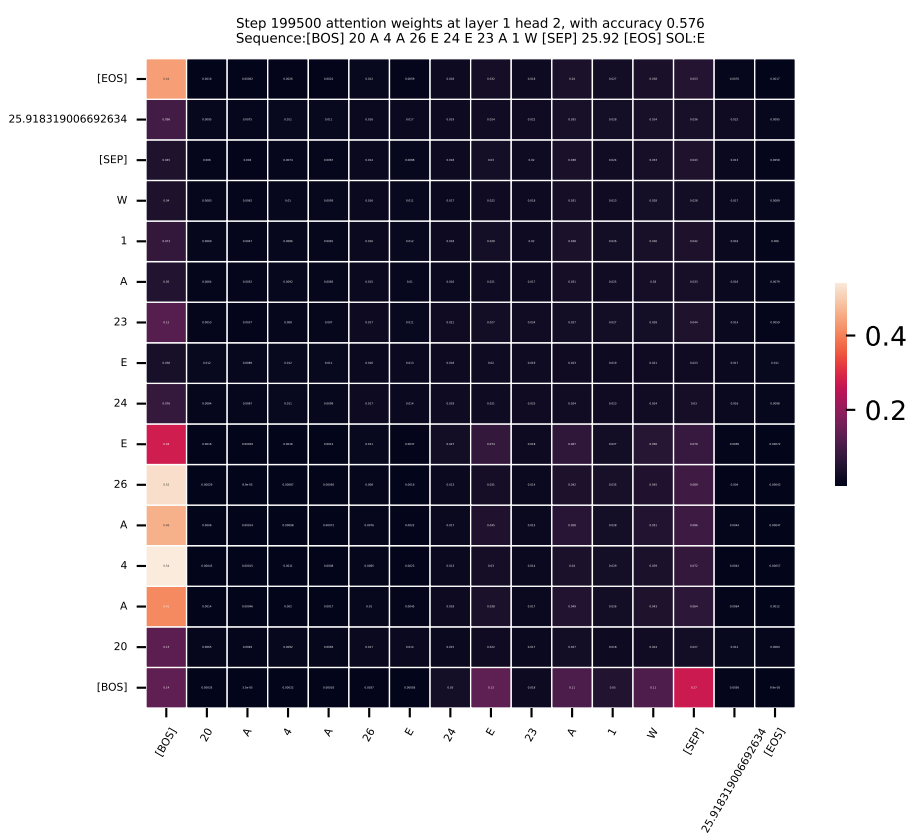

Step 199500 attention weights at layer 0 head 1, with accuracy 0.576
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

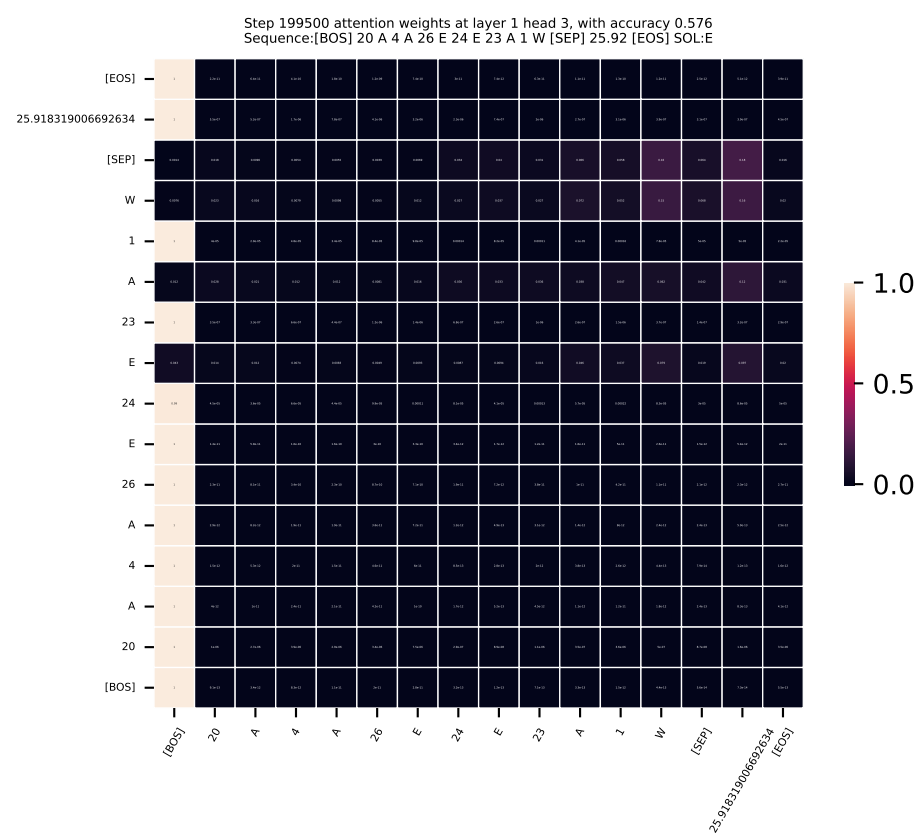

Step 199500 attention weights at layer 0 head 2, with accuracy 0.576
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

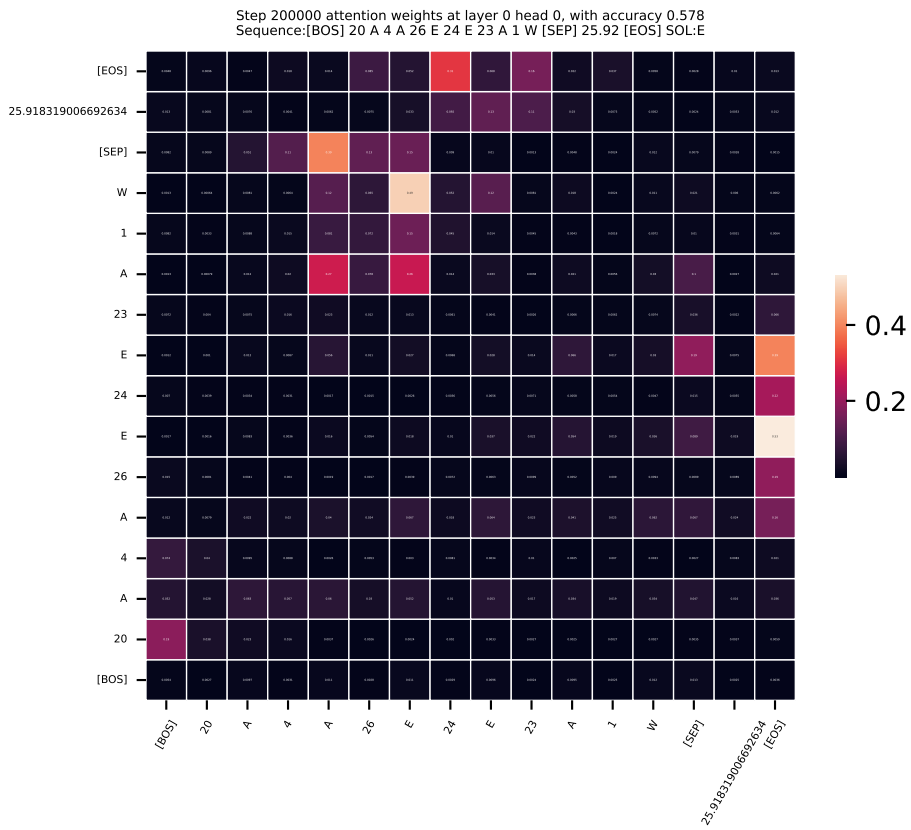

Step 199500 attention weights at layer 0 head 3, with accuracy 0.576
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

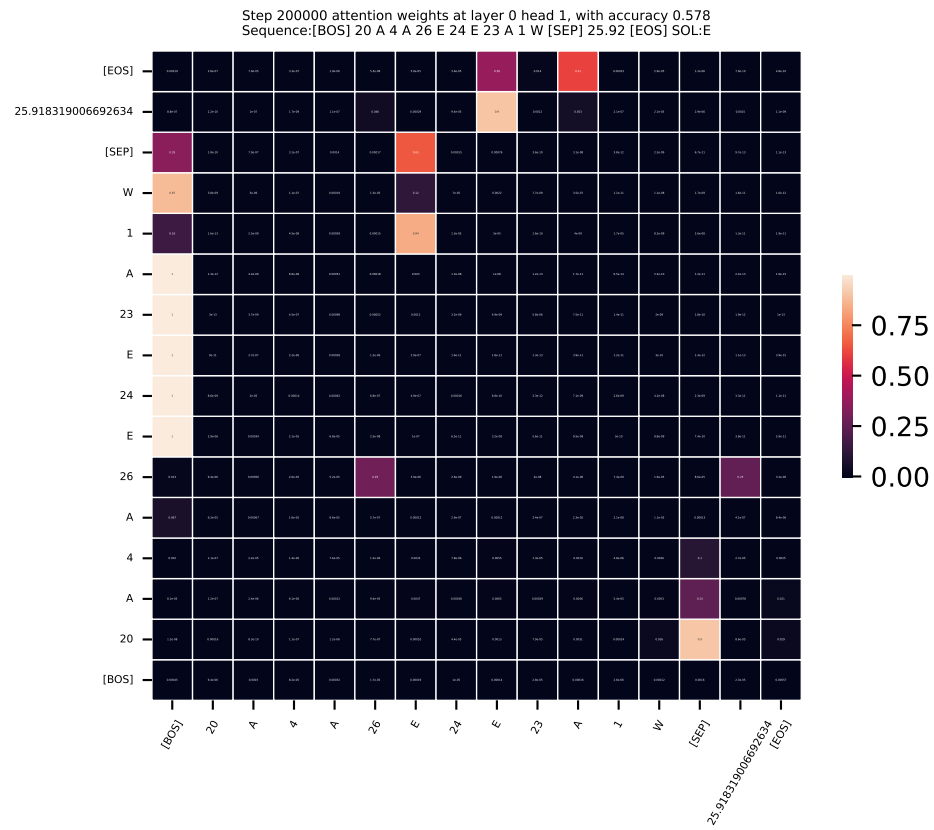

Step 199500 attention weights at layer 1 head 0, with accuracy 0.576
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

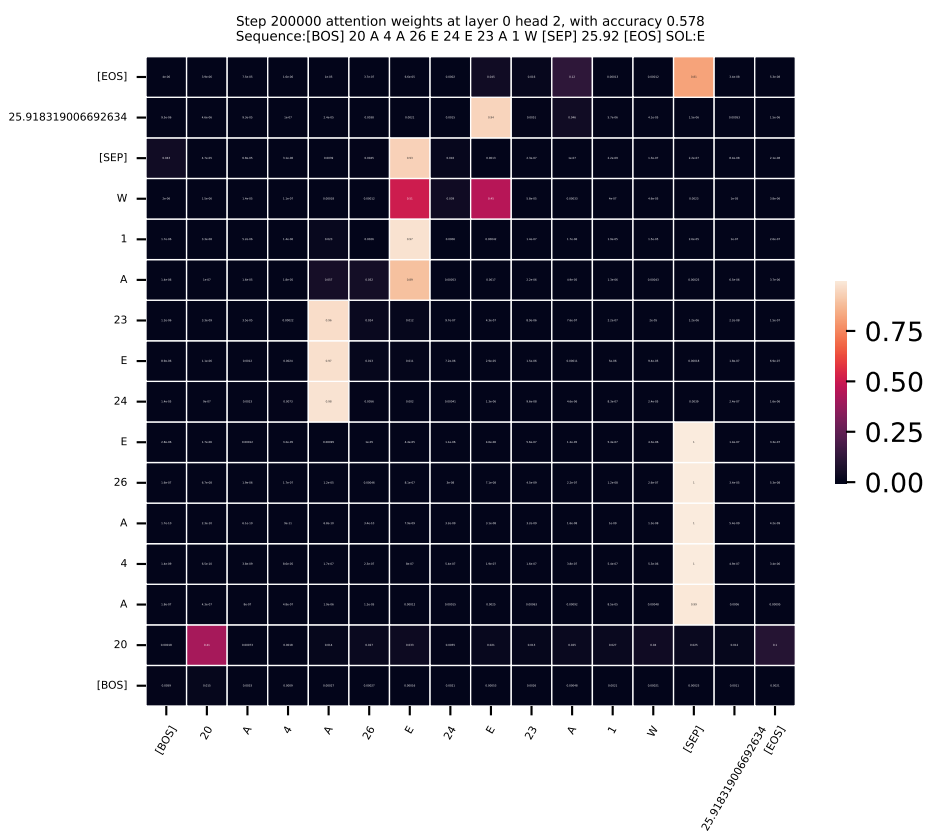

Step 199500 attention weights at layer 1 head 1, with accuracy 0.576
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.2
0.1

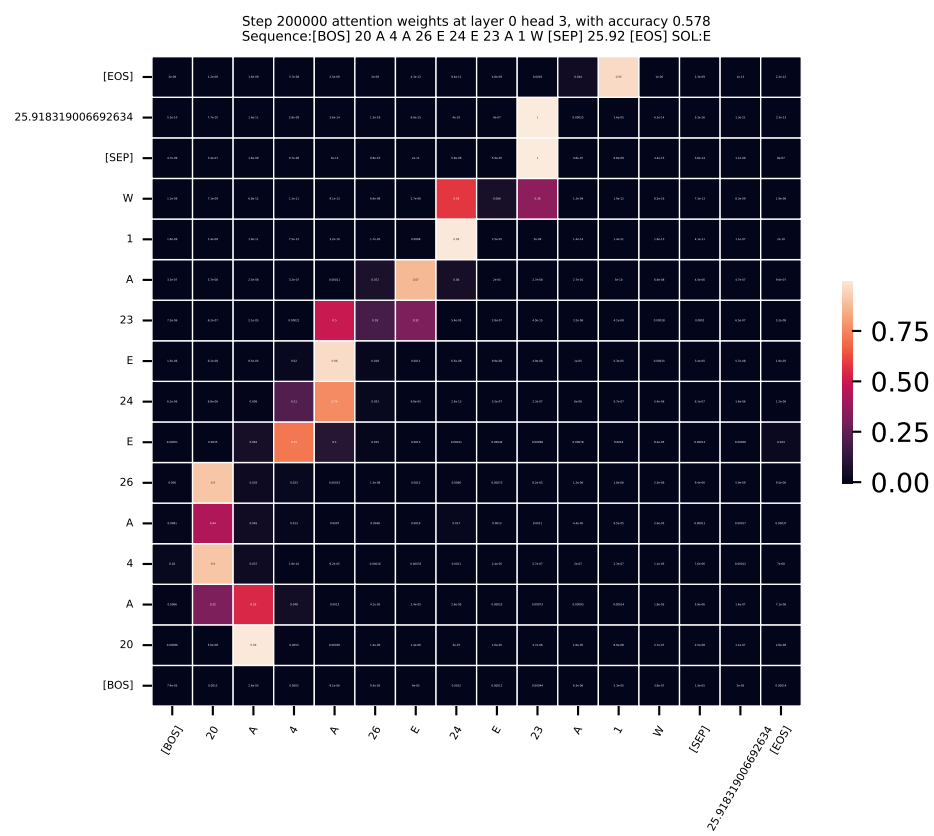

Step 199500 attention weights at layer 1 head 2, with accuracy 0.576
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

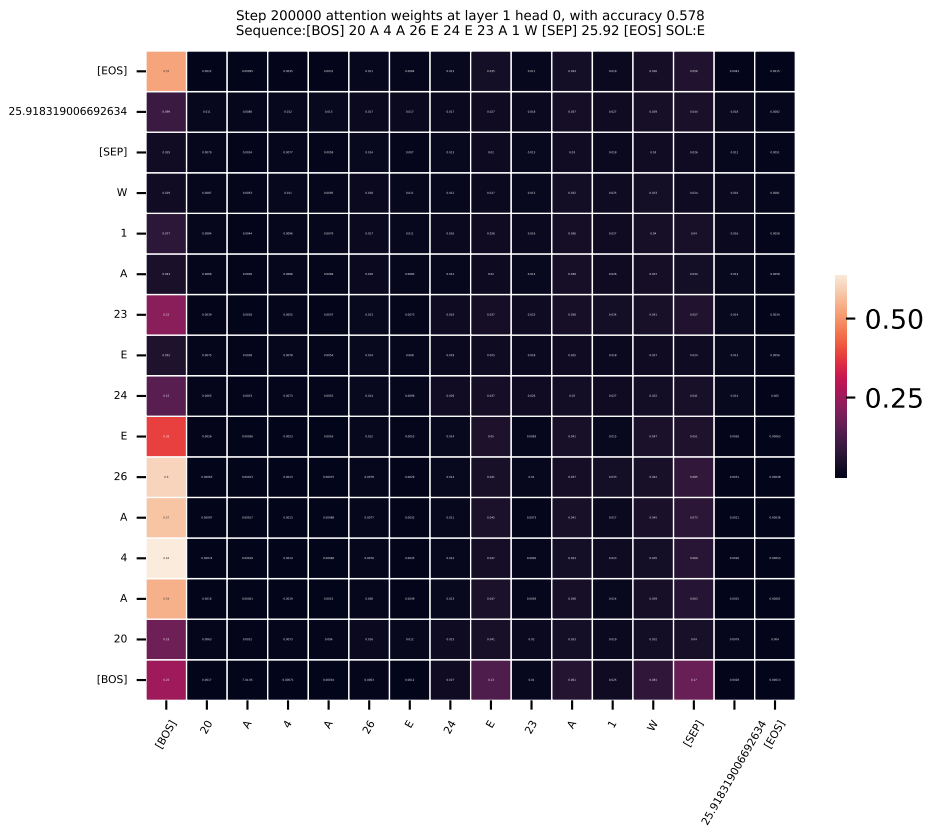

Step 199500 attention weights at layer 1 head 3, with accuracy 0.576
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
1.0
0.5
0.0

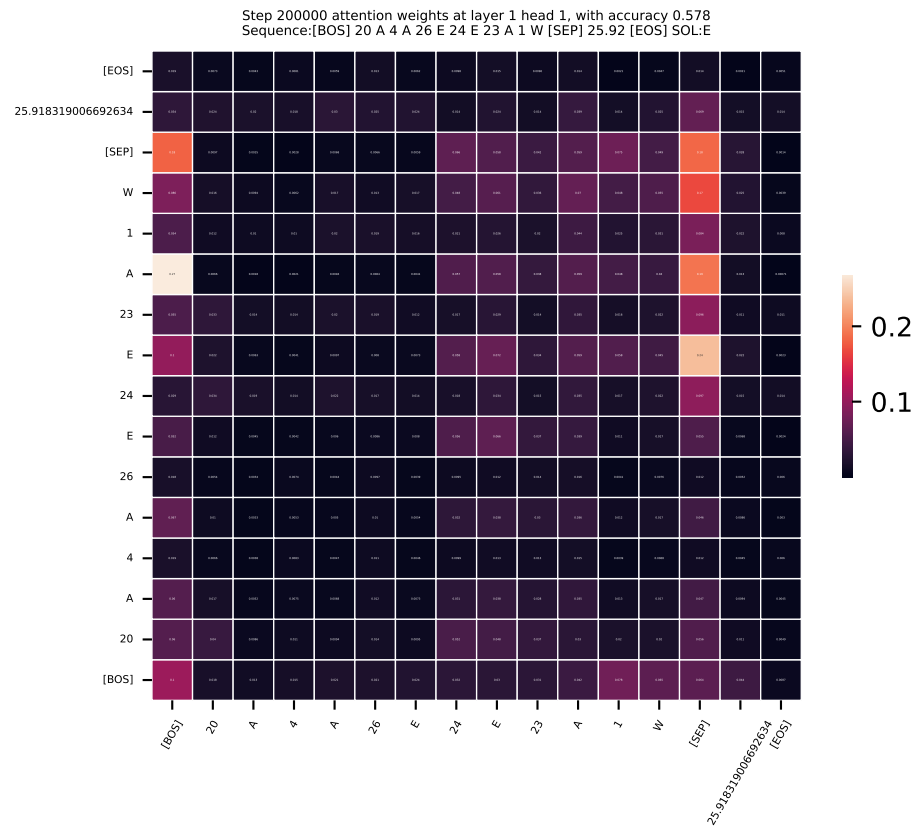

Step 200000 attention weights at layer 0 head 0, with accuracy 0.578
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

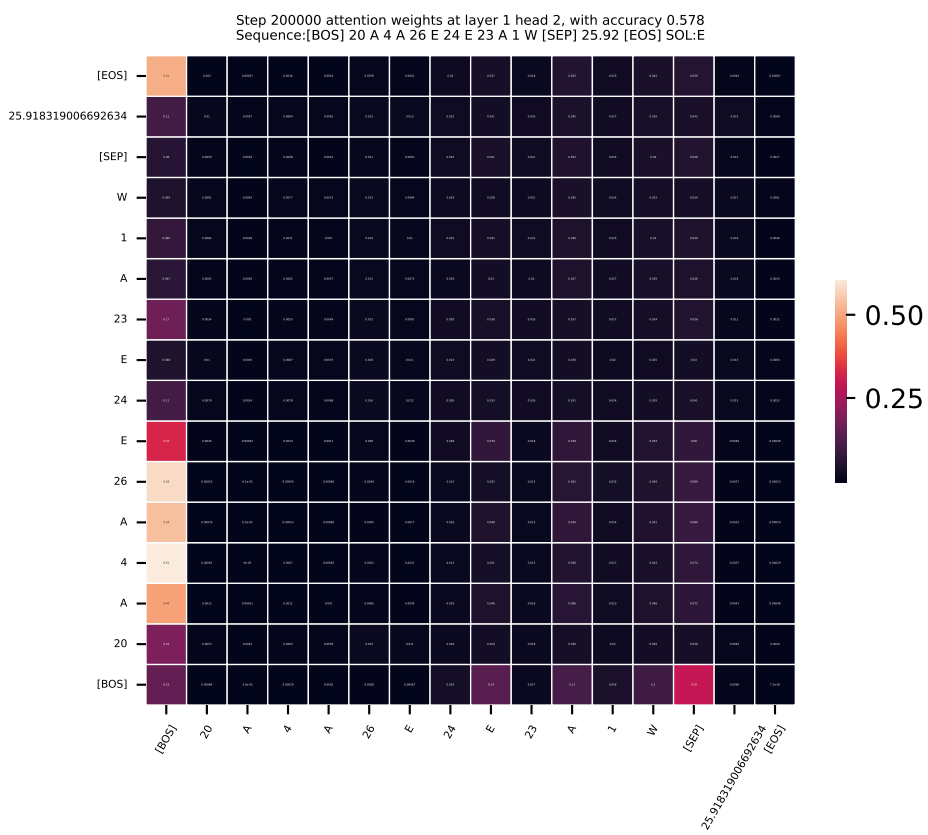

Step 200000 attention weights at layer 0 head 1, with accuracy 0.578
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

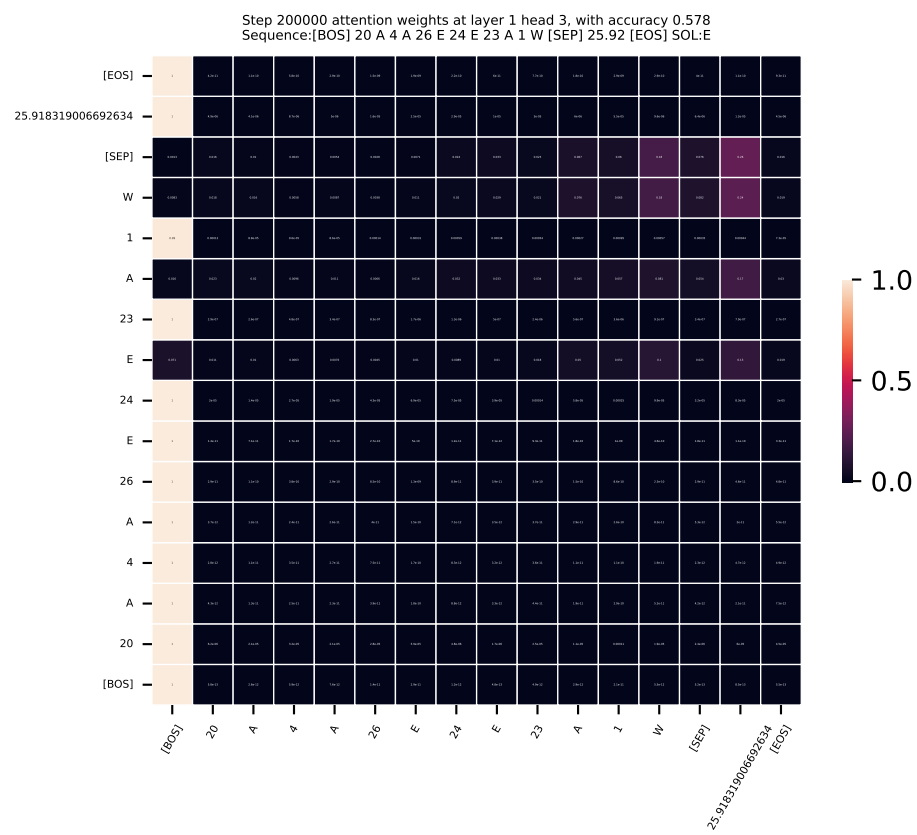

Step 200000 attention weights at layer 0 head 2, with accuracy 0.578
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

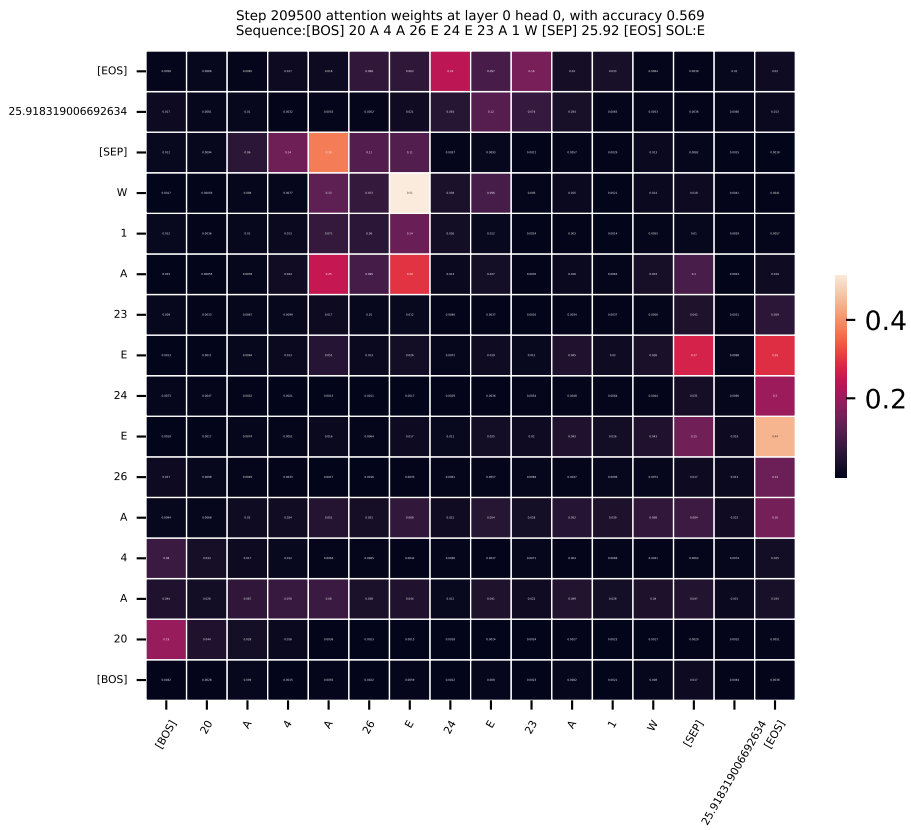

Step 200000 attention weights at layer 0 head 3, with accuracy 0.578
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

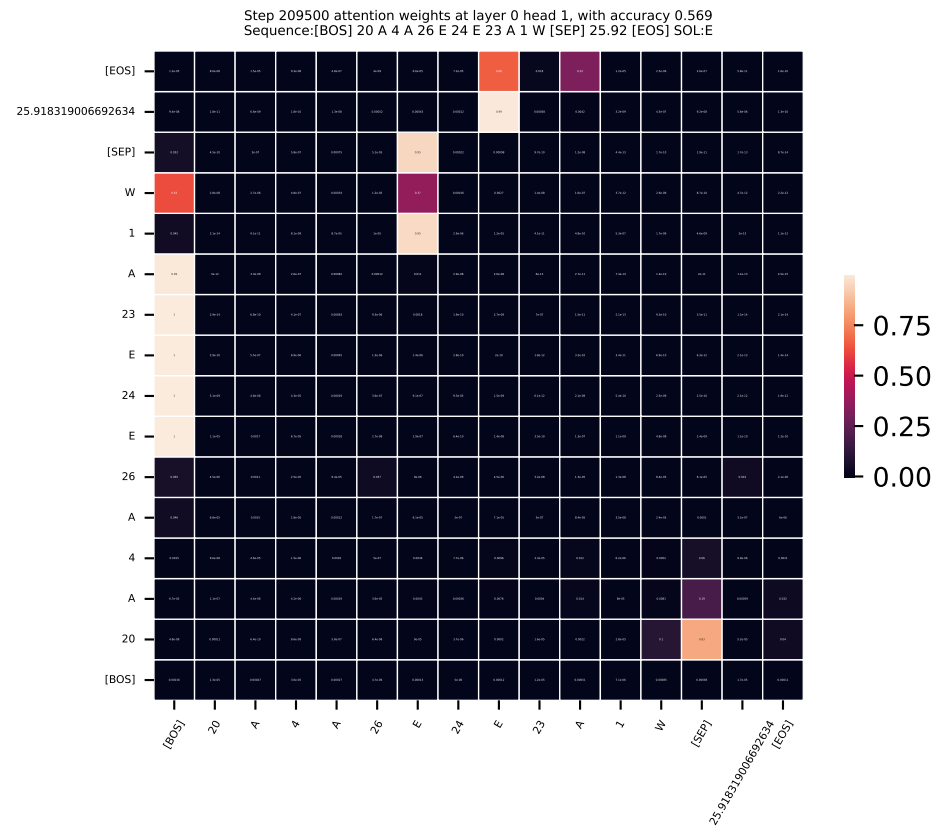

Step 200000 attention weights at layer 1 head 0, with accuracy 0.578
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

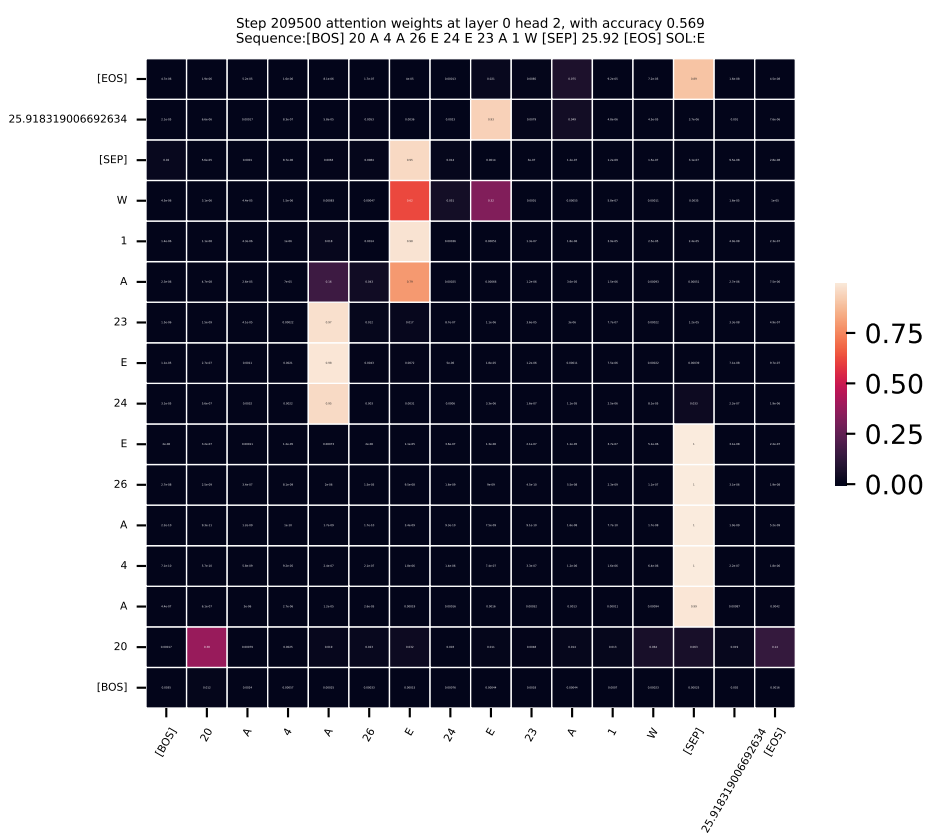

Step 200000 attention weights at layer 1 head 1, with accuracy 0.578
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.2
0.1

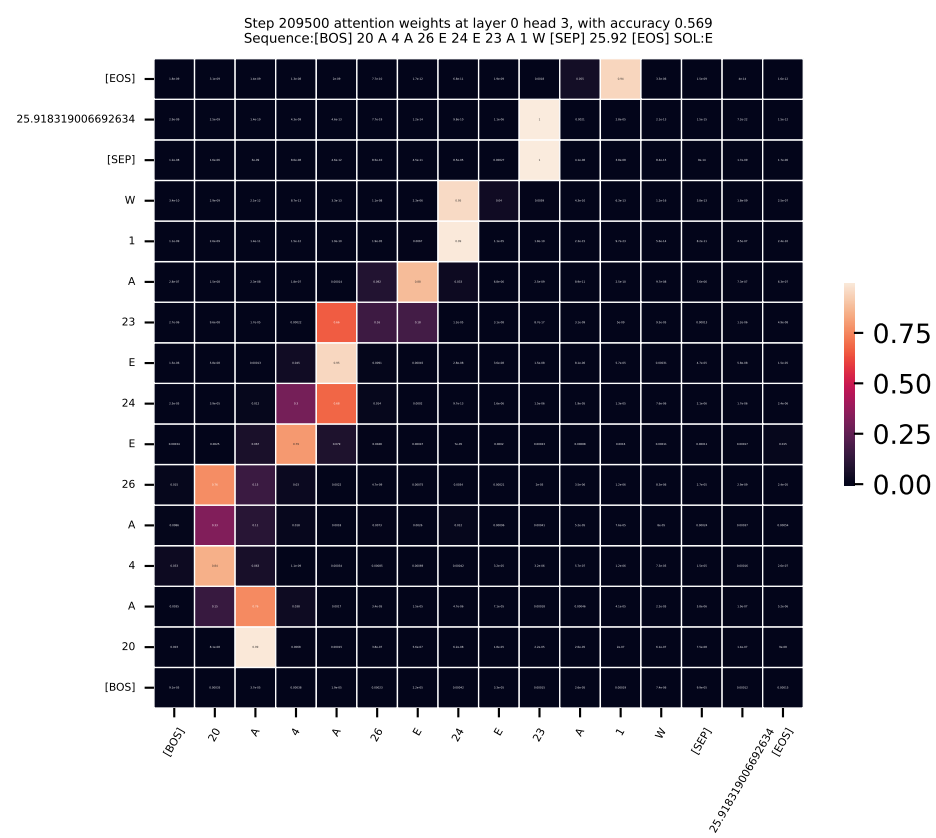

Step 200000 attention weights at layer 1 head 2, with accuracy 0.578
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

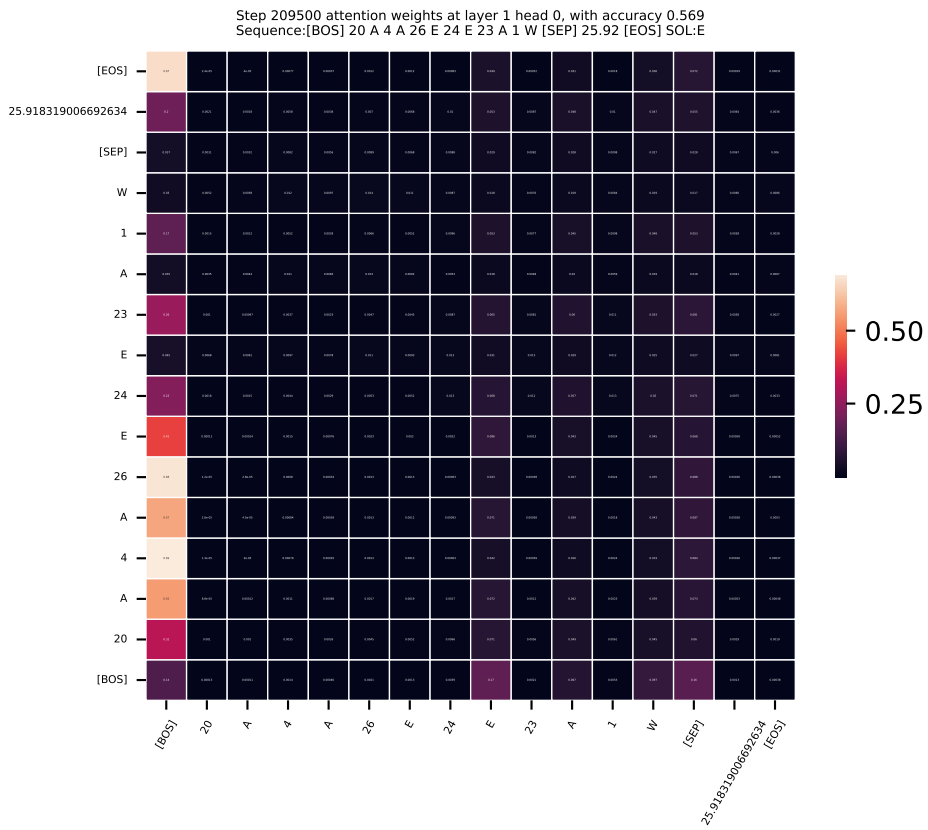

Step 200000 attention weights at layer 1 head 3, with accuracy 0.578
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
1.0
0.5
0.0

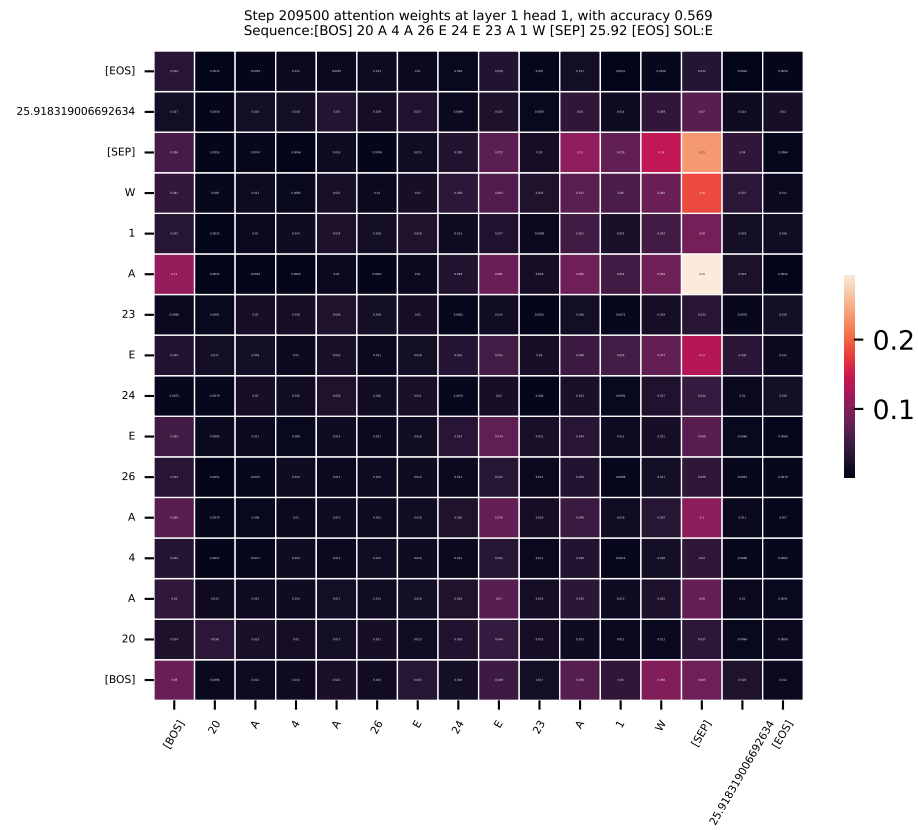

Step 209500 attention weights at layer 0 head 0, with accuracy 0.569
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.4
0.2

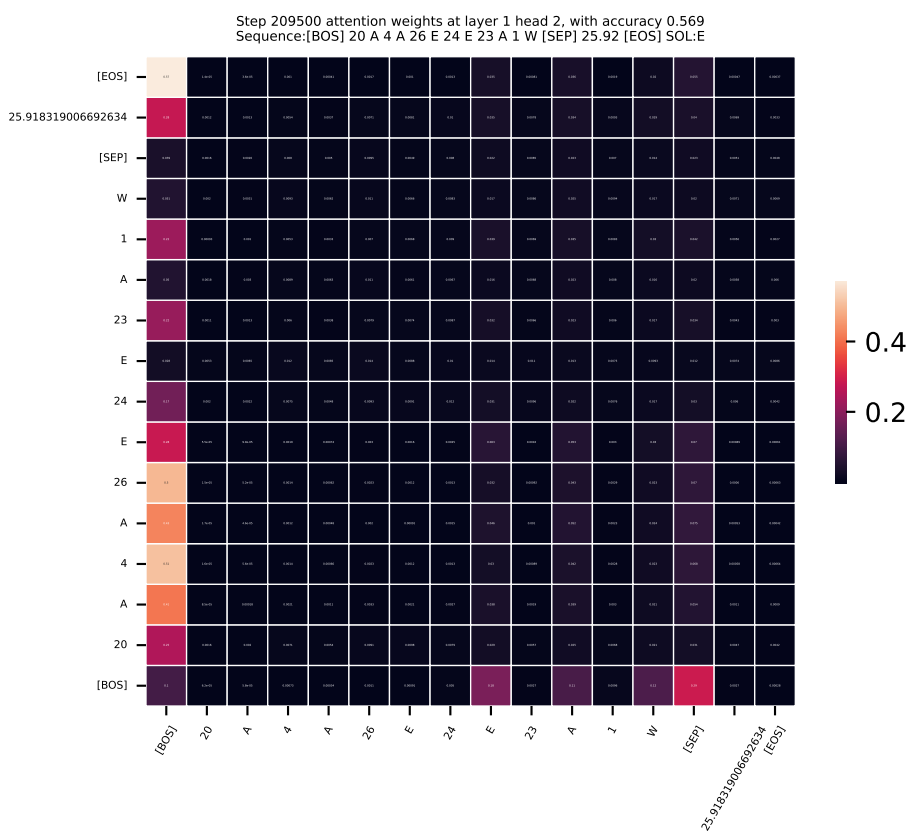

Step 209500 attention weights at layer 0 head 1, with accuracy 0.569
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

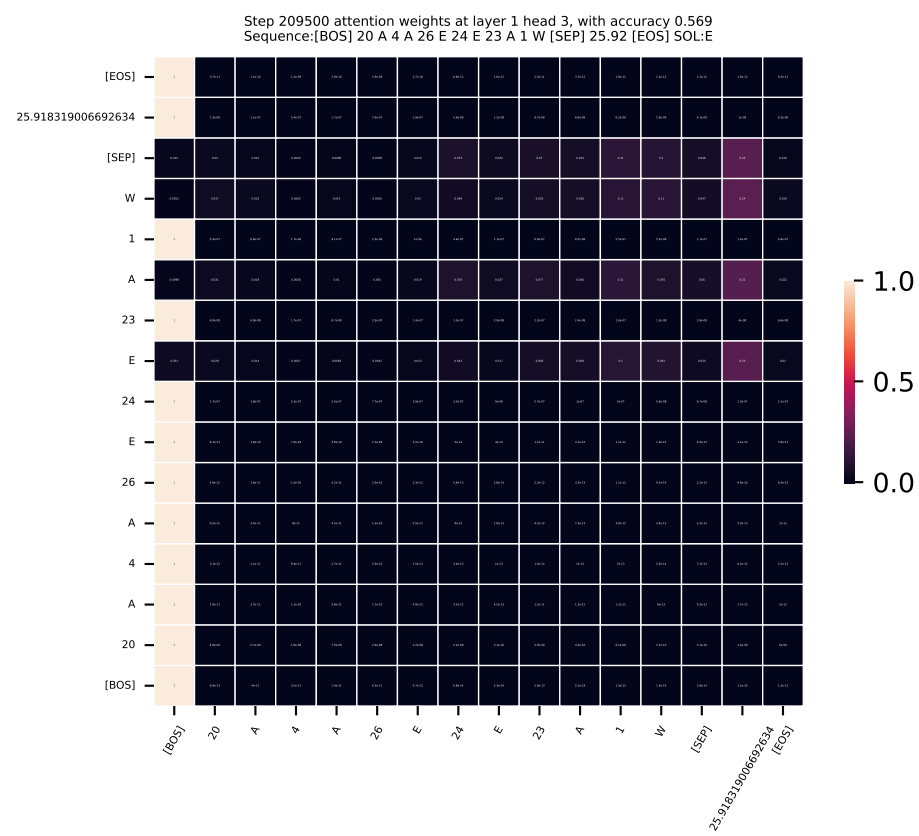

Step 209500 attention weights at layer 0 head 2, with accuracy 0.569
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

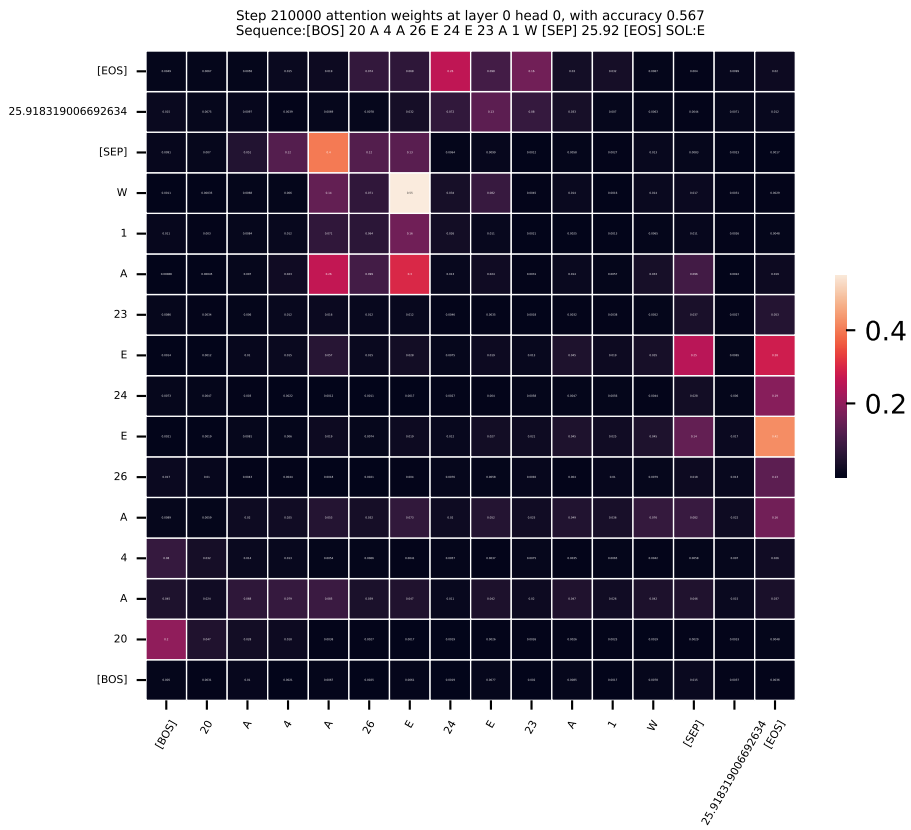

Step 209500 attention weights at layer 0 head 3, with accuracy 0.569
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

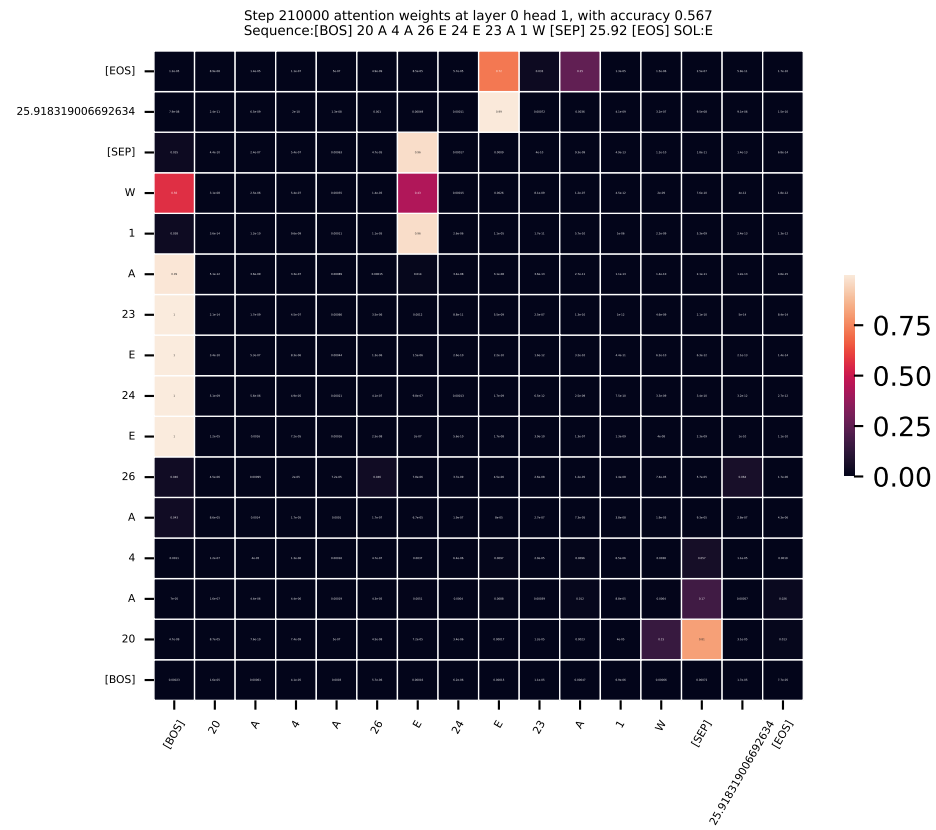

Step 209500 attention weights at layer 1 head 0, with accuracy 0.569
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

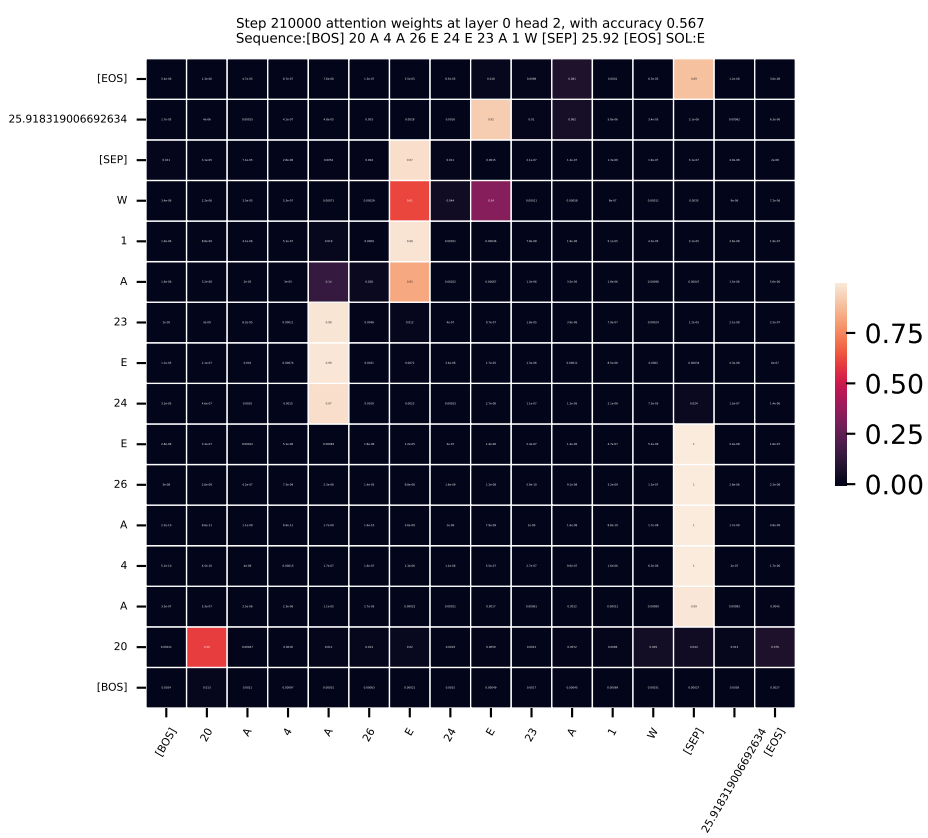

Step 209500 attention weights at layer 1 head 1, with accuracy 0.569
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.2
0.1

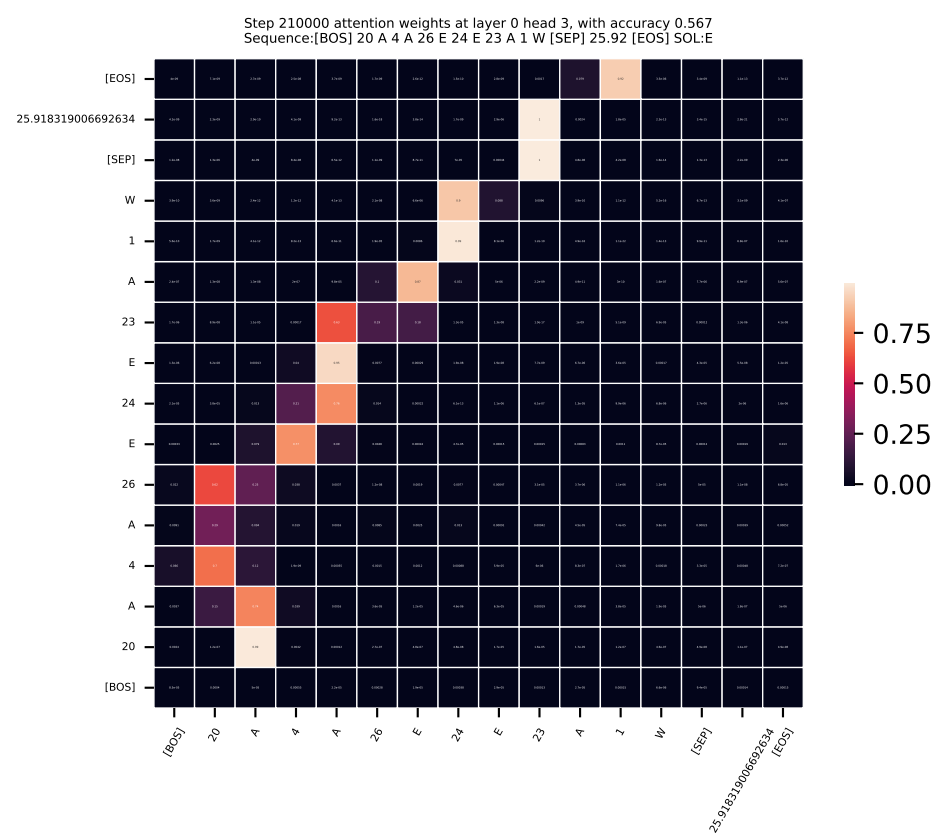

Step 209500 attention weights at layer 1 head 2, with accuracy 0.569
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

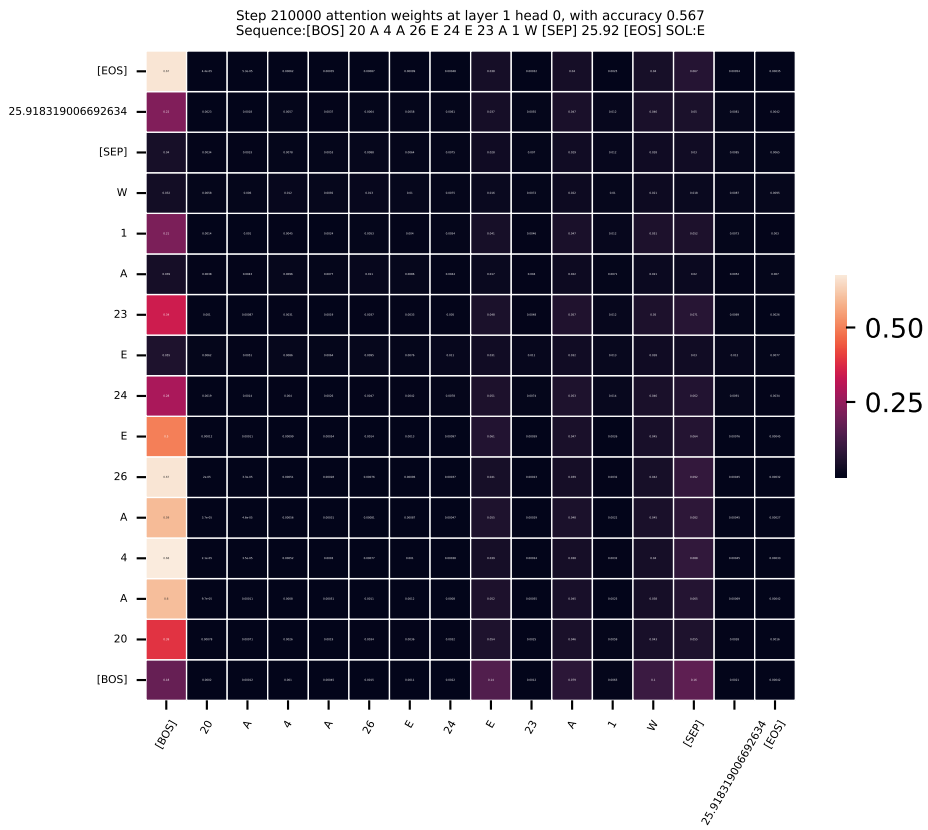

Step 209500 attention weights at layer 1 head 3, with accuracy 0.569
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
1.0
0.5
0.0

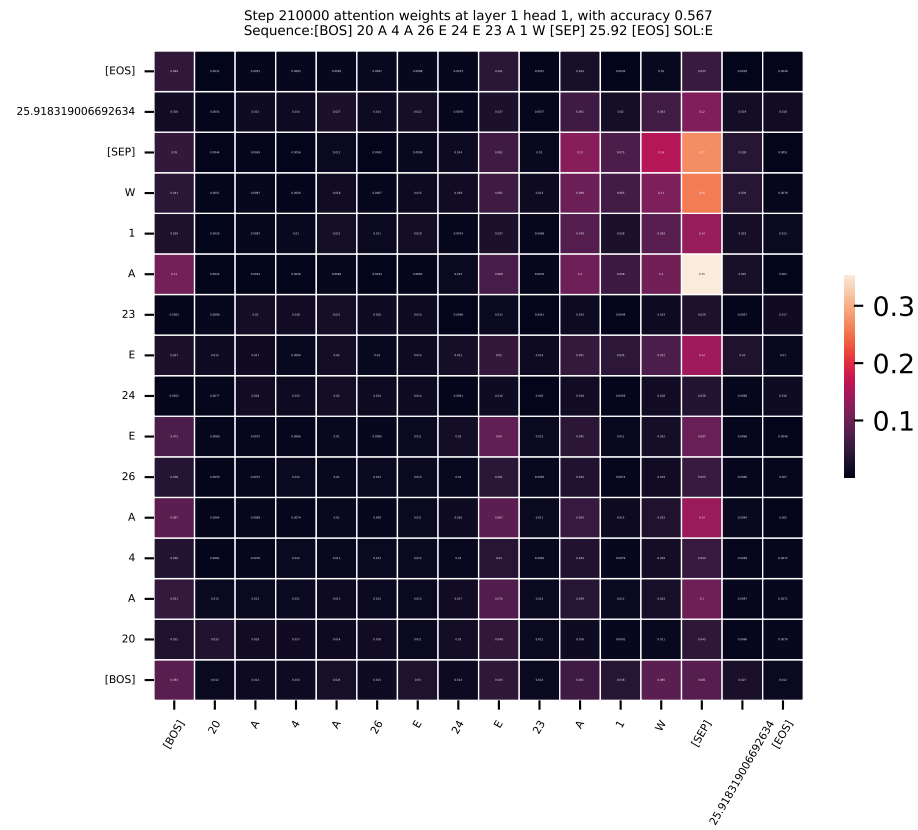

Step 210000 attention weights at layer 0 head 0, with accuracy 0.567
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

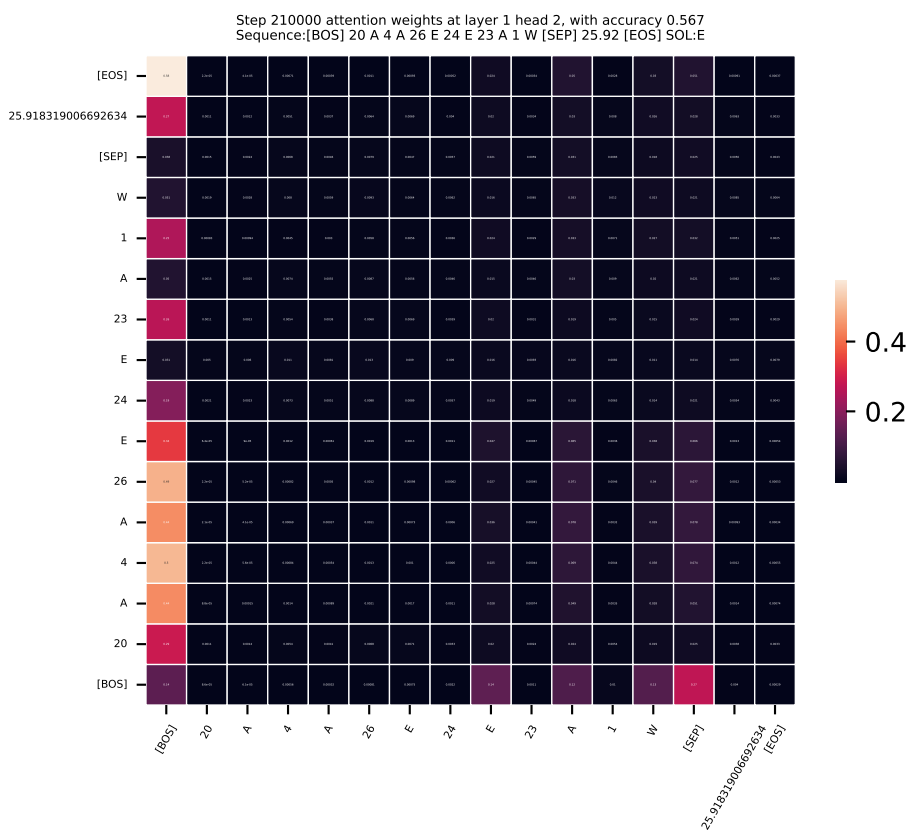

Step 210000 attention weights at layer 0 head 1, with accuracy 0.567
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

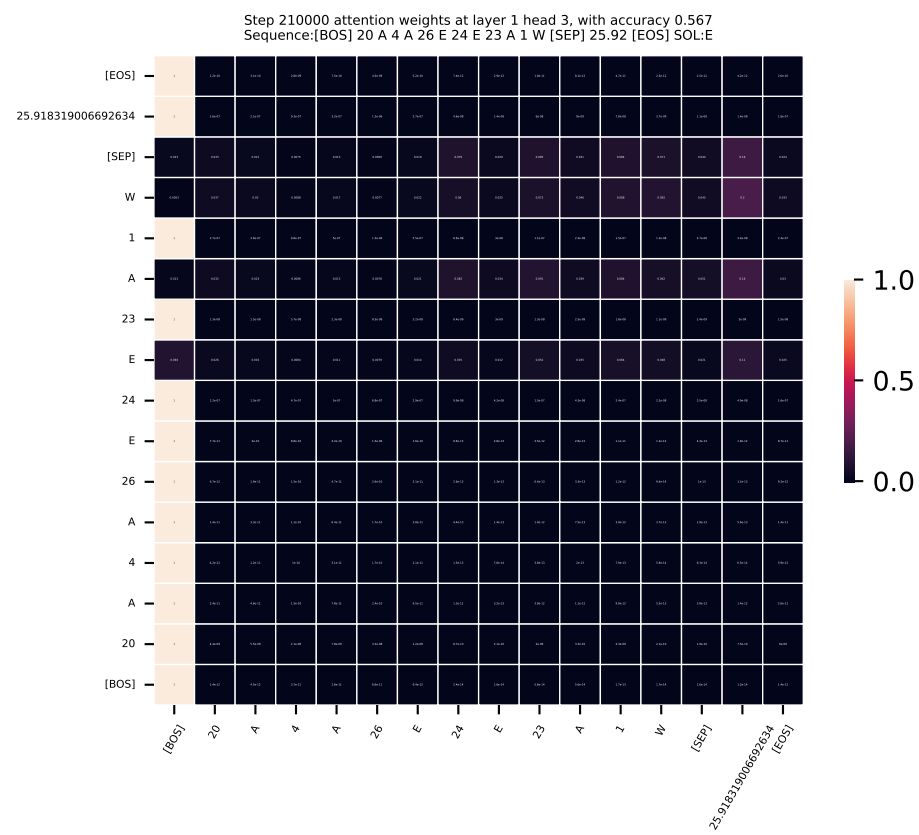

Step 210000 attention weights at layer 0 head 2, with accuracy 0.567
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

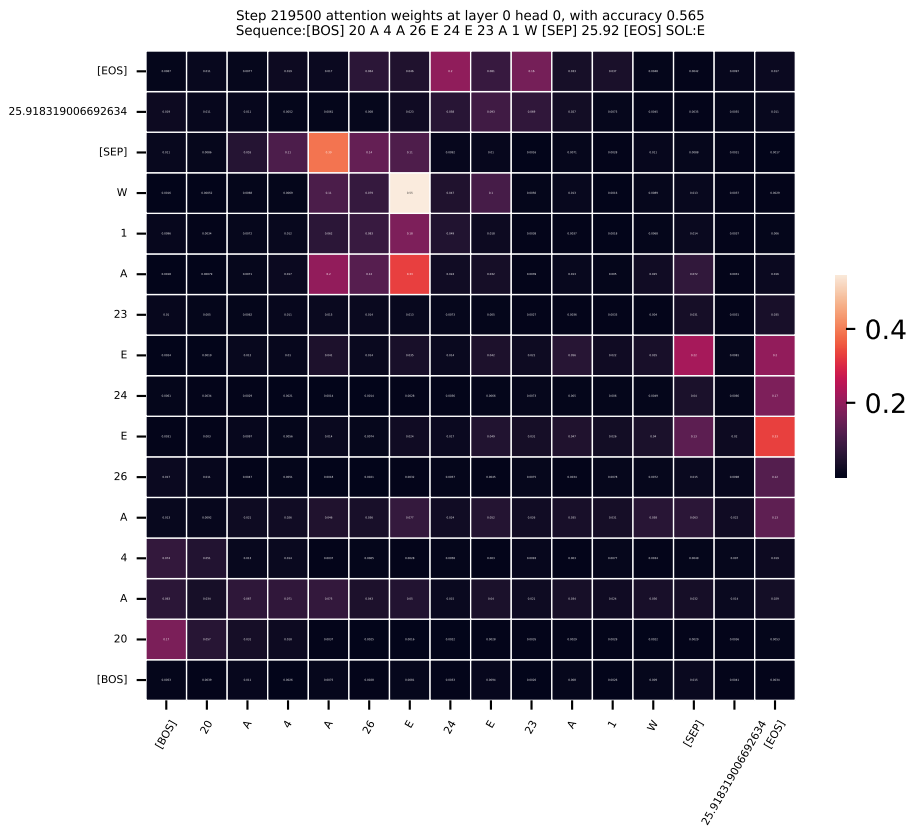

Step 210000 attention weights at layer 0 head 3, with accuracy 0.567
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

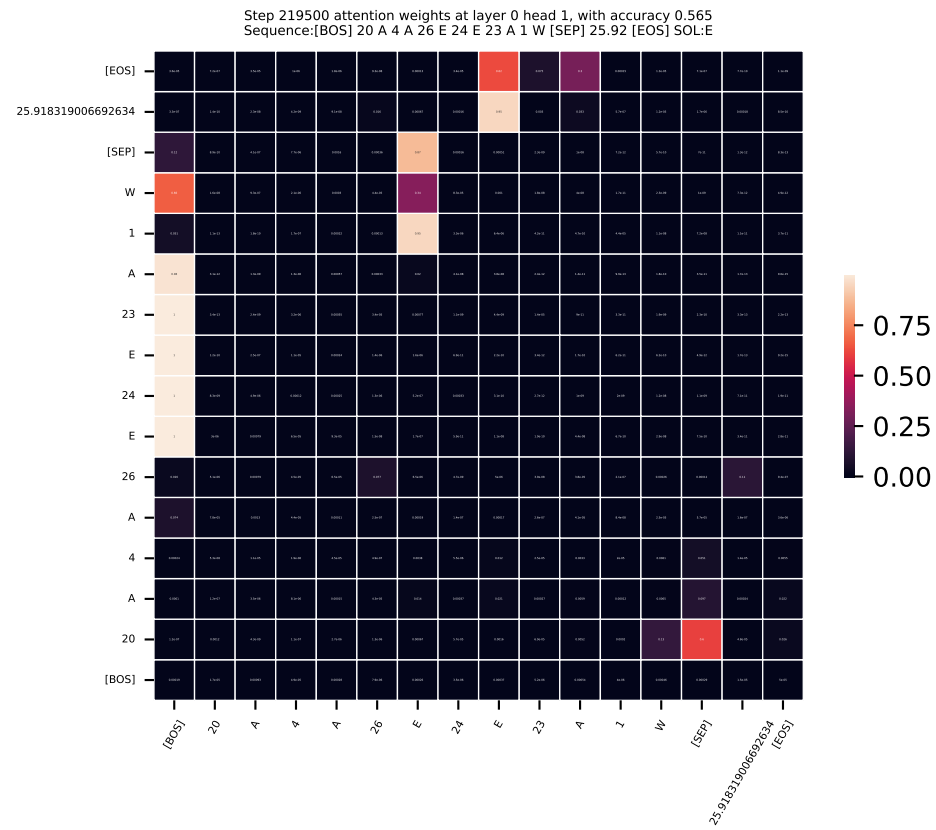

Step 210000 attention weights at layer 1 head 0, with accuracy 0.567
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

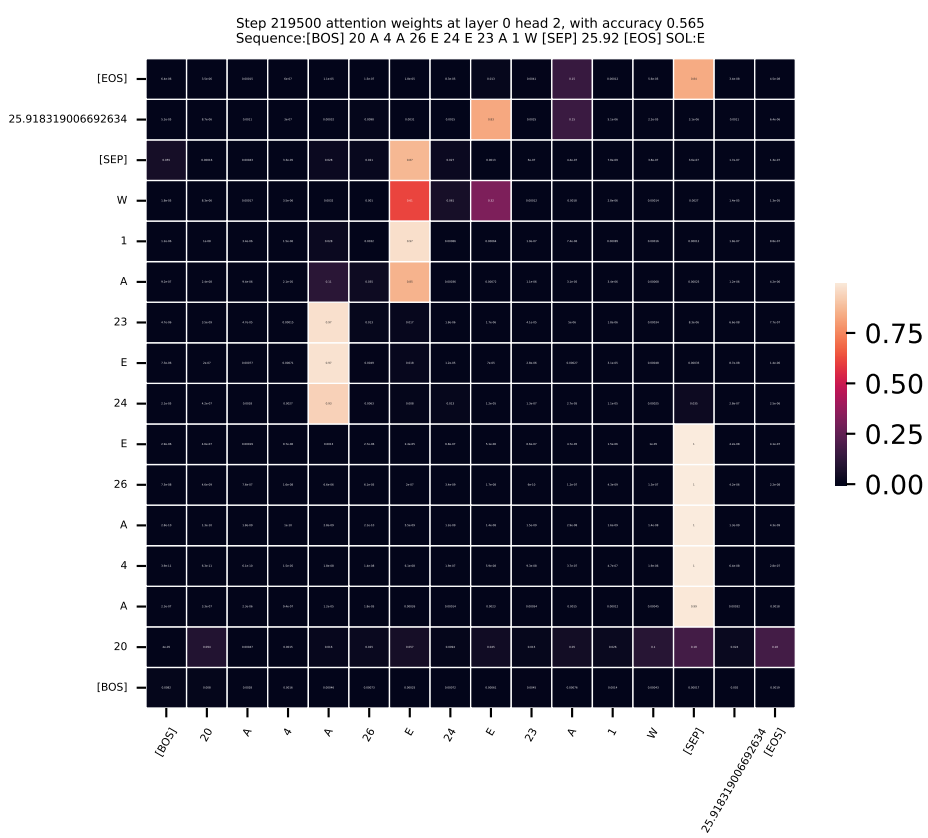

Step 210000 attention weights at layer 1 head 1, with accuracy 0.567
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.3
0.2
0.1

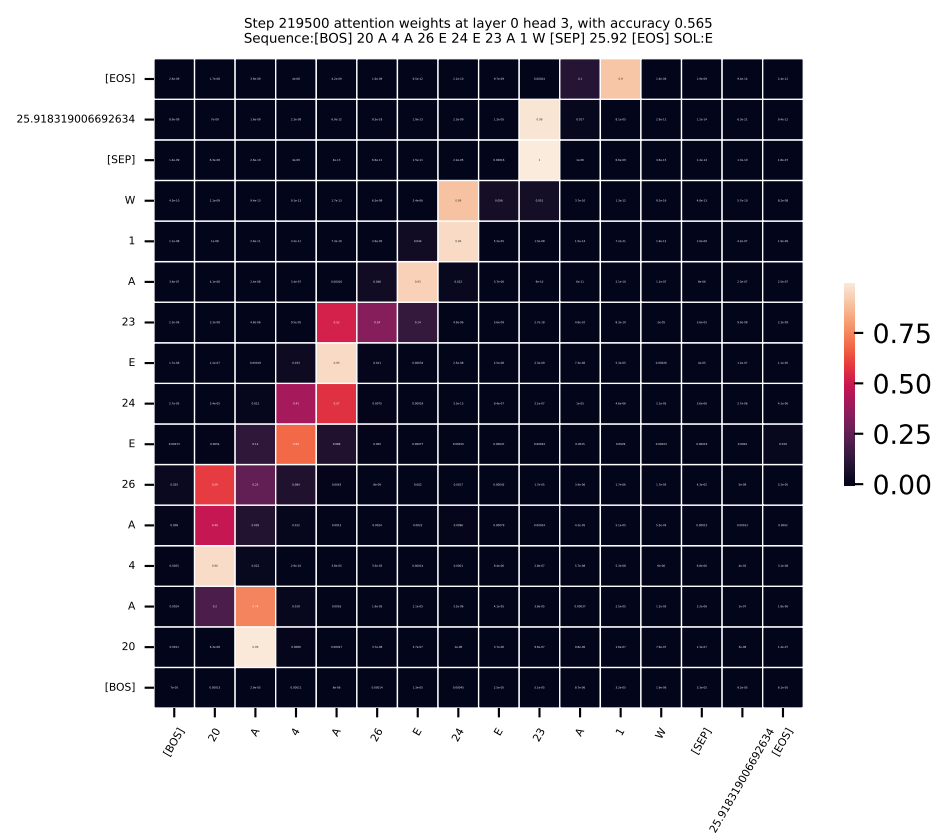

Step 210000 attention weights at layer 1 head 2, with accuracy 0.567
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

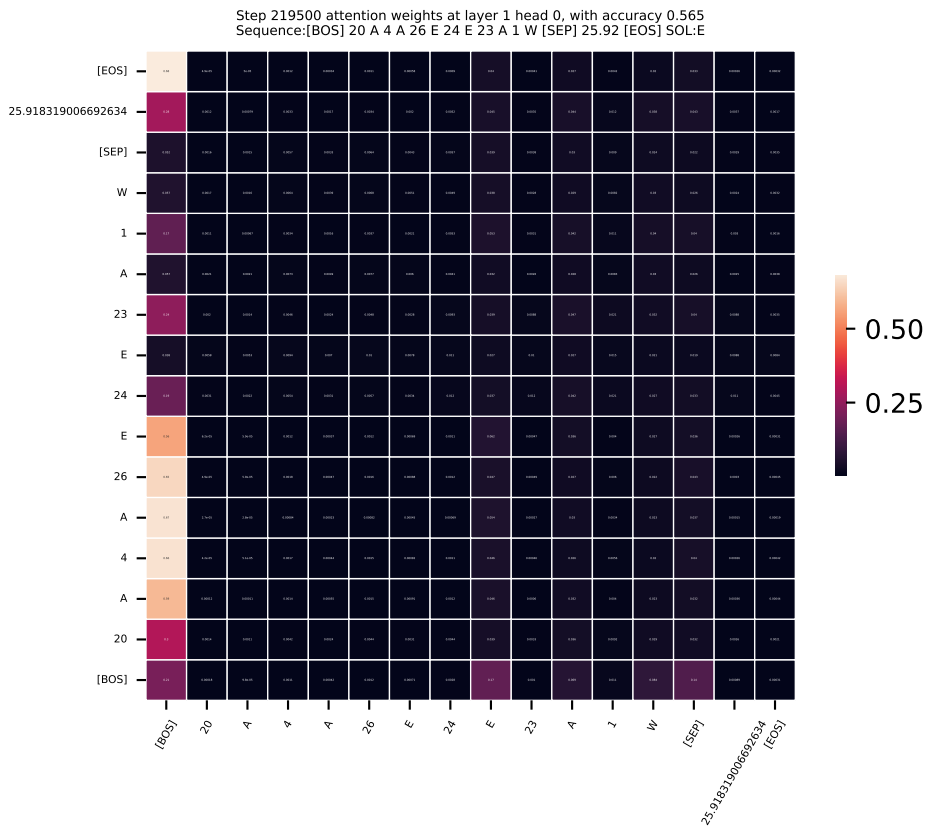

Step 210000 attention weights at layer 1 head 3, with accuracy 0.567
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
1.0
0.5
0.0

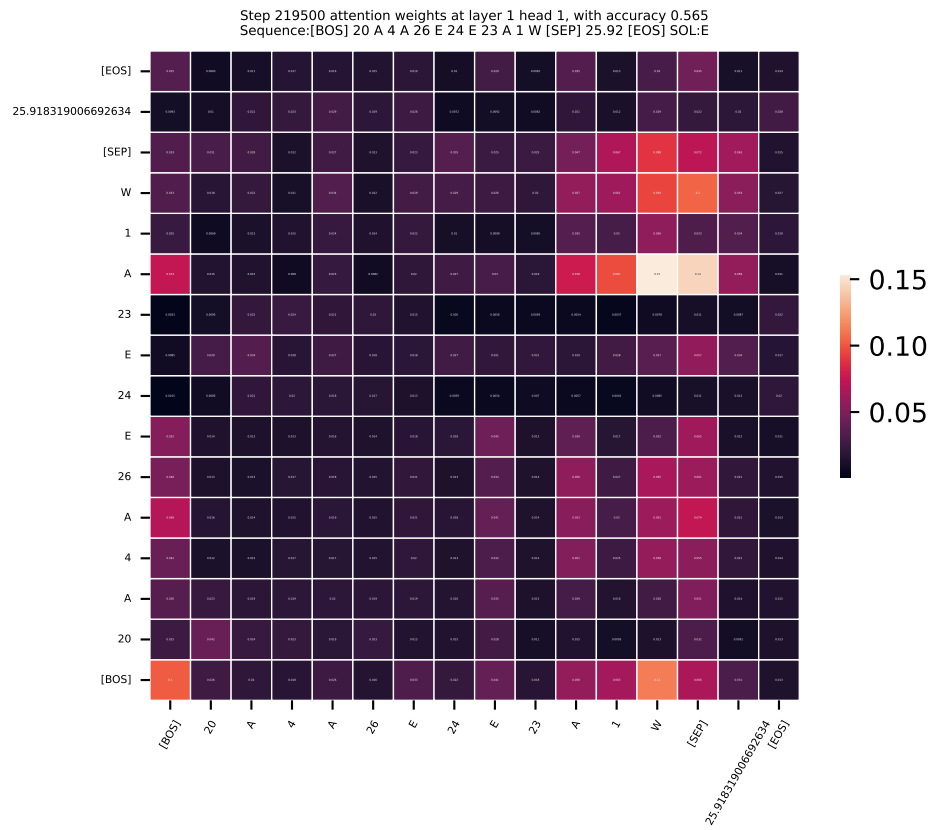

Step 219500 attention weights at layer 0 head 0, with accuracy 0.565
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

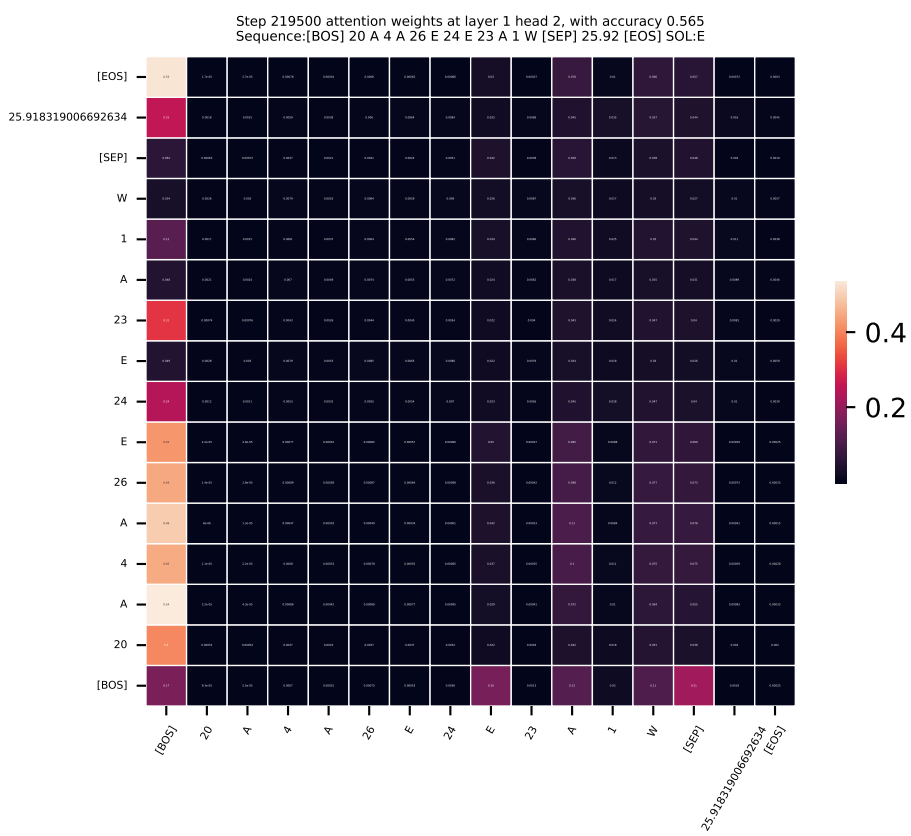

Step 219500 attention weights at layer 0 head 1, with accuracy 0.565
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

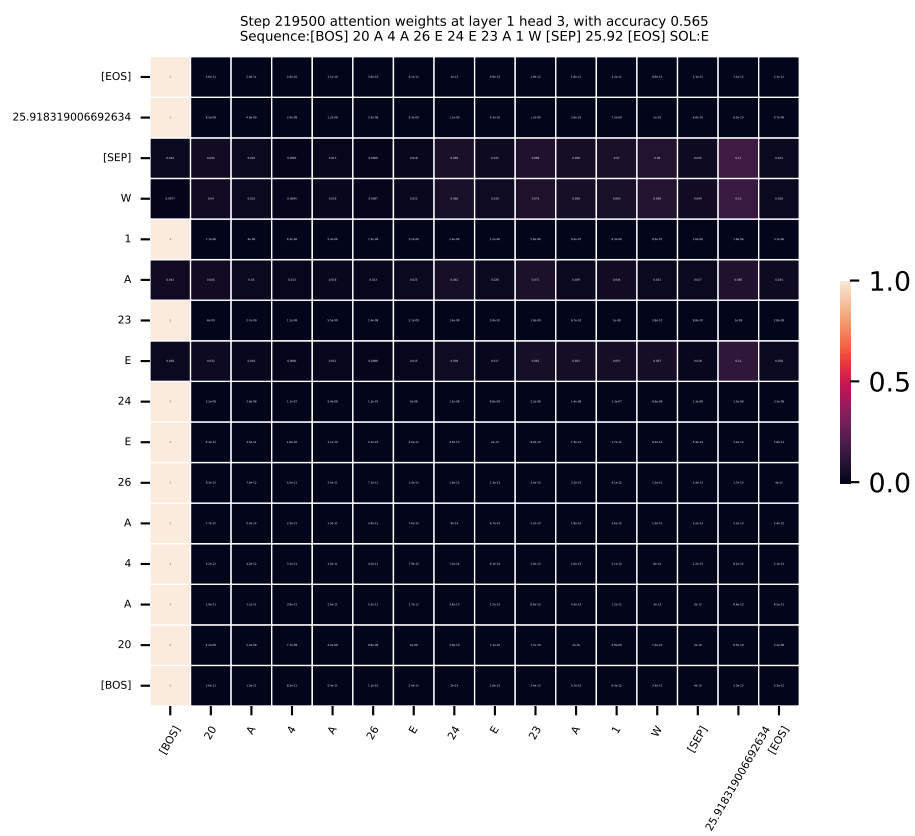

Step 219500 attention weights at layer 0 head 2, with accuracy 0.565
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

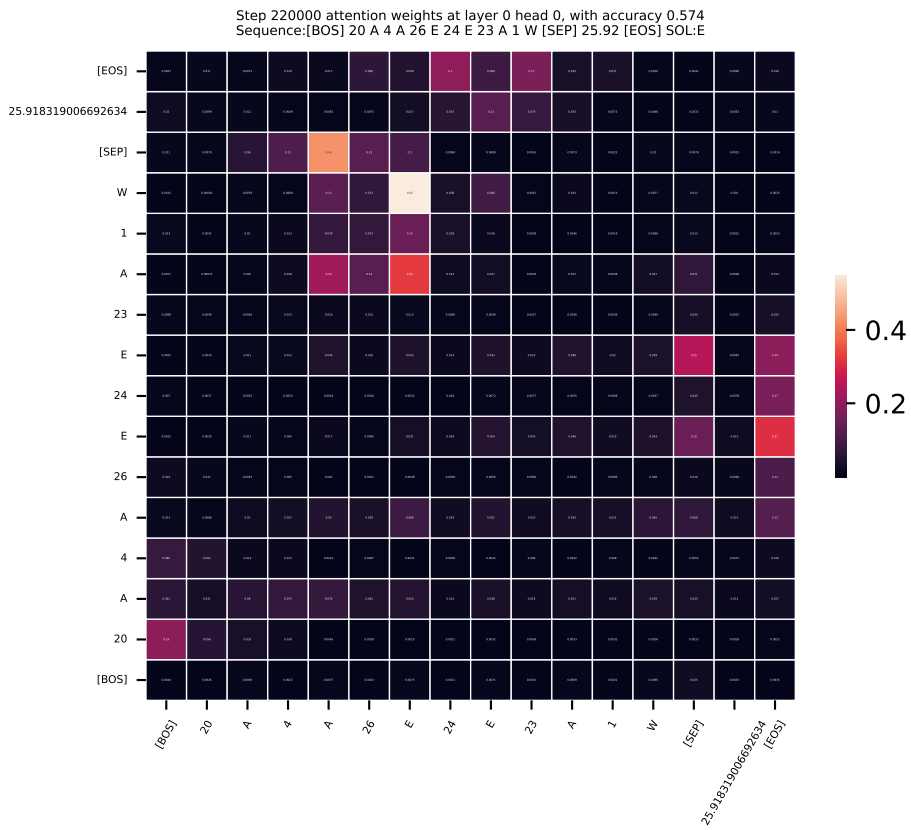

Step 219500 attention weights at layer 0 head 3, with accuracy 0.565
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

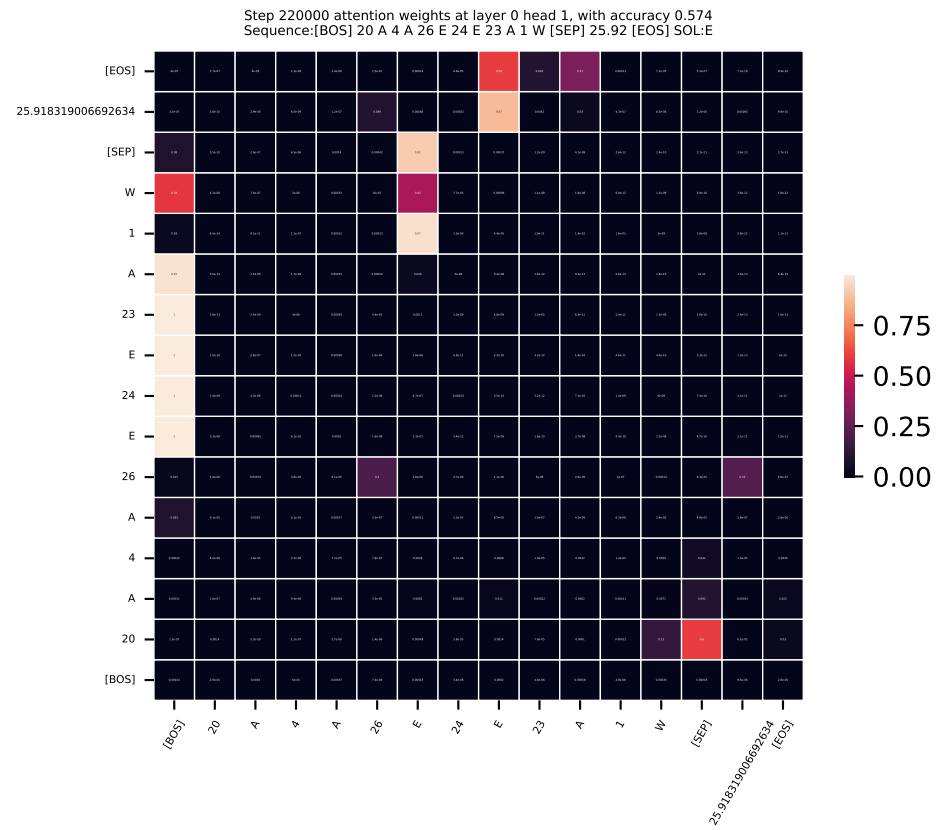

Step 219500 attention weights at layer 1 head 0, with accuracy 0.565
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

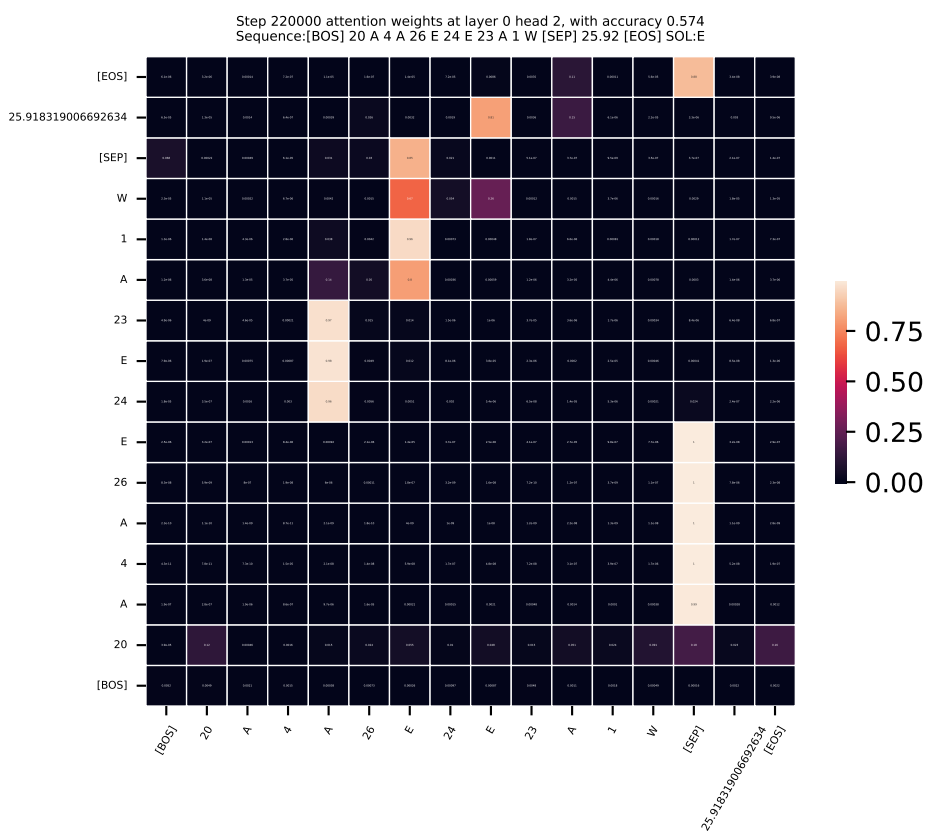

Step 219500 attention weights at layer 1 head 1, with accuracy 0.565
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.15
0.10
0.05

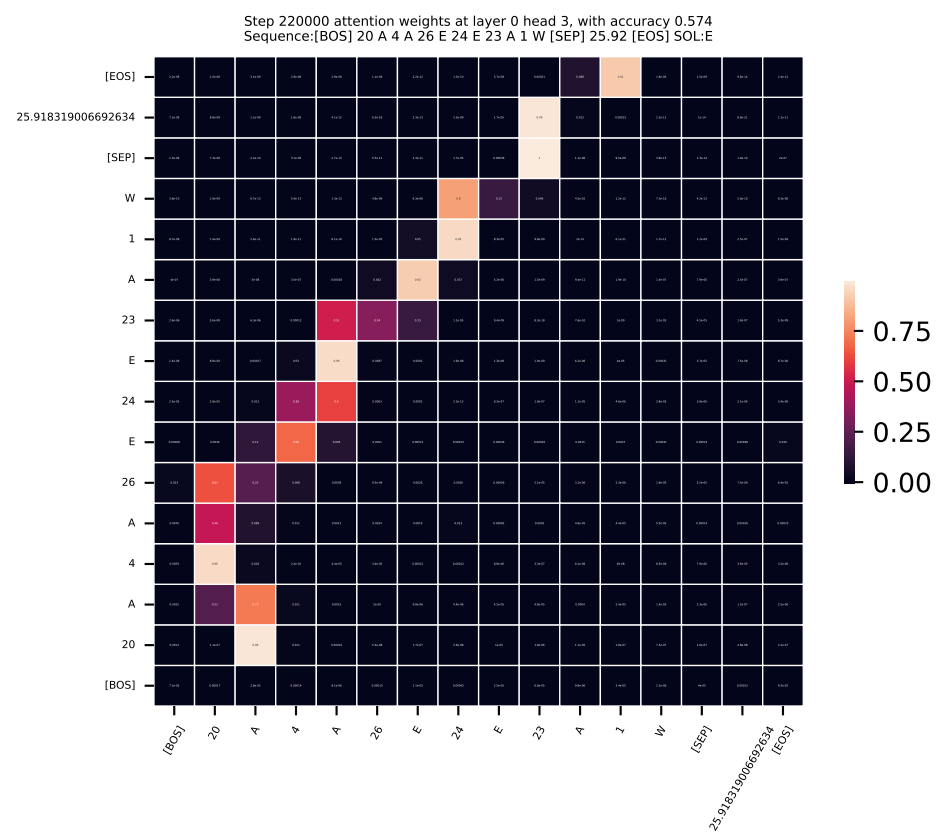

Step 219500 attention weights at layer 1 head 2, with accuracy 0.565
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

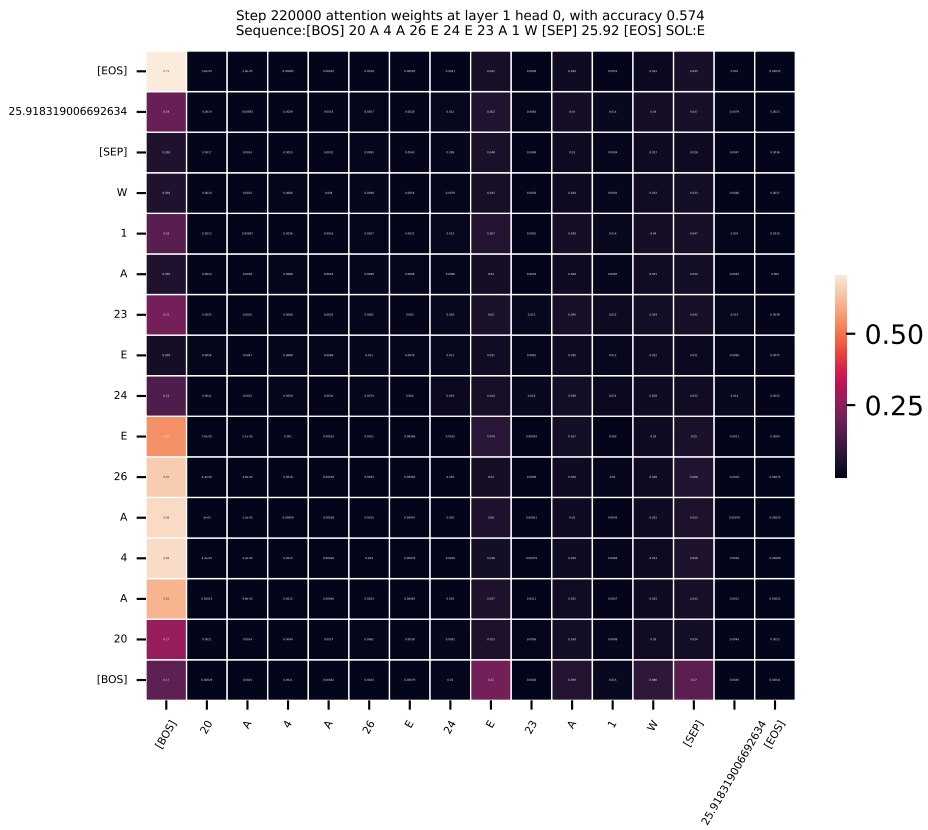

Step 219500 attention weights at layer 1 head 3, with accuracy 0.565
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
1.0
0.5
0.0

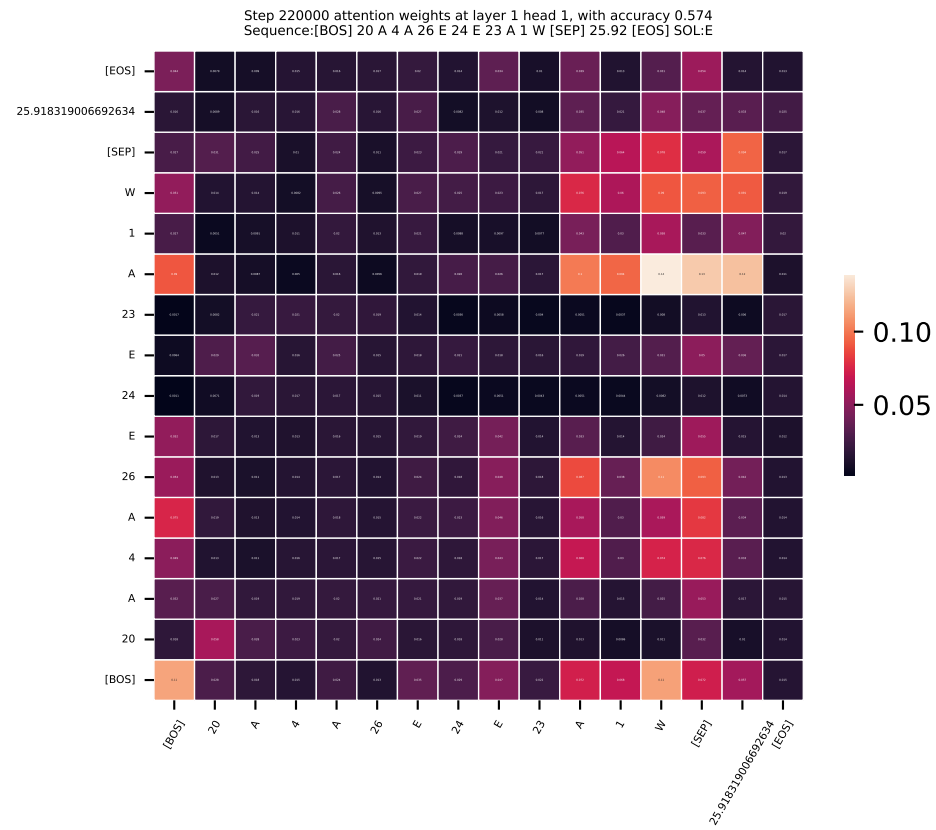

Step 220000 attention weights at layer 0 head 0, with accuracy 0.574
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

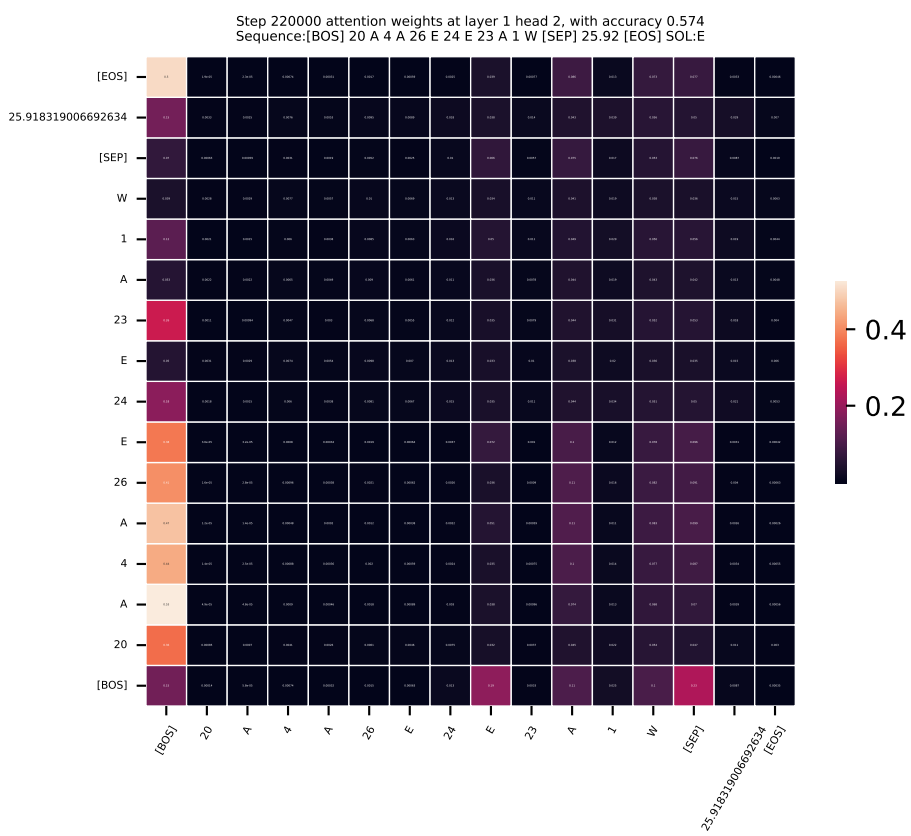

Step 220000 attention weights at layer 0 head 1, with accuracy 0.574
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

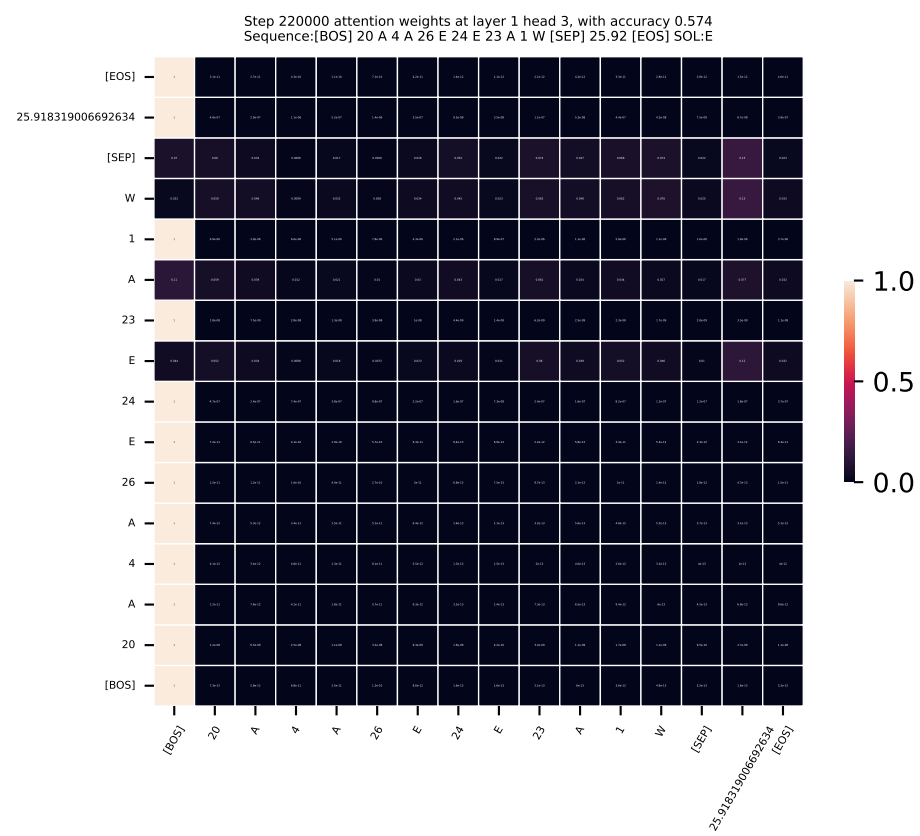

Step 220000 attention weights at layer 0 head 2, with accuracy 0.574
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

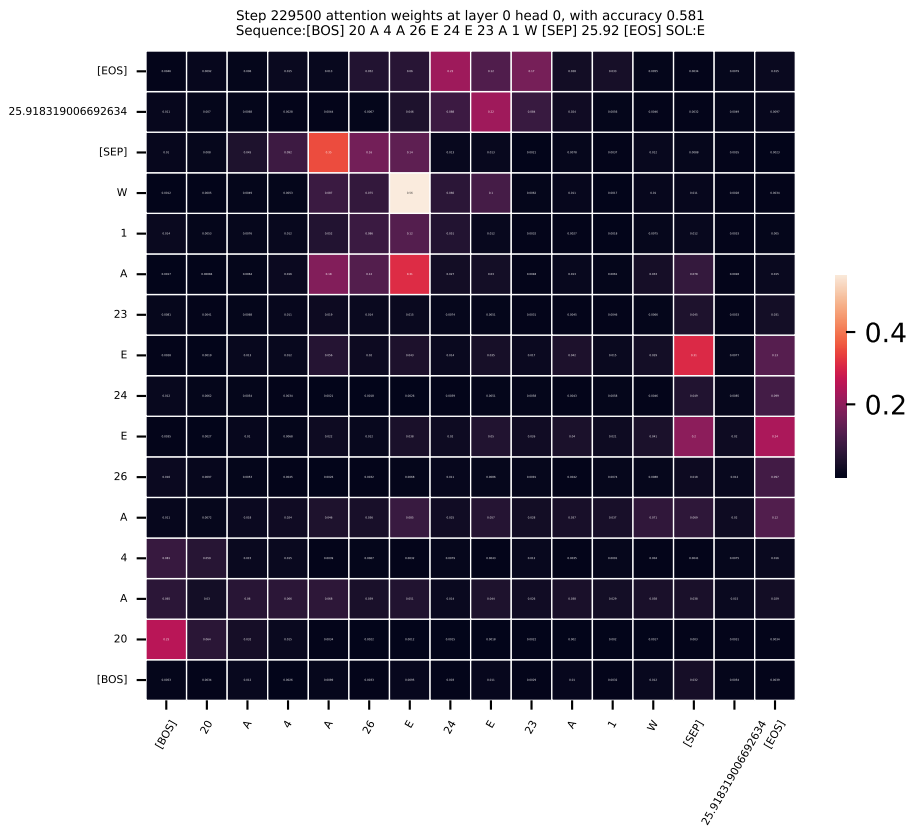

Step 220000 attention weights at layer 0 head 3, with accuracy 0.574
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

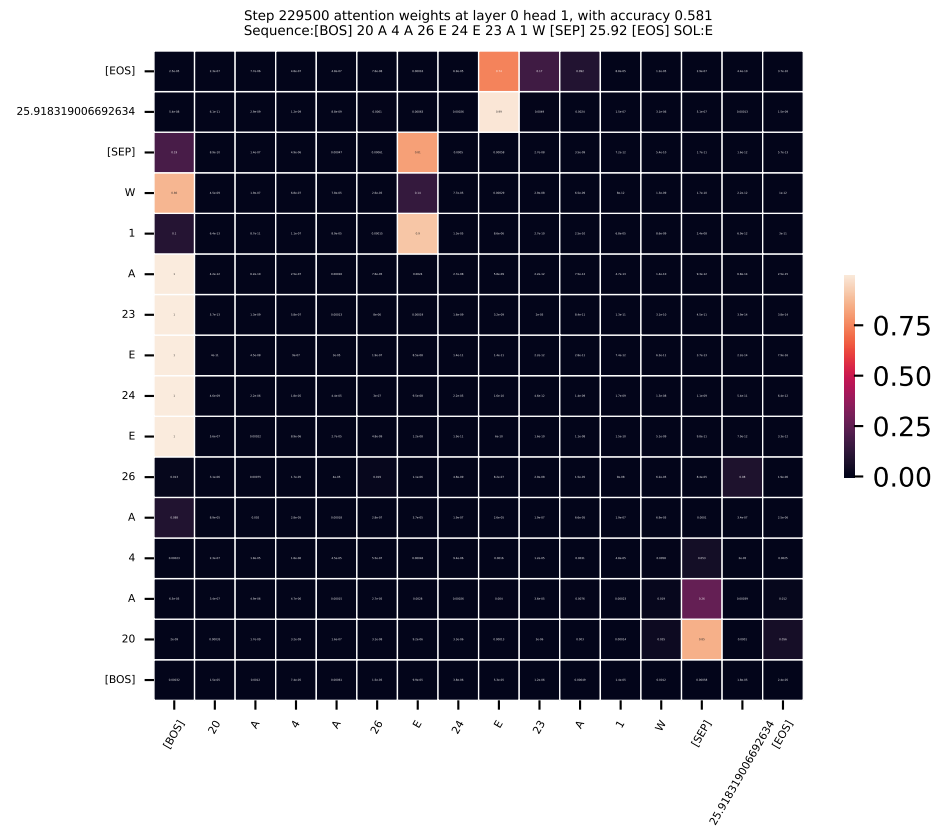

Step 220000 attention weights at layer 1 head 0, with accuracy 0.574
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

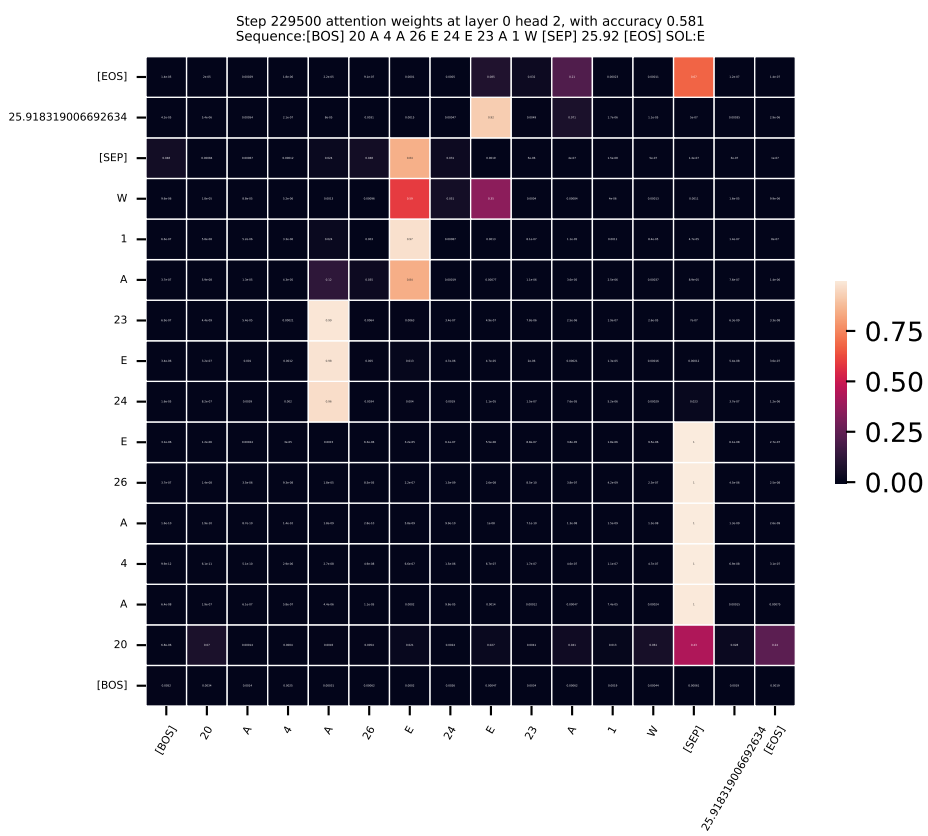

Step 220000 attention weights at layer 1 head 1, with accuracy 0.574
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.10
0.05

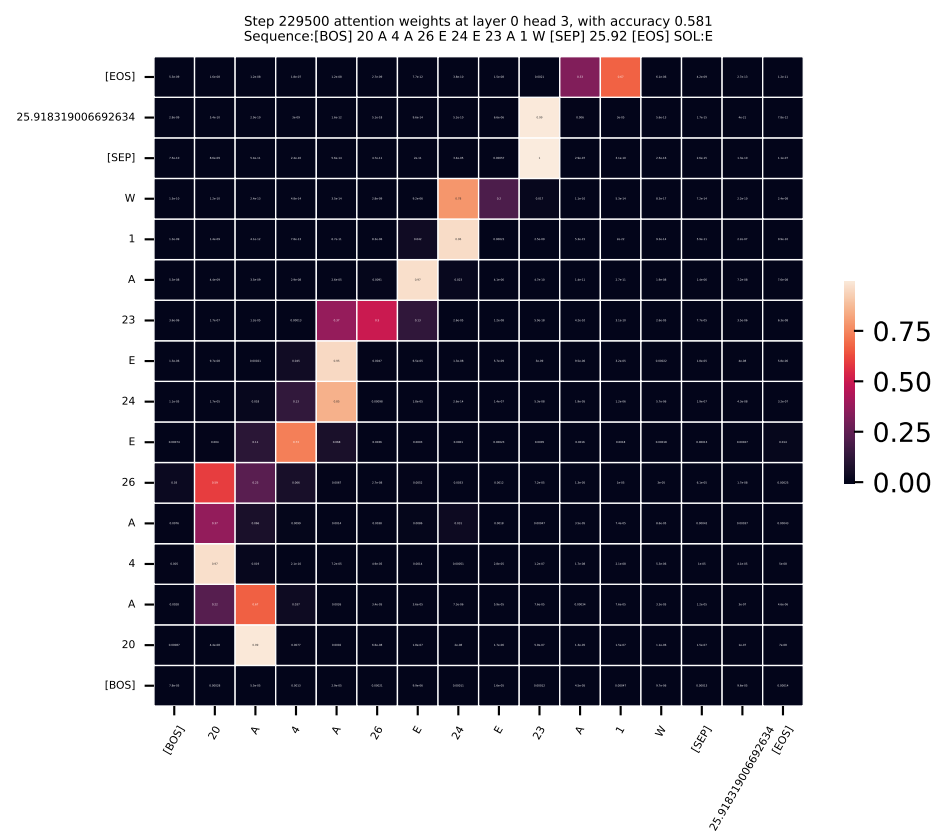

Step 220000 attention weights at layer 1 head 2, with accuracy 0.574
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

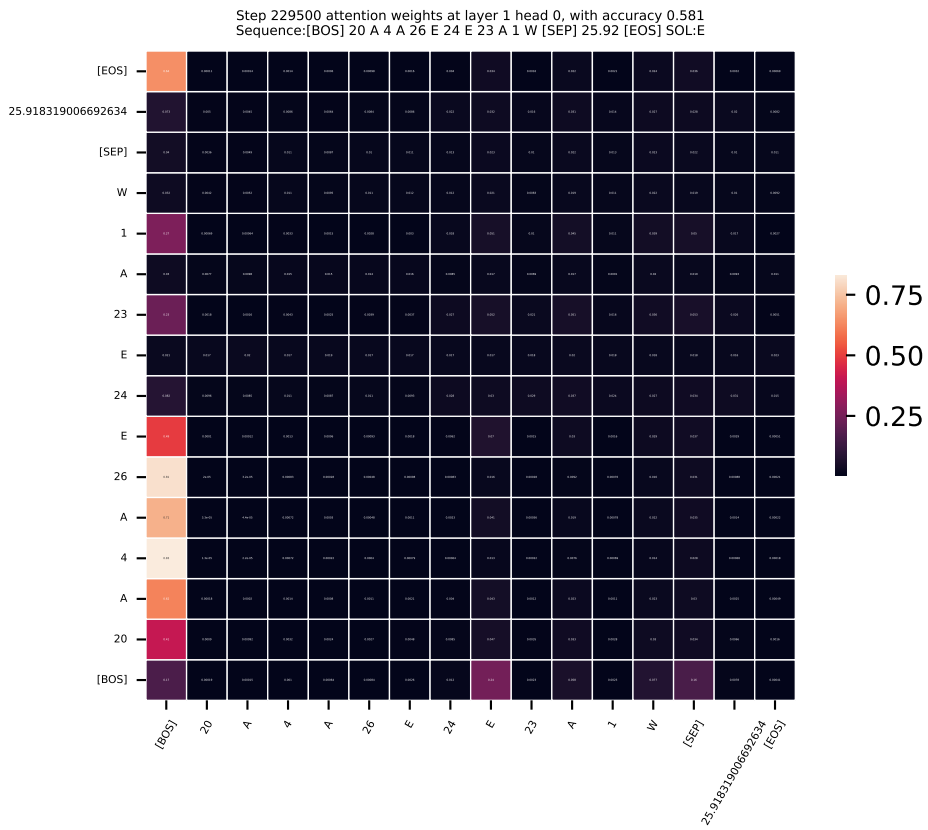

Step 220000 attention weights at layer 1 head 3, with accuracy 0.574
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
1.0
0.5
0.0

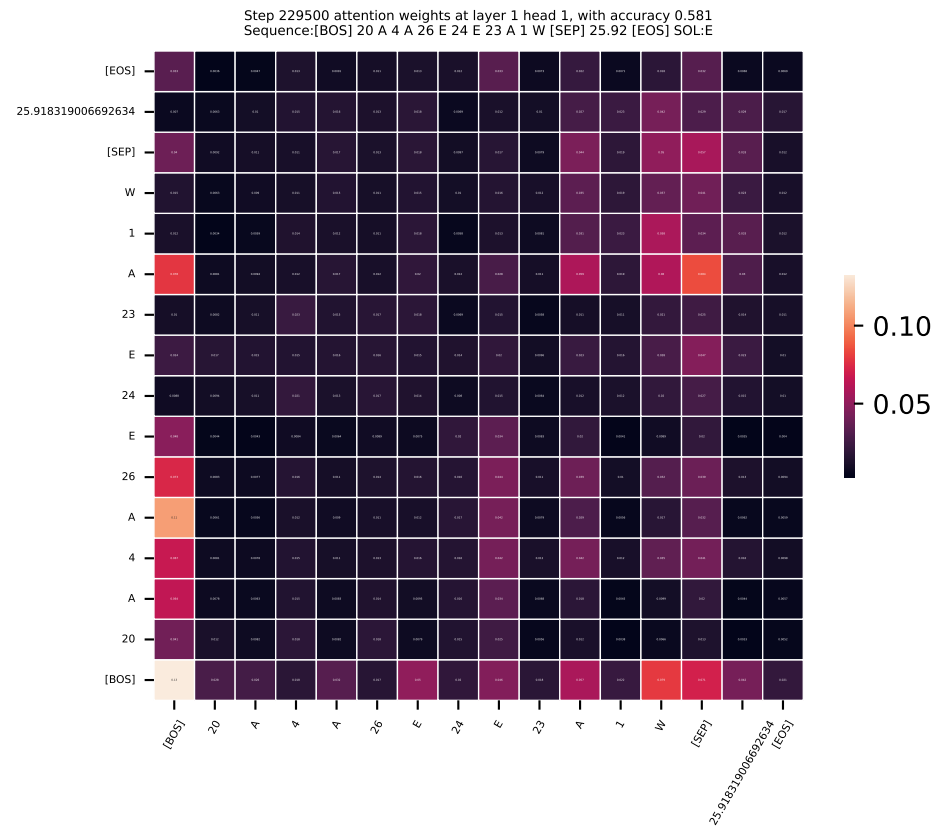

Step 229500 attention weights at layer 0 head 0, with accuracy 0.581
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

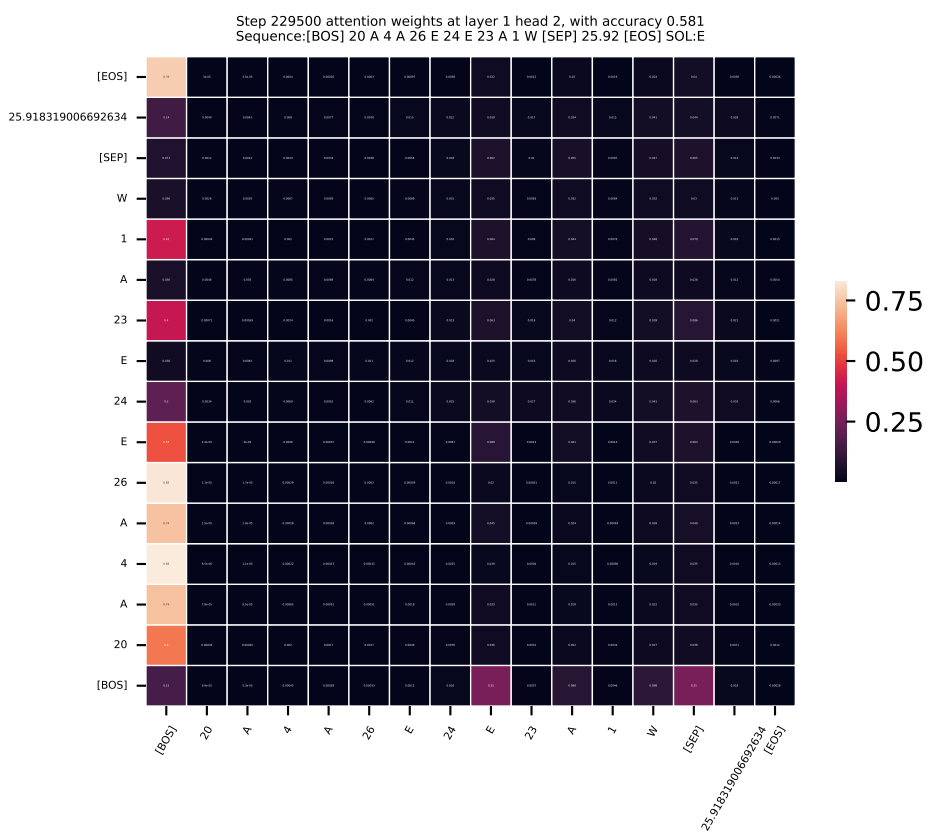

Step 229500 attention weights at layer 0 head 1, with accuracy 0.581
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

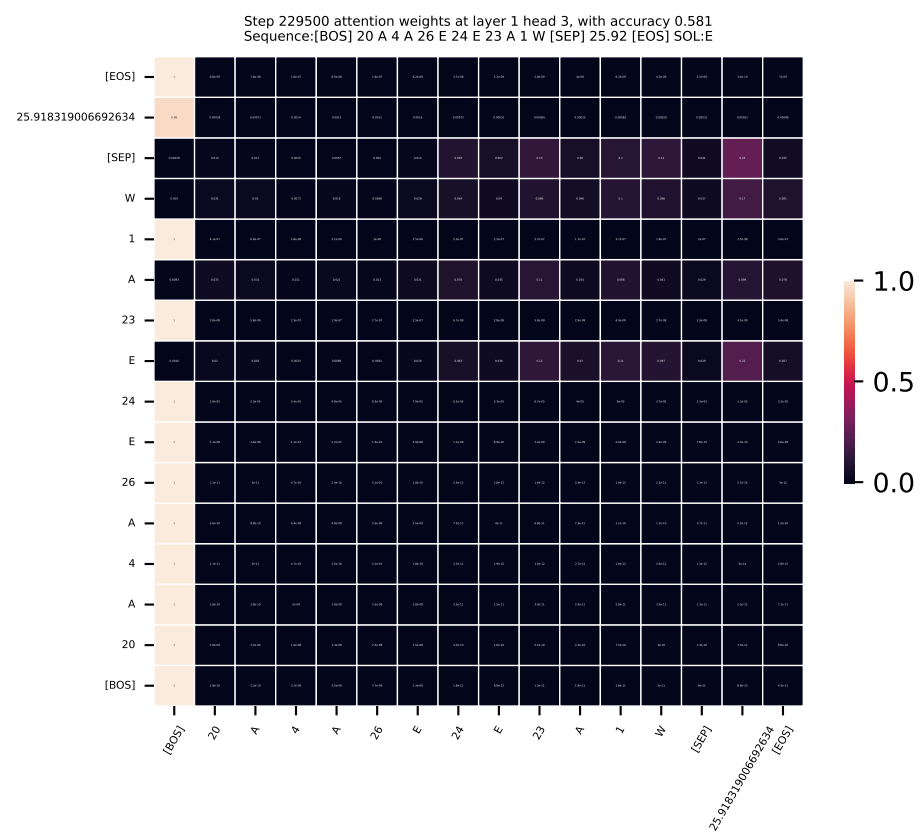

Step 229500 attention weights at layer 0 head 2, with accuracy 0.581
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

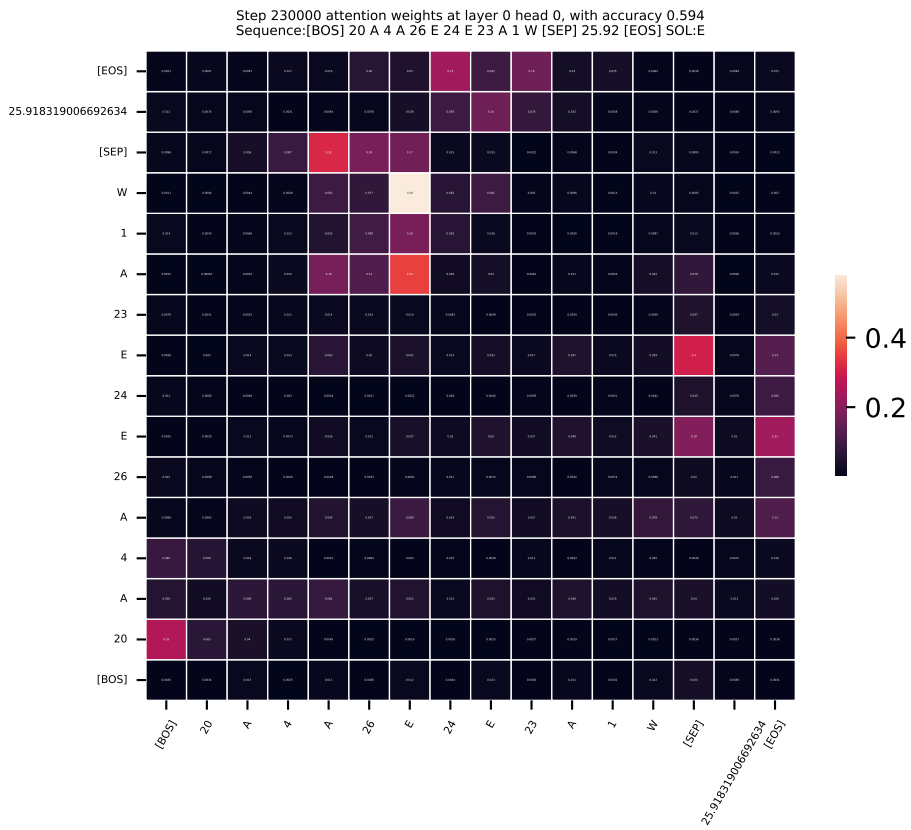

Step 229500 attention weights at layer 0 head 3, with accuracy 0.581
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

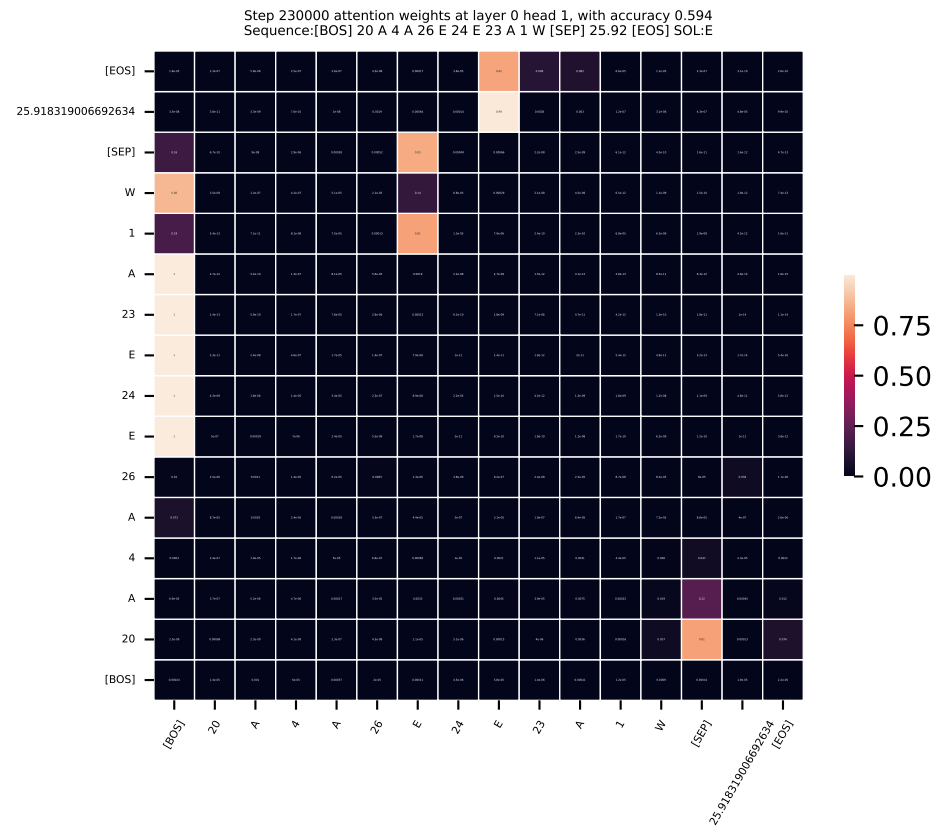

Step 229500 attention weights at layer 1 head 0, with accuracy 0.581
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

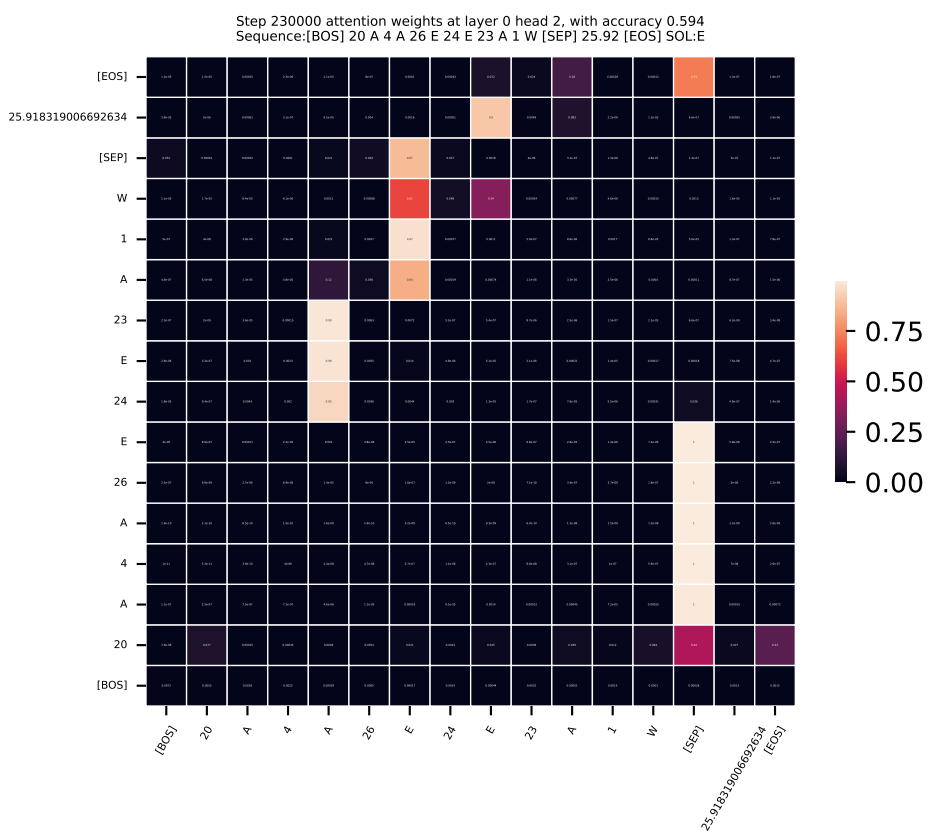

Step 229500 attention weights at layer 1 head 1, with accuracy 0.581
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.10
0.05

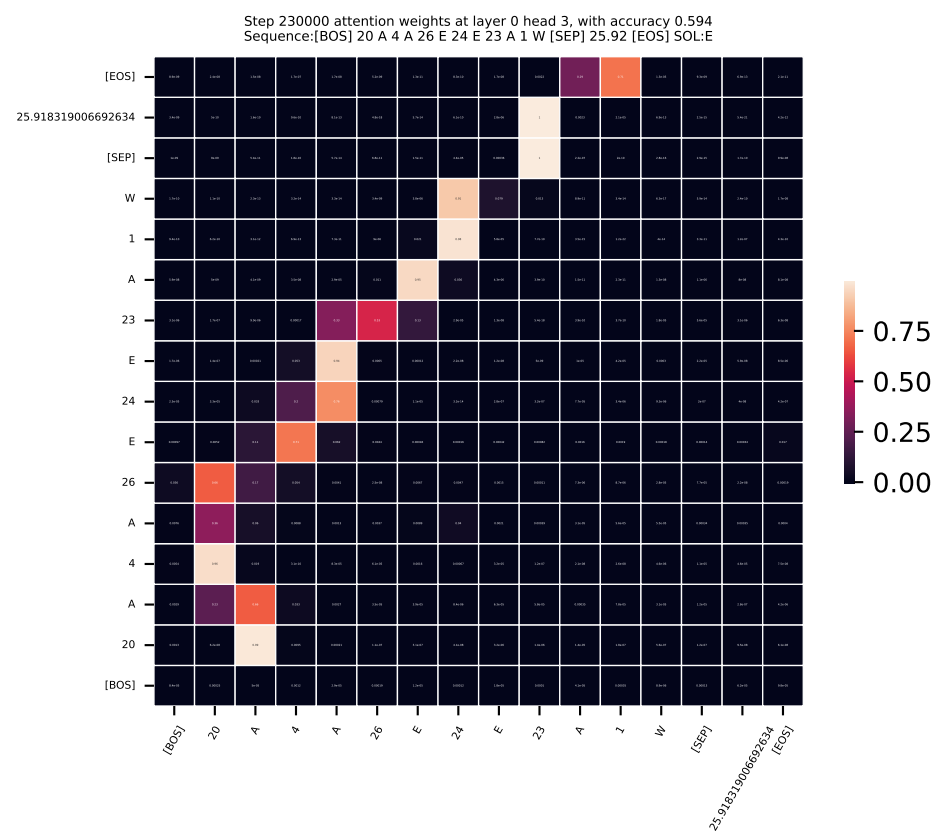

Step 229500 attention weights at layer 1 head 2, with accuracy 0.581
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

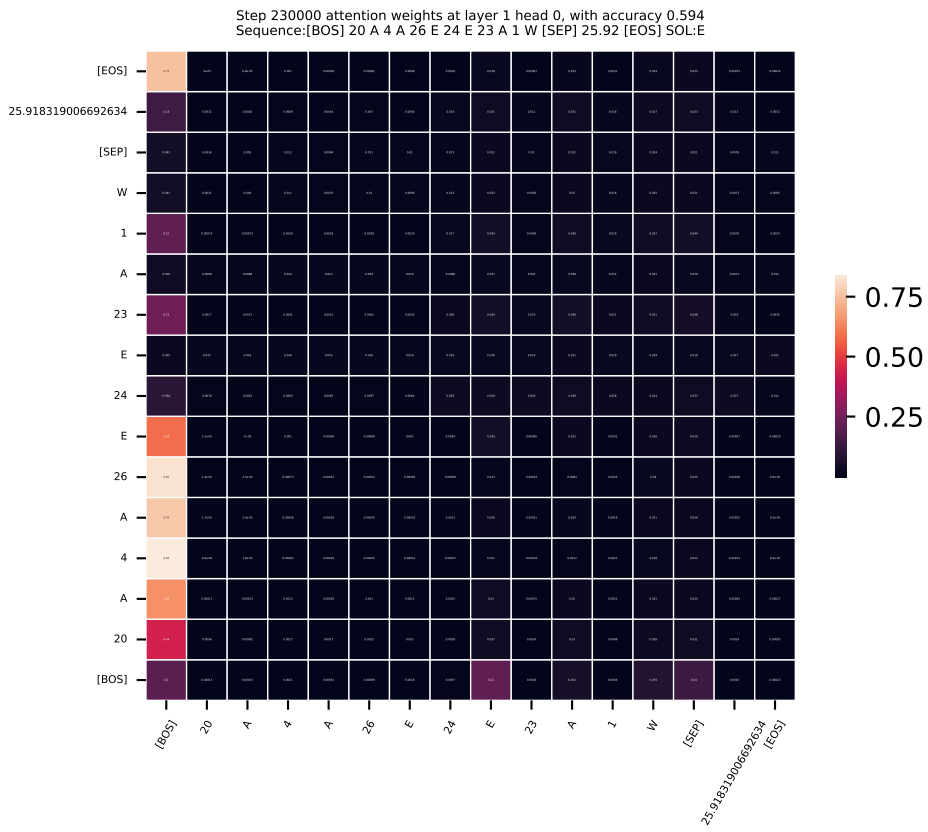

Step 229500 attention weights at layer 1 head 3, with accuracy 0.581
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
1.0
0.5
0.0

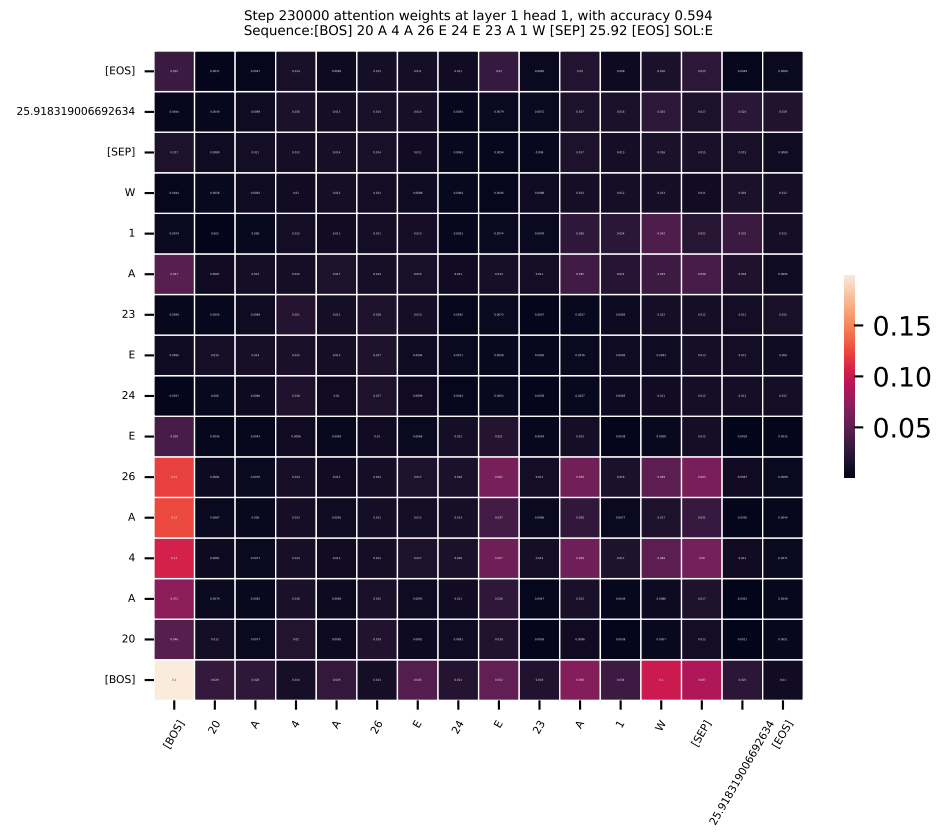

Step 230000 attention weights at layer 0 head 0, with accuracy 0.594
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

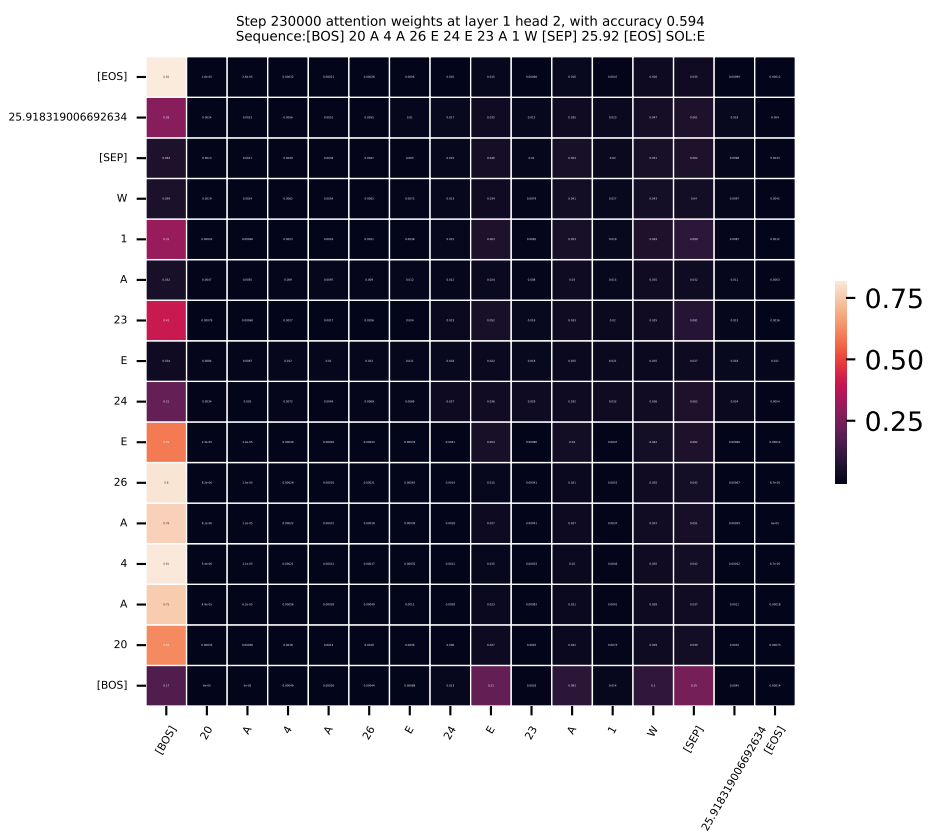

Step 230000 attention weights at layer 0 head 1, with accuracy 0.594
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

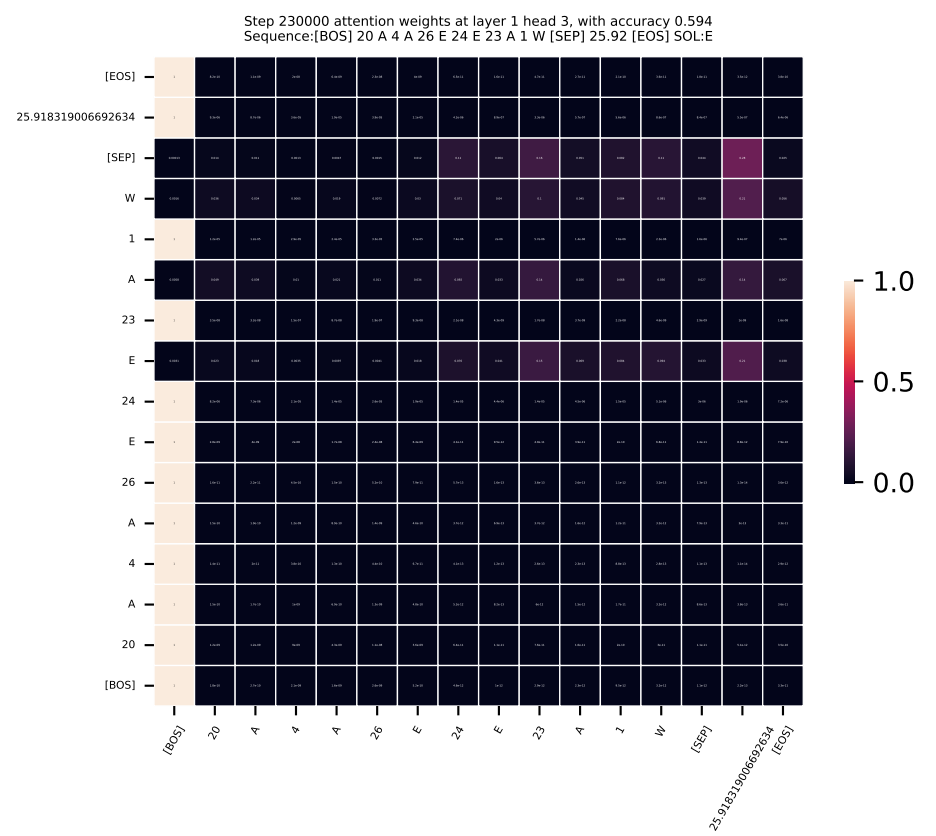

Step 230000 attention weights at layer 0 head 2, with accuracy 0.594
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

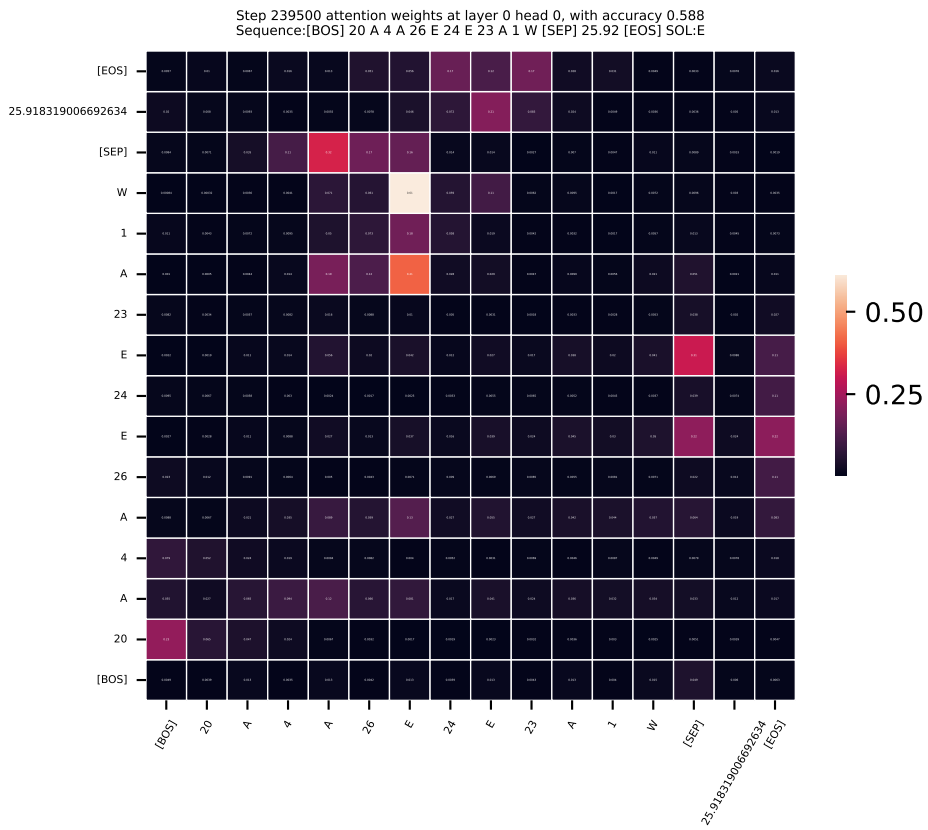

Step 230000 attention weights at layer 0 head 3, with accuracy 0.594
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

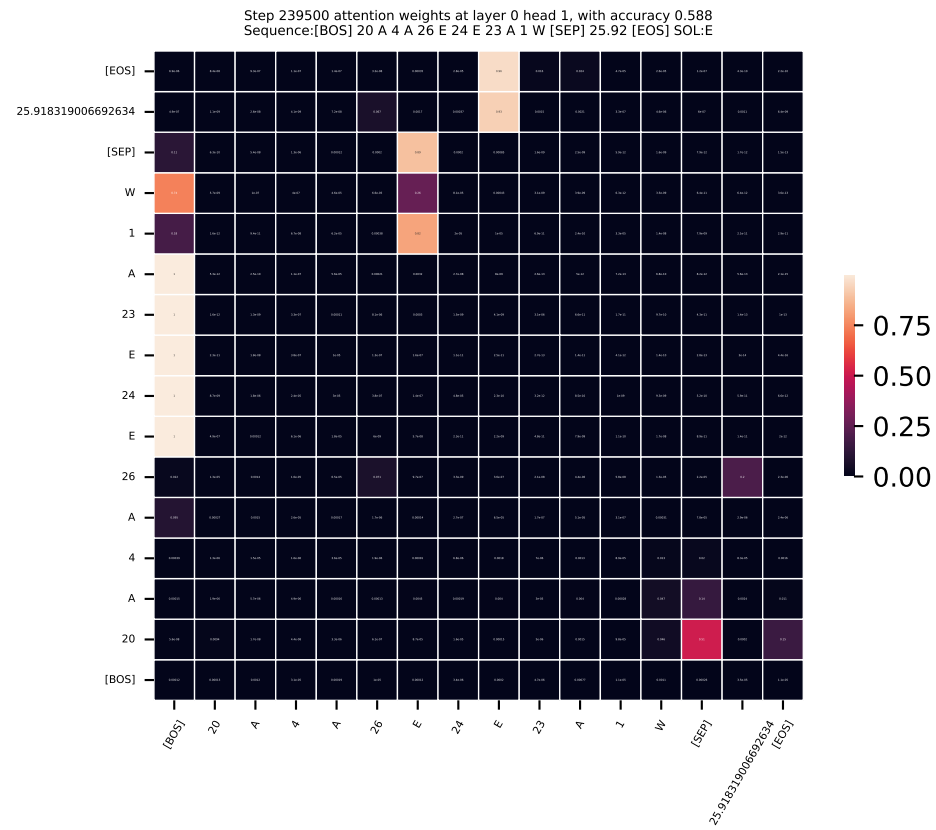

Step 230000 attention weights at layer 1 head 0, with accuracy 0.594
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

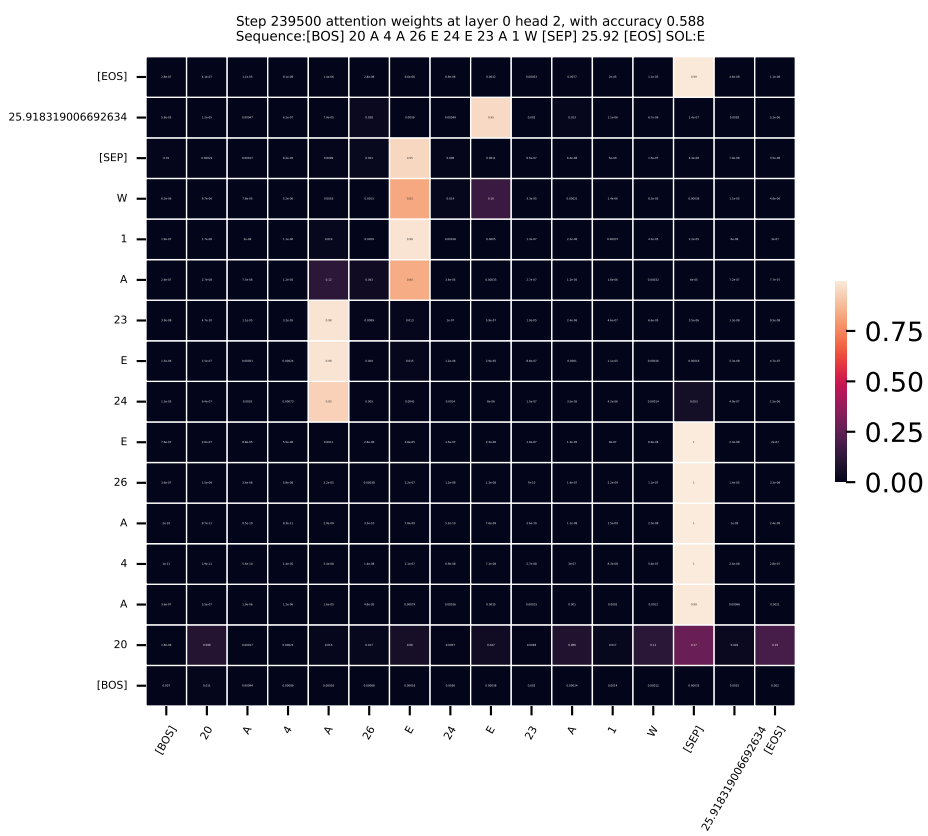

Step 230000 attention weights at layer 1 head 1, with accuracy 0.594
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.15
0.10
0.05

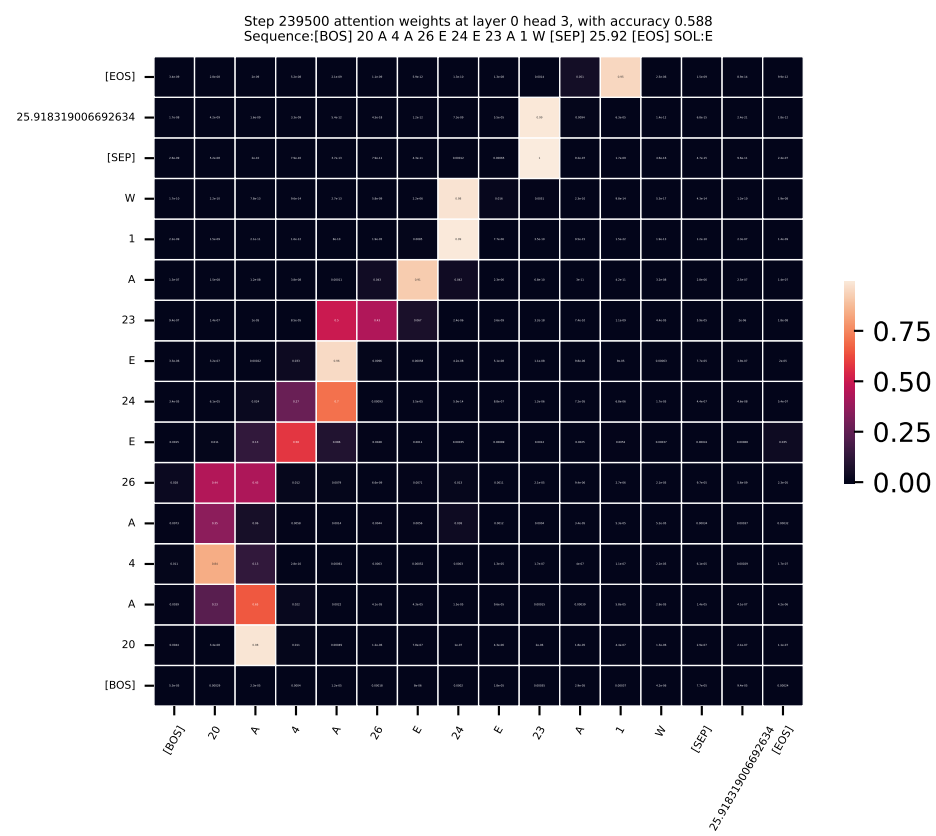

Step 230000 attention weights at layer 1 head 2, with accuracy 0.594
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

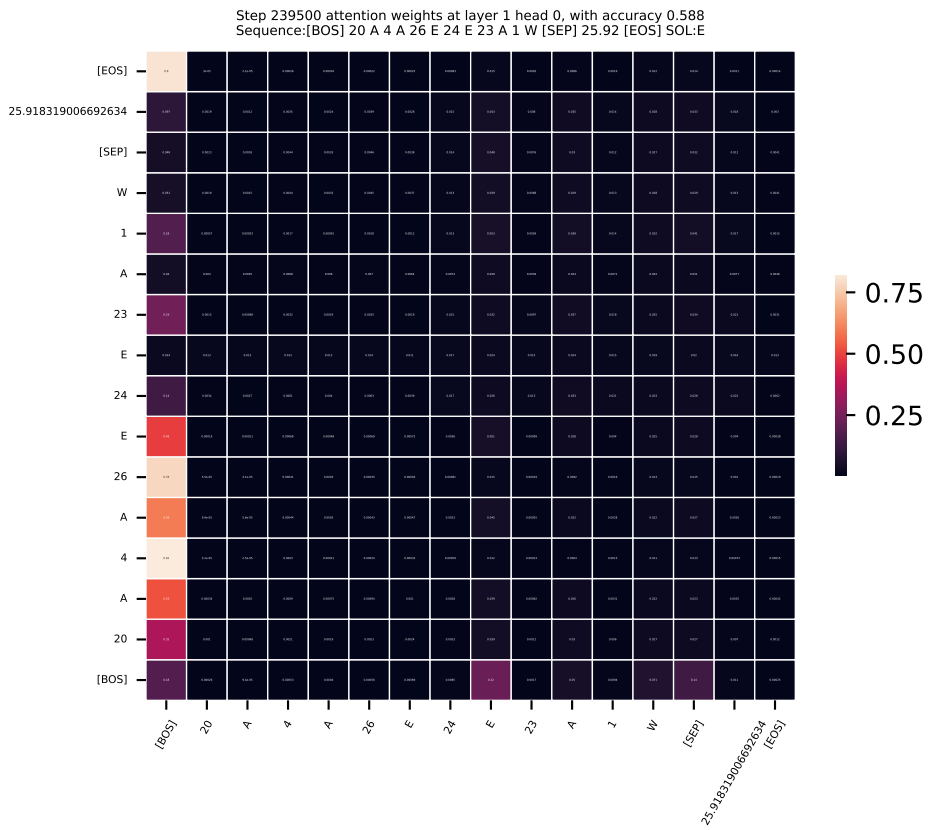

Step 230000 attention weights at layer 1 head 3, with accuracy 0.594
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
1.0
0.5
0.0

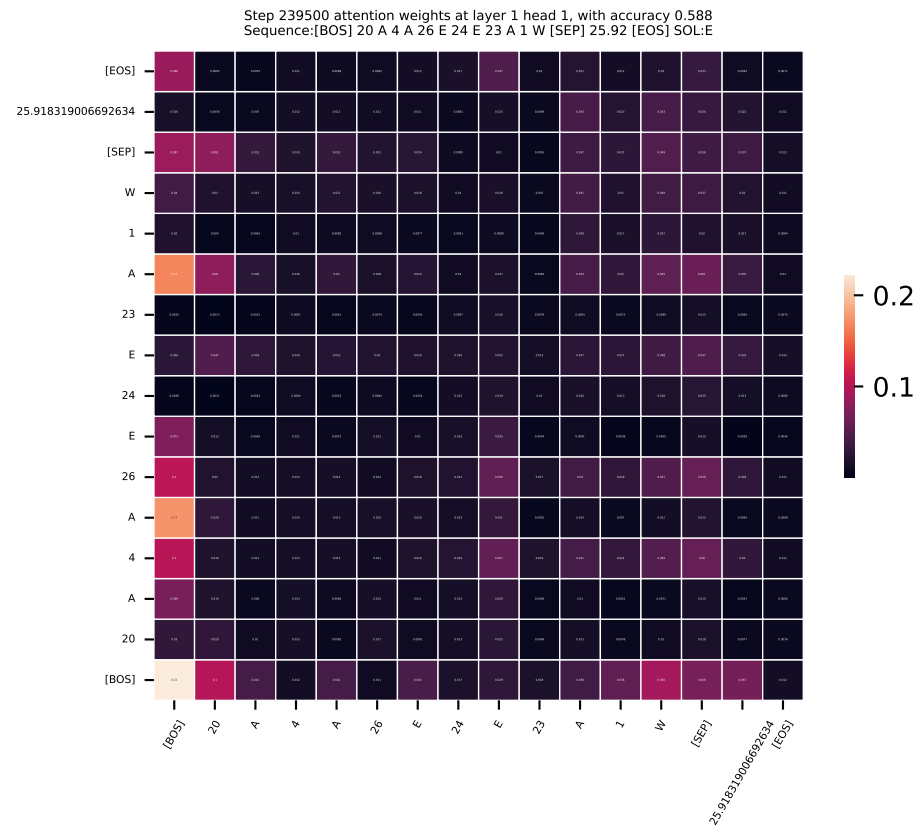

Step 239500 attention weights at layer 0 head 0, with accuracy 0.588
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

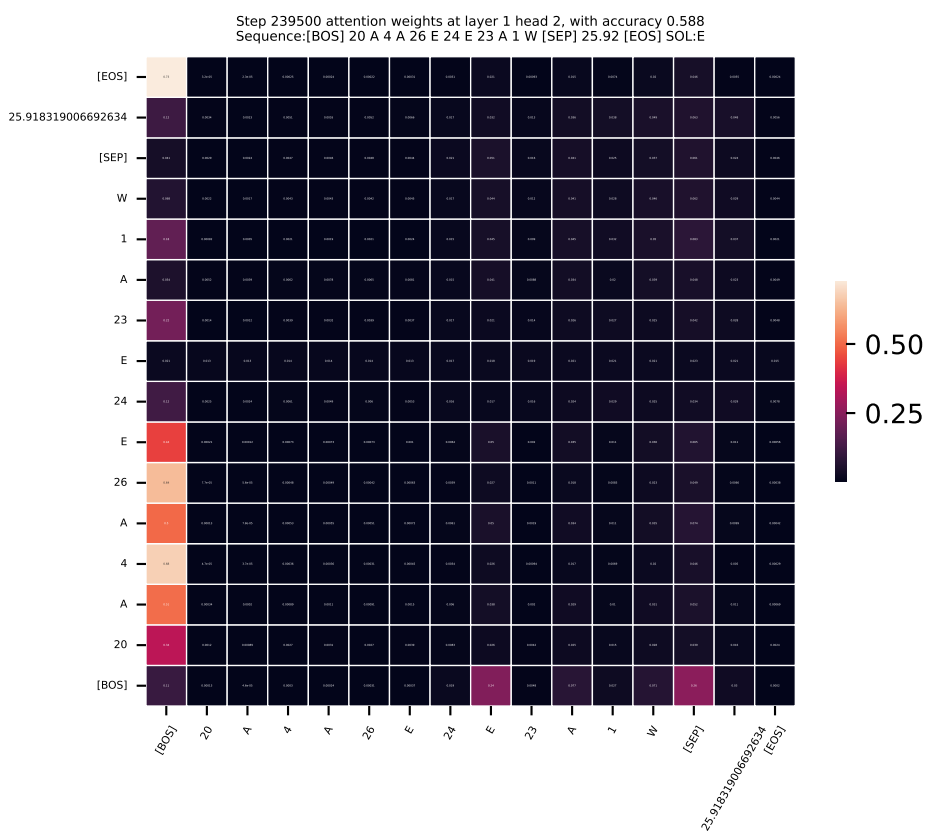

Step 239500 attention weights at layer 0 head 1, with accuracy 0.588
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

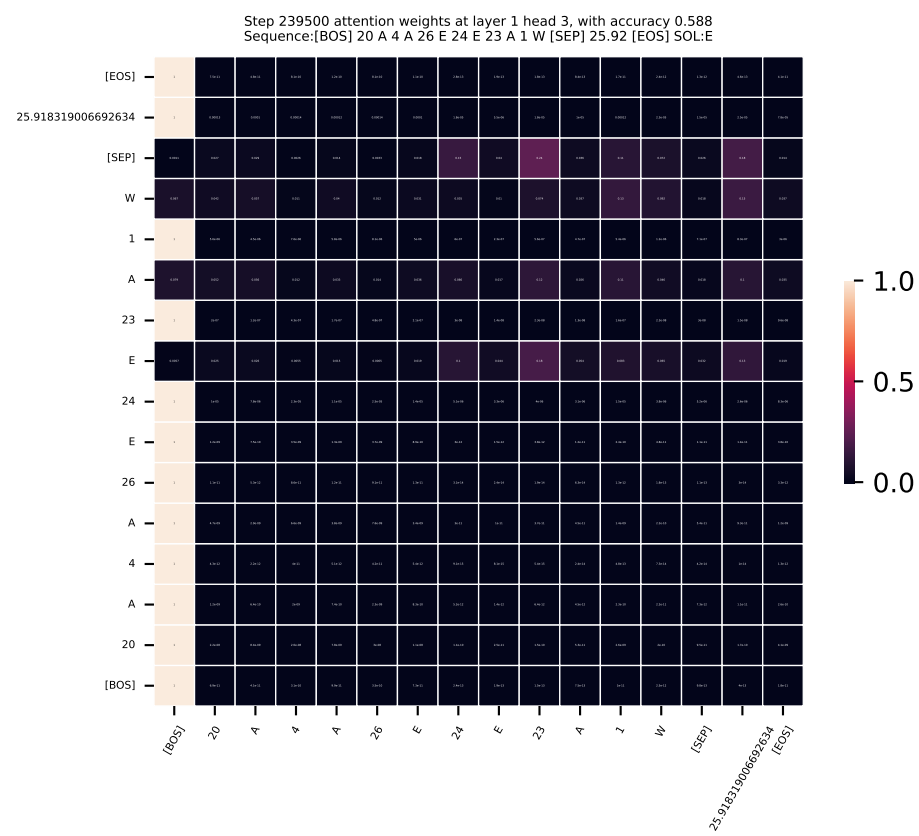

Step 239500 attention weights at layer 0 head 2, with accuracy 0.588
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

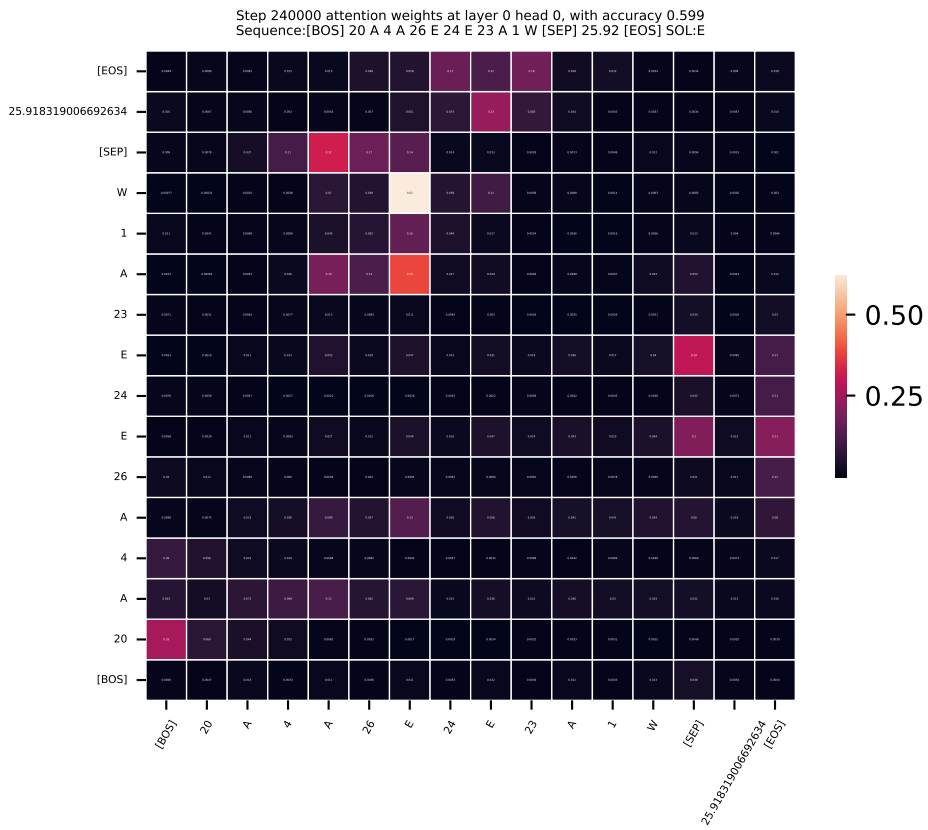

Step 239500 attention weights at layer 0 head 3, with accuracy 0.588
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

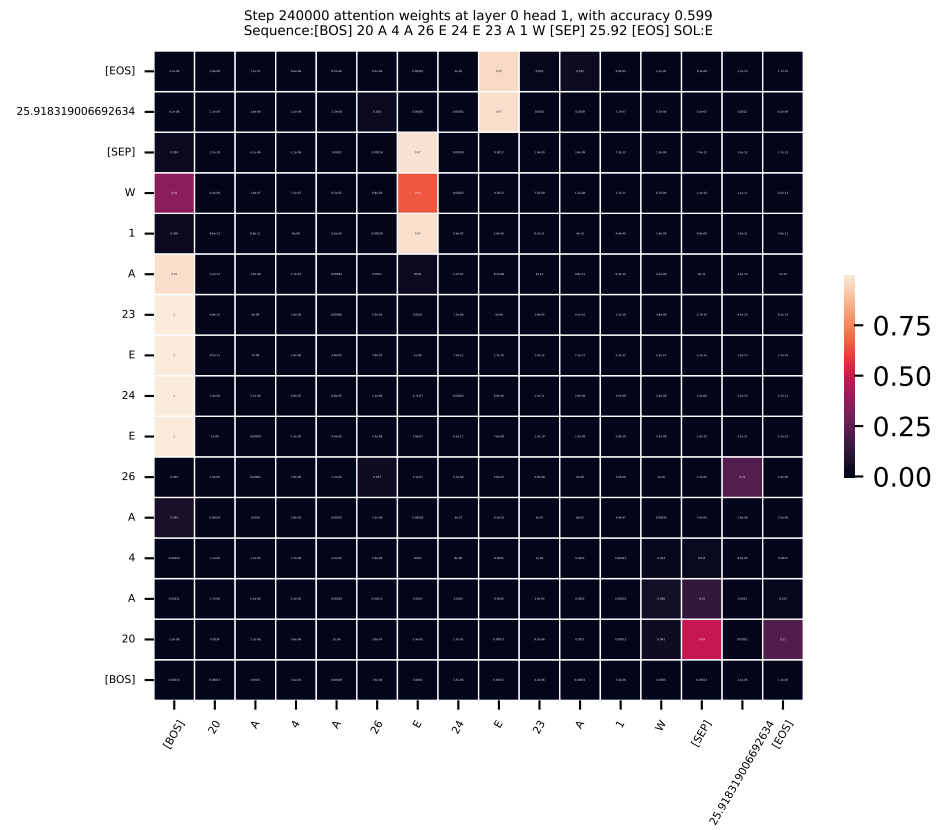

Step 239500 attention weights at layer 1 head 0, with accuracy 0.588
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

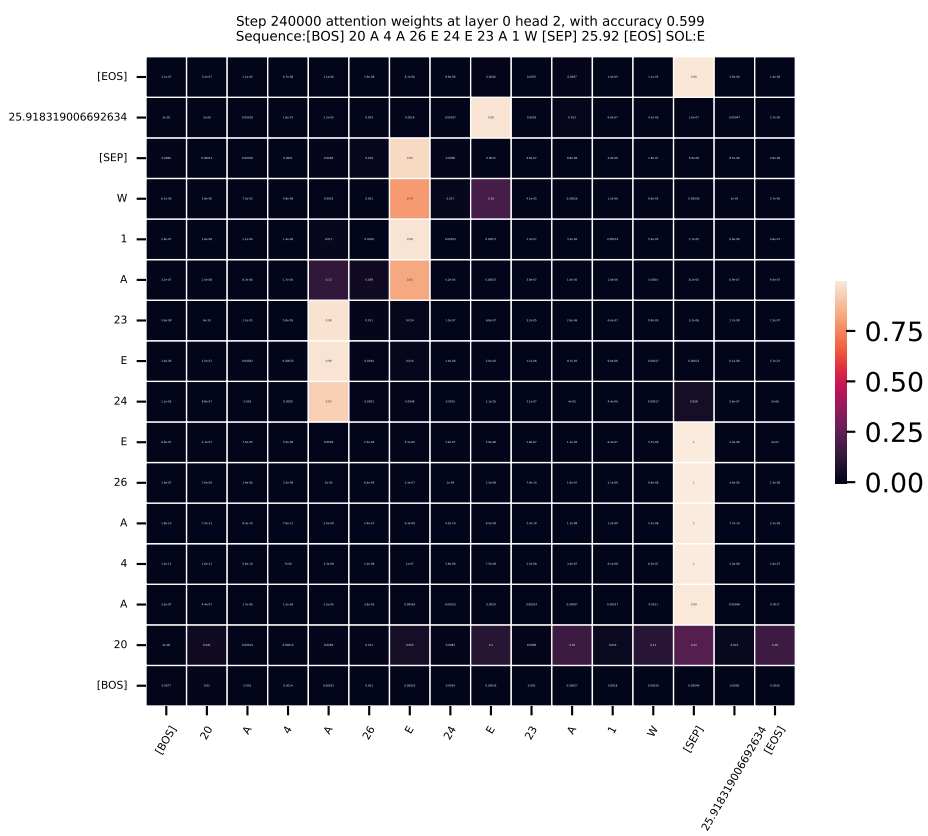

Step 239500 attention weights at layer 1 head 1, with accuracy 0.588
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.2
0.1

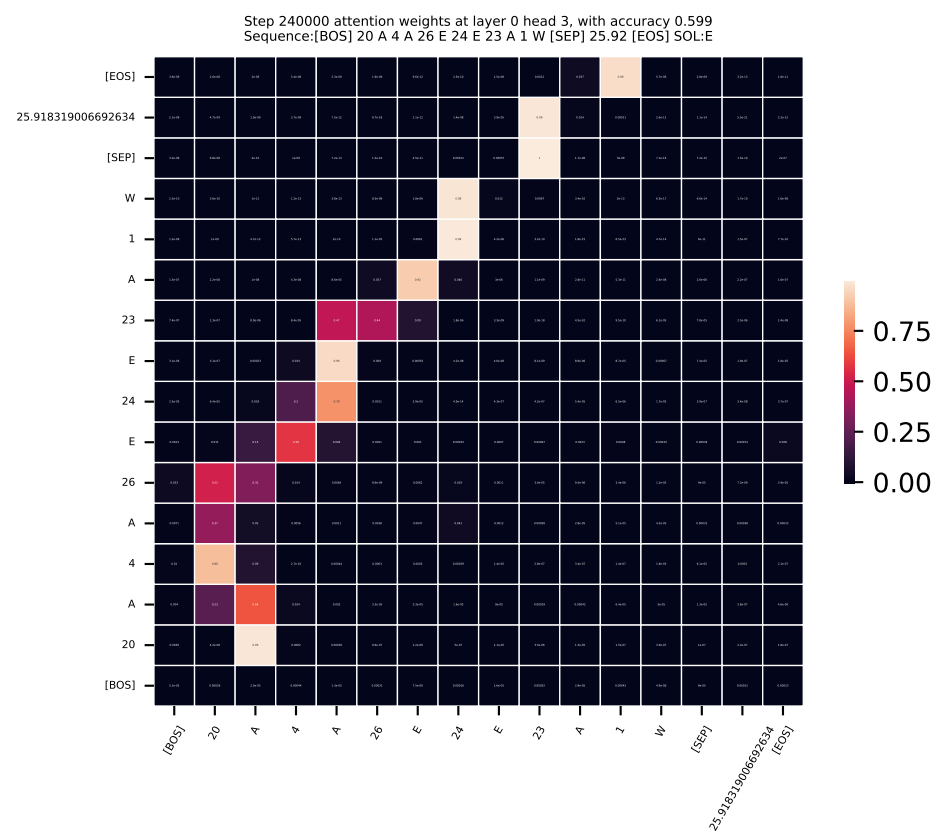

Step 239500 attention weights at layer 1 head 2, with accuracy 0.588
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

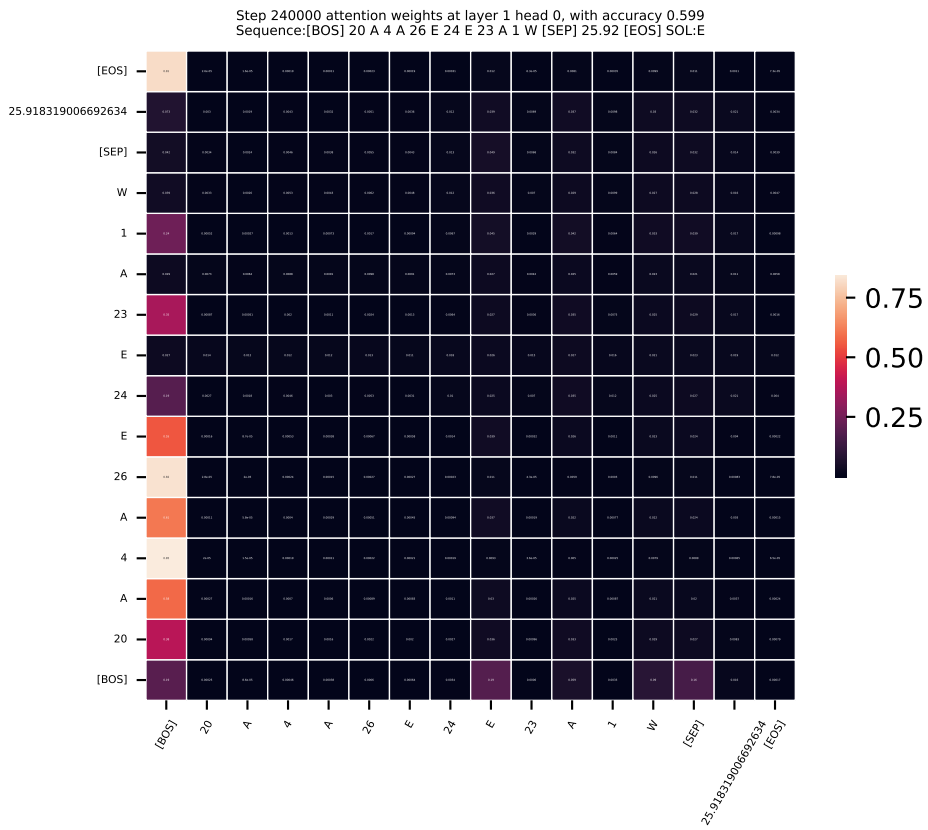

Step 239500 attention weights at layer 1 head 3, with accuracy 0.588
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
1.0
0.5
0.0

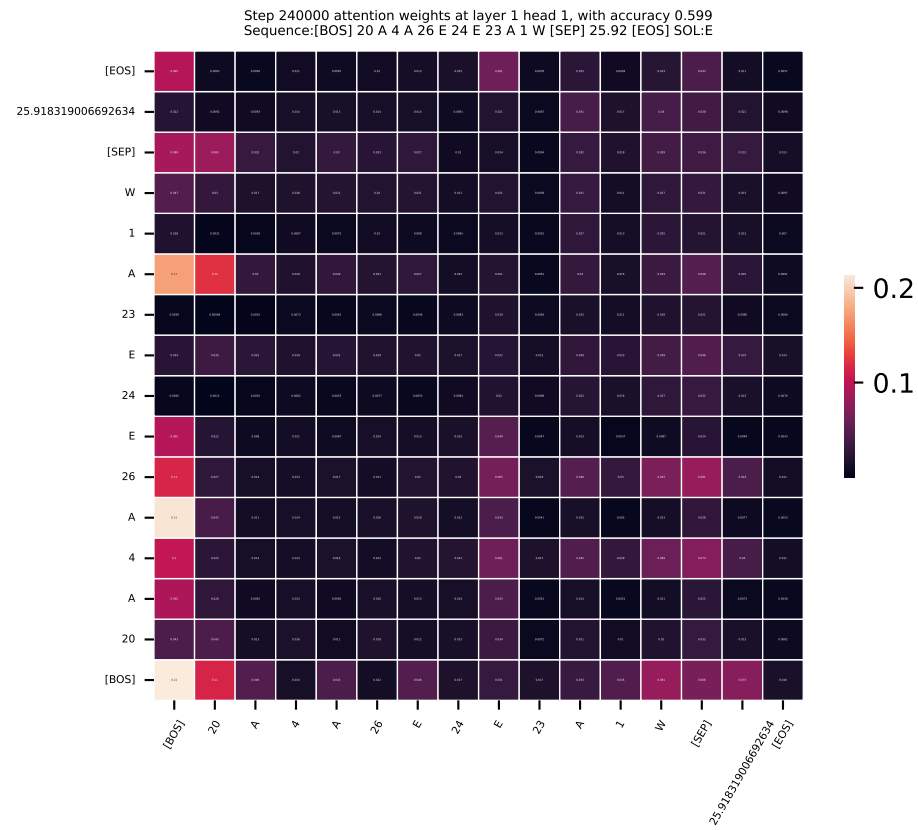

Step 240000 attention weights at layer 0 head 0, with accuracy 0.599
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

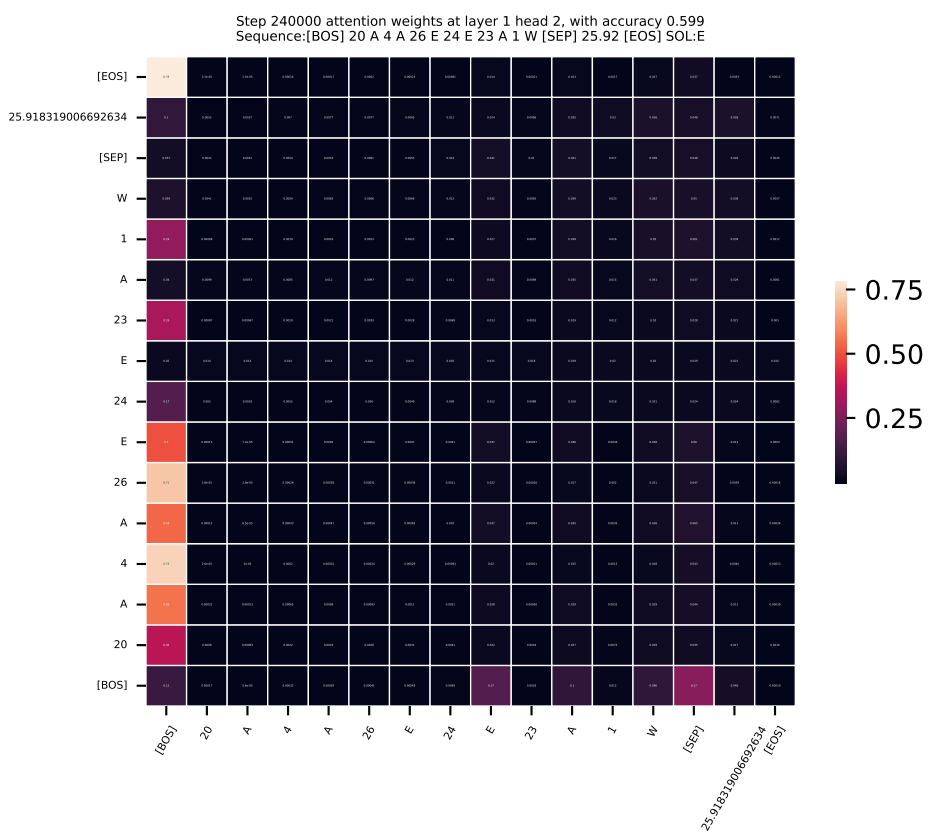

Step 240000 attention weights at layer 0 head 1, with accuracy 0.599
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

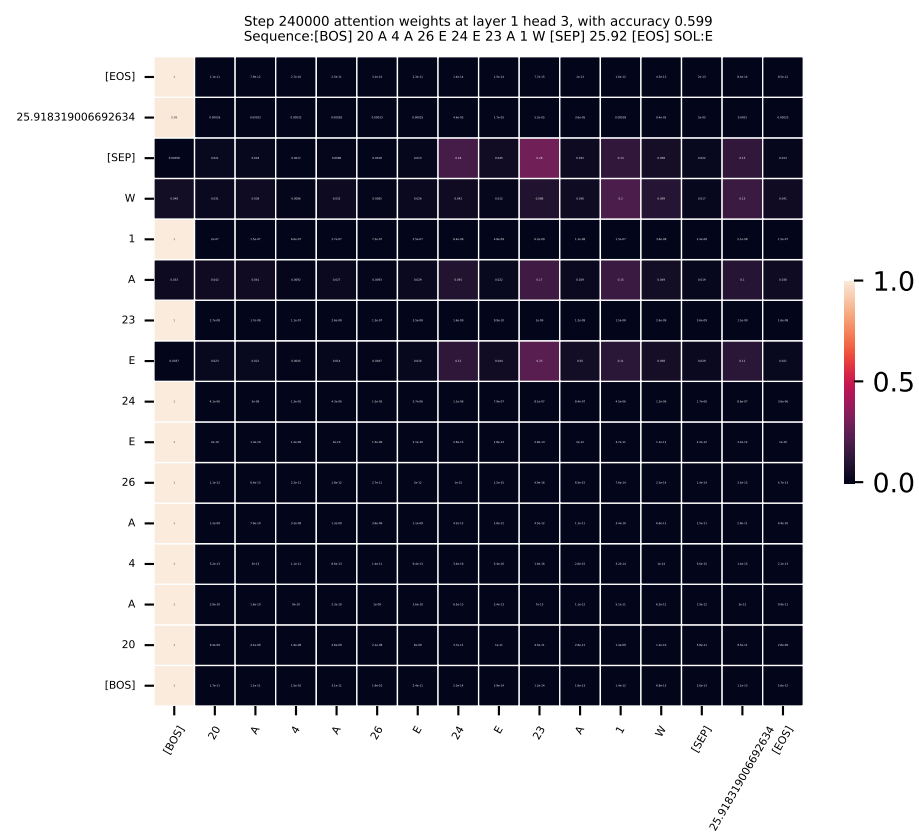

Step 240000 attention weights at layer 0 head 2, with accuracy 0.599
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

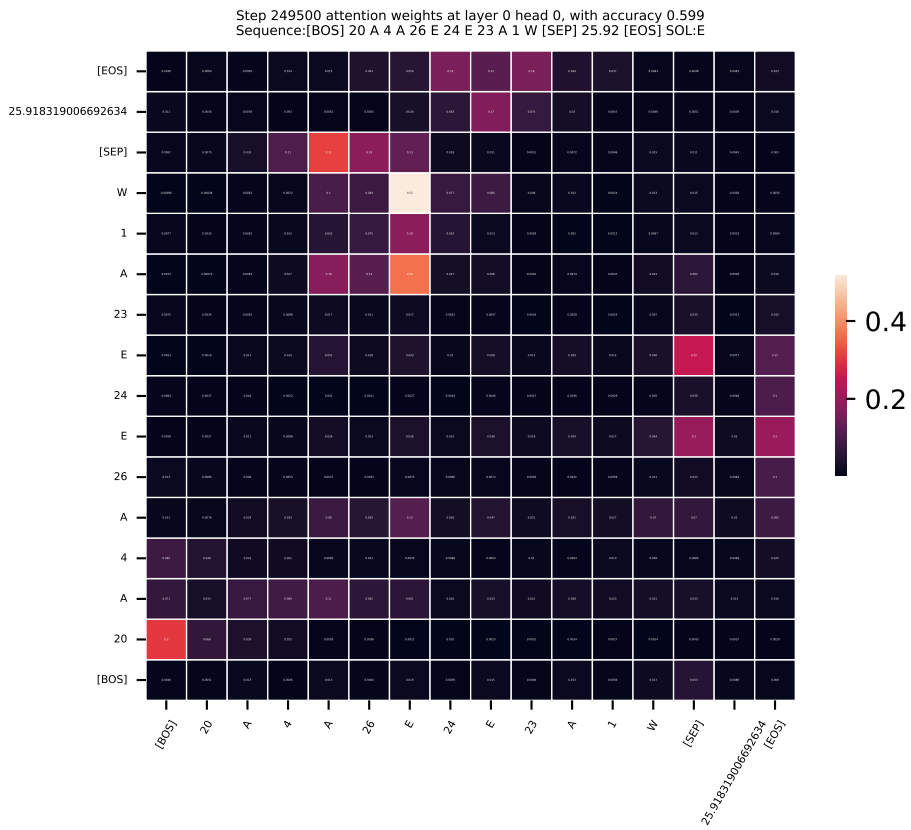

Step 240000 attention weights at layer 0 head 3, with accuracy 0.599
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

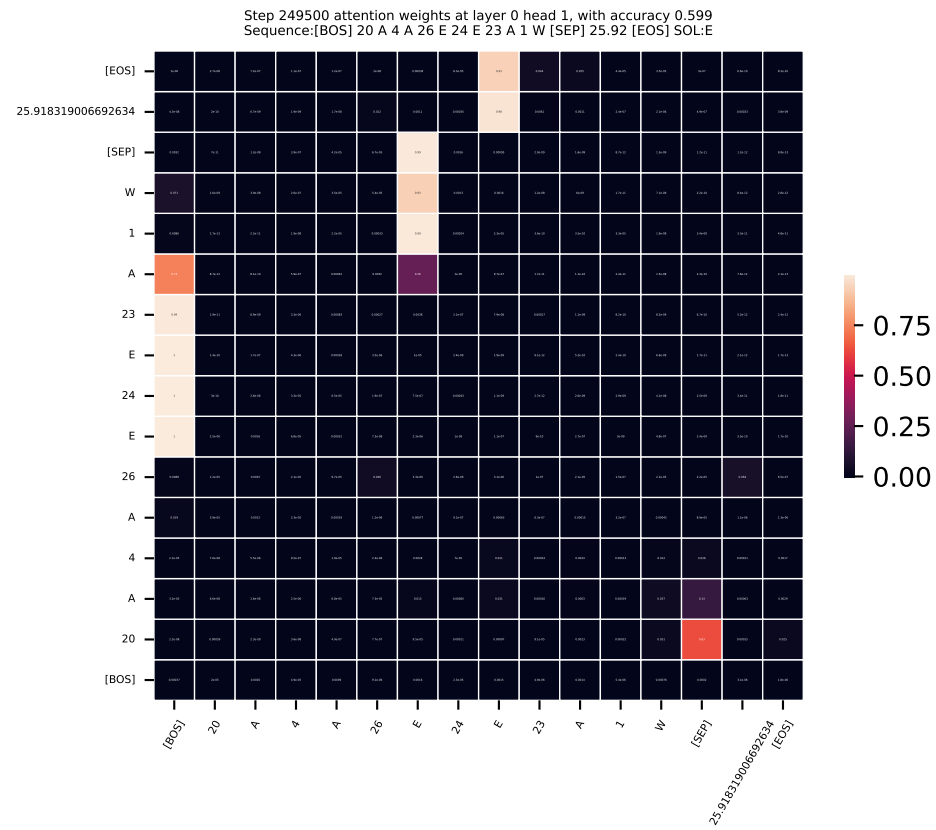

Step 240000 attention weights at layer 1 head 0, with accuracy 0.599
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

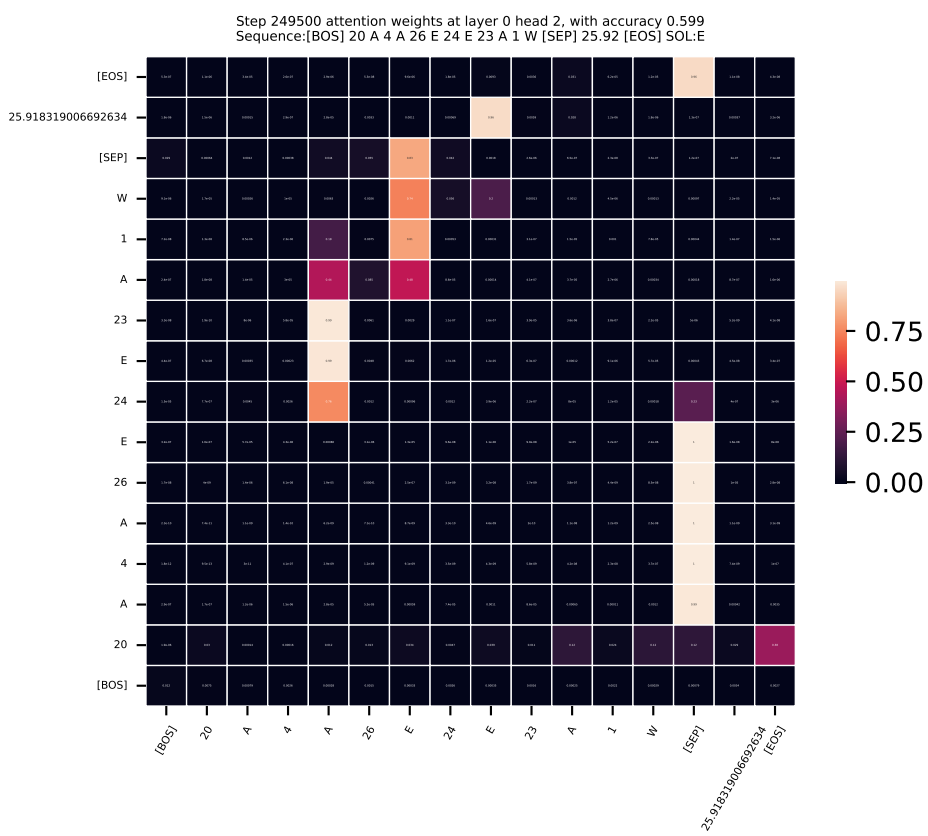

Step 240000 attention weights at layer 1 head 1, with accuracy 0.599
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.2
0.1

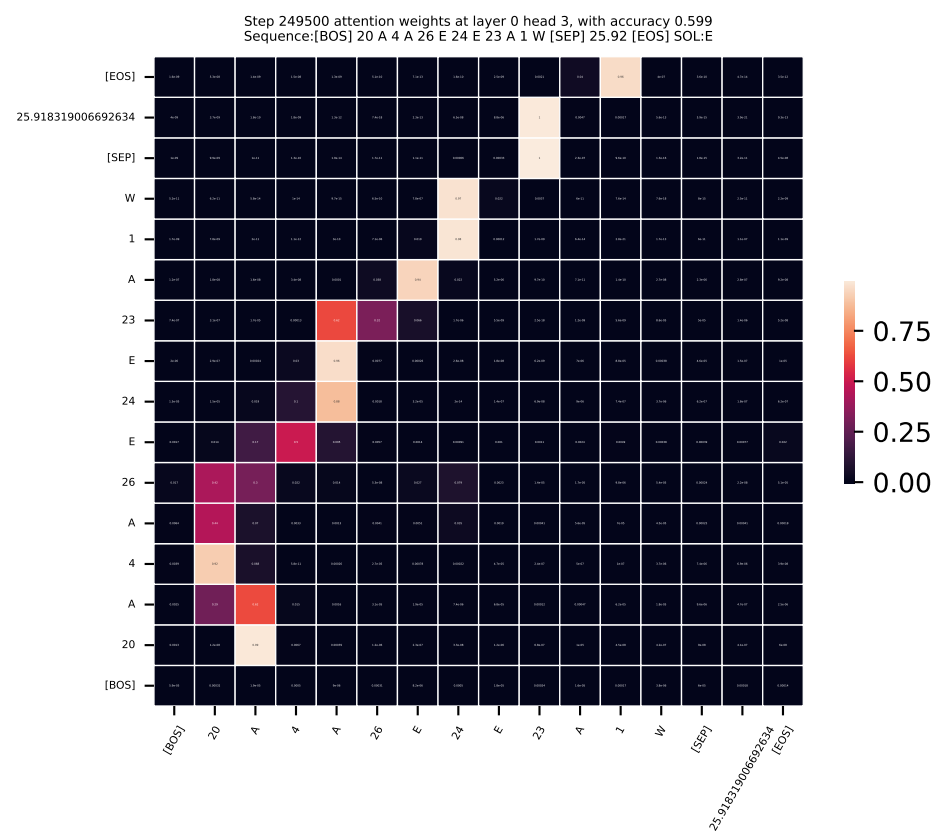

Step 240000 attention weights at layer 1 head 2, with accuracy 0.599
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

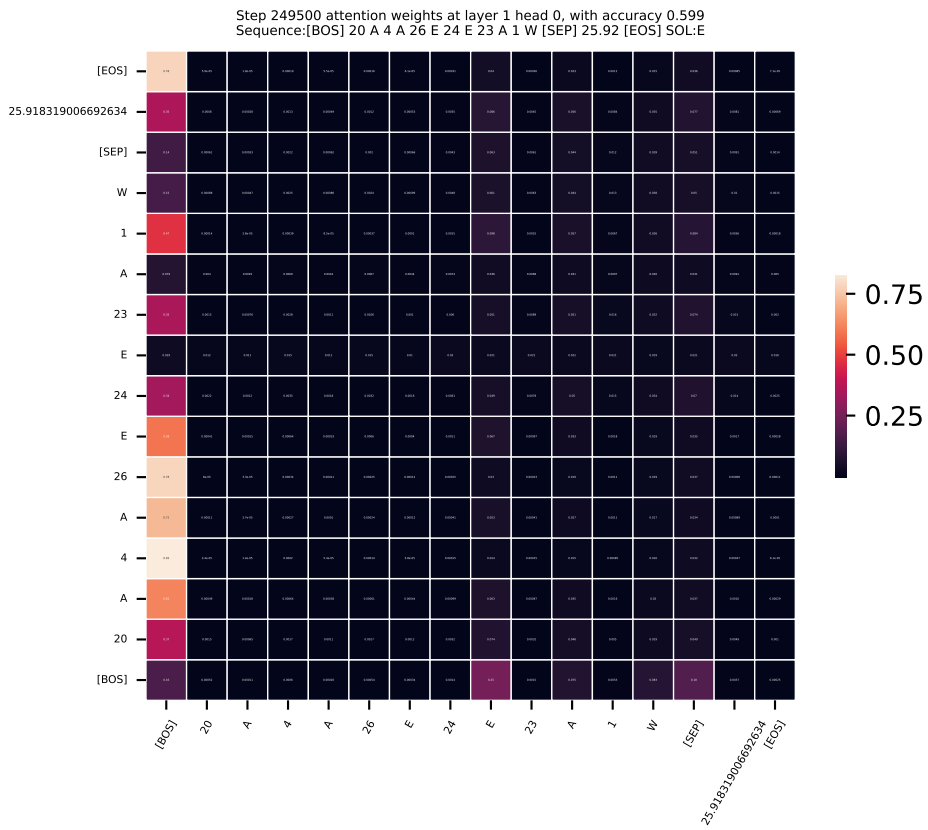

Step 240000 attention weights at layer 1 head 3, with accuracy 0.599
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
1.0
0.5
0.0

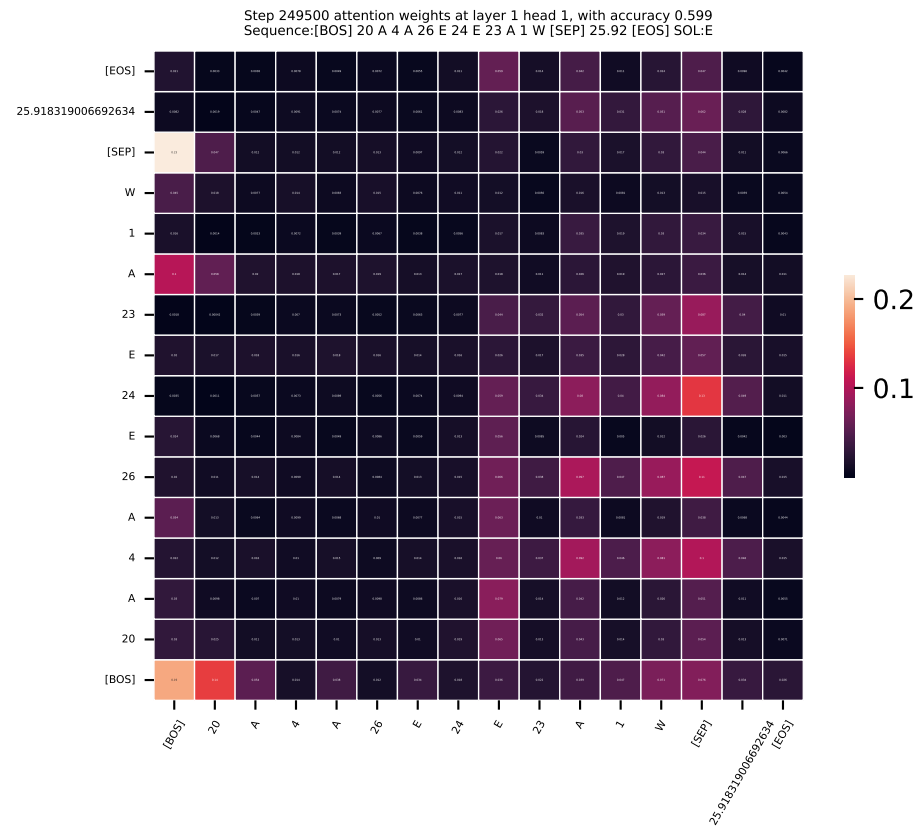

Step 249500 attention weights at layer 0 head 0, with accuracy 0.599
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

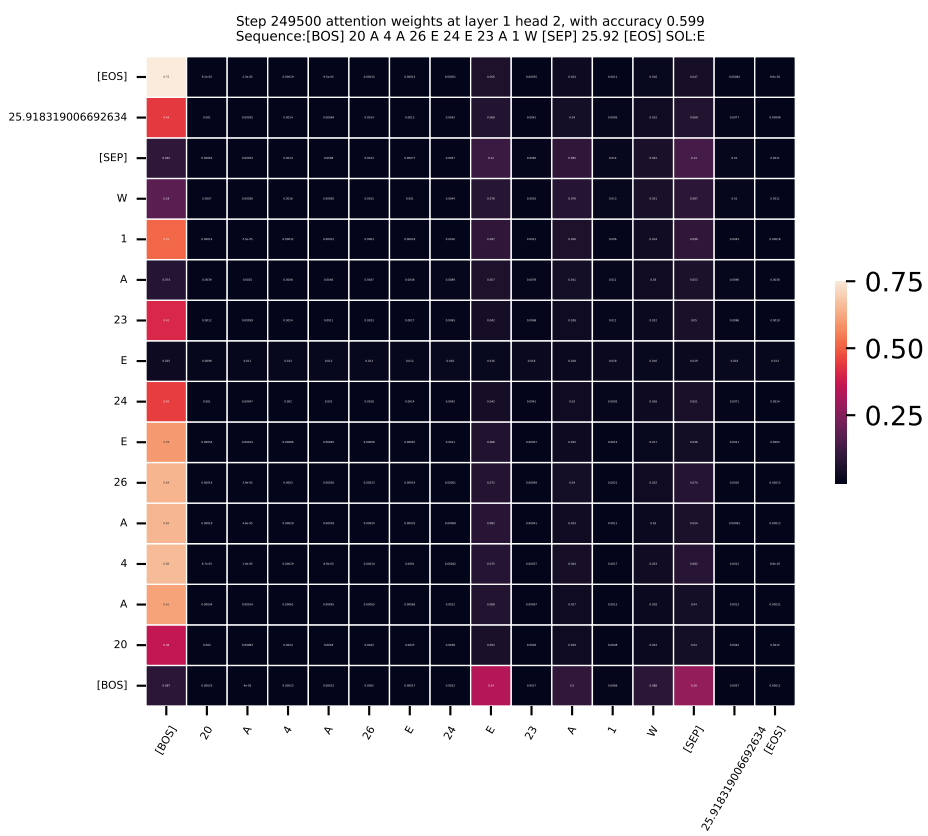

Step 249500 attention weights at layer 0 head 1, with accuracy 0.599
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

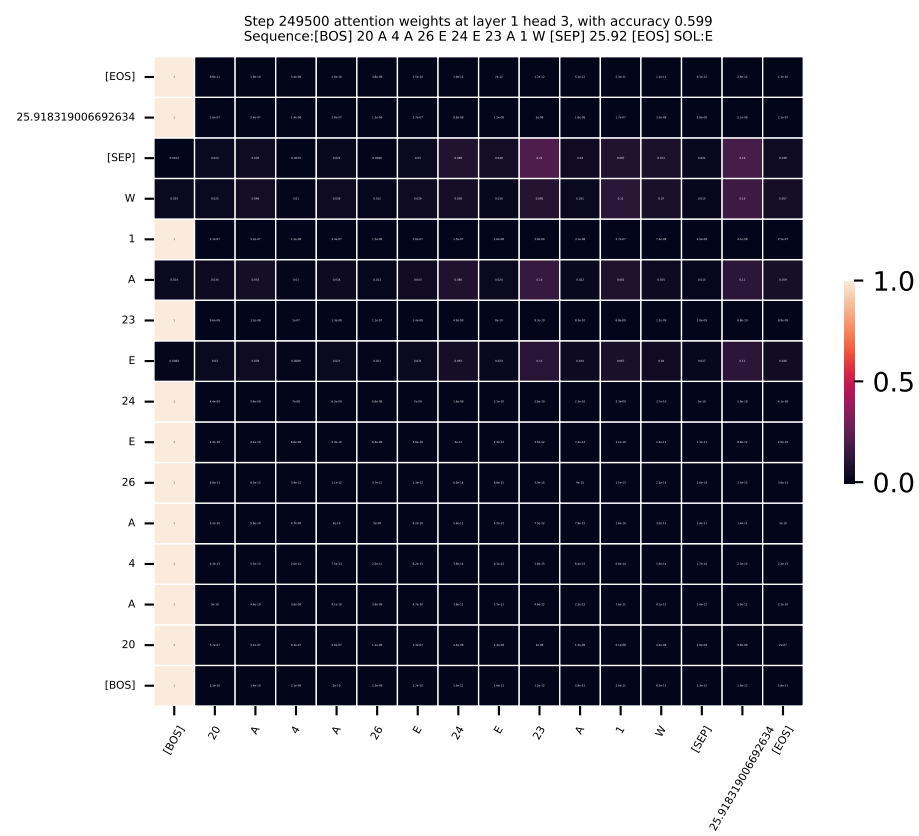

Step 249500 attention weights at layer 0 head 2, with accuracy 0.599
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

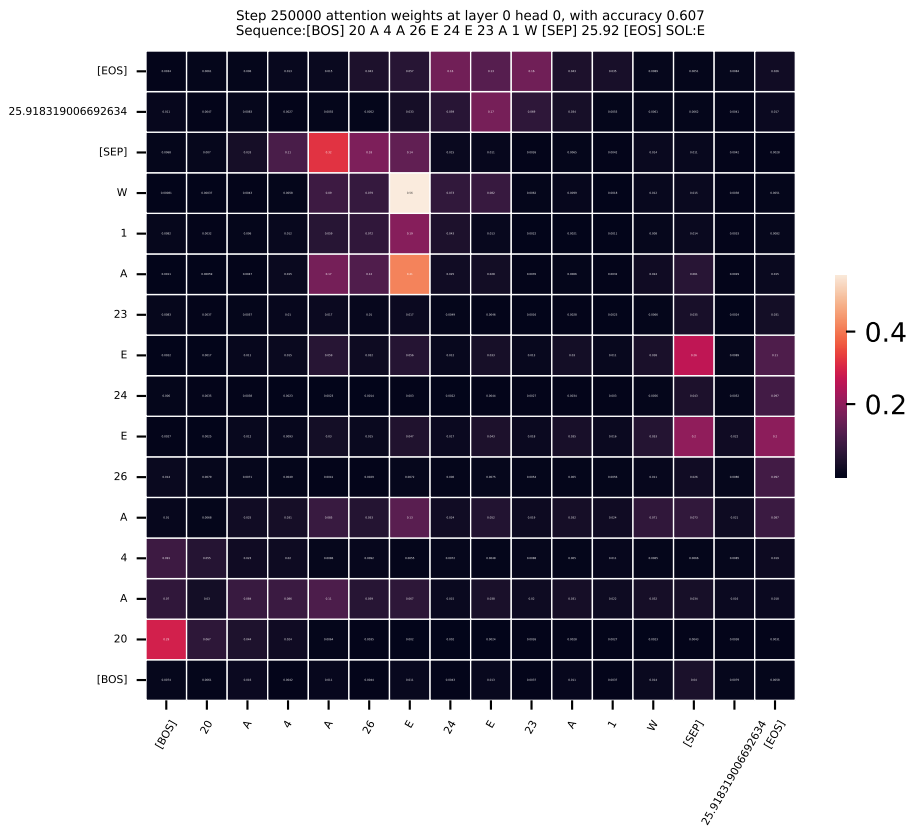

Step 249500 attention weights at layer 0 head 3, with accuracy 0.599
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

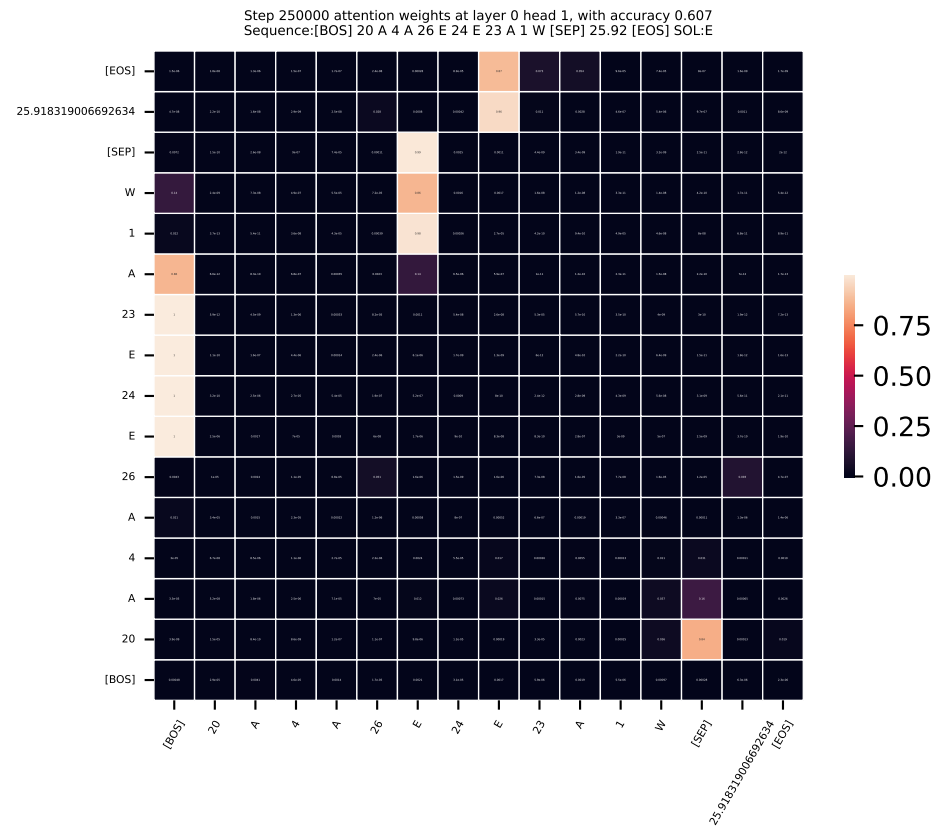

Step 249500 attention weights at layer 1 head 0, with accuracy 0.599
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

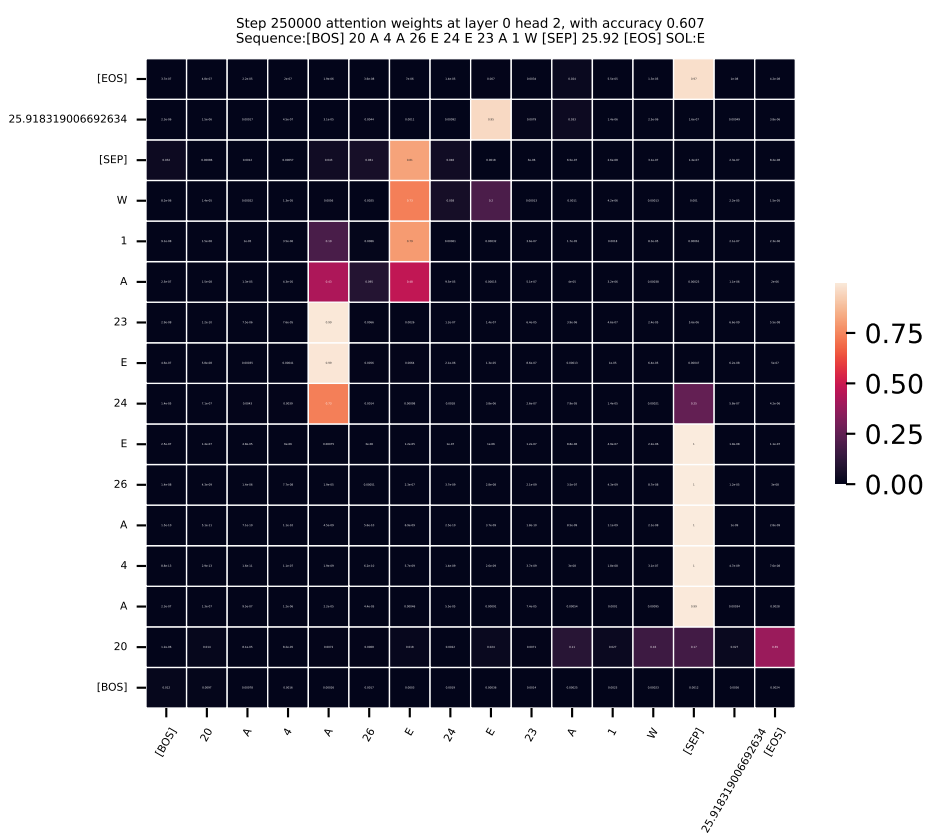

Step 249500 attention weights at layer 1 head 1, with accuracy 0.599
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.2
0.1

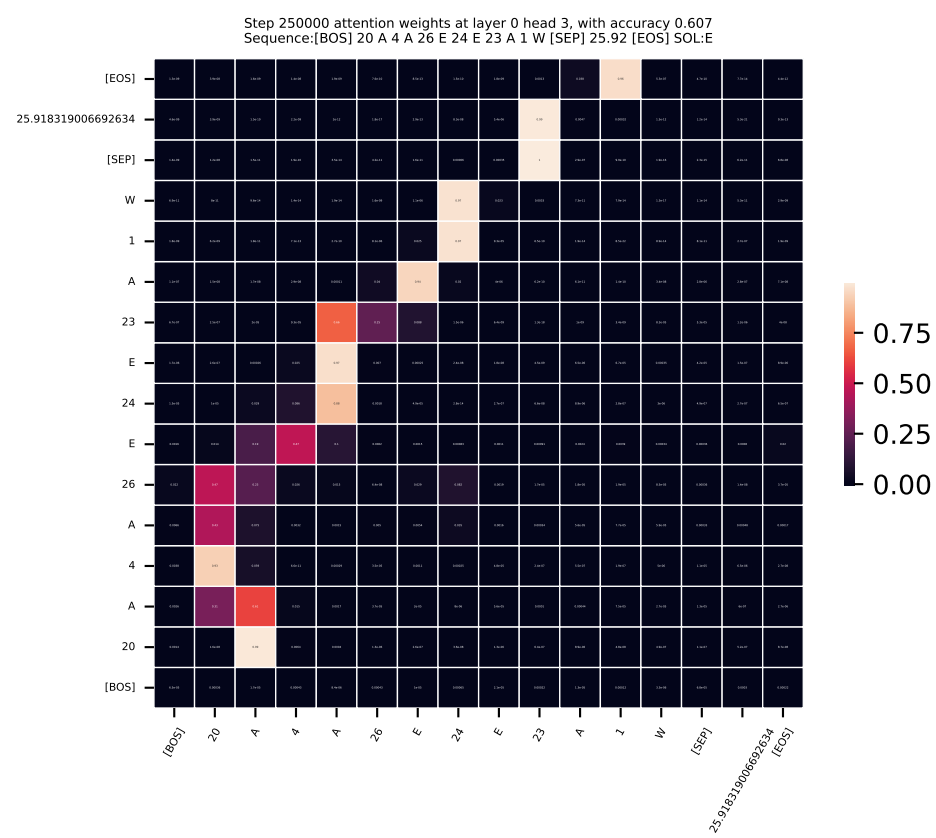

Step 249500 attention weights at layer 1 head 2, with accuracy 0.599
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

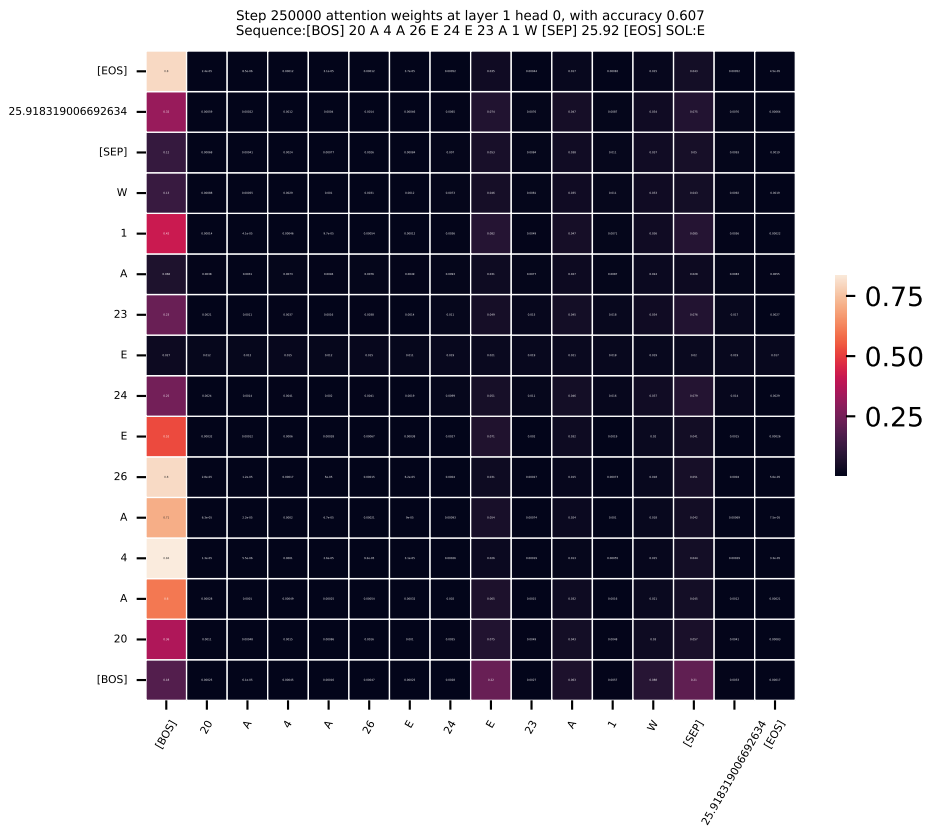

Step 249500 attention weights at layer 1 head 3, with accuracy 0.599
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
1.0
0.5
0.0

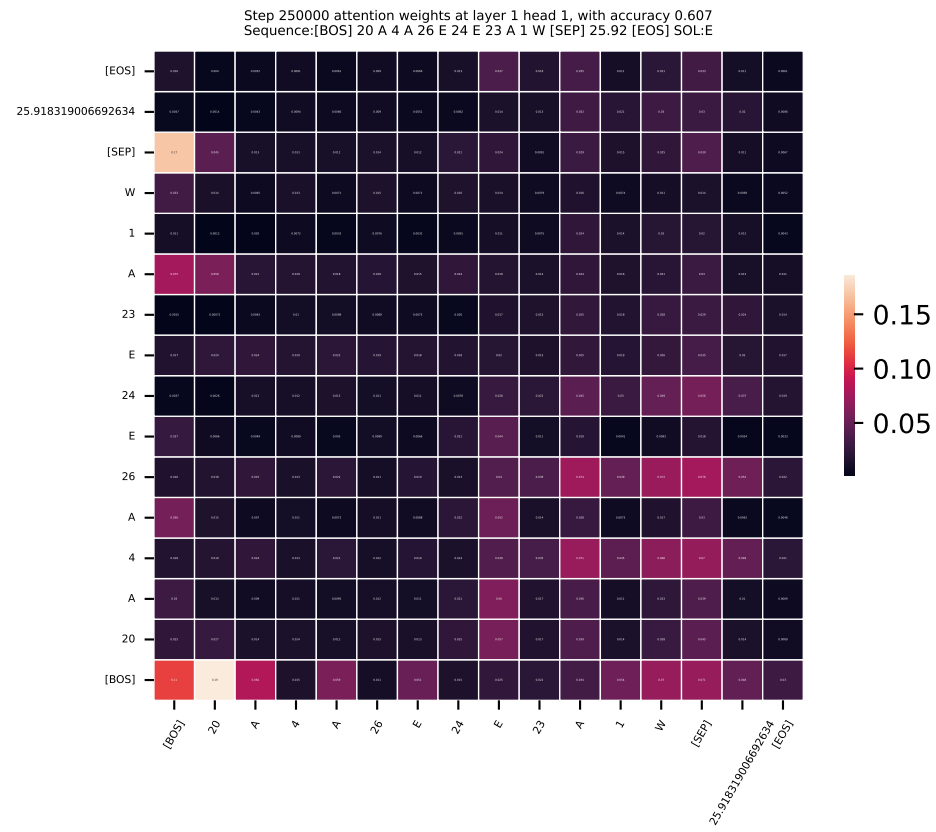

Step 250000 attention weights at layer 0 head 0, with accuracy 0.607
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

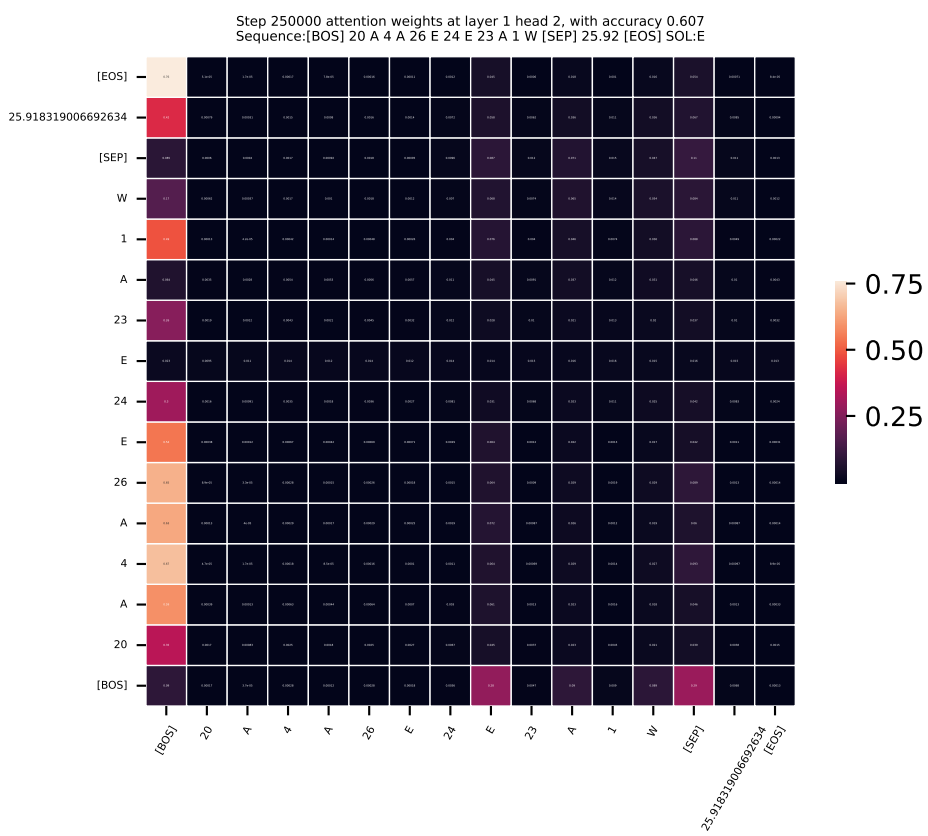

Step 250000 attention weights at layer 0 head 1, with accuracy 0.607
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

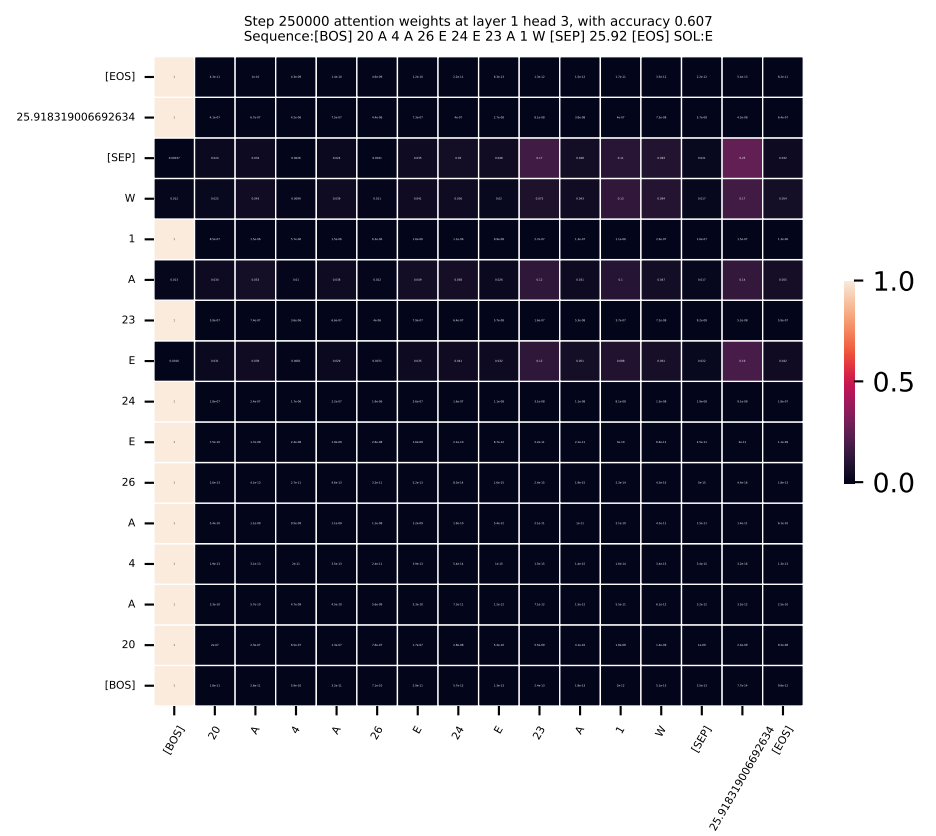

Step 250000 attention weights at layer 0 head 2, with accuracy 0.607
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

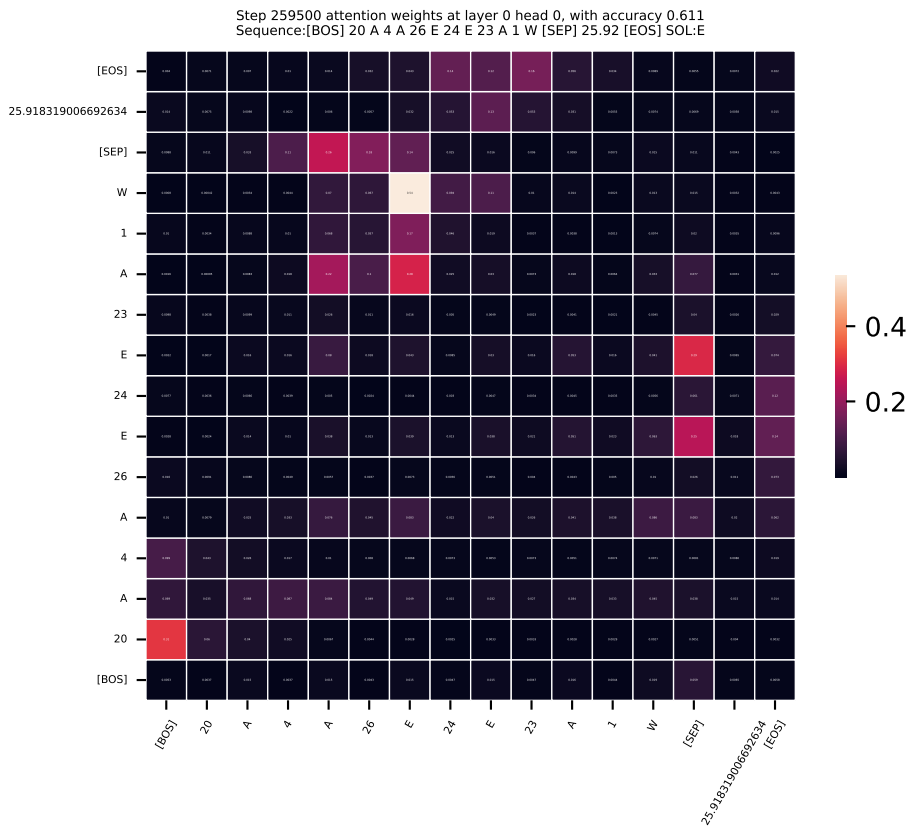

Step 250000 attention weights at layer 0 head 3, with accuracy 0.607
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

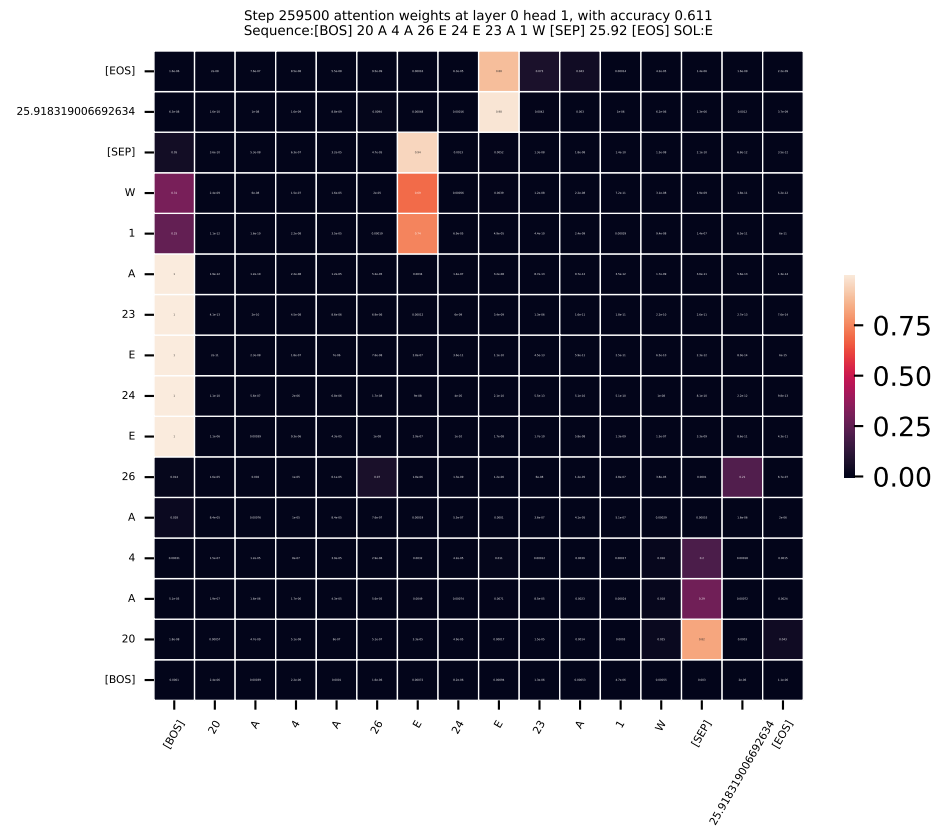

Step 250000 attention weights at layer 1 head 0, with accuracy 0.607
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

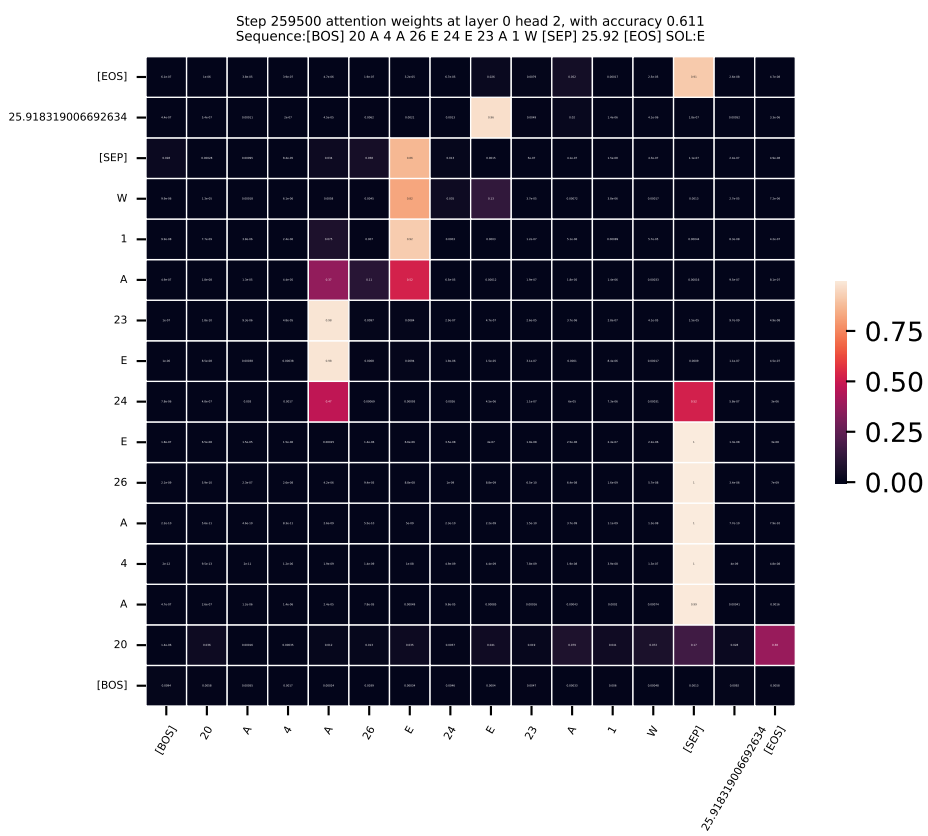

Step 250000 attention weights at layer 1 head 1, with accuracy 0.607
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.15
0.10
0.05

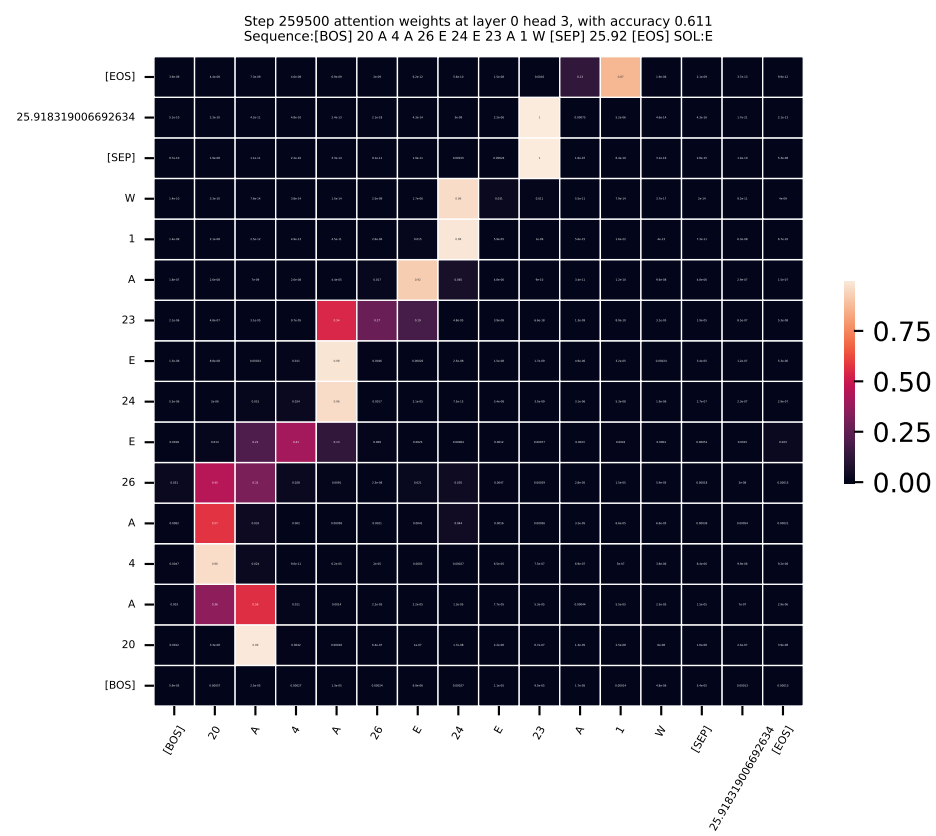

Step 250000 attention weights at layer 1 head 2, with accuracy 0.607
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

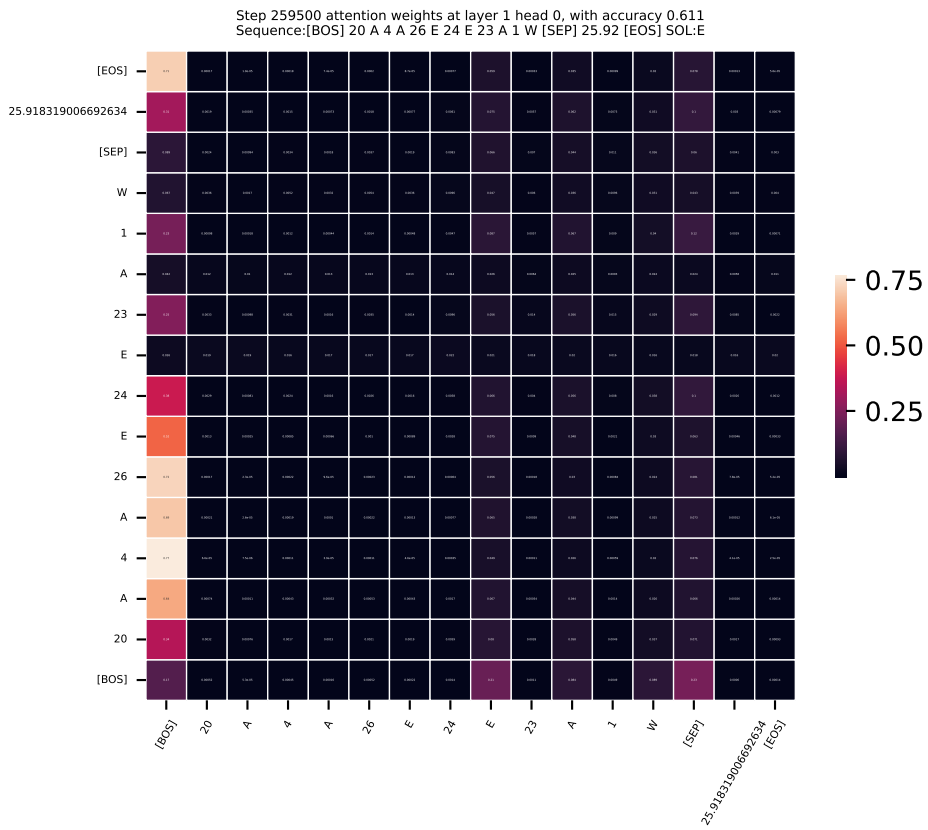

Step 250000 attention weights at layer 1 head 3, with accuracy 0.607
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
1.0
0.5
0.0

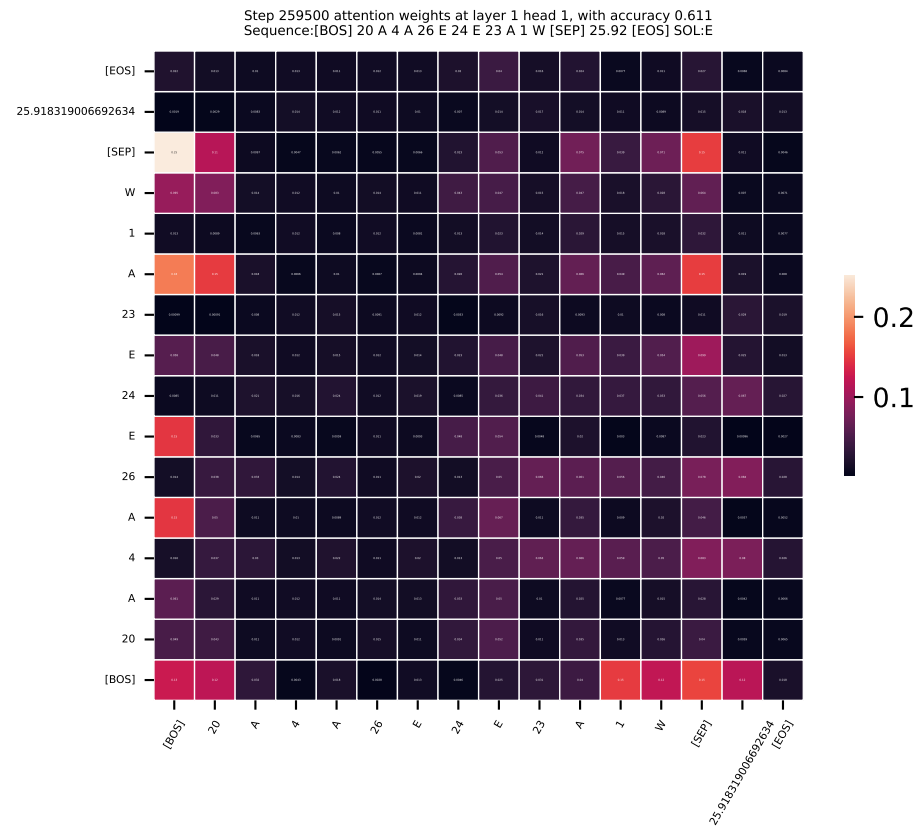

Step 259500 attention weights at layer 0 head 0, with accuracy 0.611
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

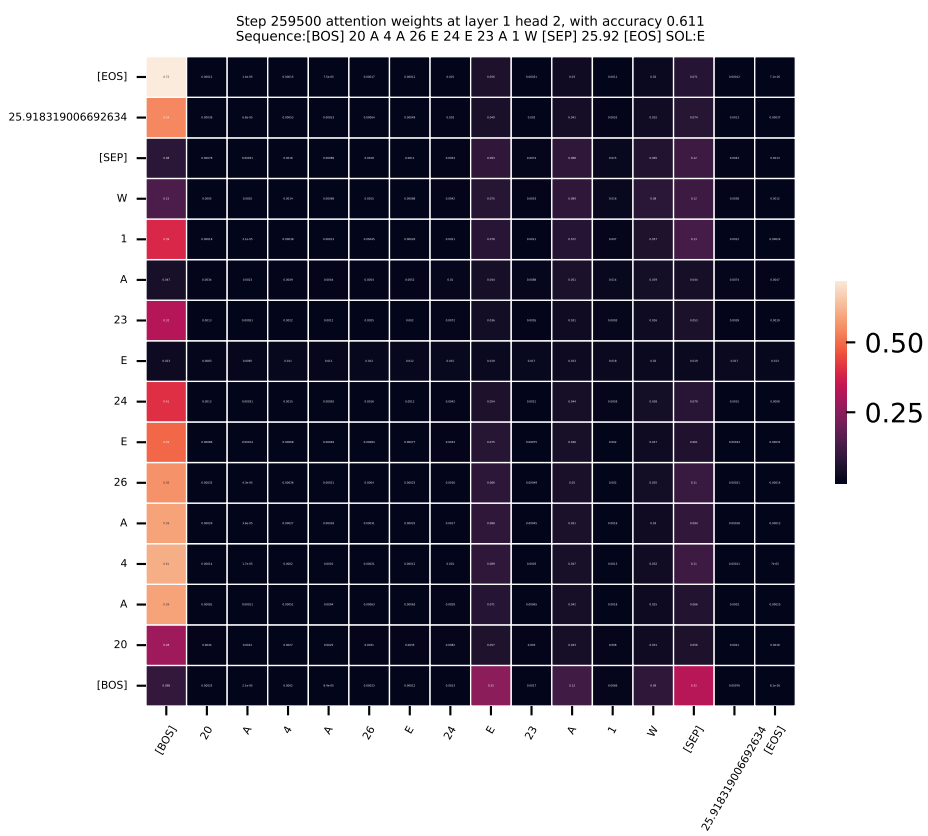

Step 259500 attention weights at layer 0 head 1, with accuracy 0.611
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

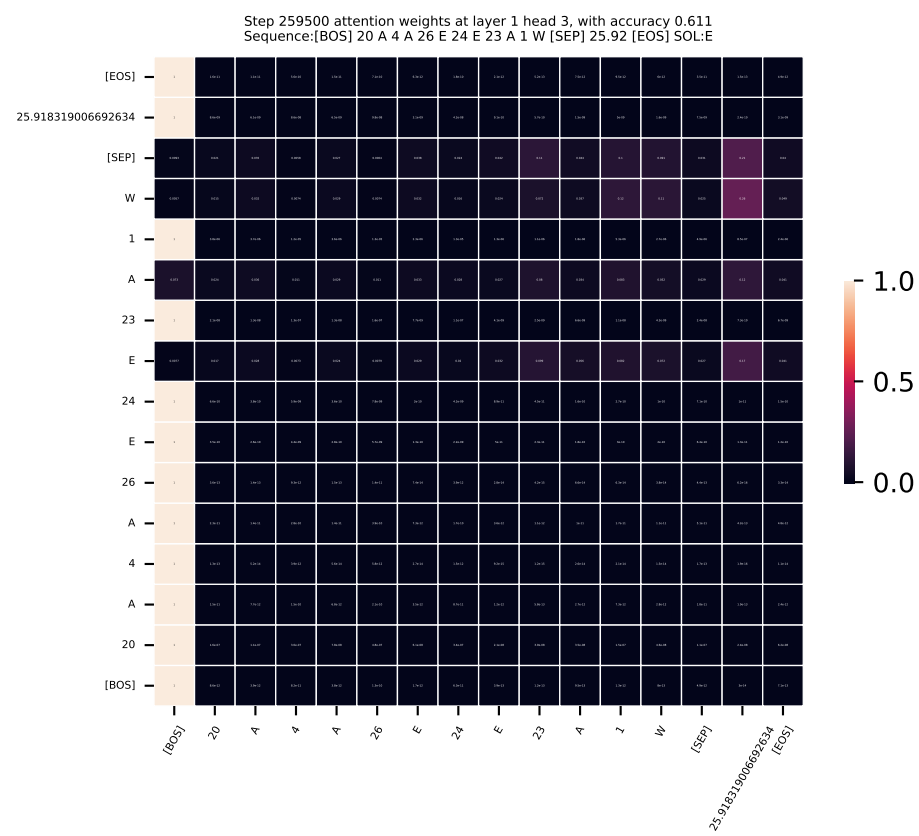

Step 259500 attention weights at layer 0 head 2, with accuracy 0.611
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

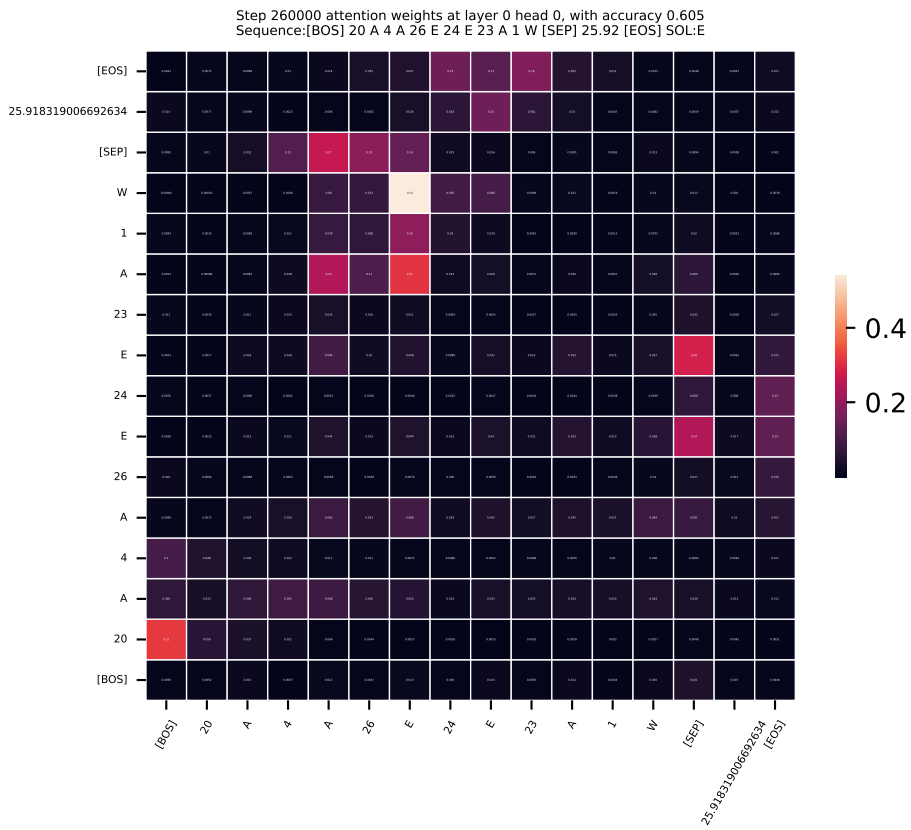

Step 259500 attention weights at layer 0 head 3, with accuracy 0.611
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

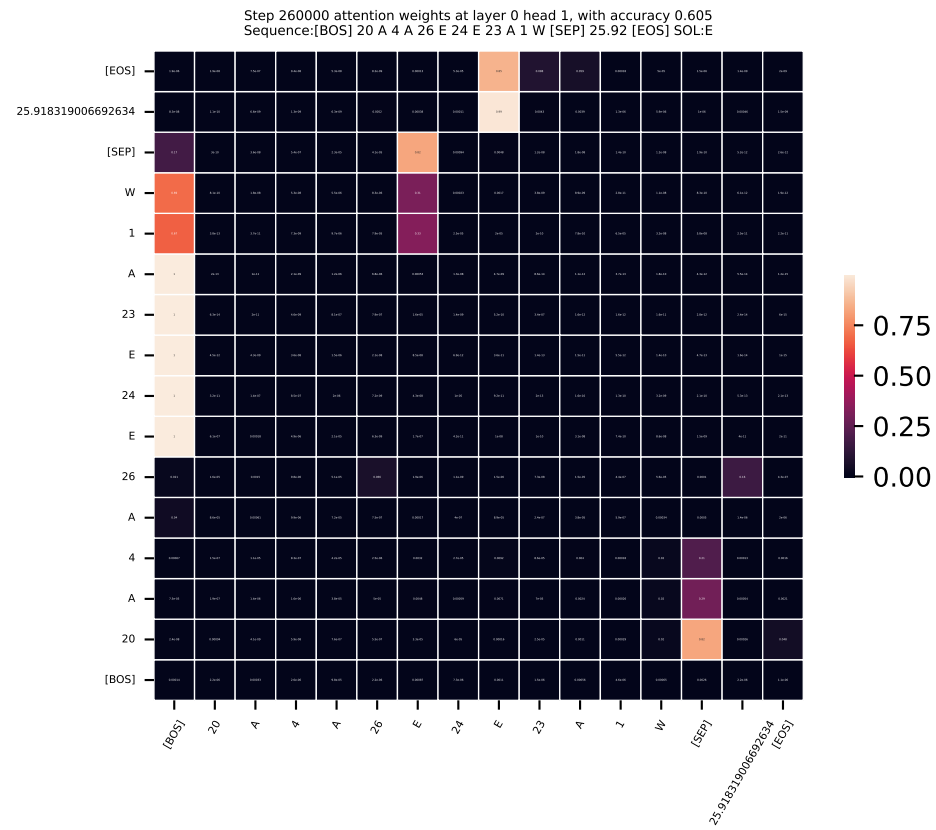

Step 259500 attention weights at layer 1 head 0, with accuracy 0.611
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

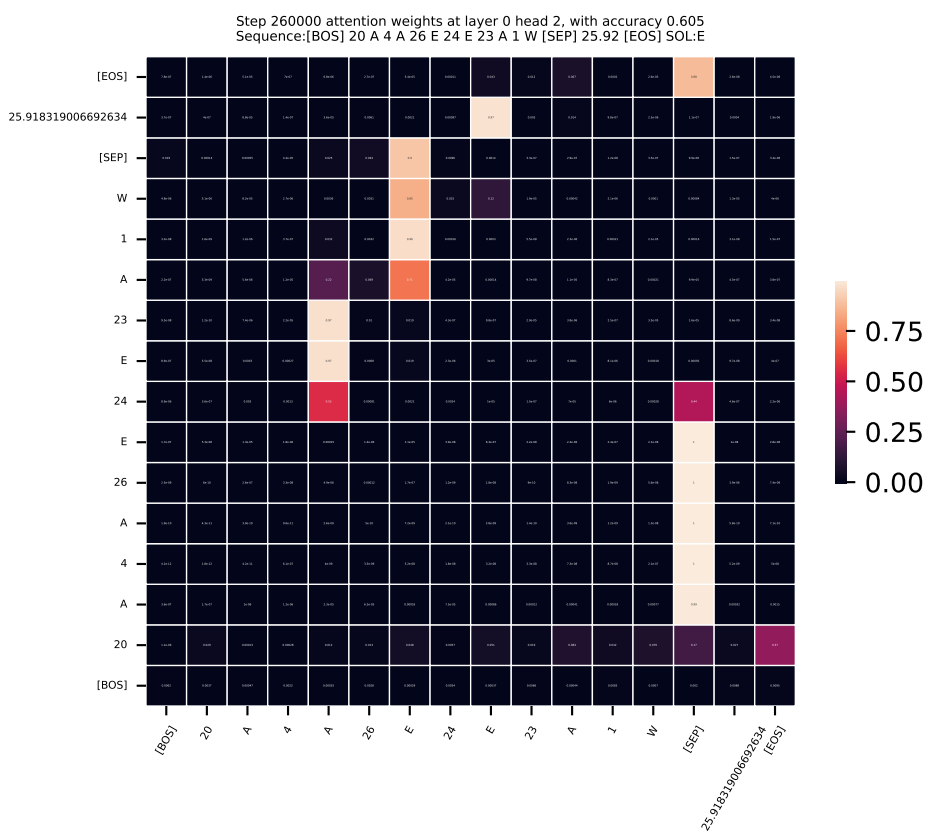

Step 259500 attention weights at layer 1 head 1, with accuracy 0.611
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.2
0.1

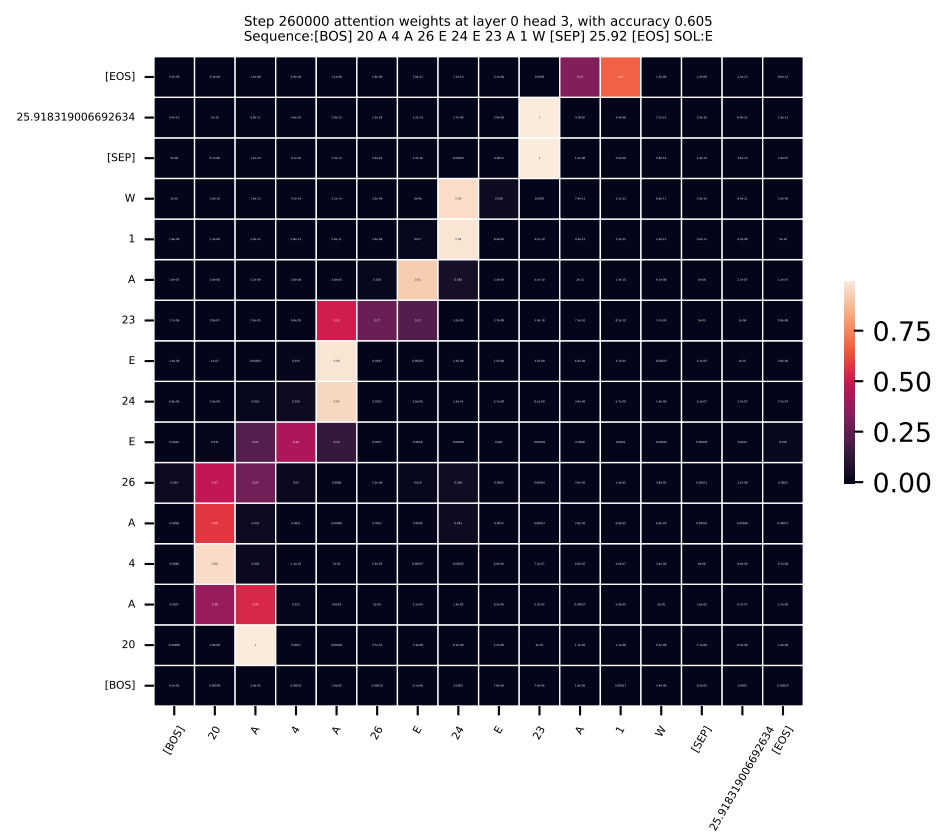

Step 259500 attention weights at layer 1 head 2, with accuracy 0.611
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

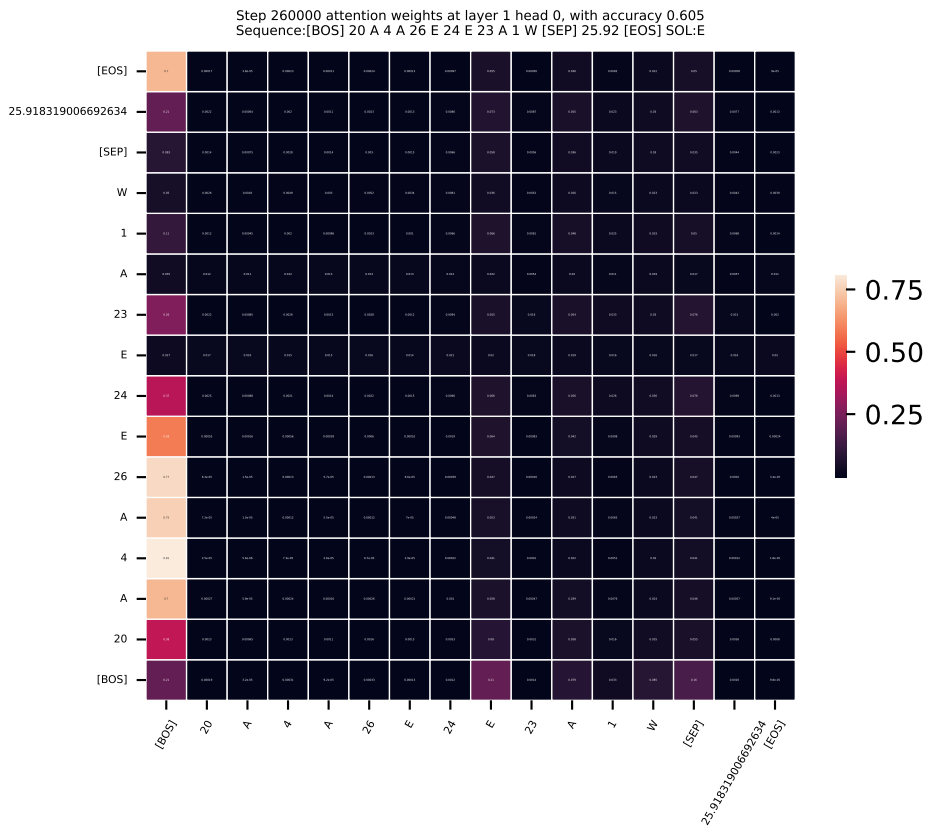

Step 259500 attention weights at layer 1 head 3, with accuracy 0.611
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
1.0
0.5
0.0

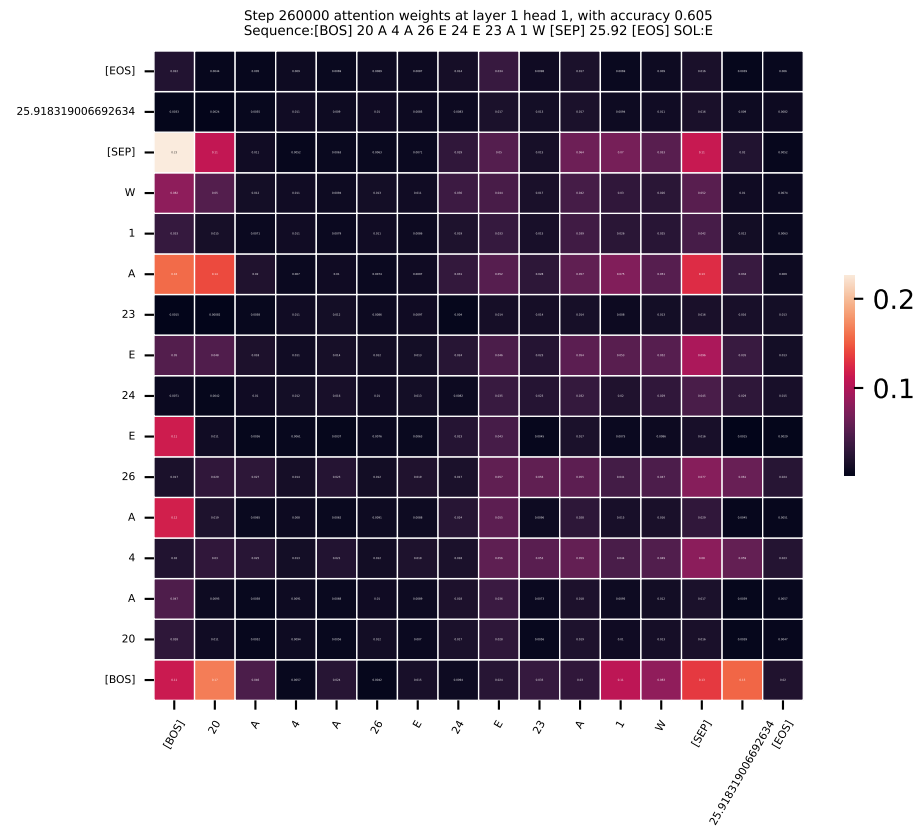

Step 260000 attention weights at layer 0 head 0, with accuracy 0.605
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

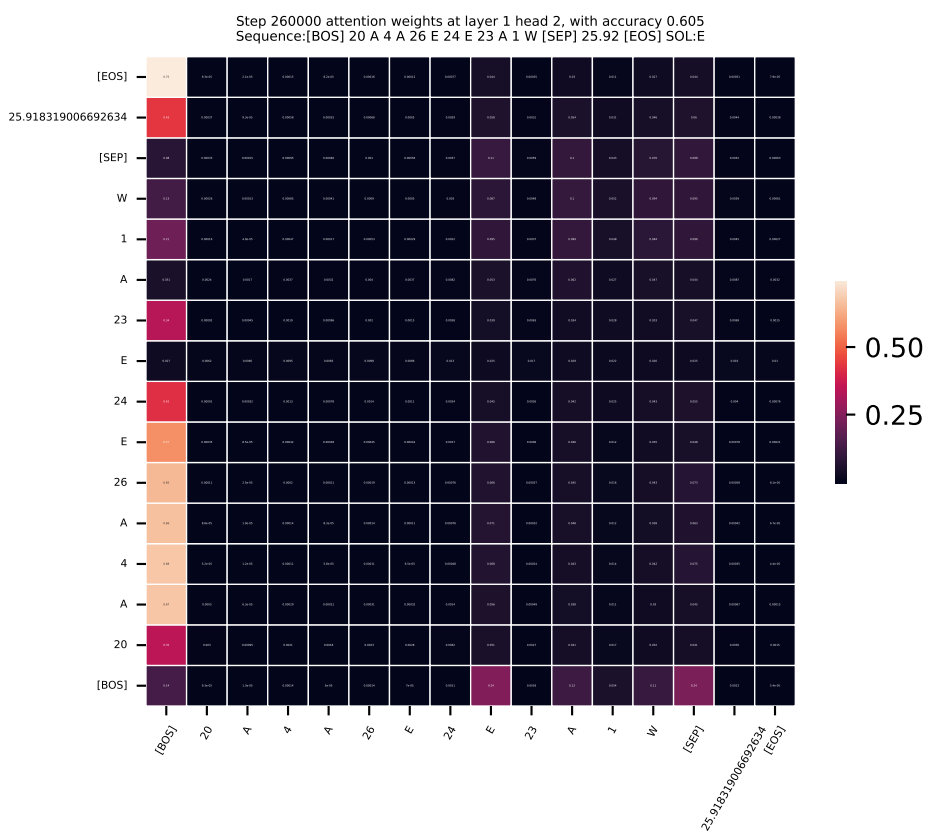

Step 260000 attention weights at layer 0 head 1, with accuracy 0.605
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

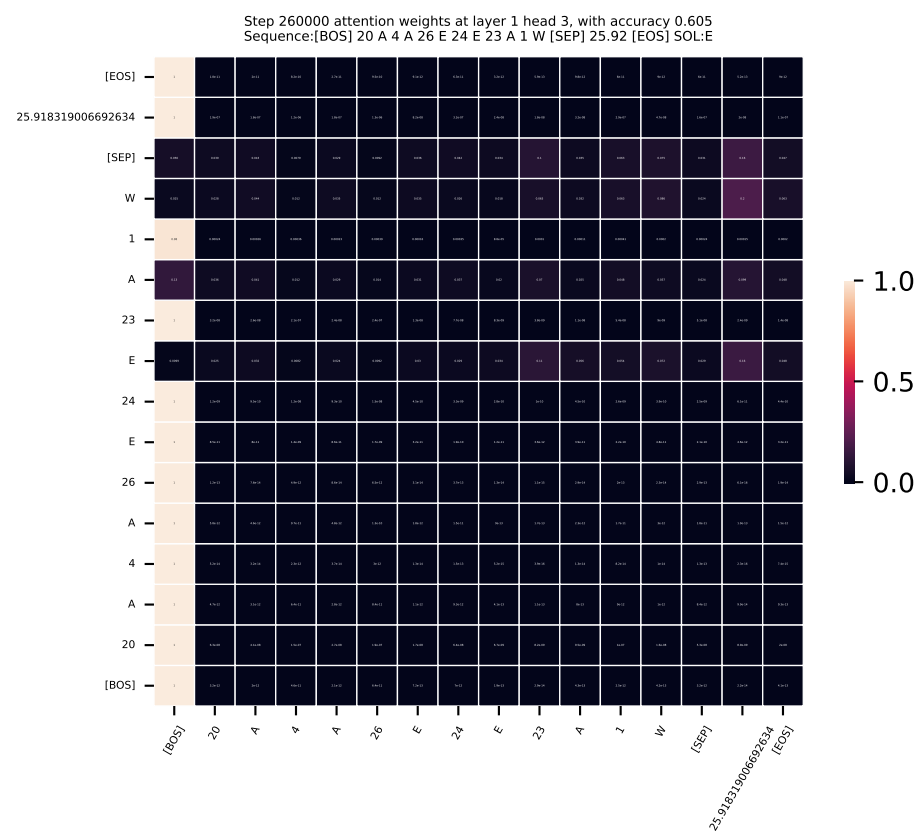

Step 260000 attention weights at layer 0 head 2, with accuracy 0.605
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

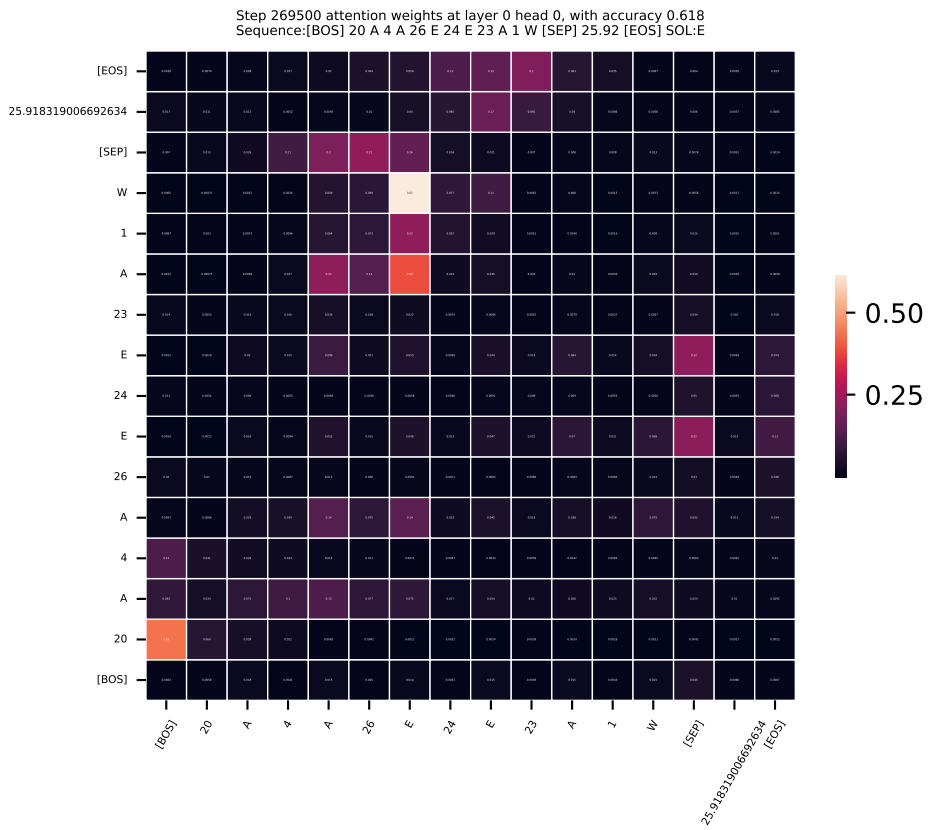

Step 260000 attention weights at layer 0 head 3, with accuracy 0.605
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

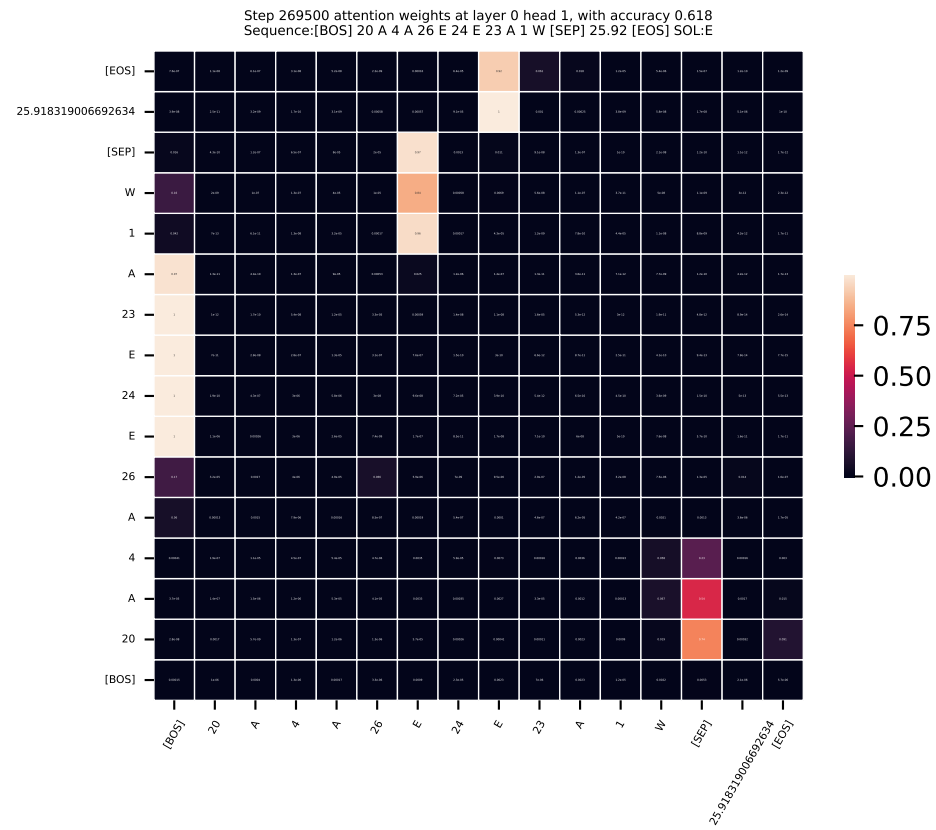

Step 260000 attention weights at layer 1 head 0, with accuracy 0.605
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

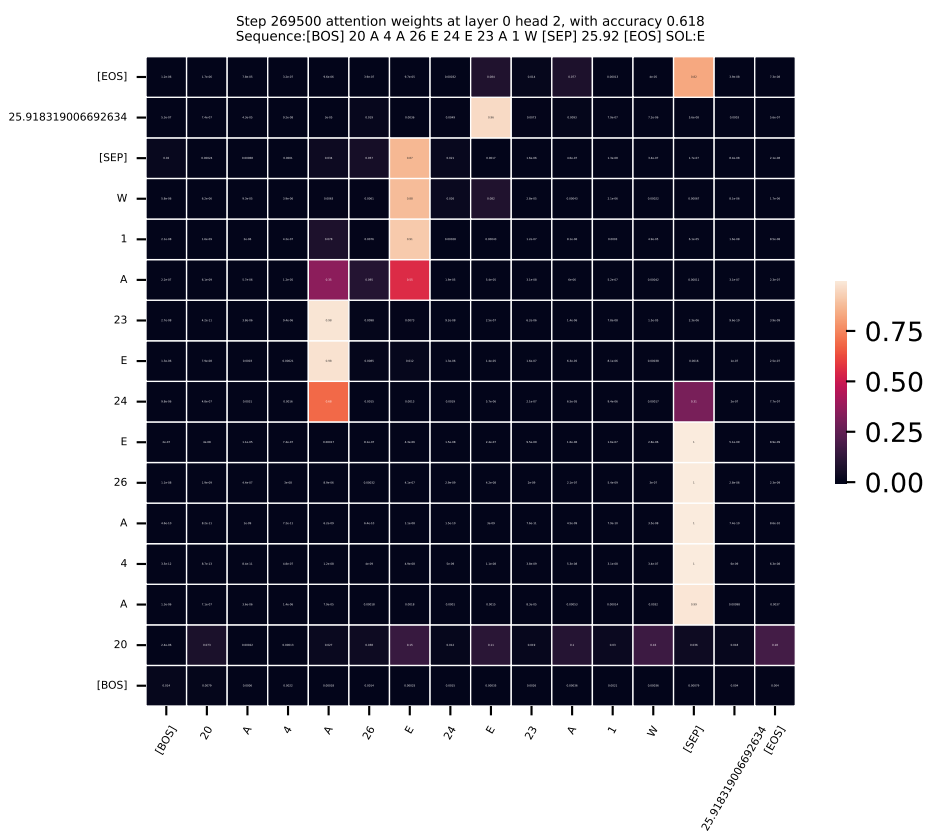

Step 260000 attention weights at layer 1 head 1, with accuracy 0.605
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.2
0.1

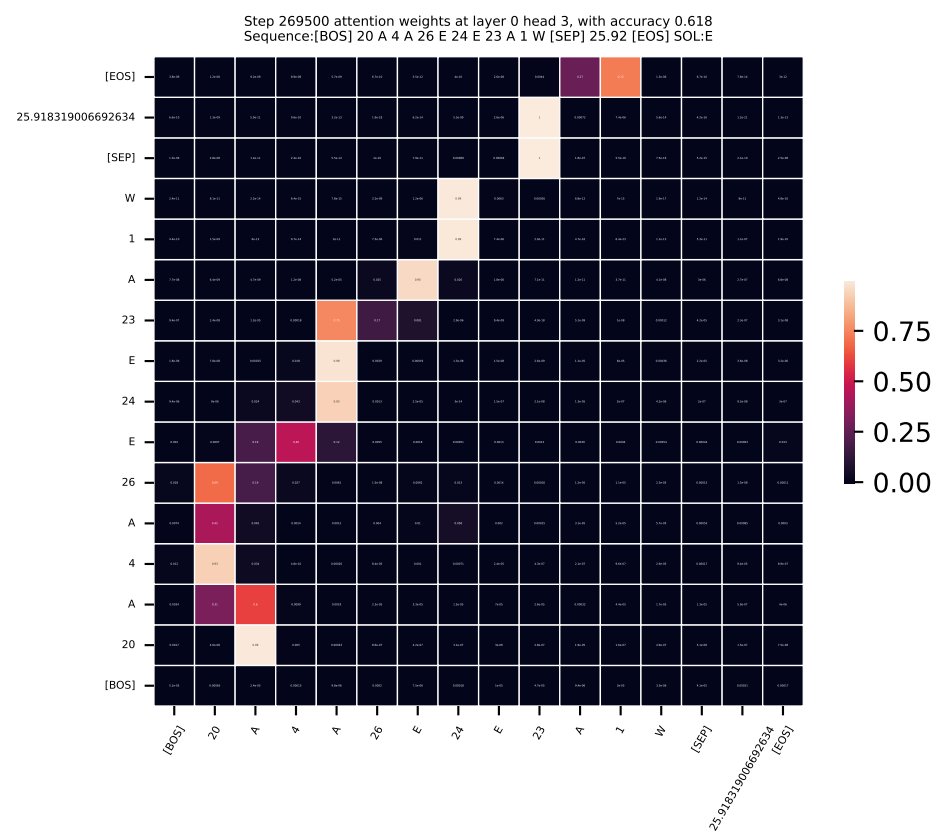

Step 260000 attention weights at layer 1 head 2, with accuracy 0.605
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

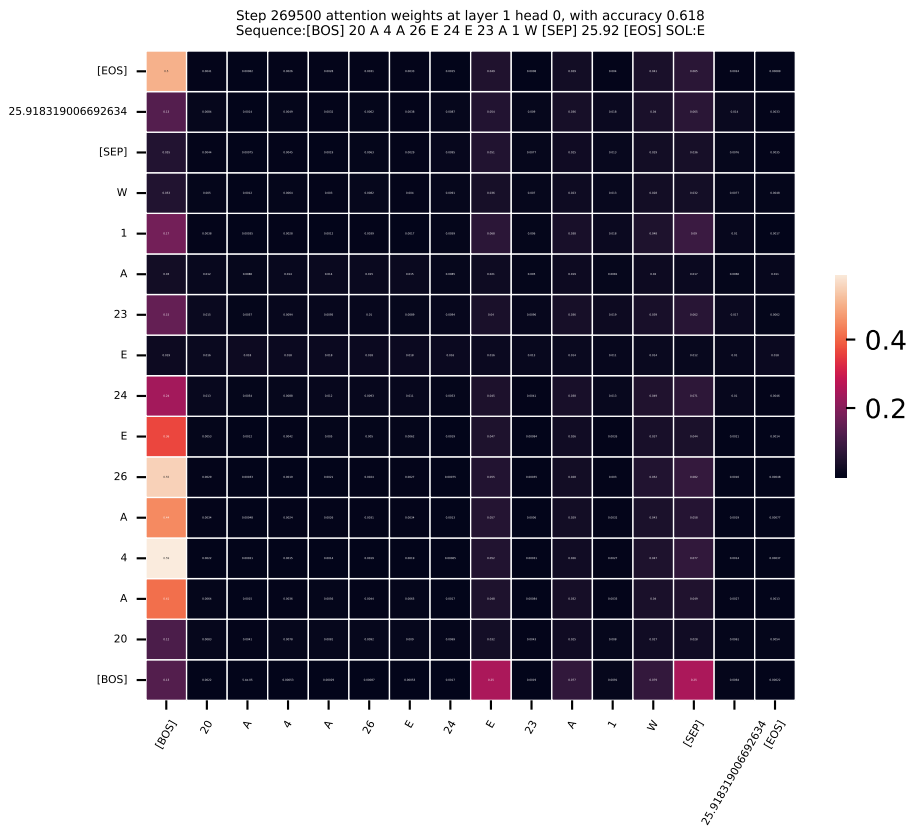

Step 260000 attention weights at layer 1 head 3, with accuracy 0.605
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
1.0
0.5
0.0

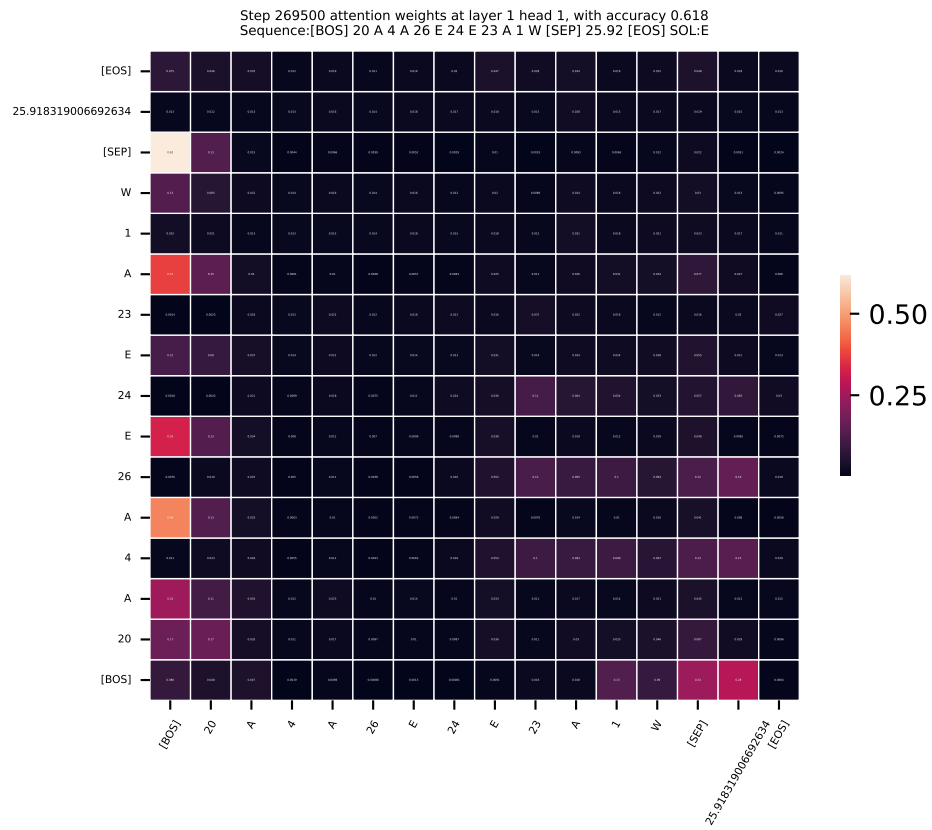

Step 269500 attention weights at layer 0 head 0, with accuracy 0.618
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

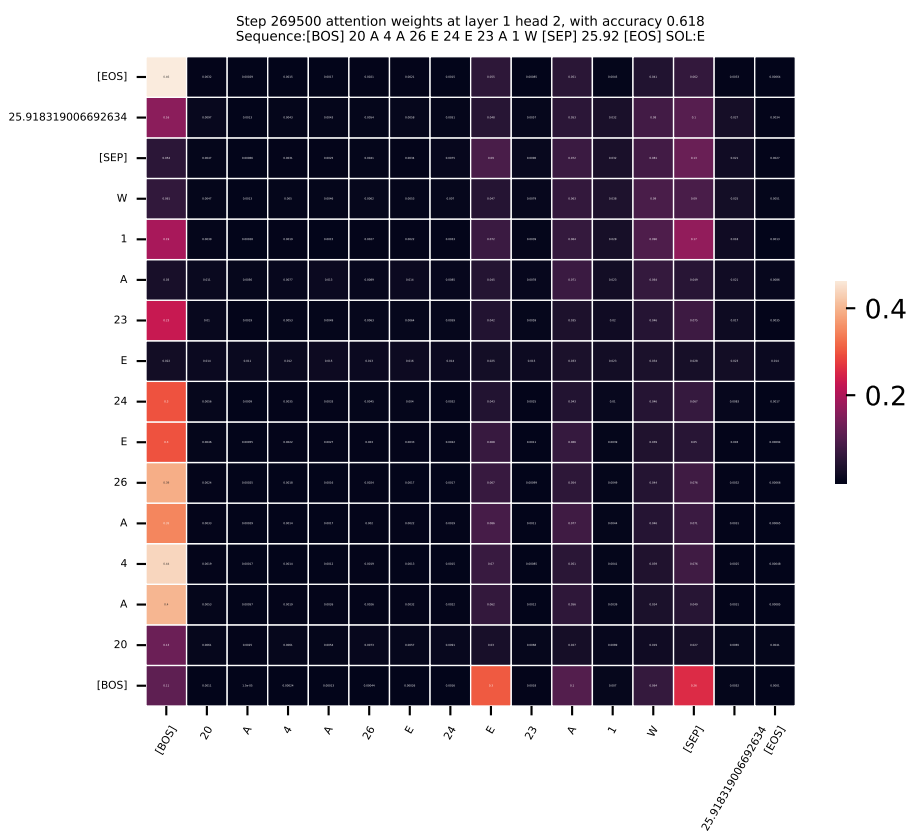

Step 269500 attention weights at layer 0 head 1, with accuracy 0.618
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

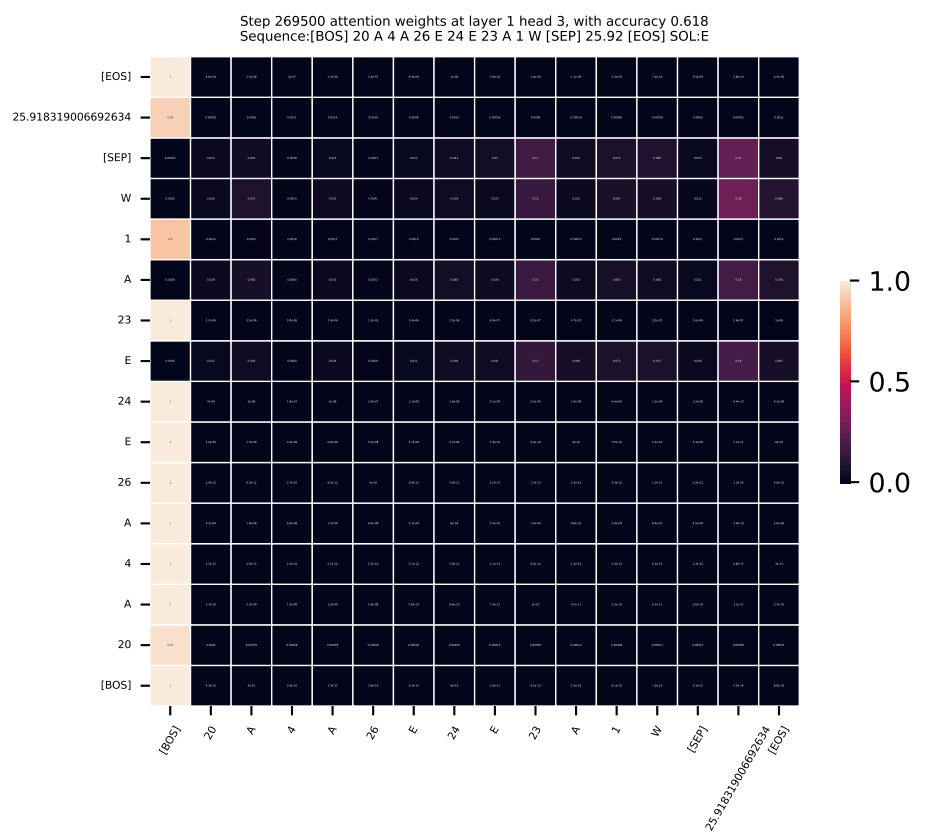

Step 269500 attention weights at layer 0 head 2, with accuracy 0.618
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

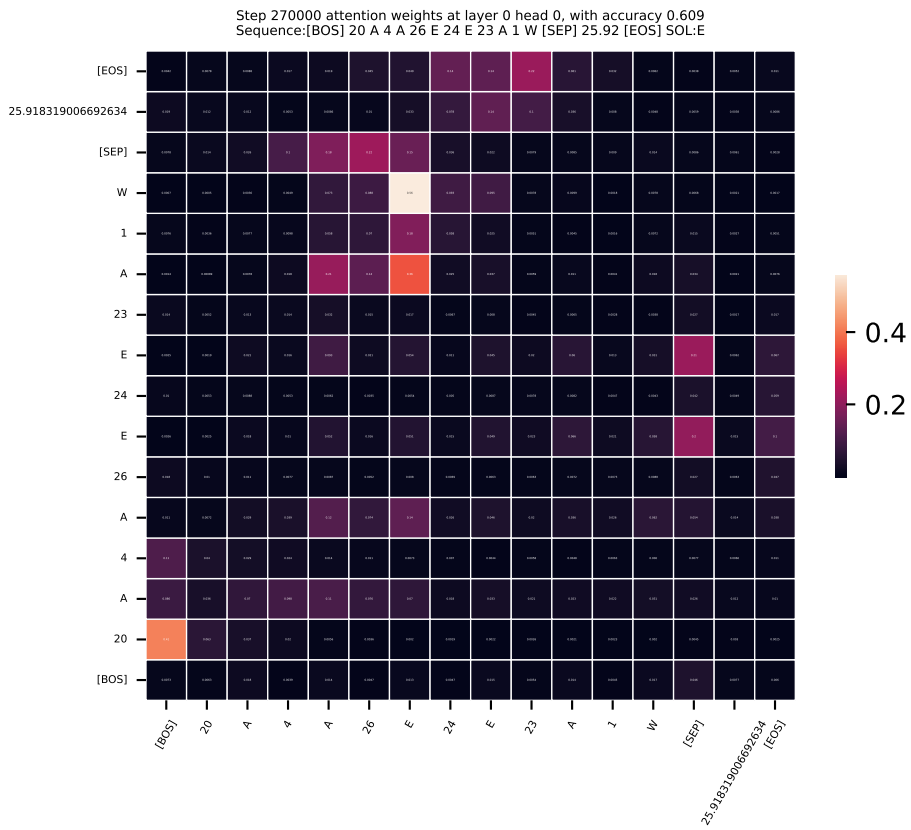

Step 269500 attention weights at layer 0 head 3, with accuracy 0.618
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

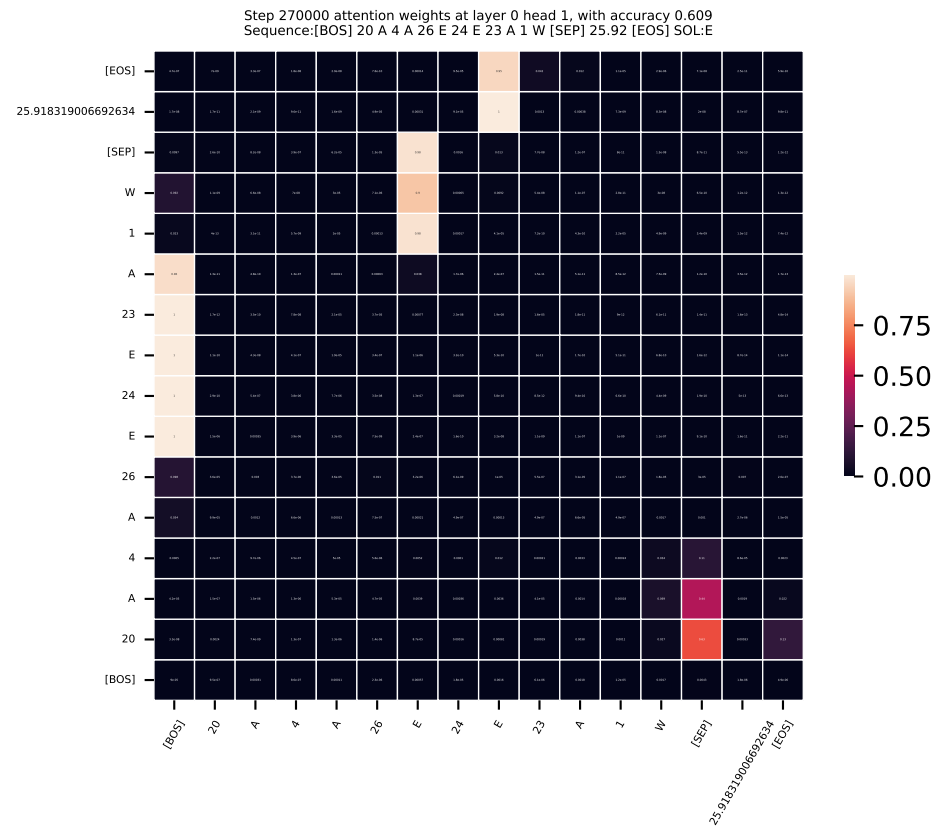

Step 269500 attention weights at layer 1 head 0, with accuracy 0.618
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.4
0.2

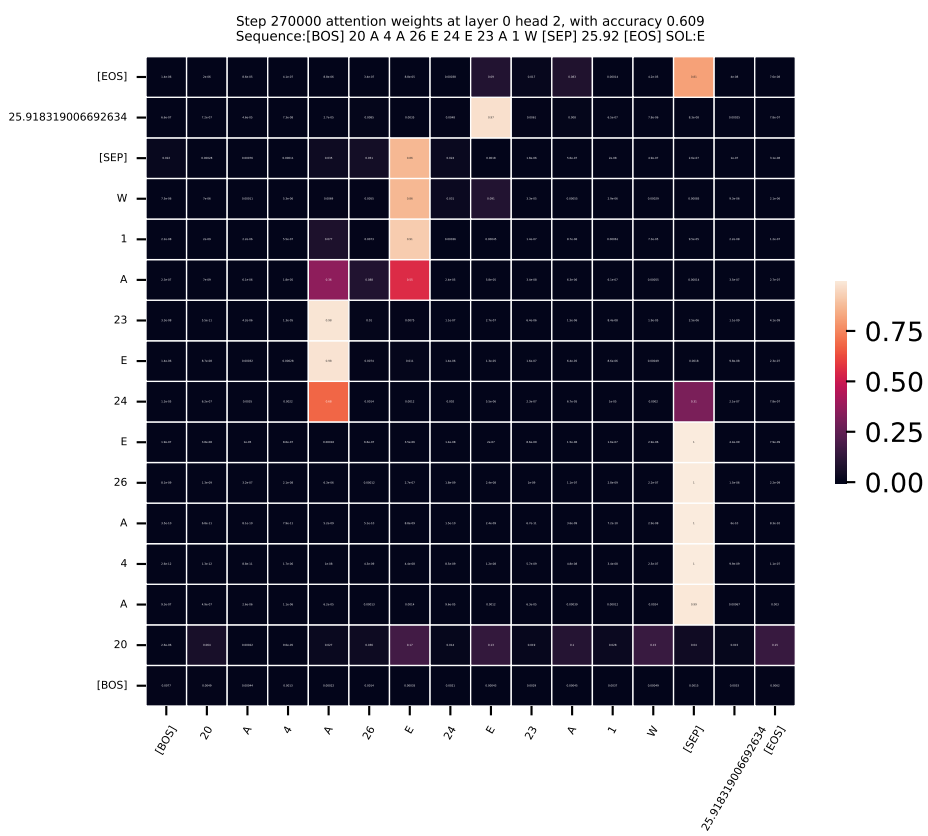

Step 269500 attention weights at layer 1 head 1, with accuracy 0.618
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

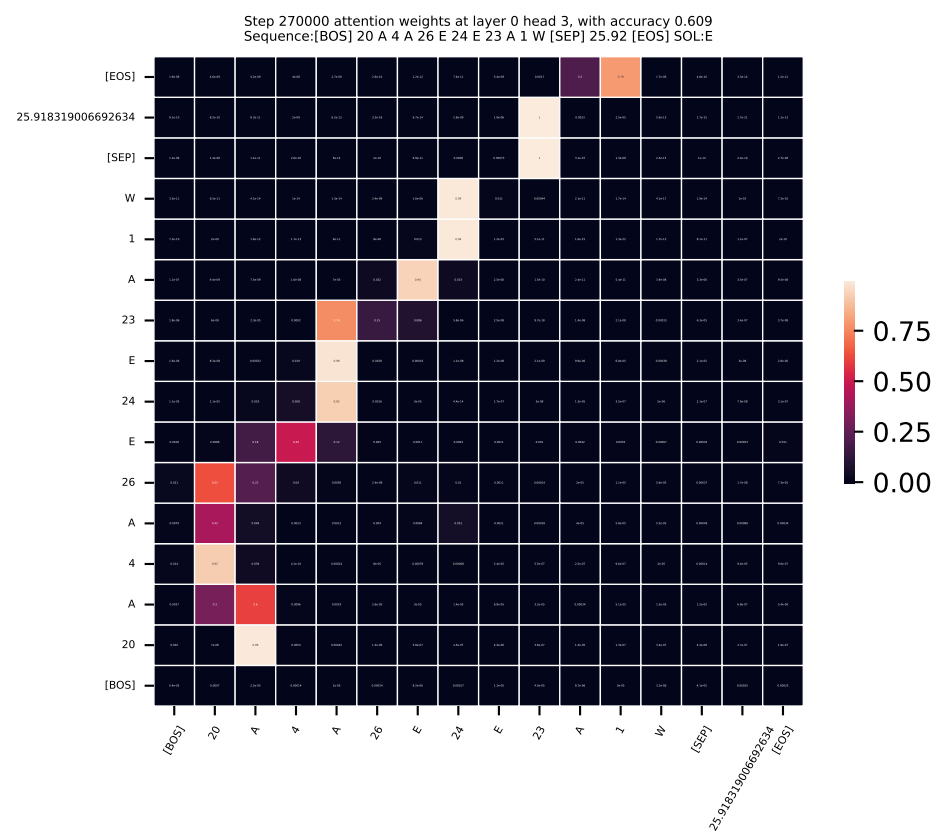

Step 269500 attention weights at layer 1 head 2, with accuracy 0.618
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

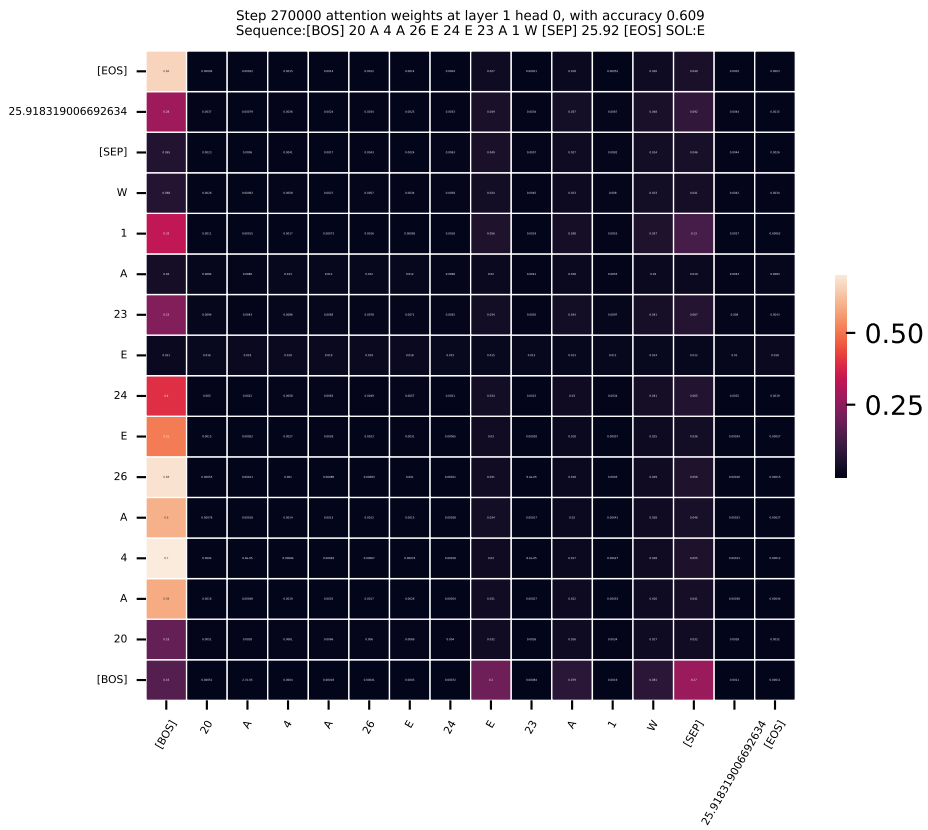

Step 269500 attention weights at layer 1 head 3, with accuracy 0.618
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
1.0
0.5
0.0

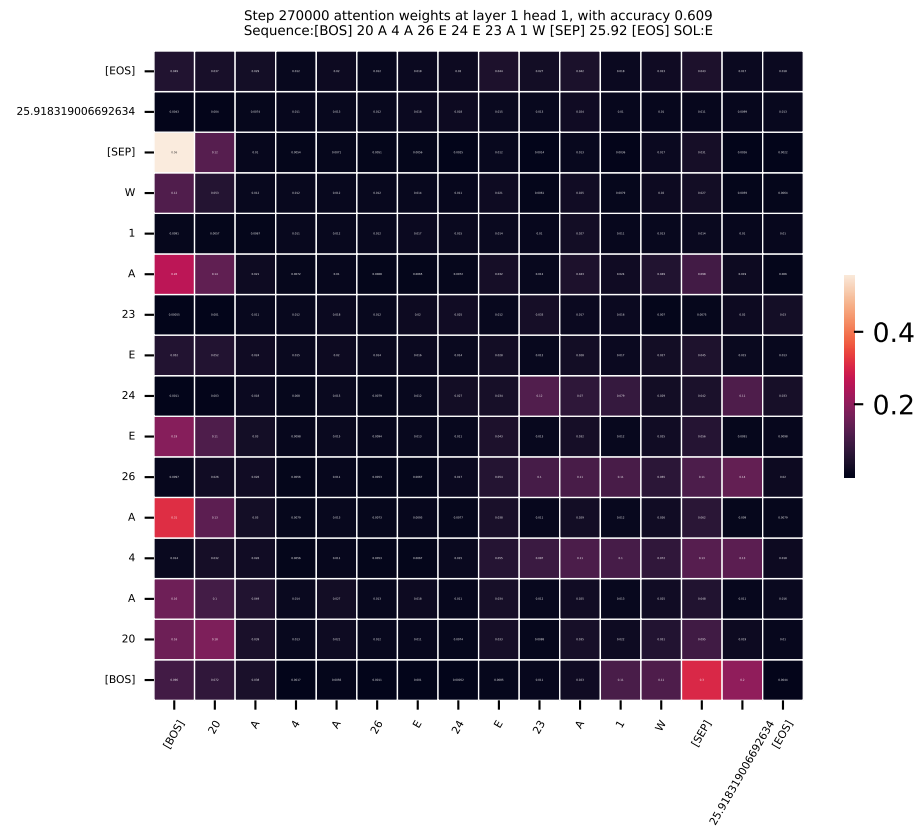

Step 270000 attention weights at layer 0 head 0, with accuracy 0.609
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

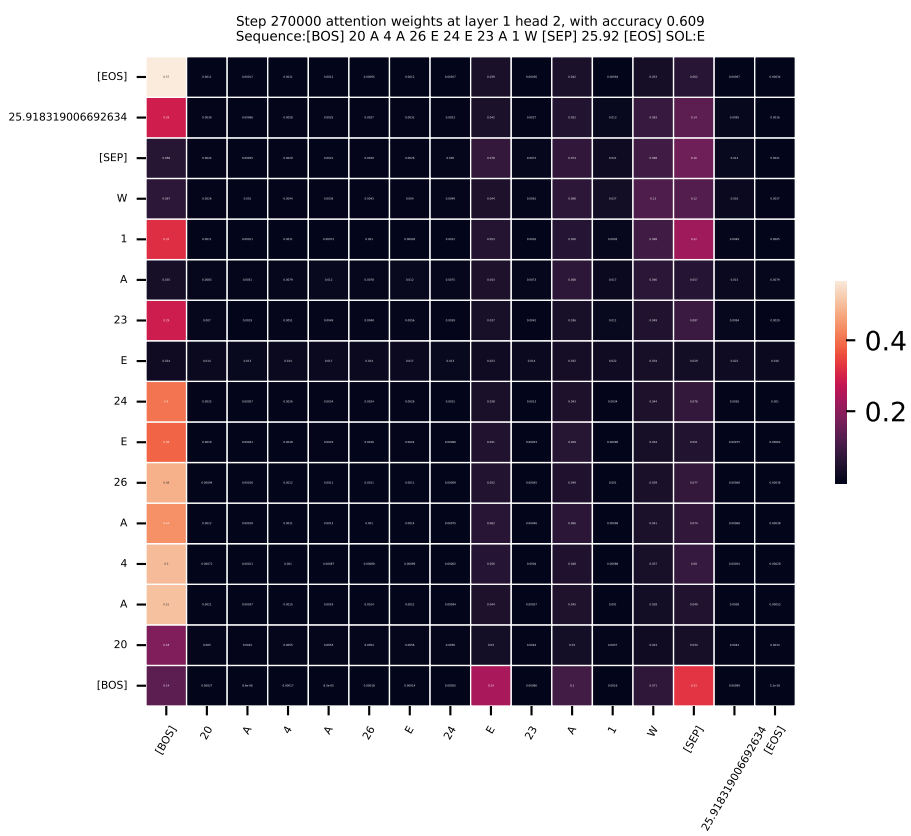

Step 270000 attention weights at layer 0 head 1, with accuracy 0.609
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

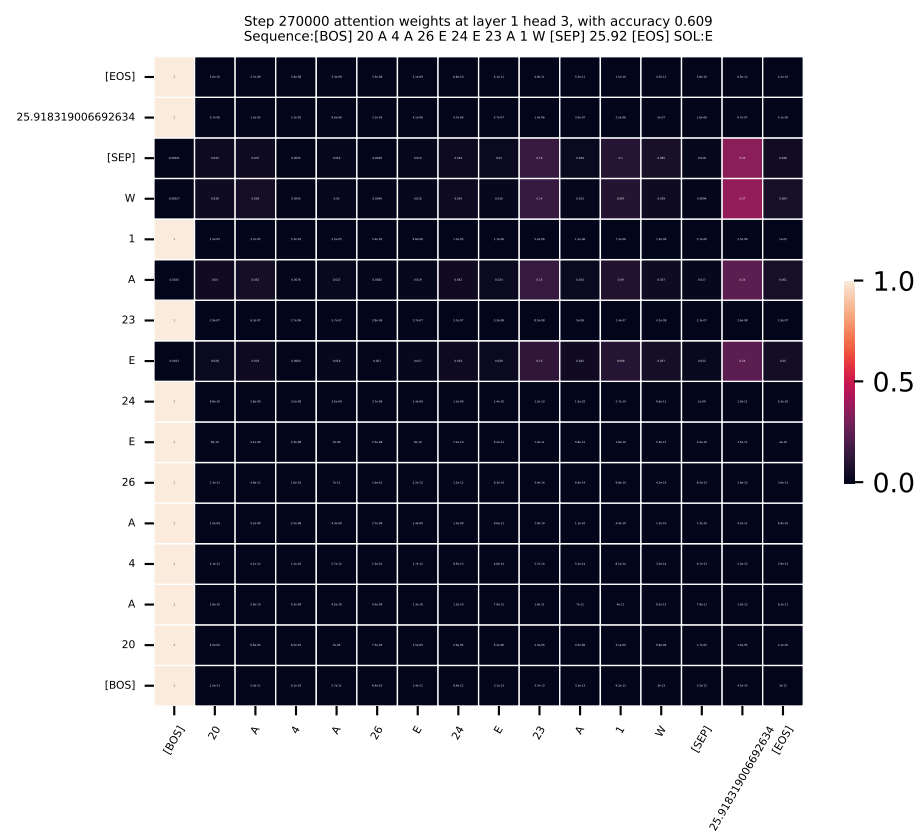

Step 270000 attention weights at layer 0 head 2, with accuracy 0.609
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

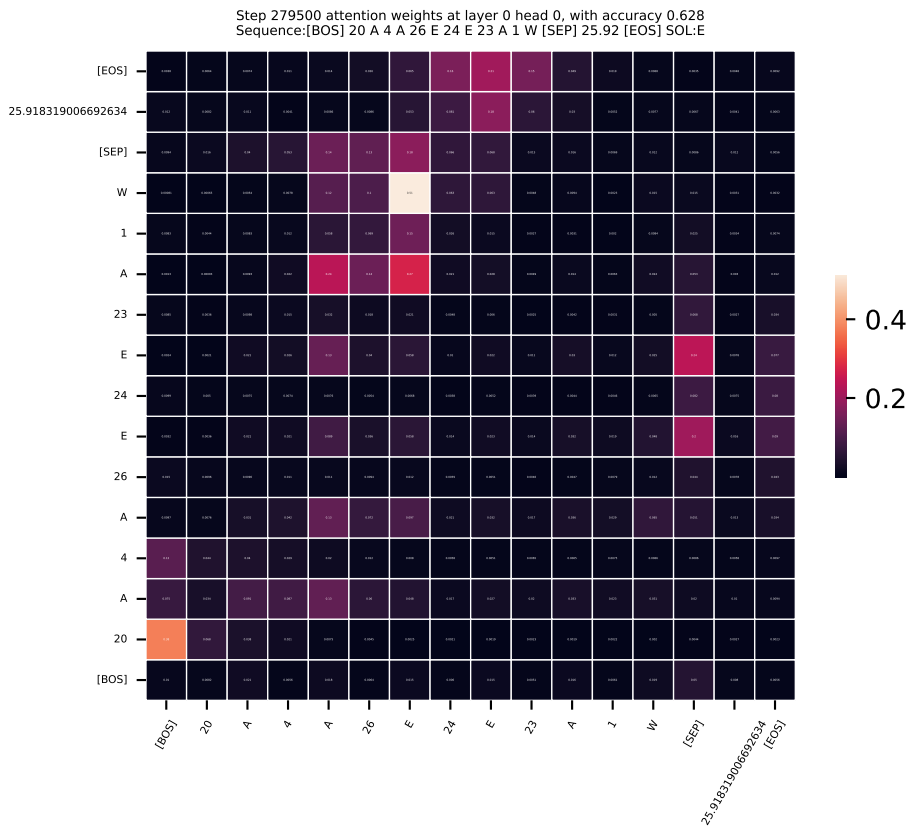

Step 270000 attention weights at layer 0 head 3, with accuracy 0.609
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

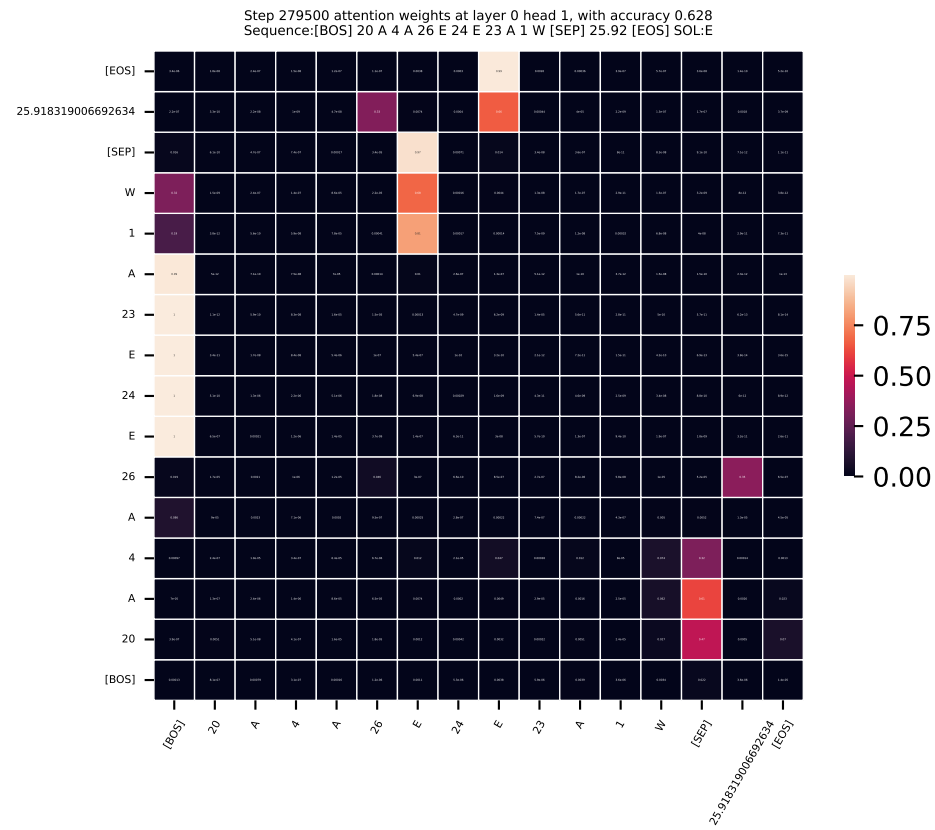

Step 270000 attention weights at layer 1 head 0, with accuracy 0.609
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

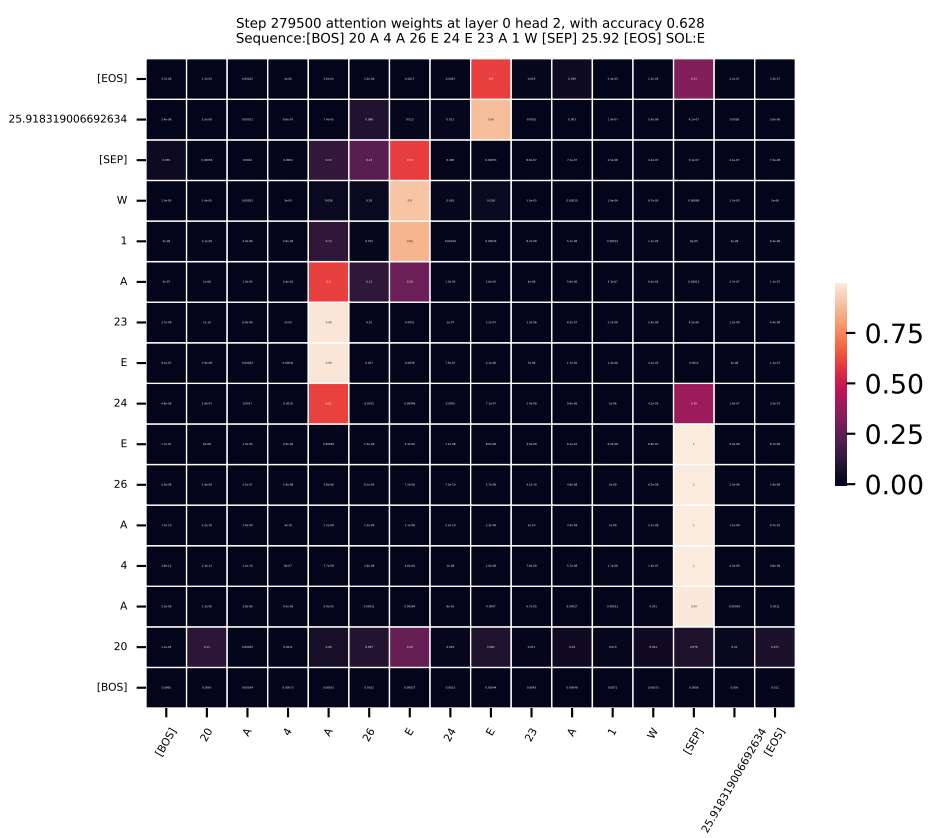

Step 270000 attention weights at layer 1 head 1, with accuracy 0.609
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

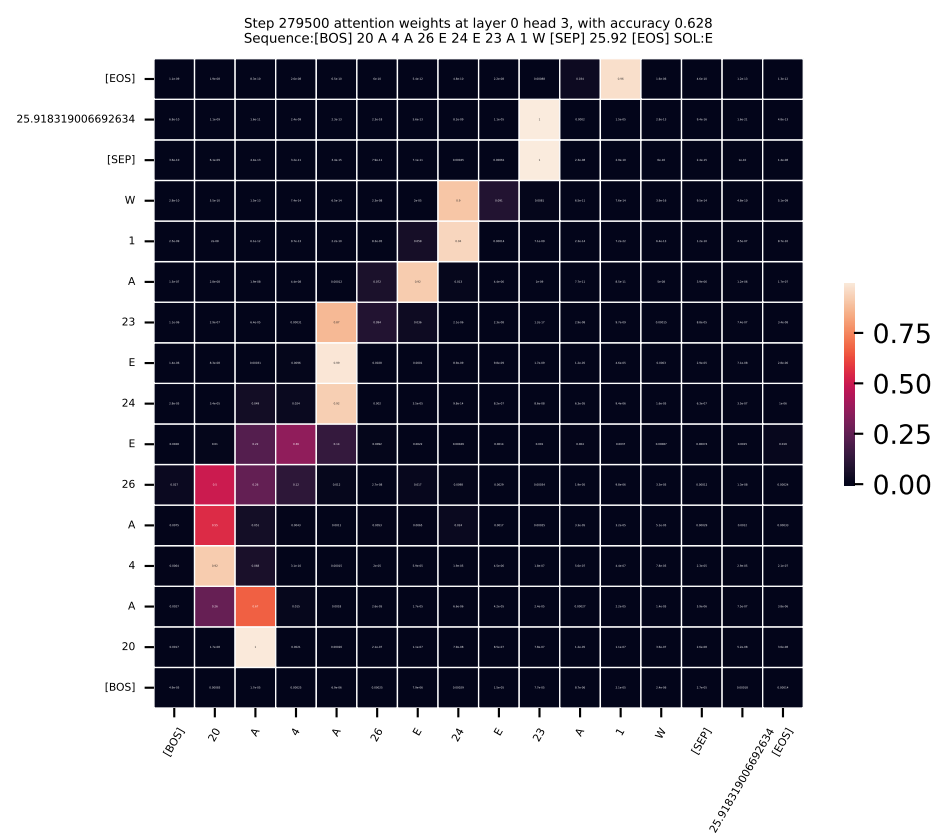

Step 270000 attention weights at layer 1 head 2, with accuracy 0.609
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

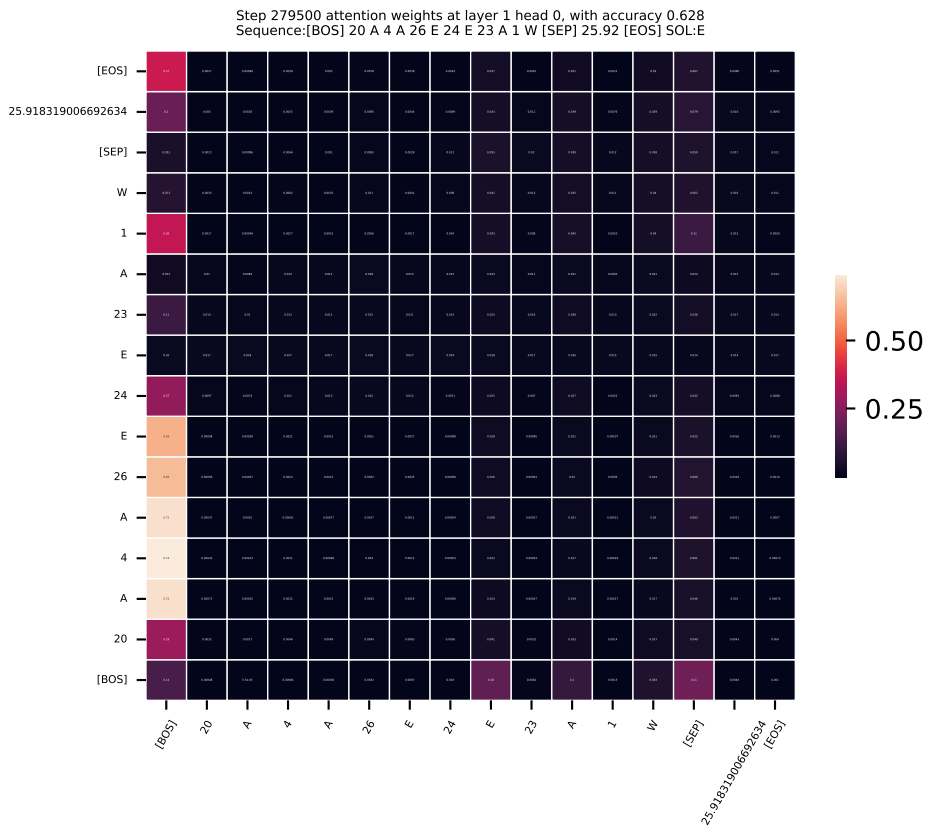

Step 270000 attention weights at layer 1 head 3, with accuracy 0.609
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
1.0
0.5
0.0

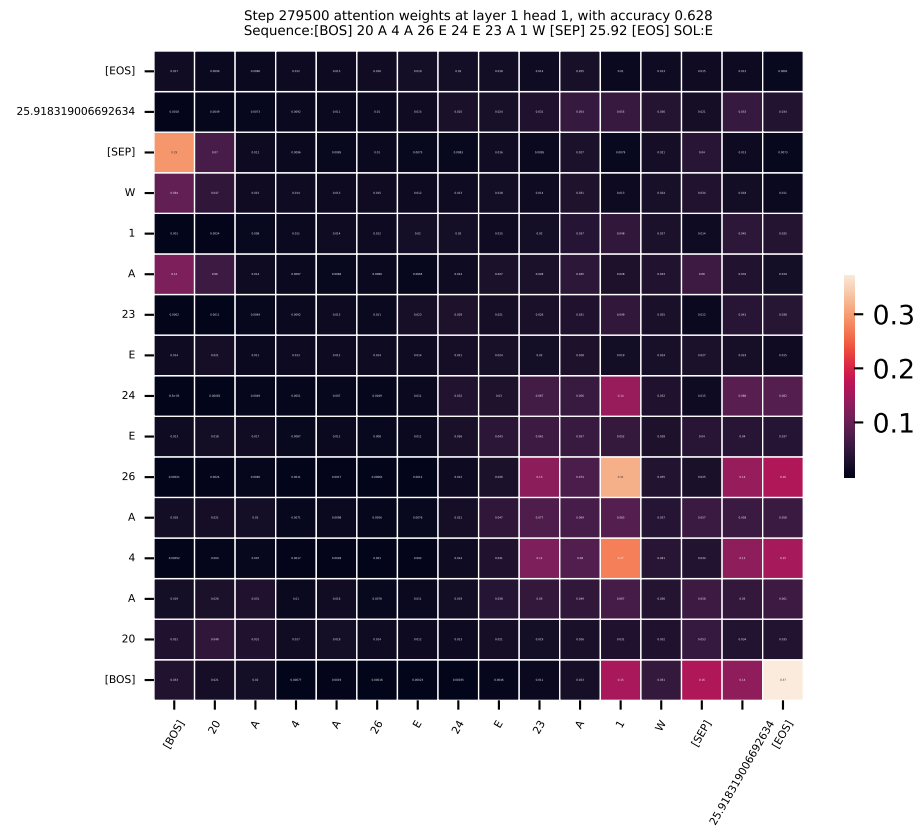

Step 279500 attention weights at layer 0 head 0, with accuracy 0.628
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

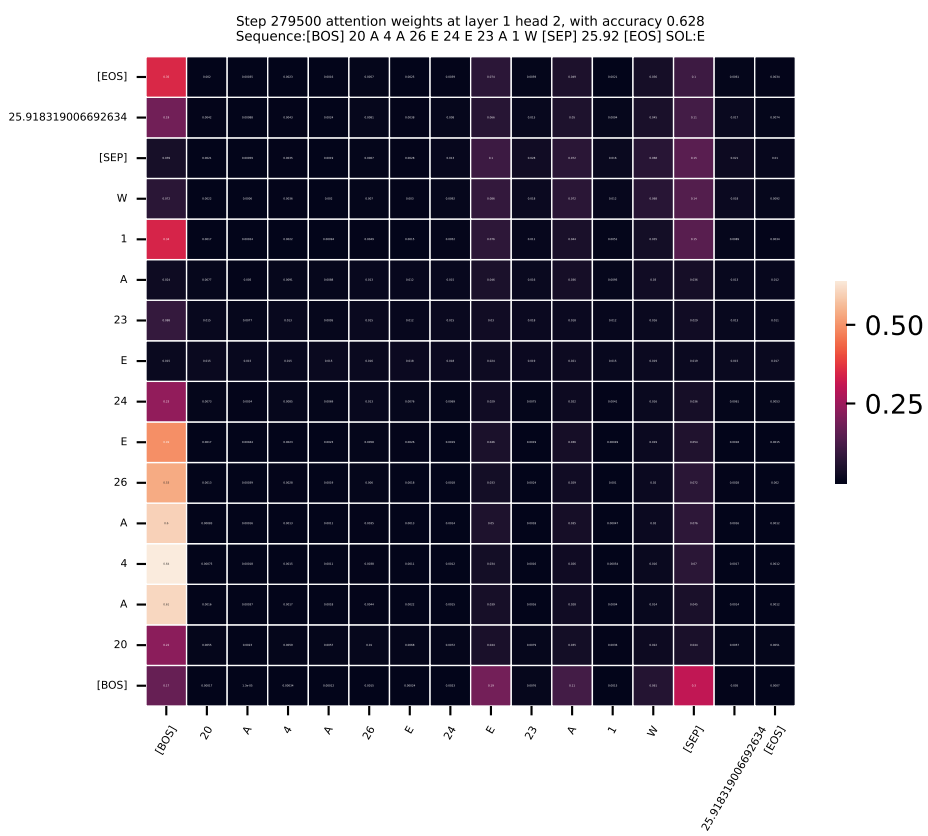

Step 279500 attention weights at layer 0 head 1, with accuracy 0.628
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

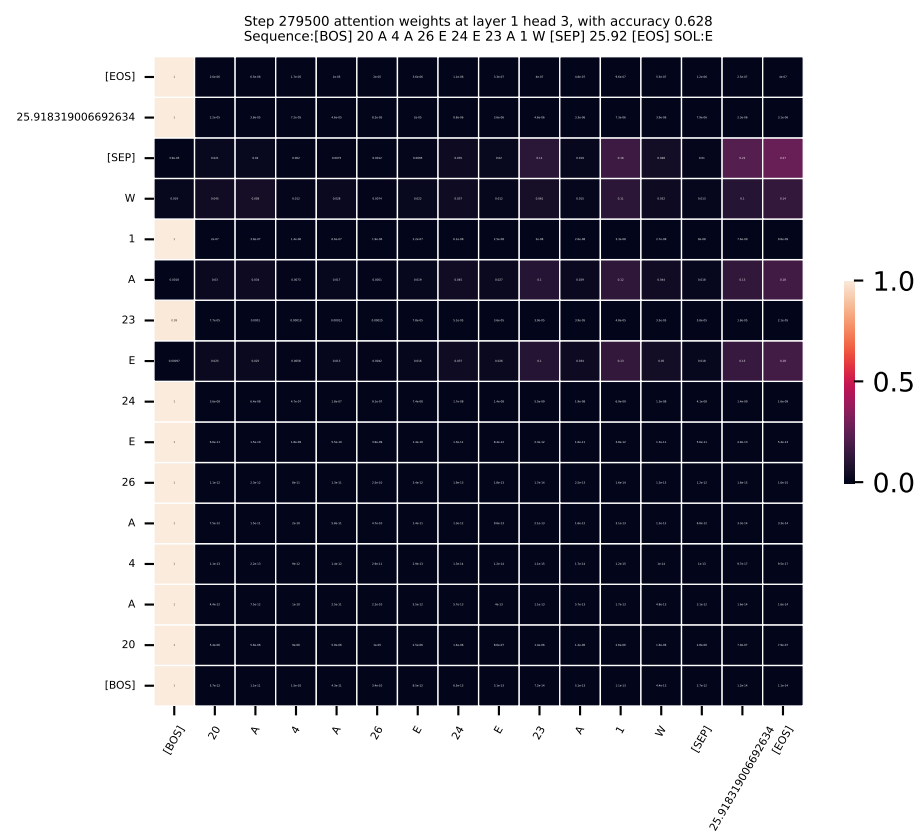

Step 279500 attention weights at layer 0 head 2, with accuracy 0.628
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

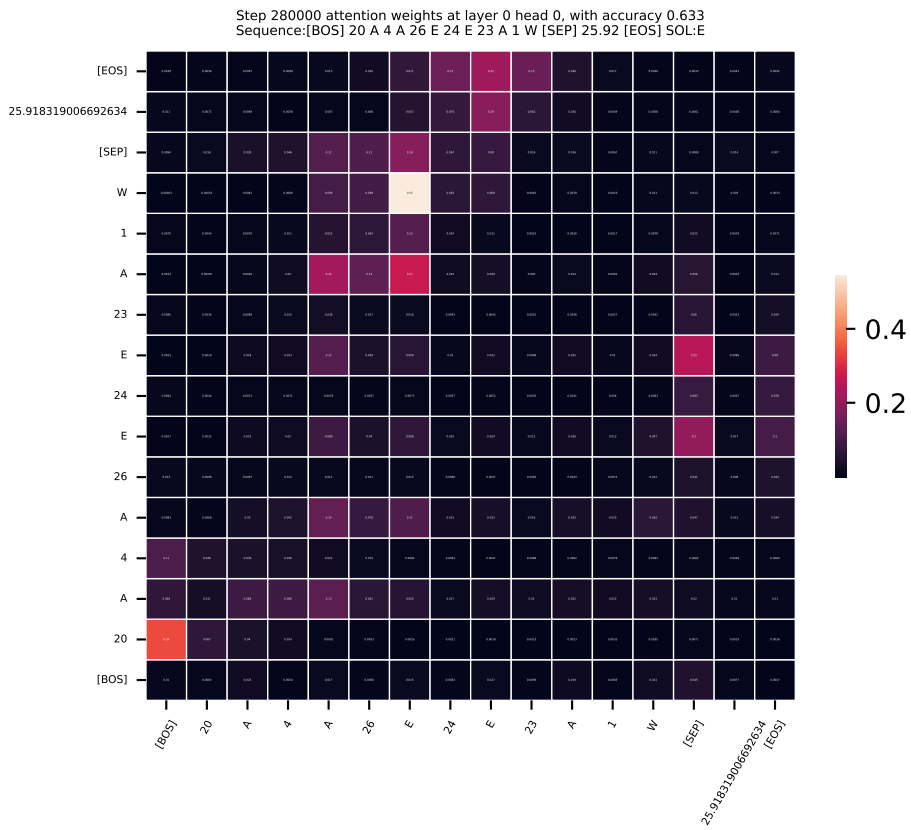

Step 279500 attention weights at layer 0 head 3, with accuracy 0.628
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

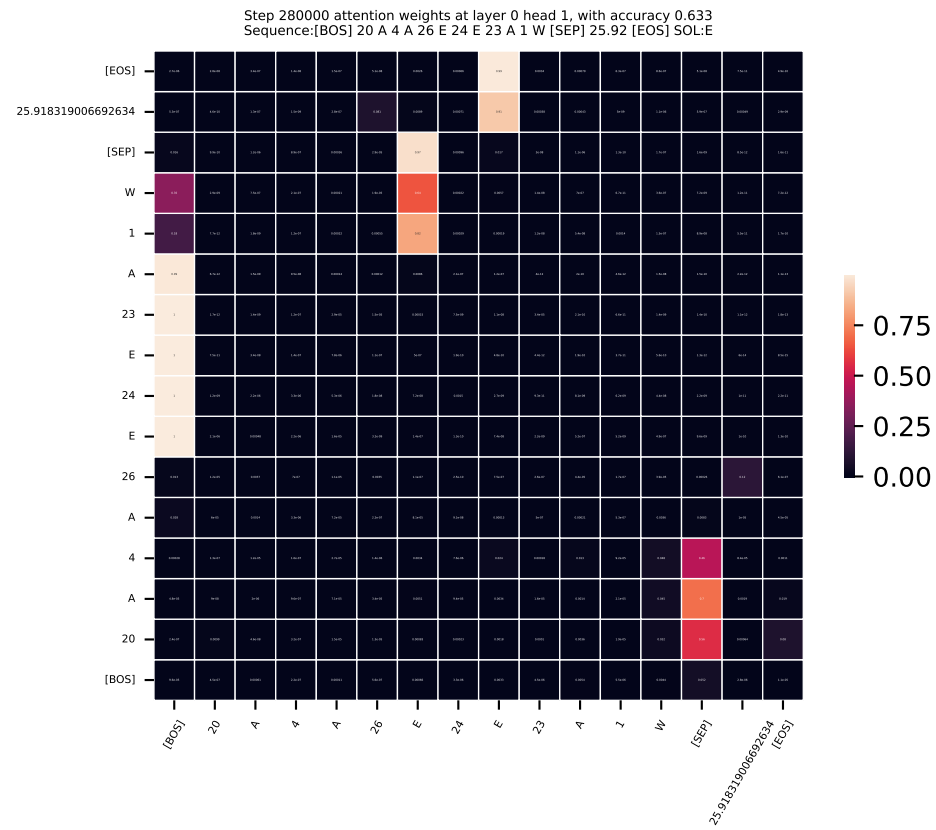

Step 279500 attention weights at layer 1 head 0, with accuracy 0.628
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

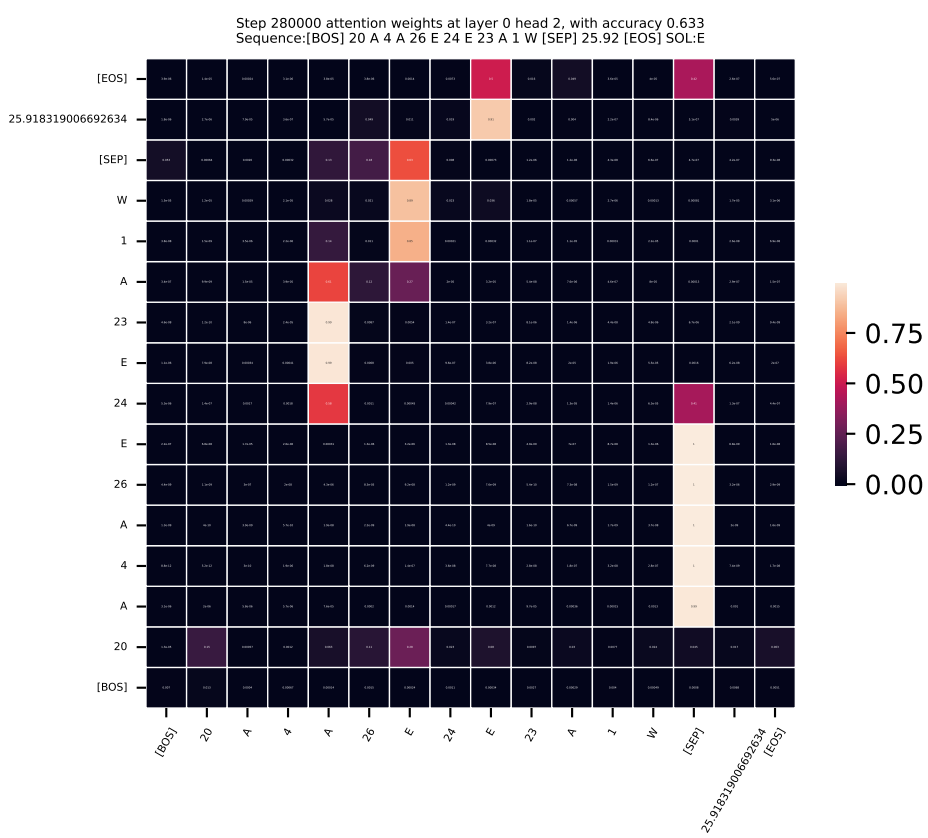

Step 279500 attention weights at layer 1 head 1, with accuracy 0.628
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.3
0.2
0.1

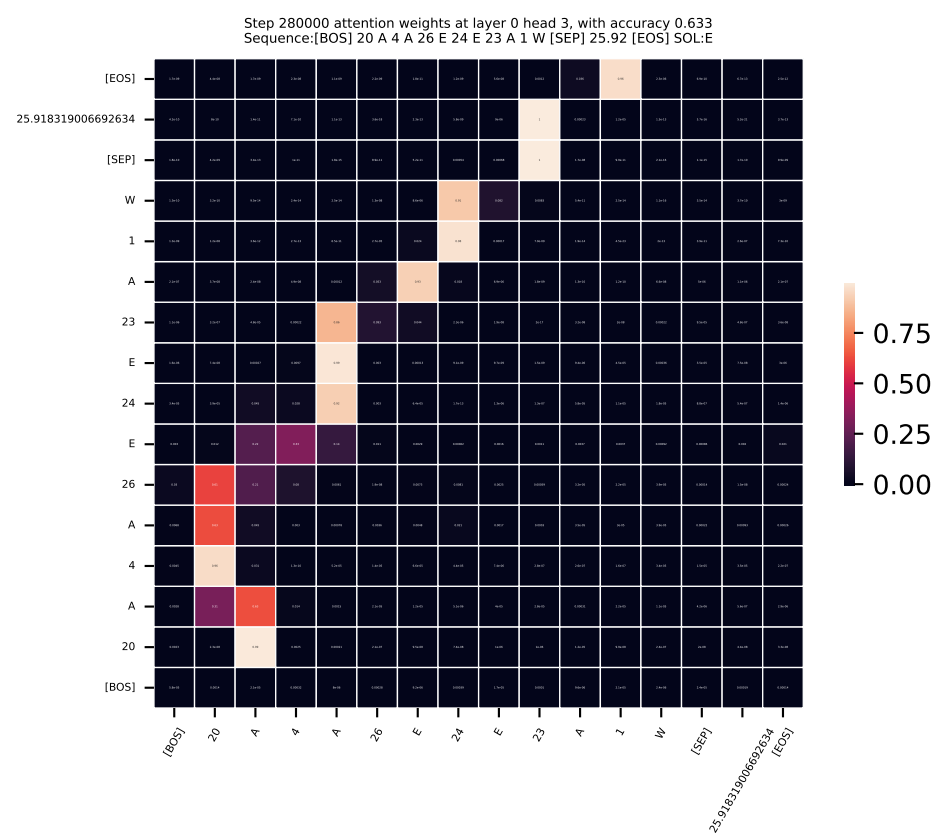

Step 279500 attention weights at layer 1 head 2, with accuracy 0.628
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

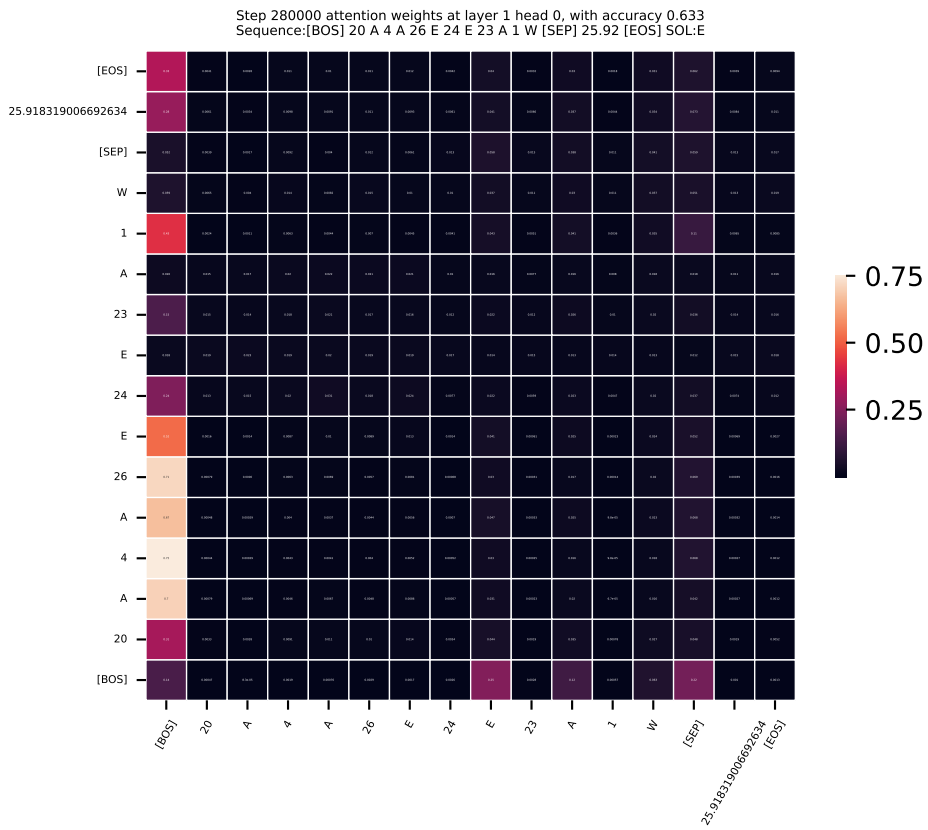

Step 279500 attention weights at layer 1 head 3, with accuracy 0.628
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
1.0
0.5
0.0

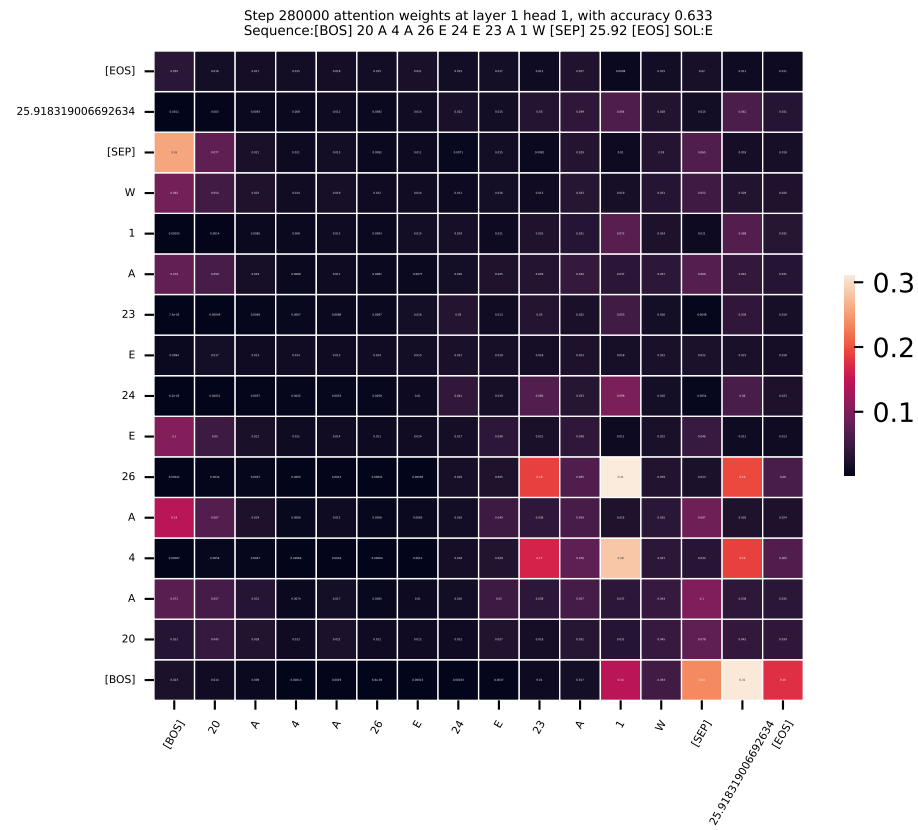

Step 280000 attention weights at layer 0 head 0, with accuracy 0.633
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

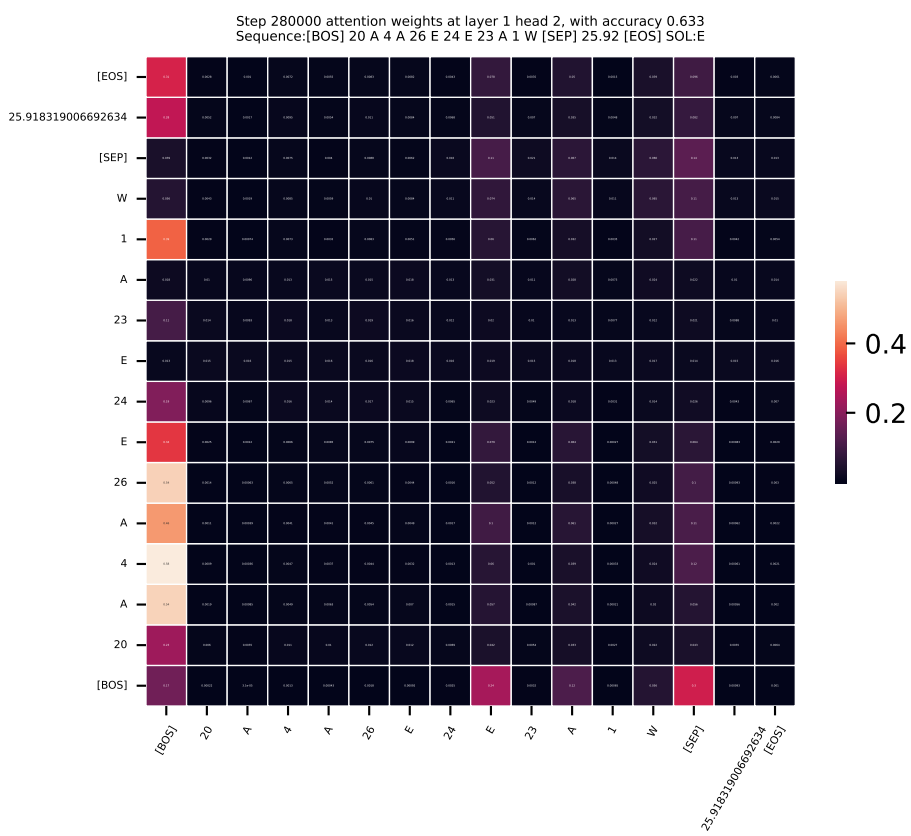

Step 280000 attention weights at layer 0 head 1, with accuracy 0.633
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

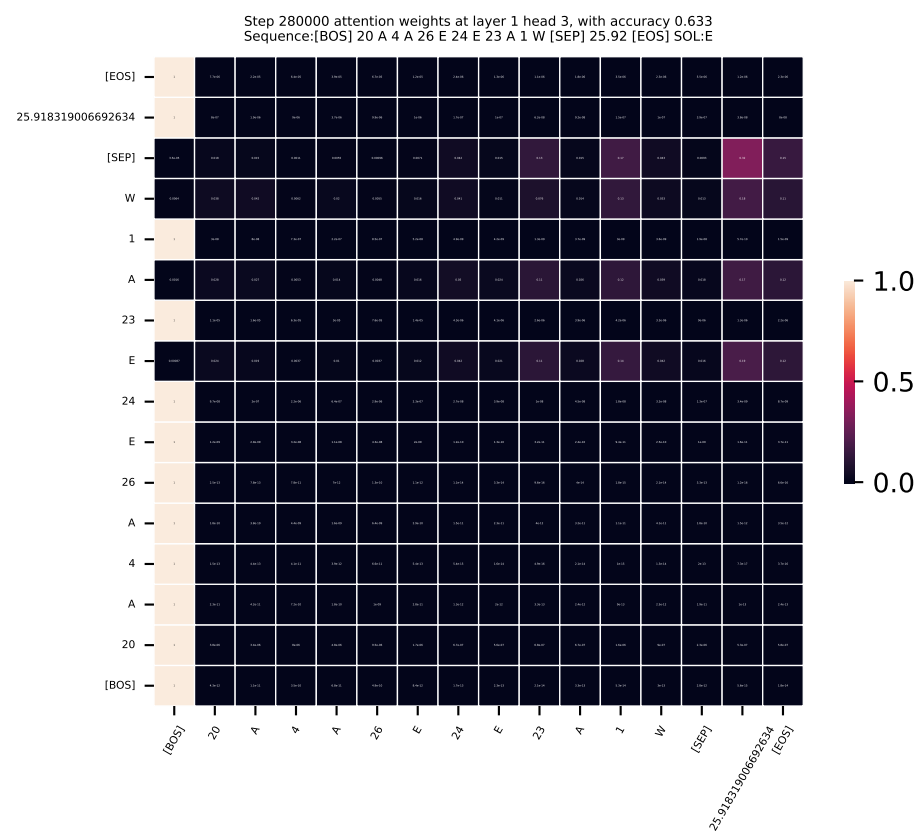

Step 280000 attention weights at layer 0 head 2, with accuracy 0.633
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

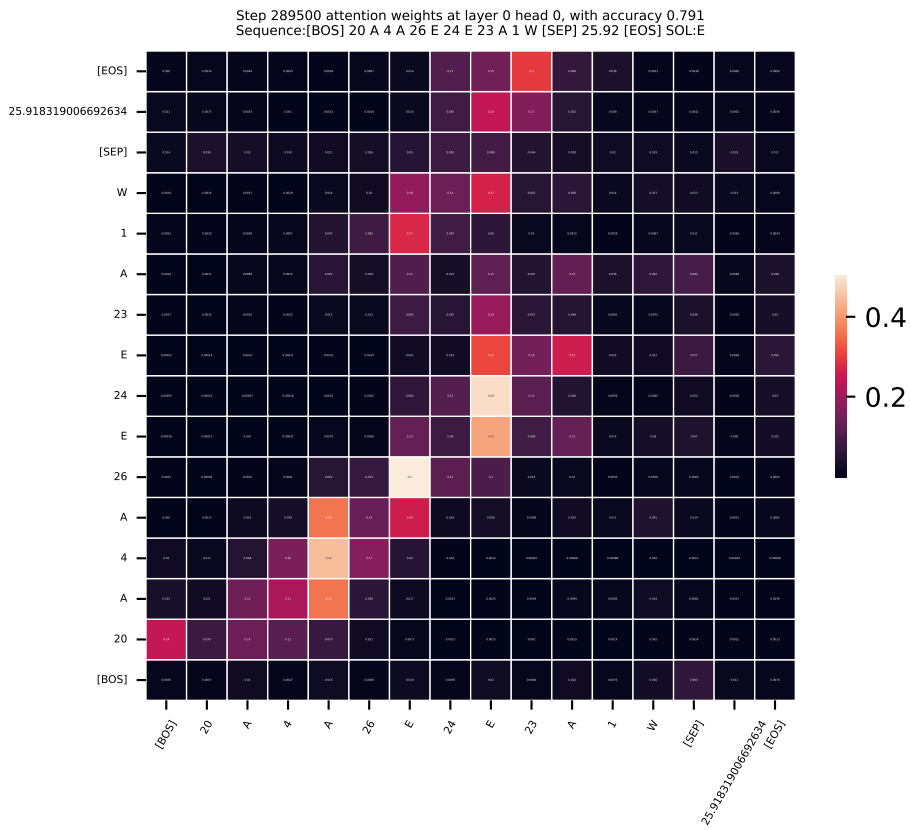

Step 280000 attention weights at layer 0 head 3, with accuracy 0.633
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

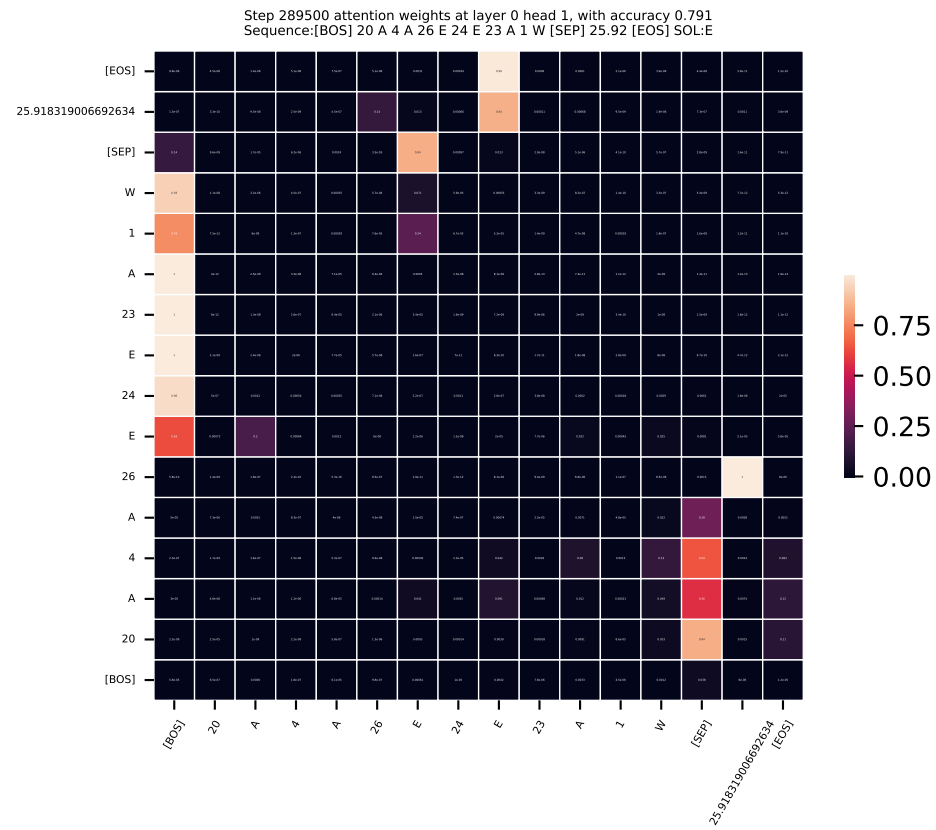

Step 280000 attention weights at layer 1 head 0, with accuracy 0.633
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

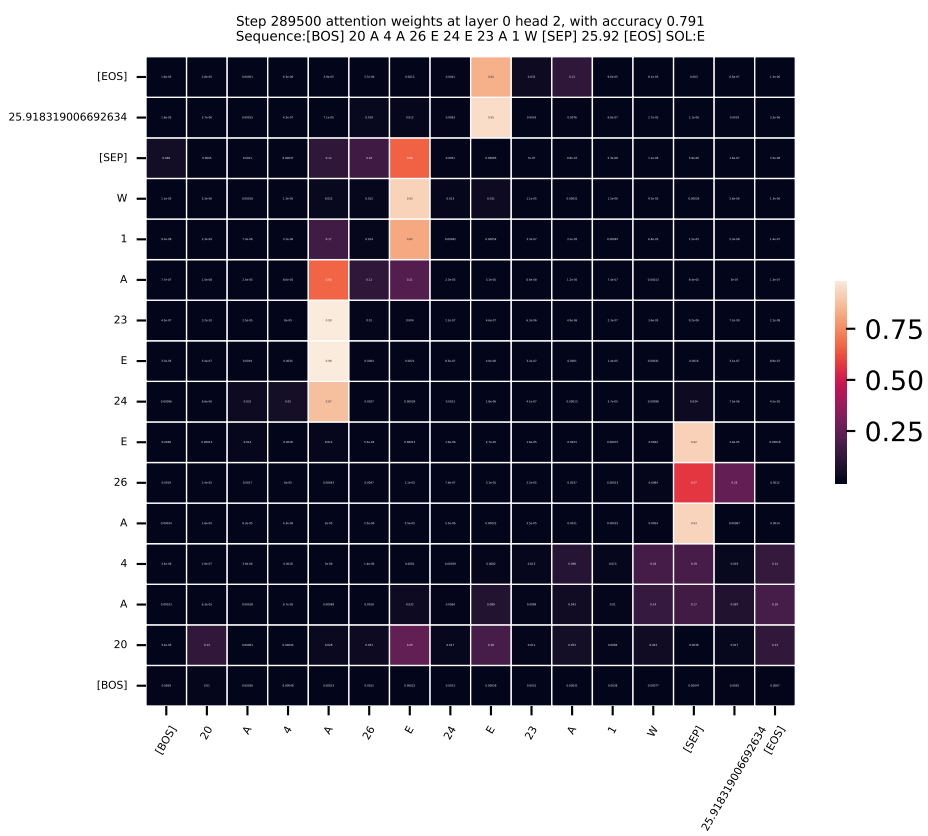

Step 280000 attention weights at layer 1 head 1, with accuracy 0.633
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.3
0.2
0.1

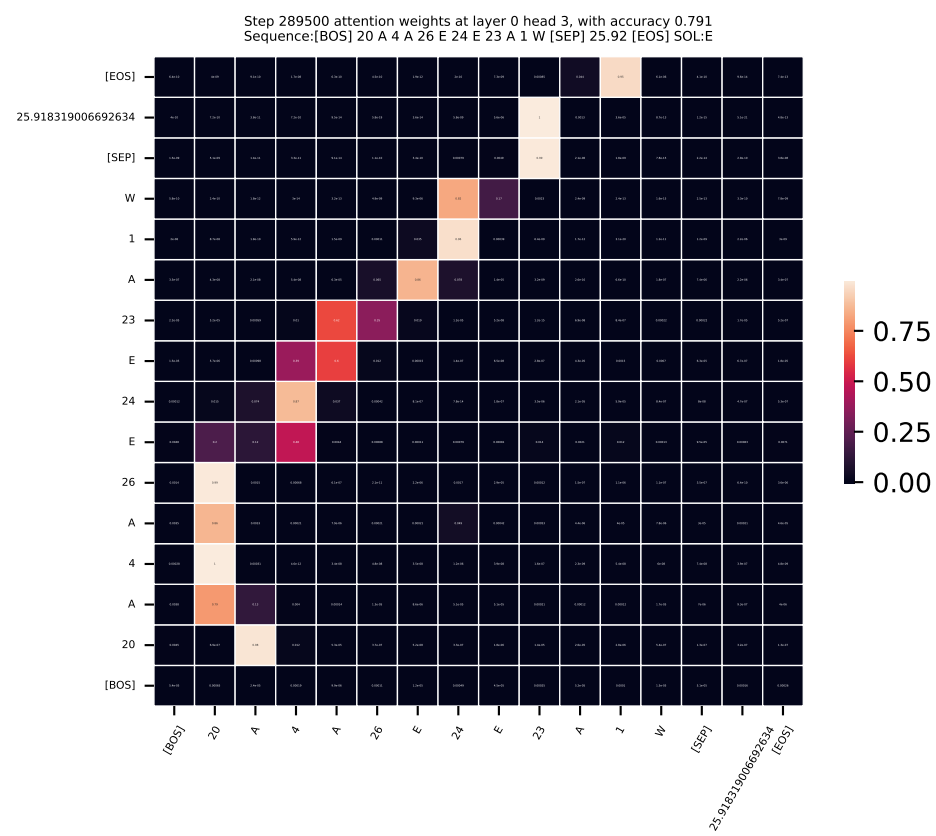

Step 280000 attention weights at layer 1 head 2, with accuracy 0.633
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

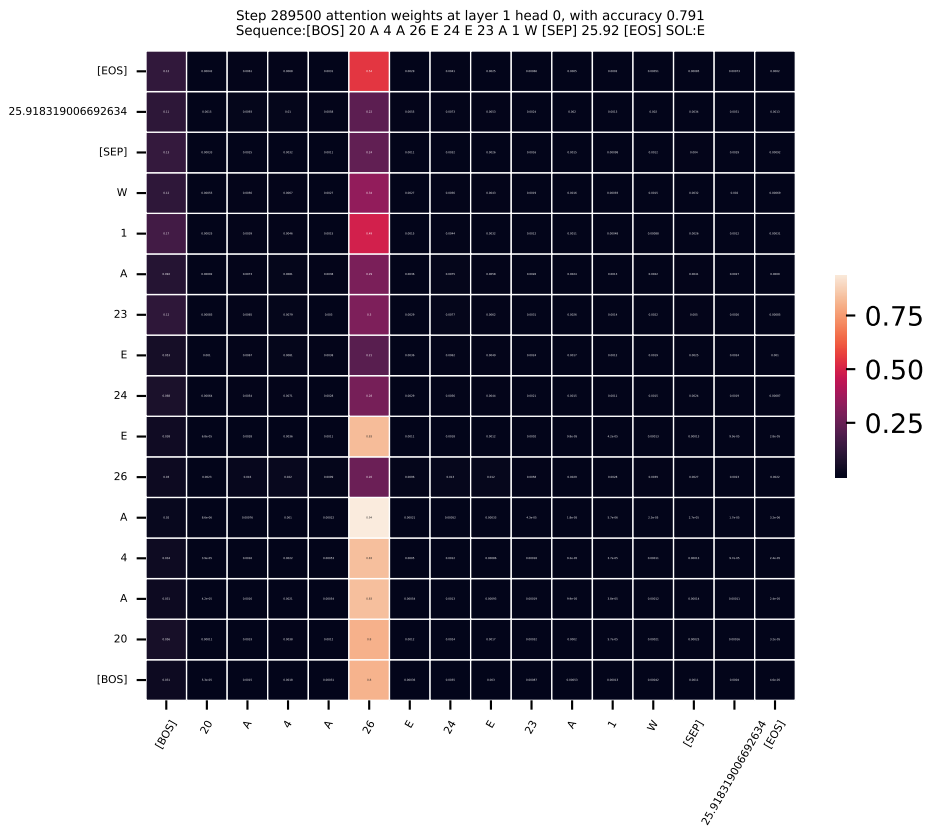

Step 280000 attention weights at layer 1 head 3, with accuracy 0.633
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
1.0
0.5
0.0

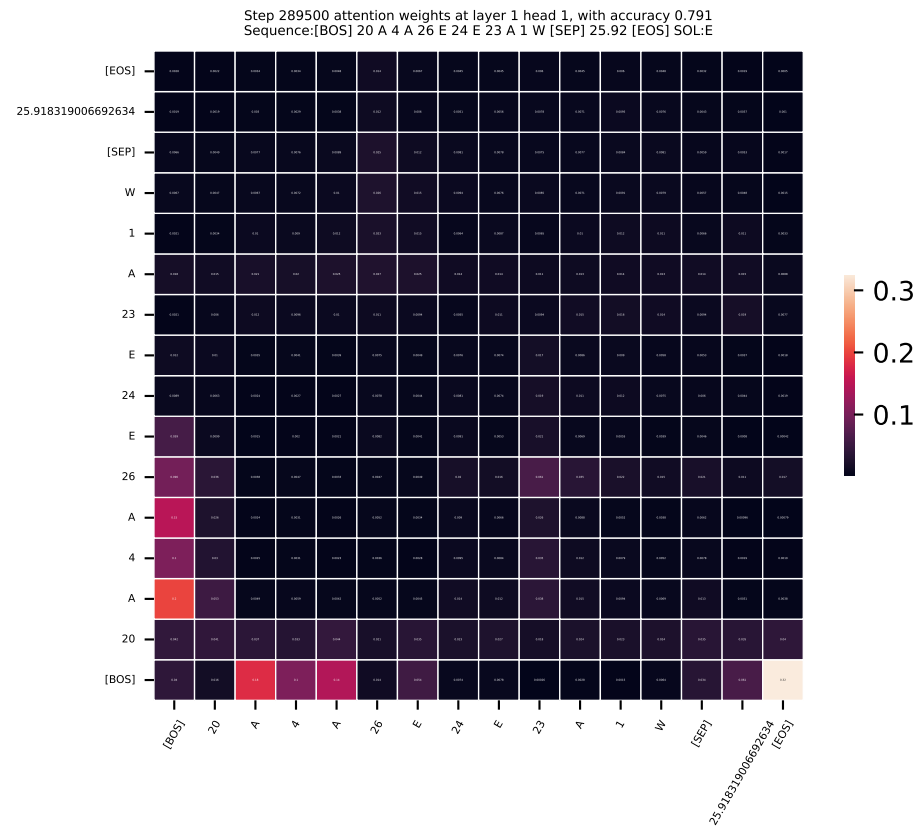

Step 289500 attention weights at layer 0 head 0, with accuracy 0.791
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

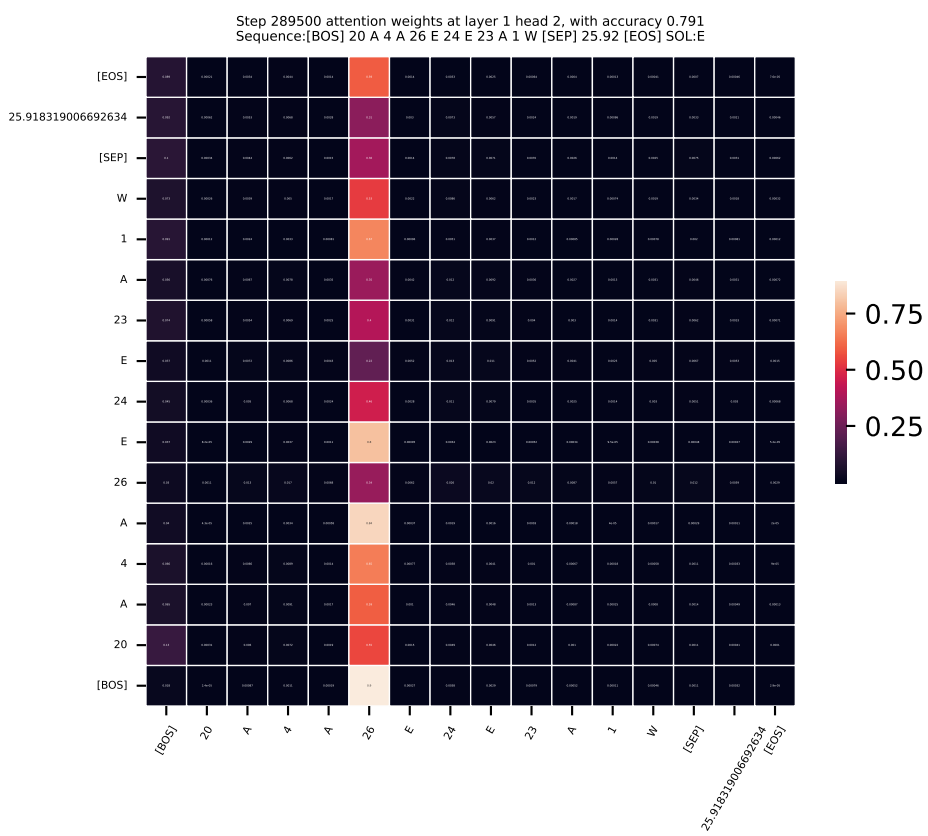

Step 289500 attention weights at layer 0 head 1, with accuracy 0.791
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

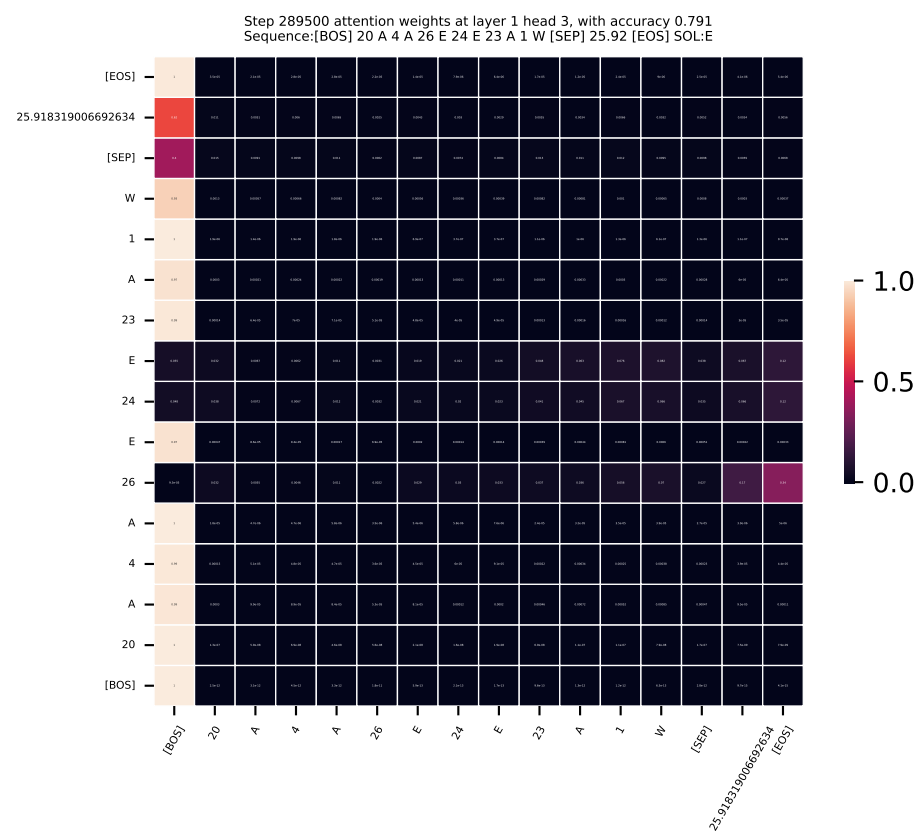

Step 289500 attention weights at layer 0 head 2, with accuracy 0.791
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

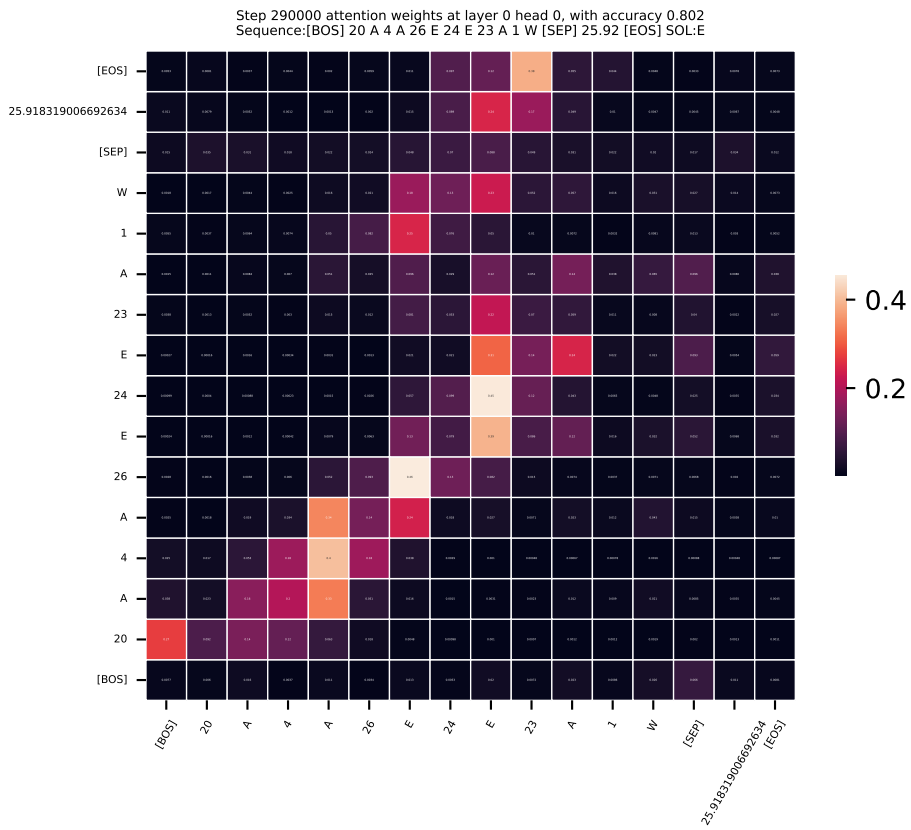

Step 289500 attention weights at layer 0 head 3, with accuracy 0.791
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

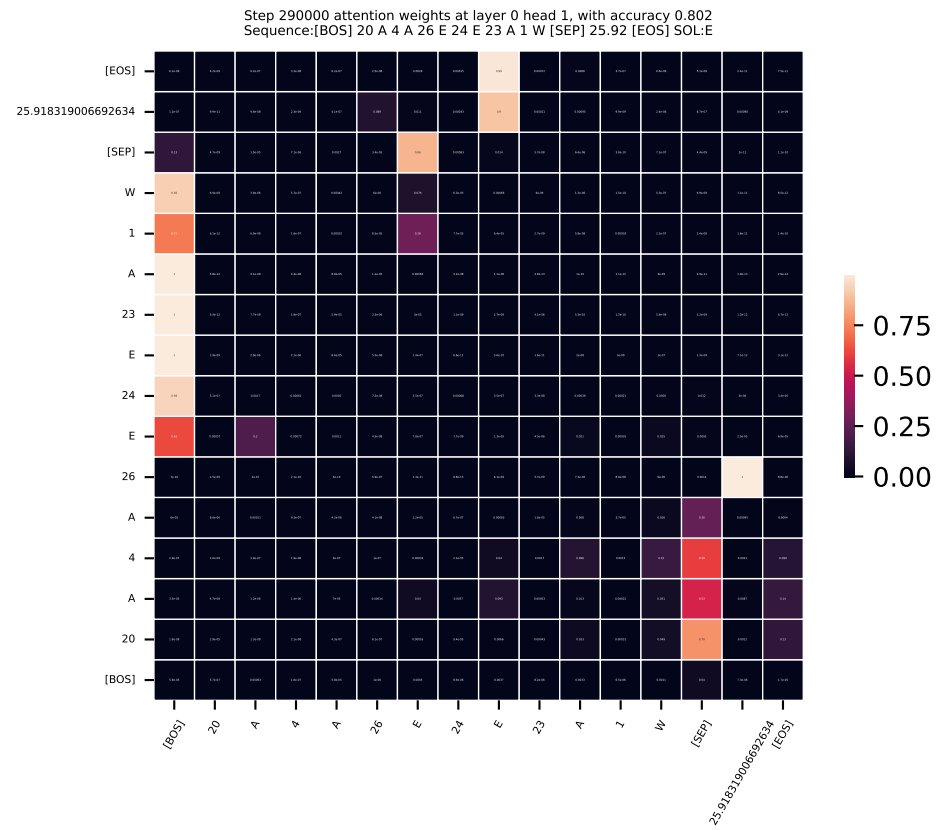

Step 289500 attention weights at layer 1 head 0, with accuracy 0.791
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

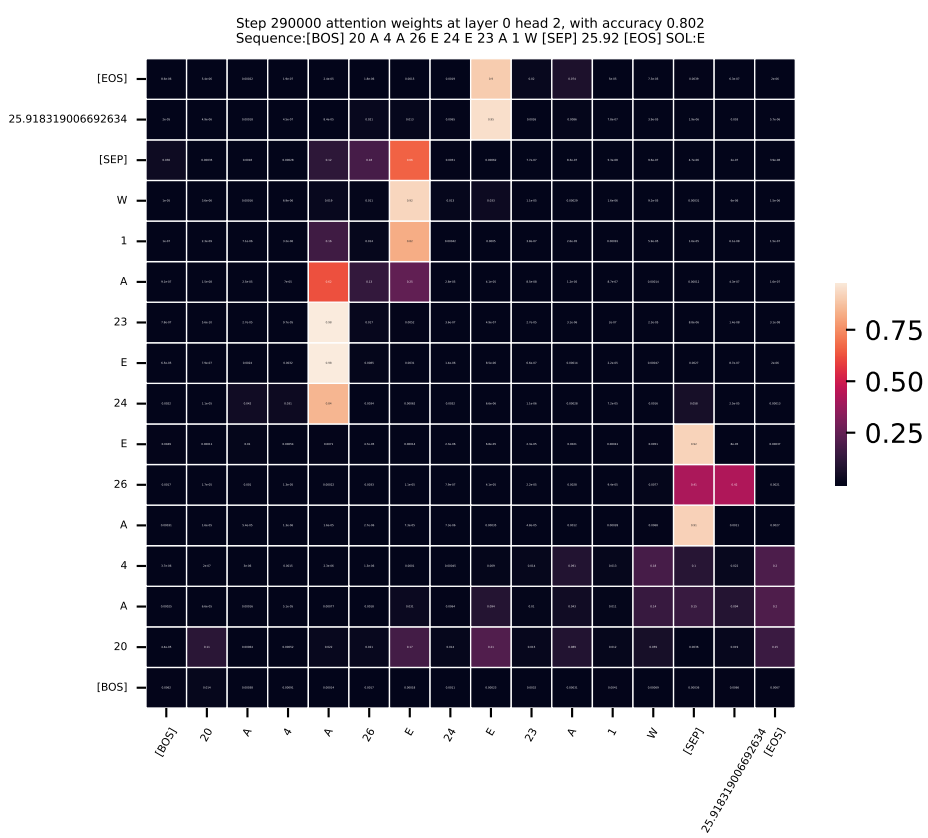

Step 289500 attention weights at layer 1 head 1, with accuracy 0.791
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

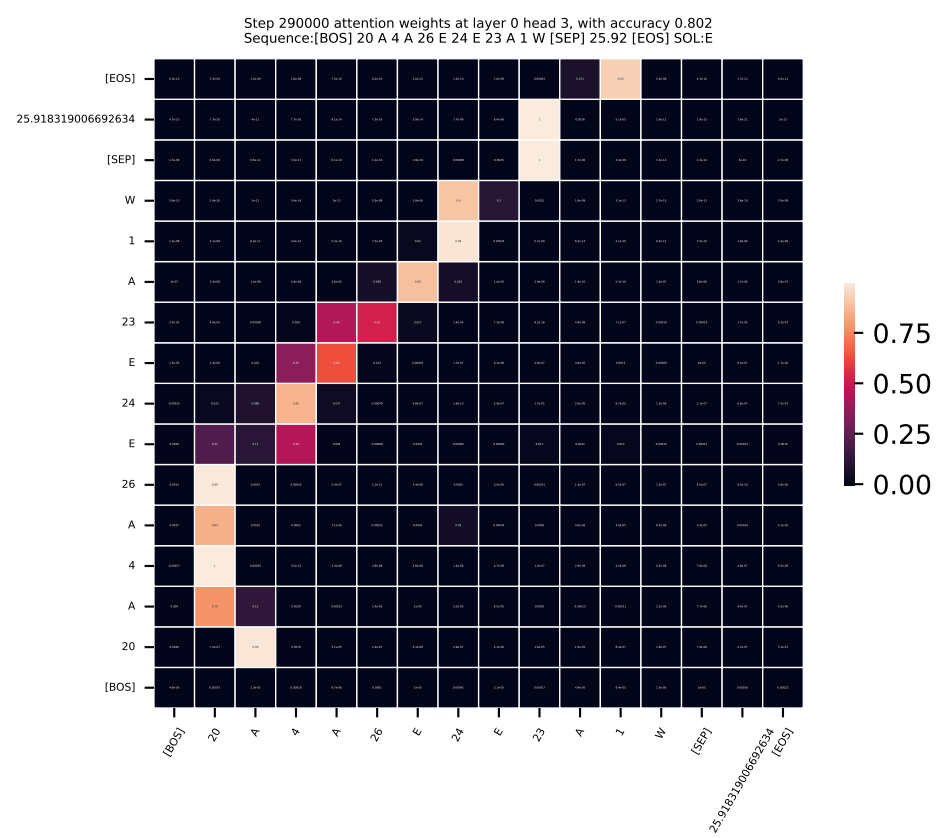

Step 289500 attention weights at layer 1 head 2, with accuracy 0.791
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

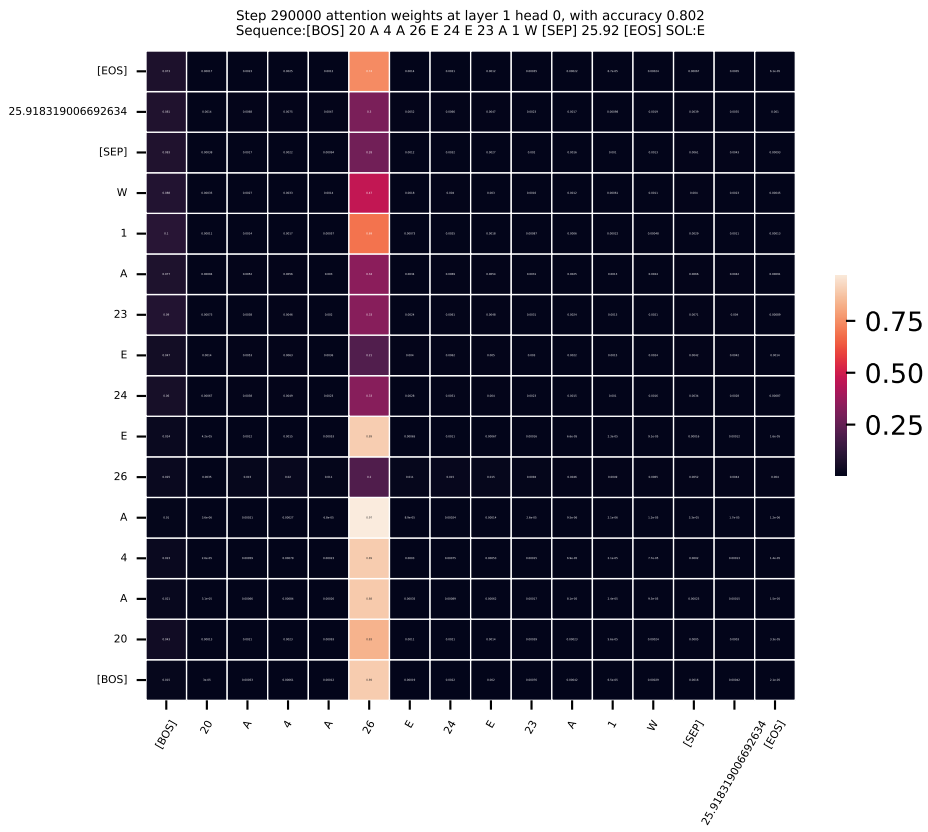

Step 289500 attention weights at layer 1 head 3, with accuracy 0.791
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
1.0
0.5
0.0

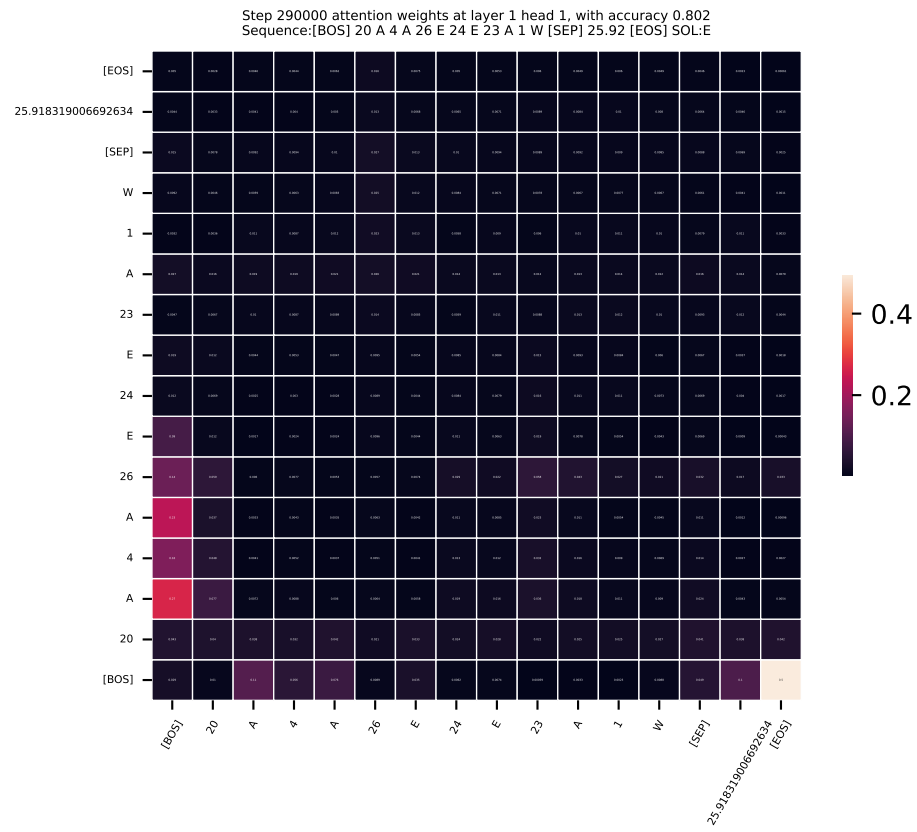

Step 290000 attention weights at layer 0 head 0, with accuracy 0.802
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

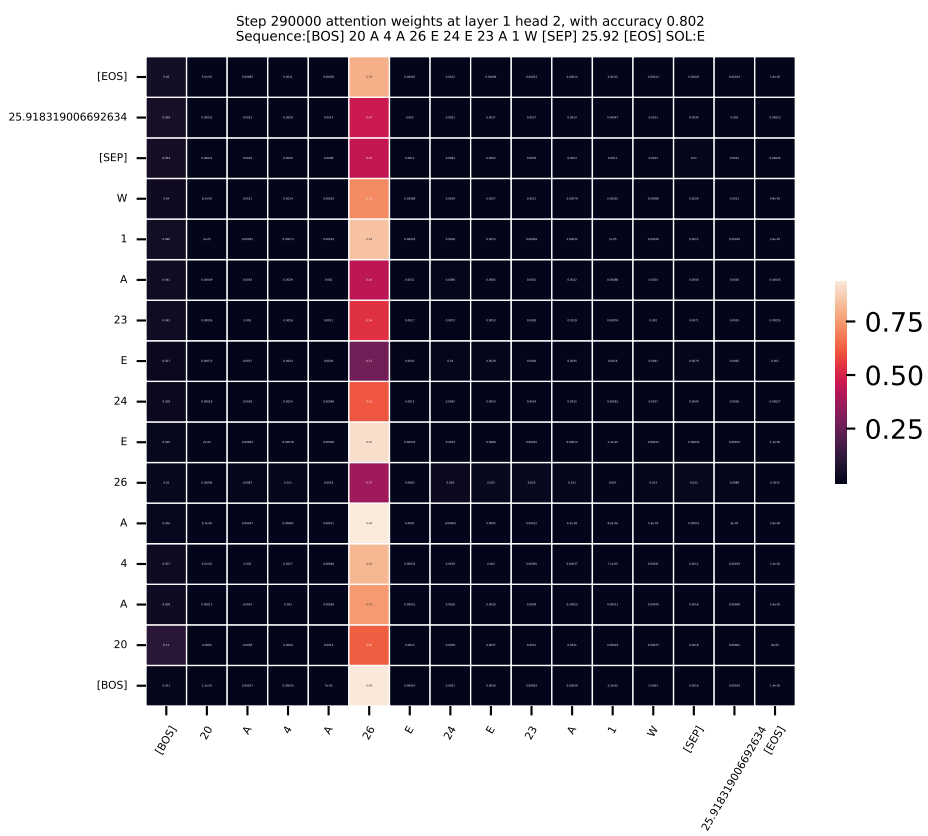

Step 290000 attention weights at layer 0 head 1, with accuracy 0.802
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

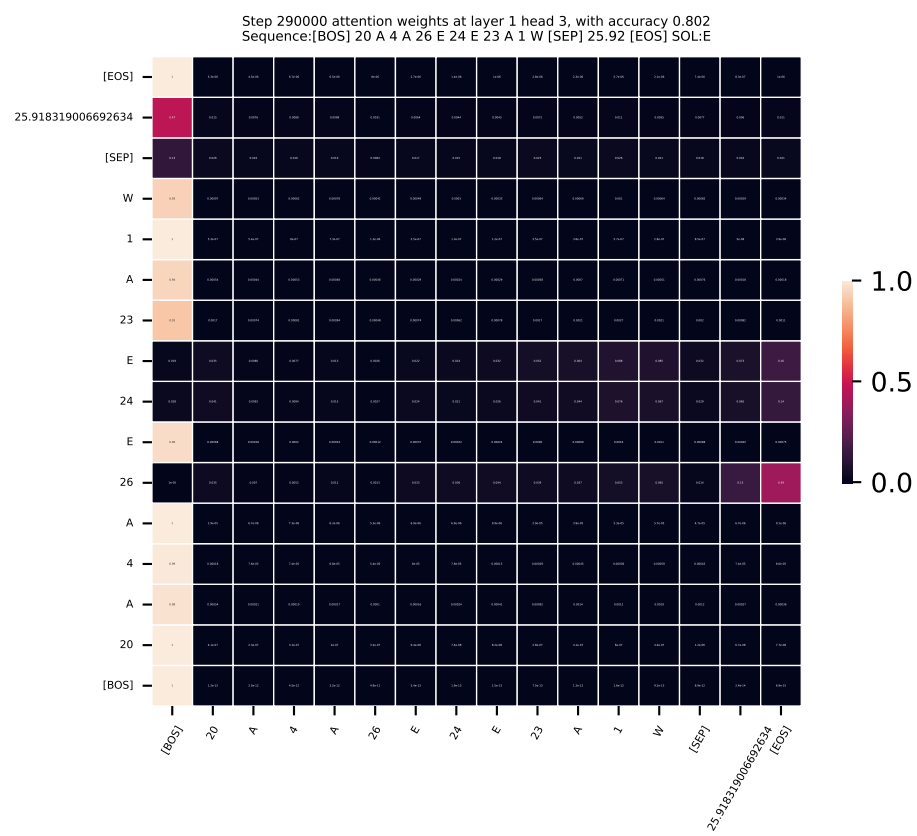

Step 290000 attention weights at layer 0 head 2, with accuracy 0.802
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

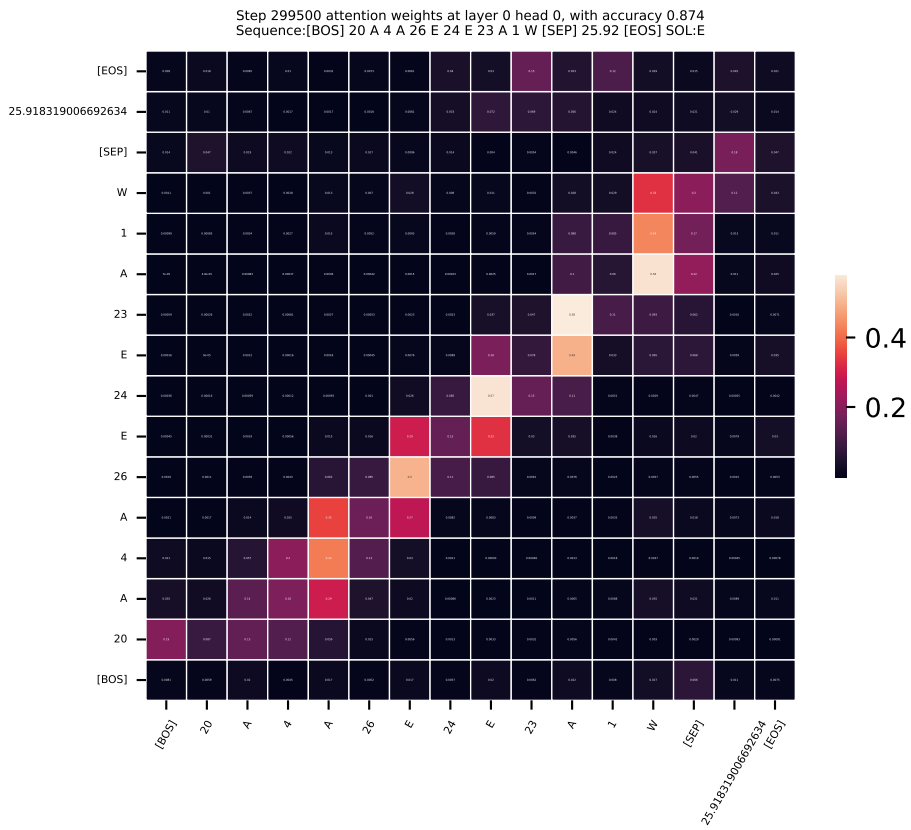

Step 290000 attention weights at layer 0 head 3, with accuracy 0.802
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

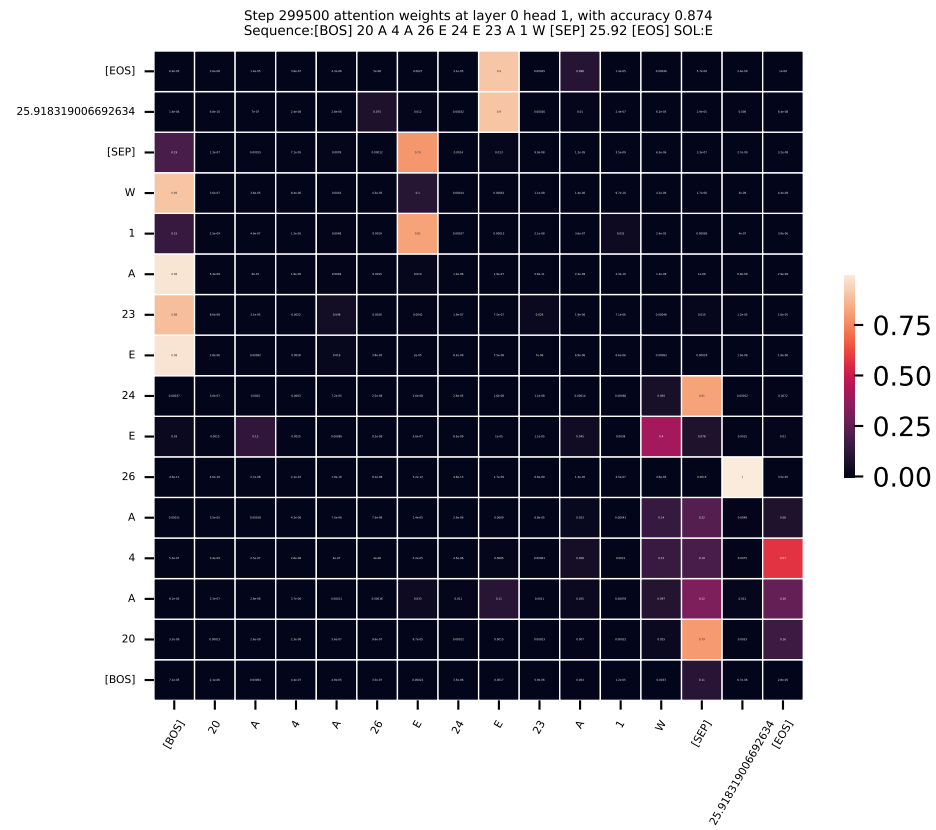

Step 290000 attention weights at layer 1 head 0, with accuracy 0.802
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

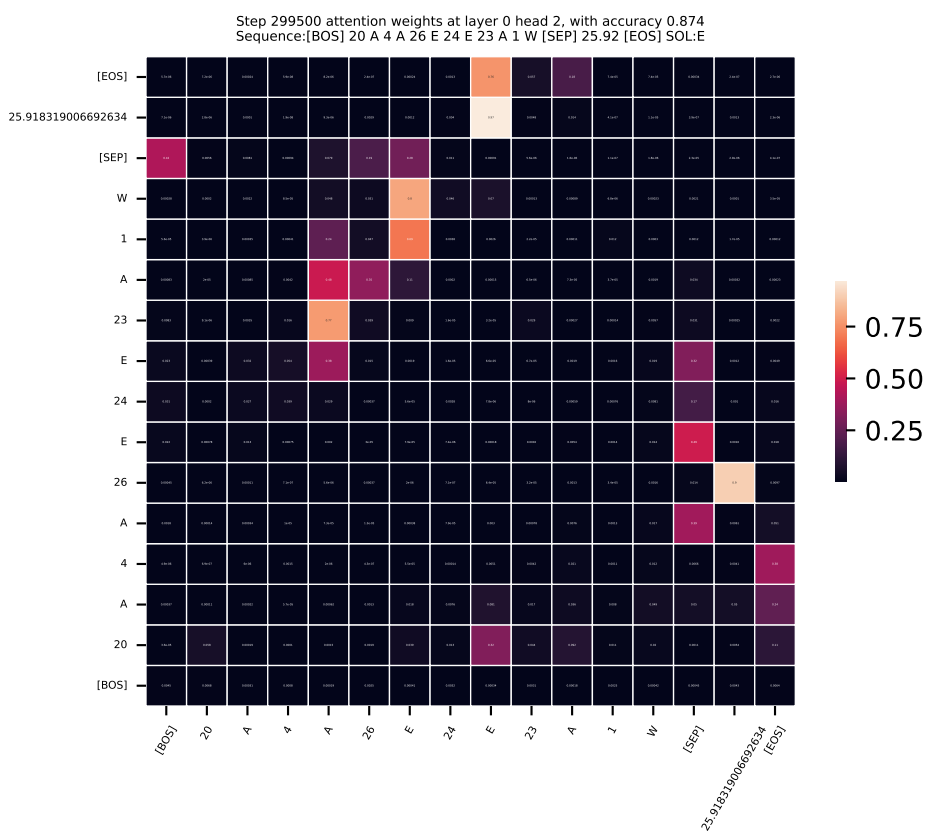

Step 290000 attention weights at layer 1 head 1, with accuracy 0.802
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

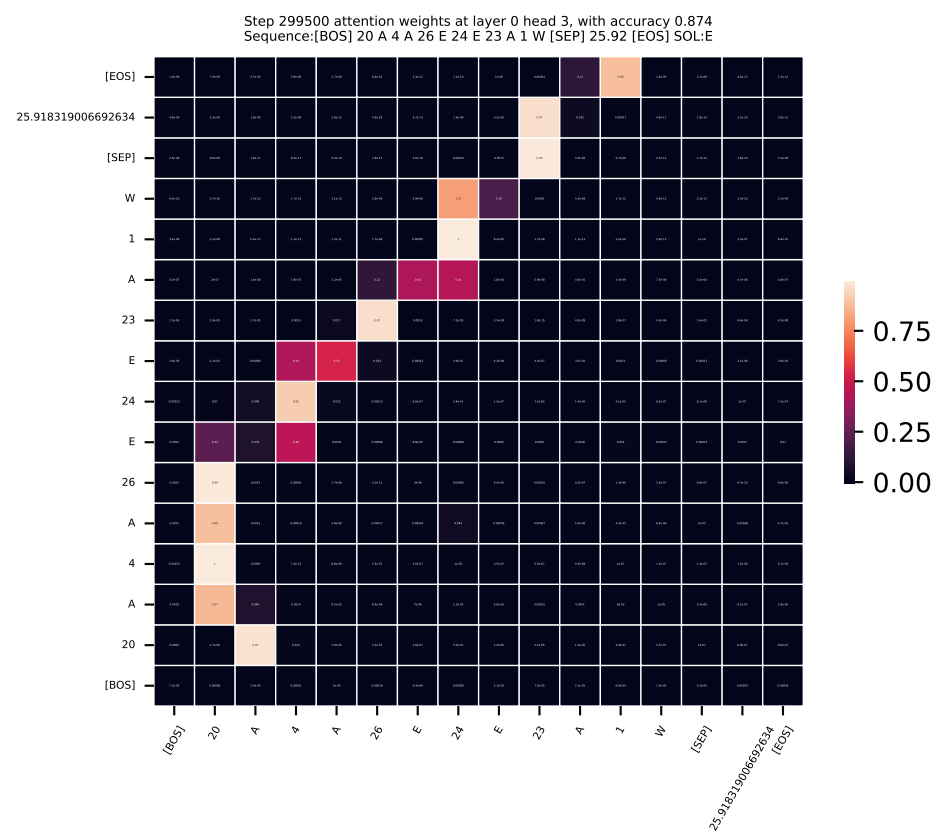

Step 290000 attention weights at layer 1 head 2, with accuracy 0.802
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

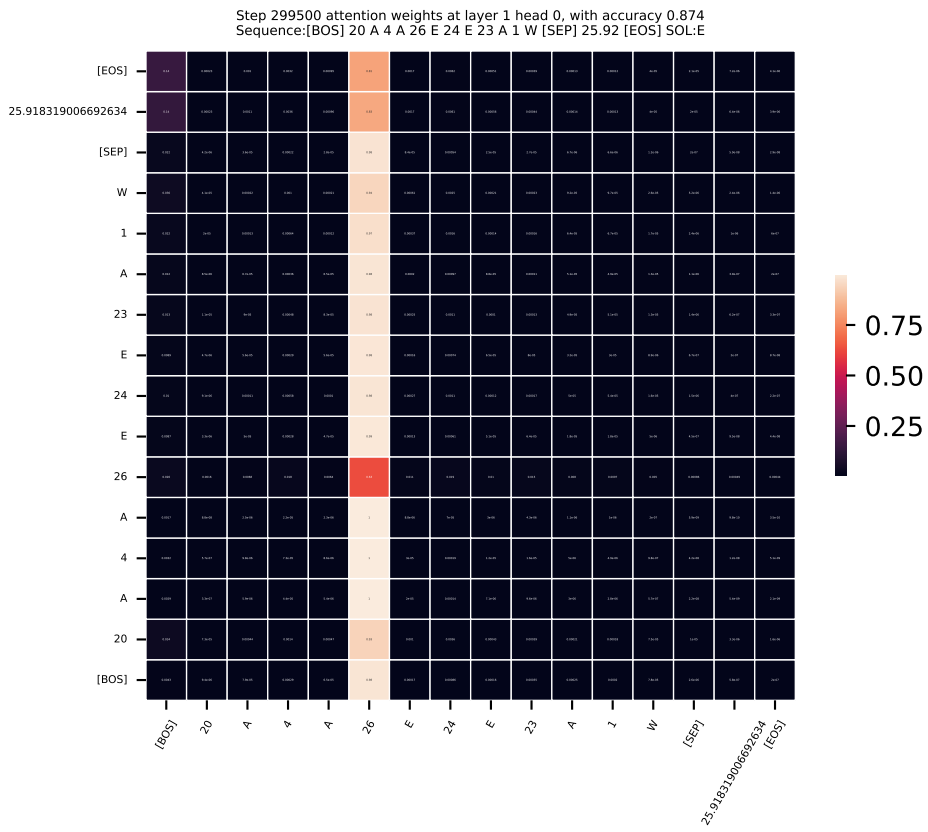

Step 290000 attention weights at layer 1 head 3, with accuracy 0.802
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
1.0
0.5
0.0

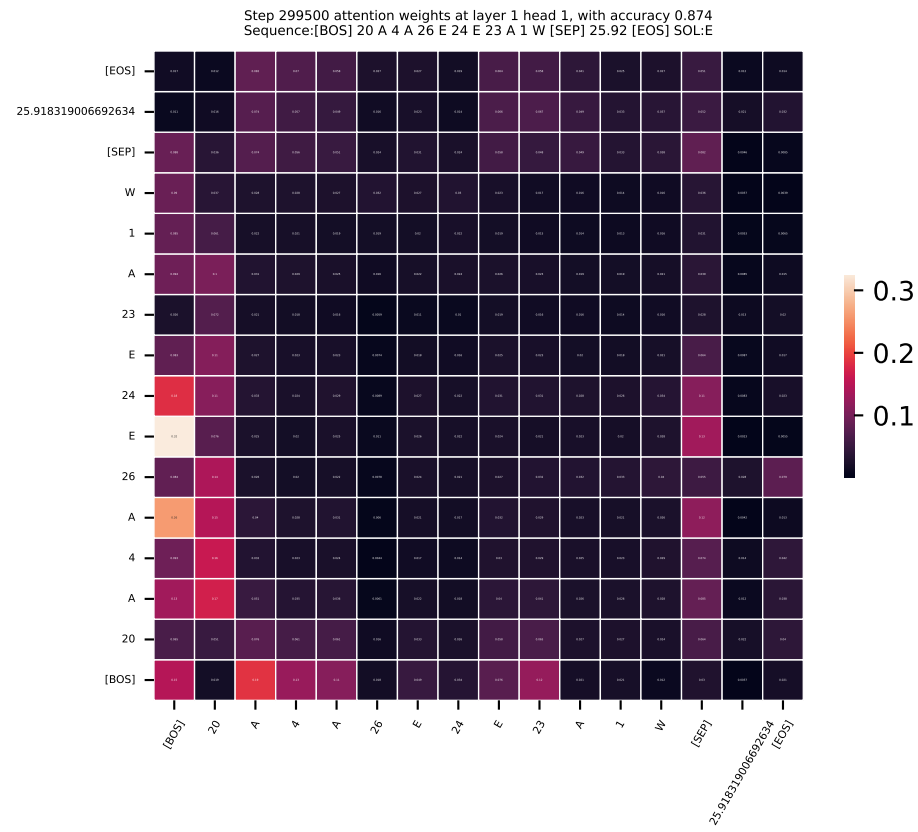

Step 299500 attention weights at layer 0 head 0, with accuracy 0.874
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

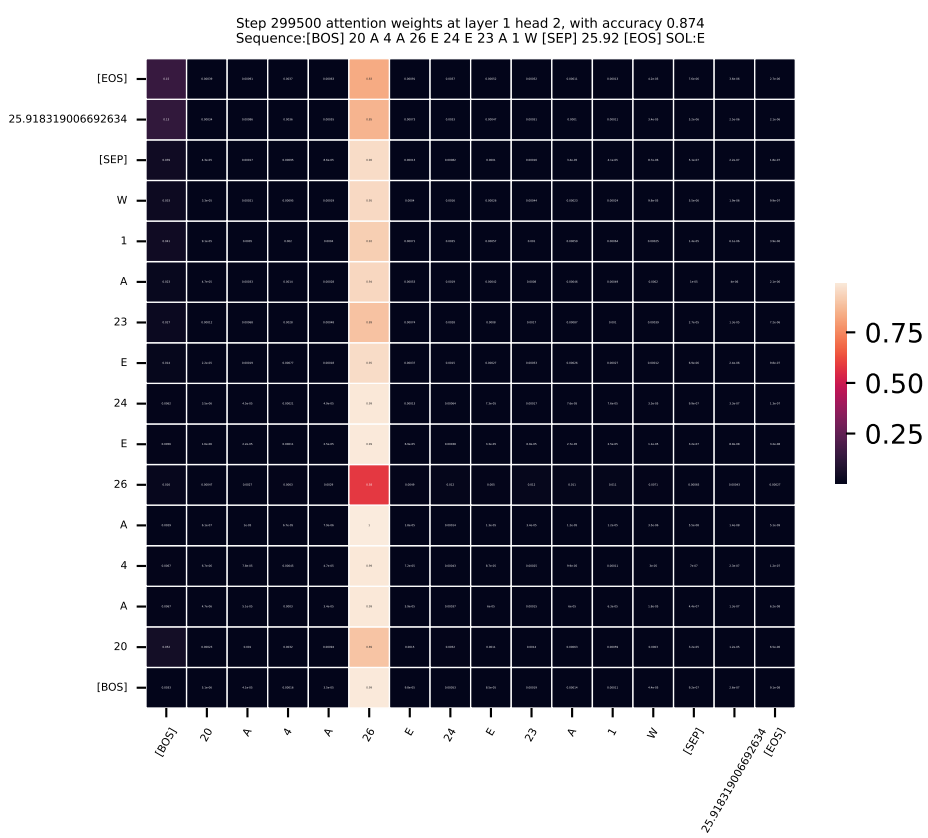

Step 299500 attention weights at layer 0 head 1, with accuracy 0.874
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

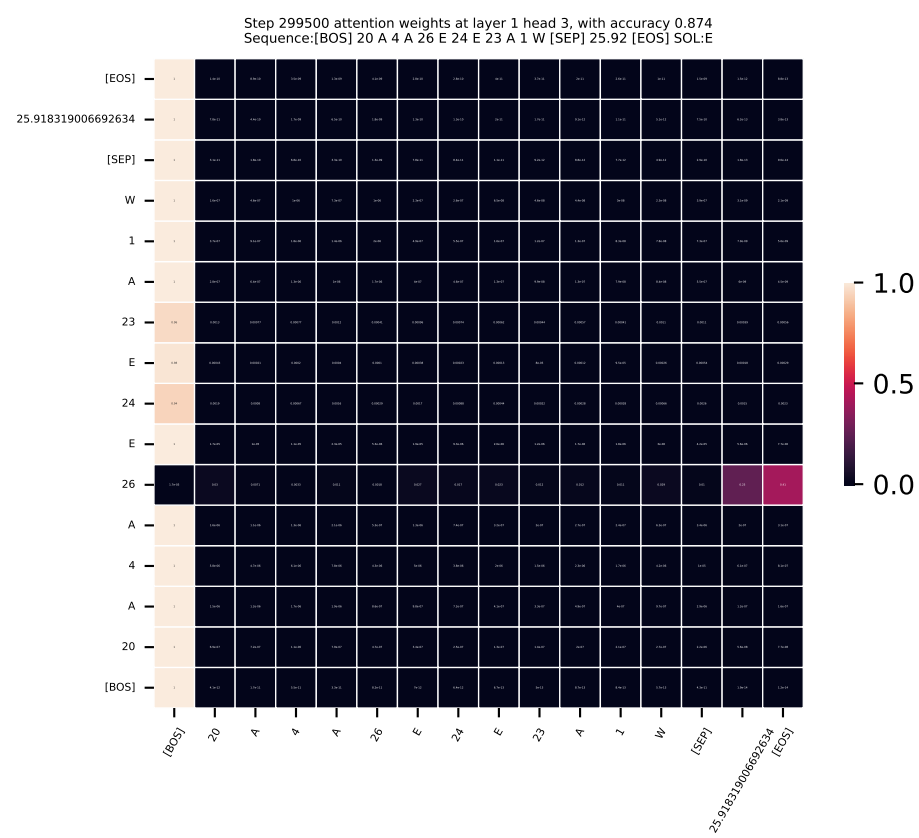

Step 299500 attention weights at layer 0 head 2, with accuracy 0.874
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

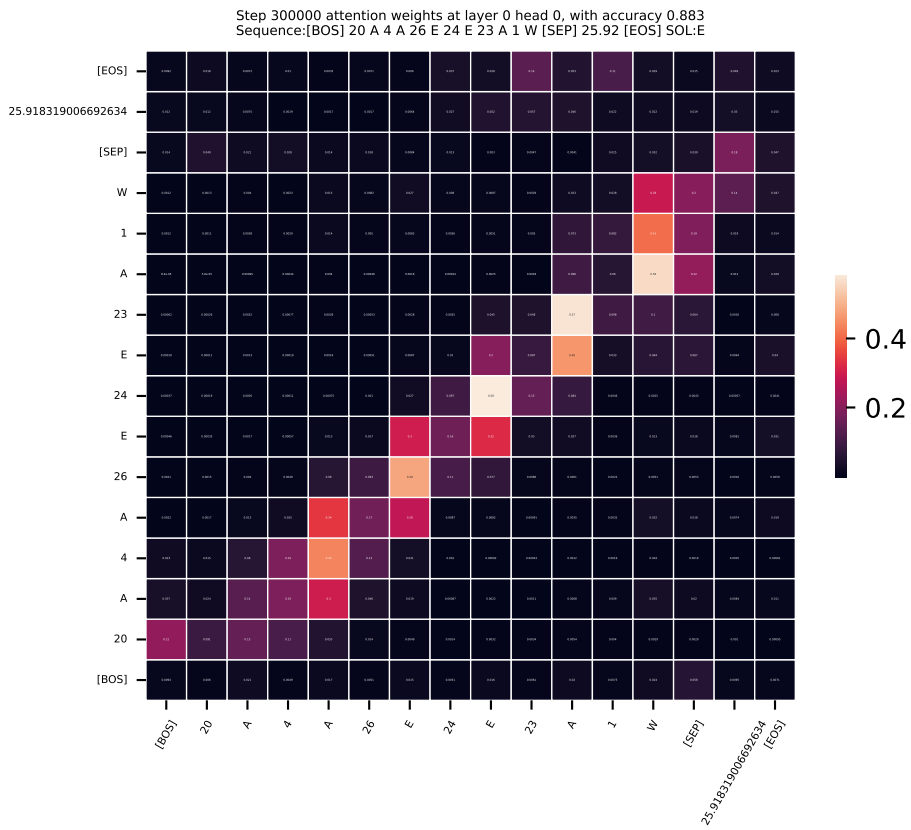

Step 299500 attention weights at layer 0 head 3, with accuracy 0.874
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

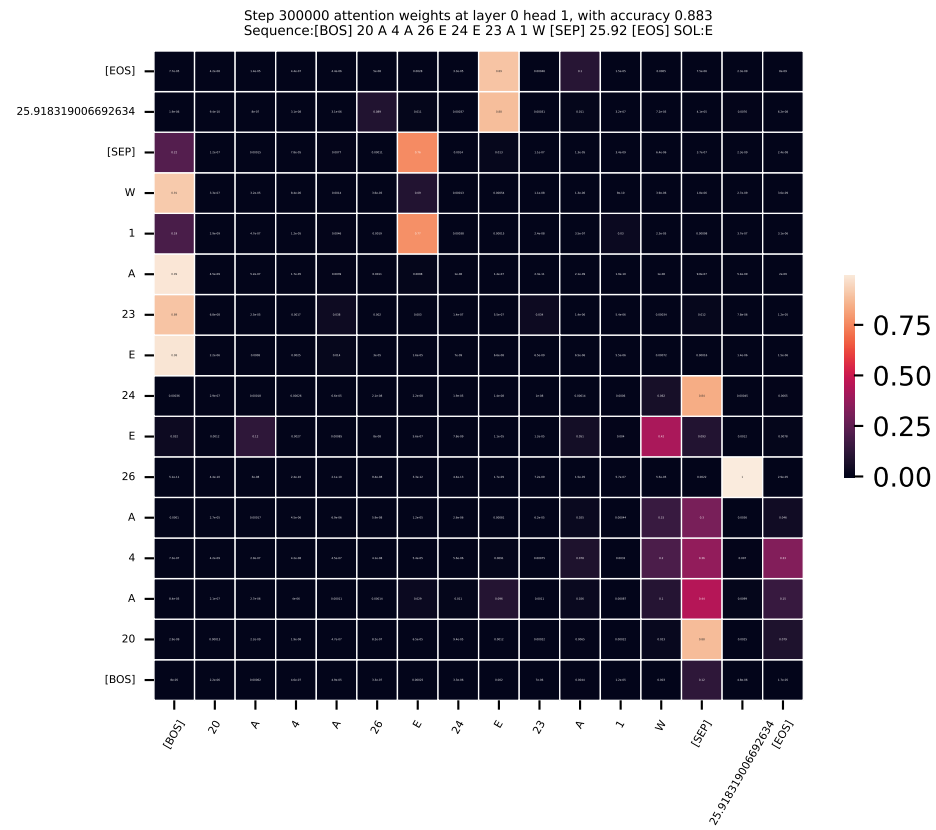

Step 299500 attention weights at layer 1 head 0, with accuracy 0.874
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

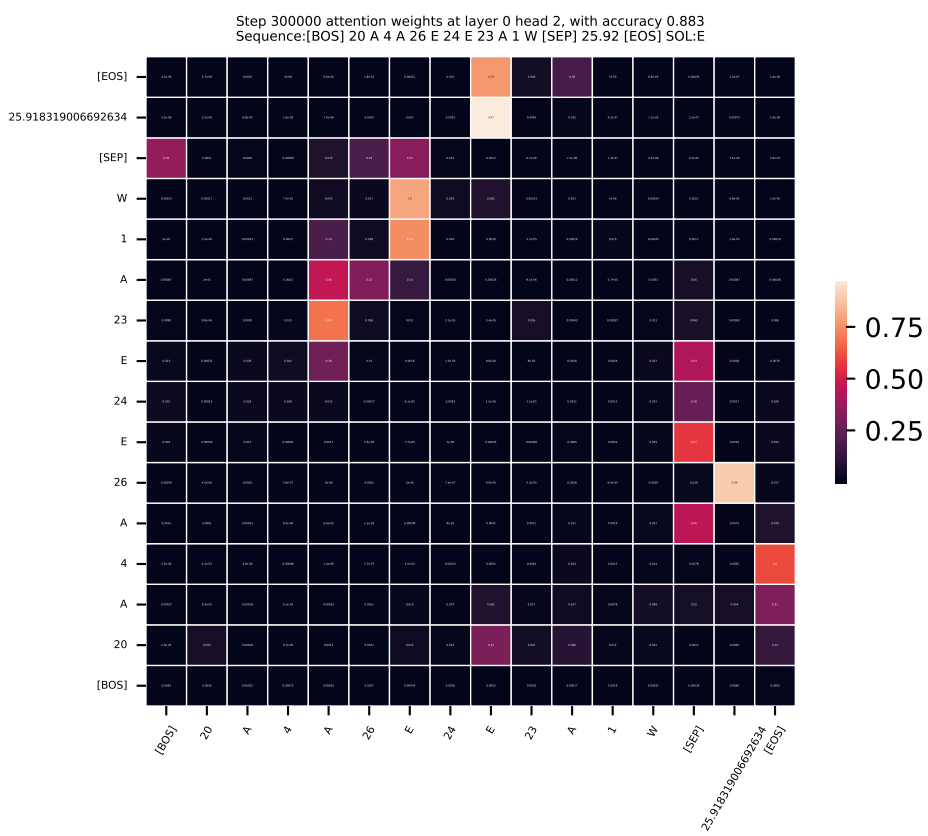

Step 299500 attention weights at layer 1 head 1, with accuracy 0.874
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

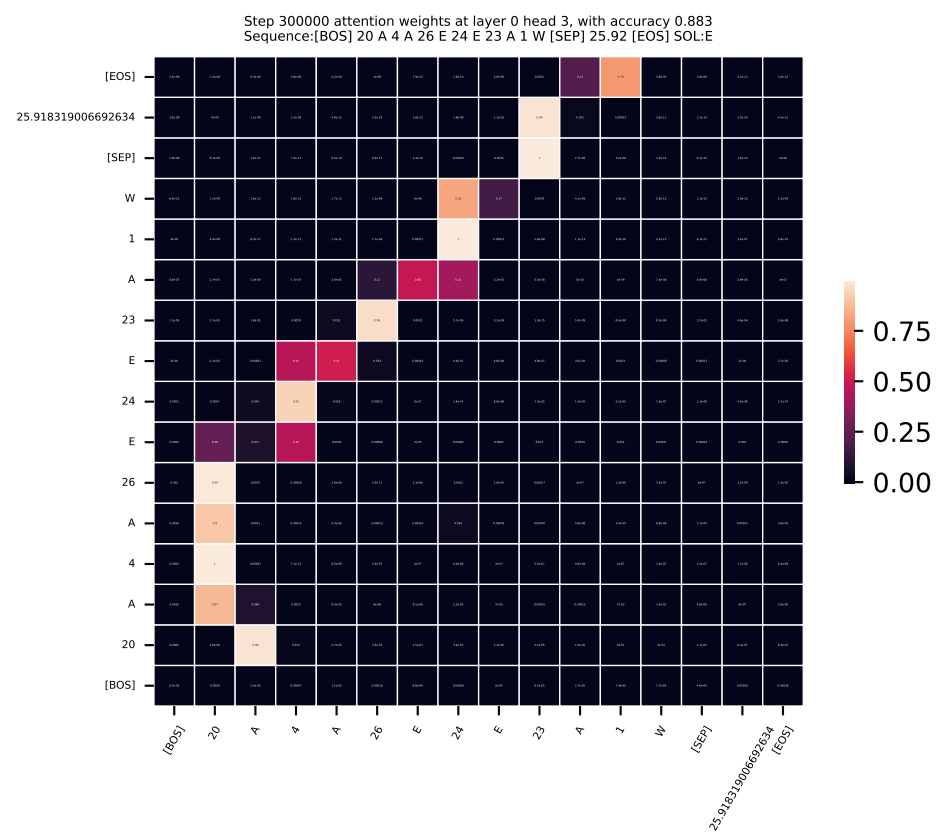

Step 299500 attention weights at layer 1 head 2, with accuracy 0.874
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

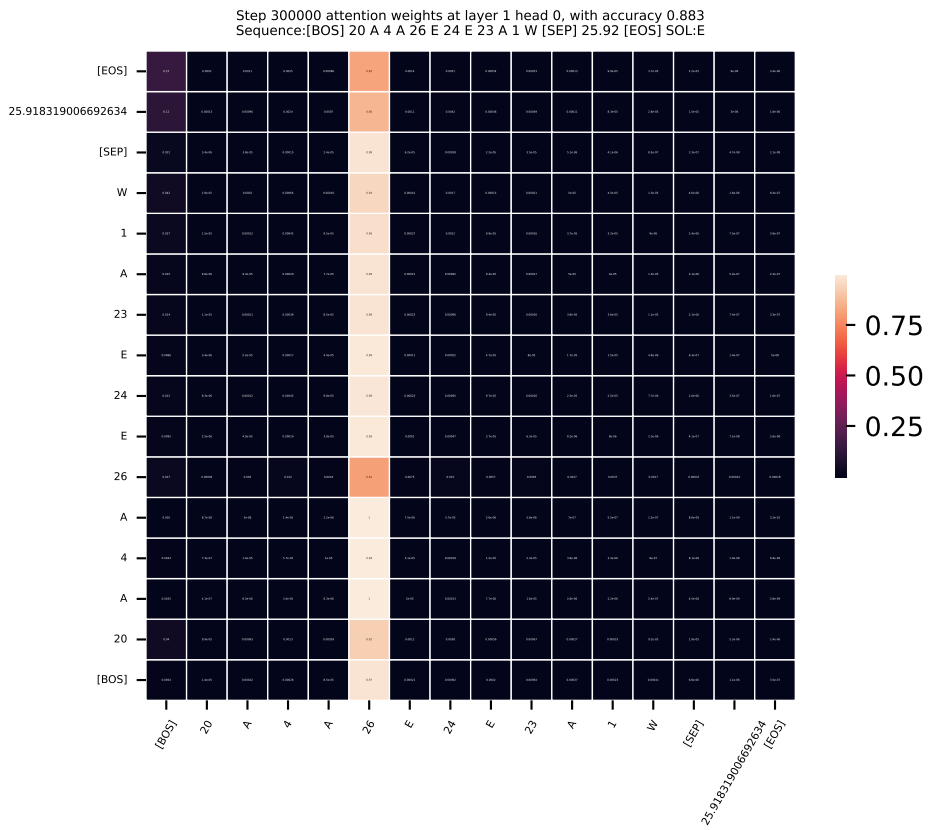

Step 299500 attention weights at layer 1 head 3, with accuracy 0.874
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
1.0
0.5
0.0

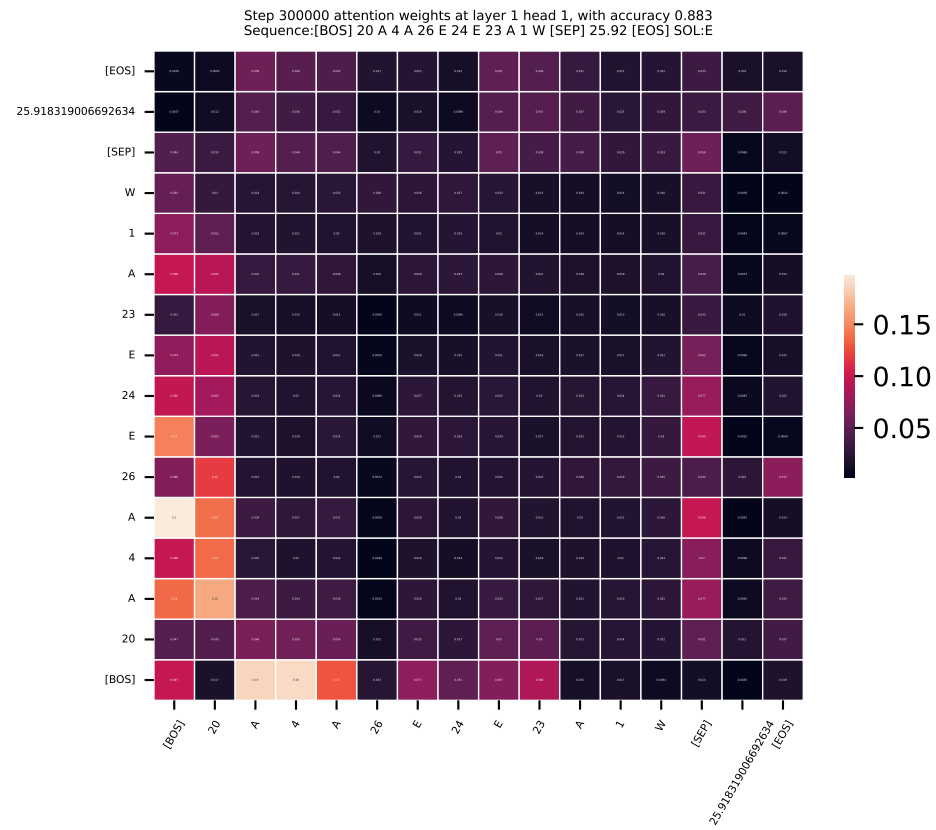

Step 300000 attention weights at layer 0 head 0, with accuracy 0.883
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

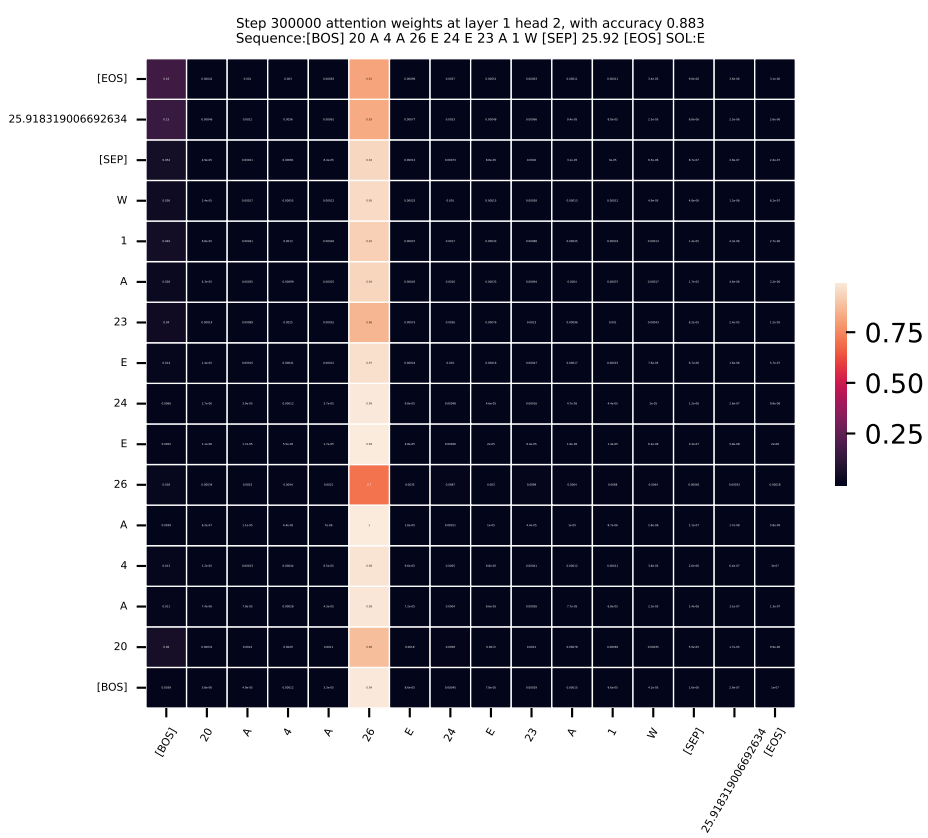

Step 300000 attention weights at layer 0 head 1, with accuracy 0.883
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

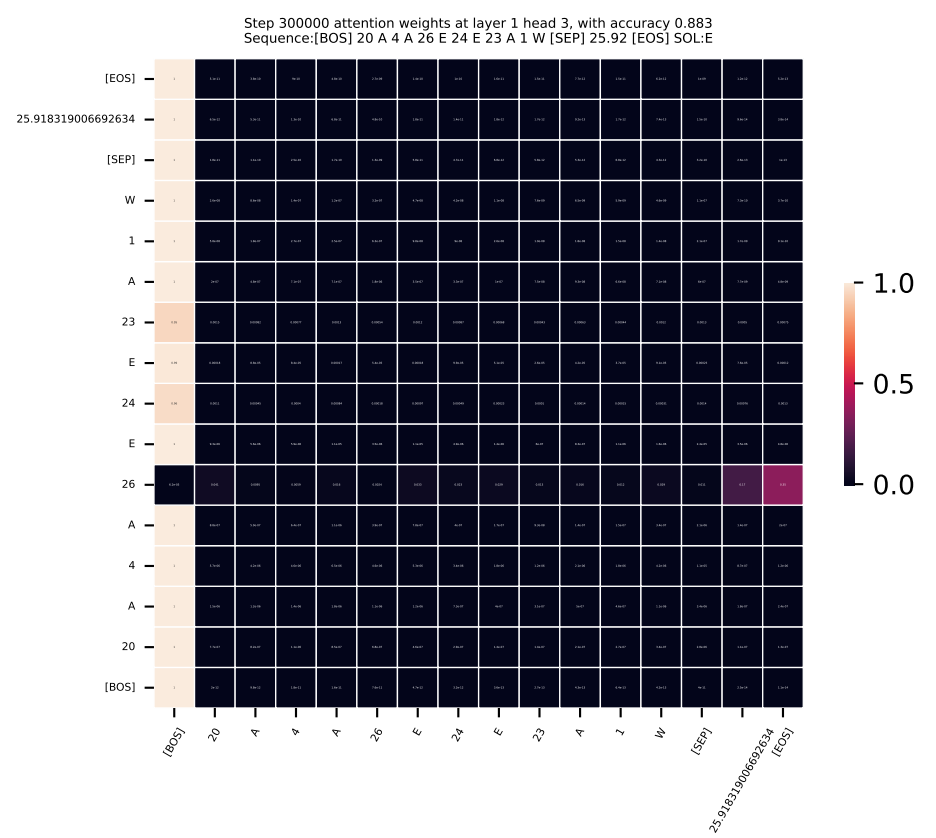

Step 300000 attention weights at layer 0 head 2, with accuracy 0.883
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

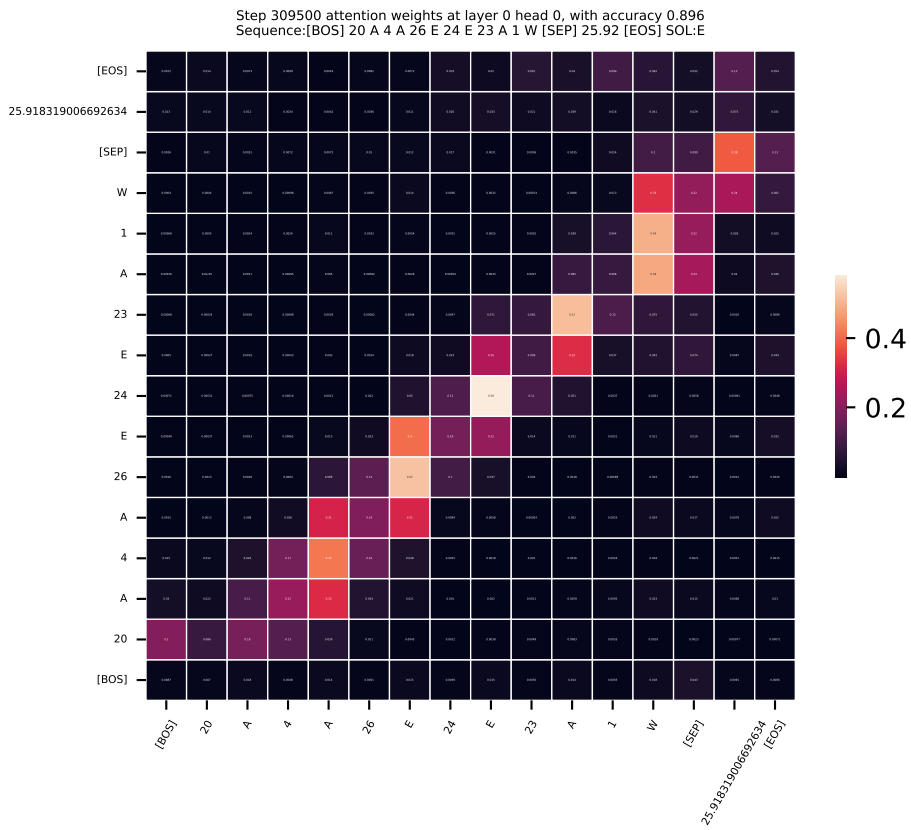

Step 300000 attention weights at layer 0 head 3, with accuracy 0.883
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

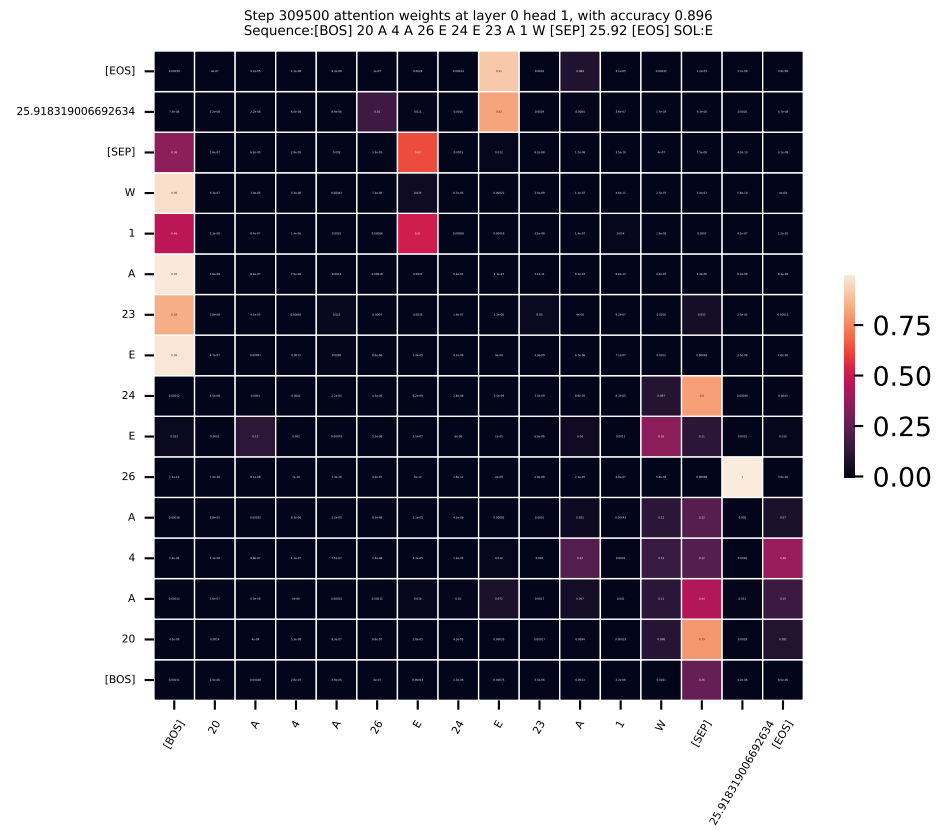

Step 300000 attention weights at layer 1 head 0, with accuracy 0.883
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

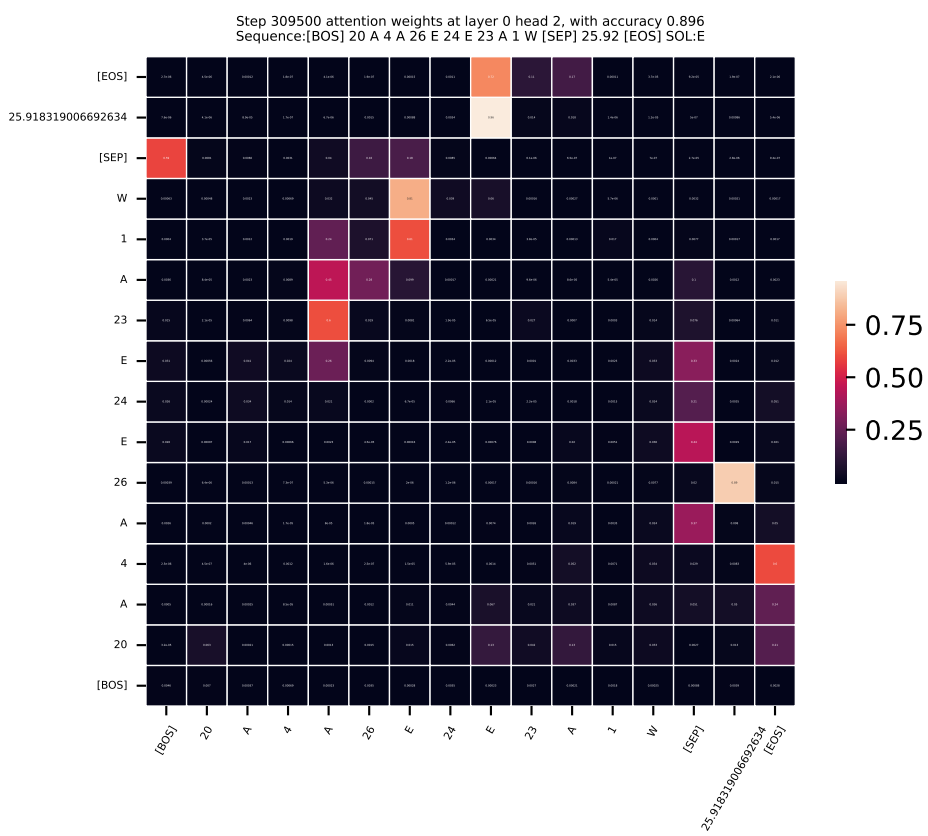

Step 300000 attention weights at layer 1 head 1, with accuracy 0.883
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.15
0.10
0.05

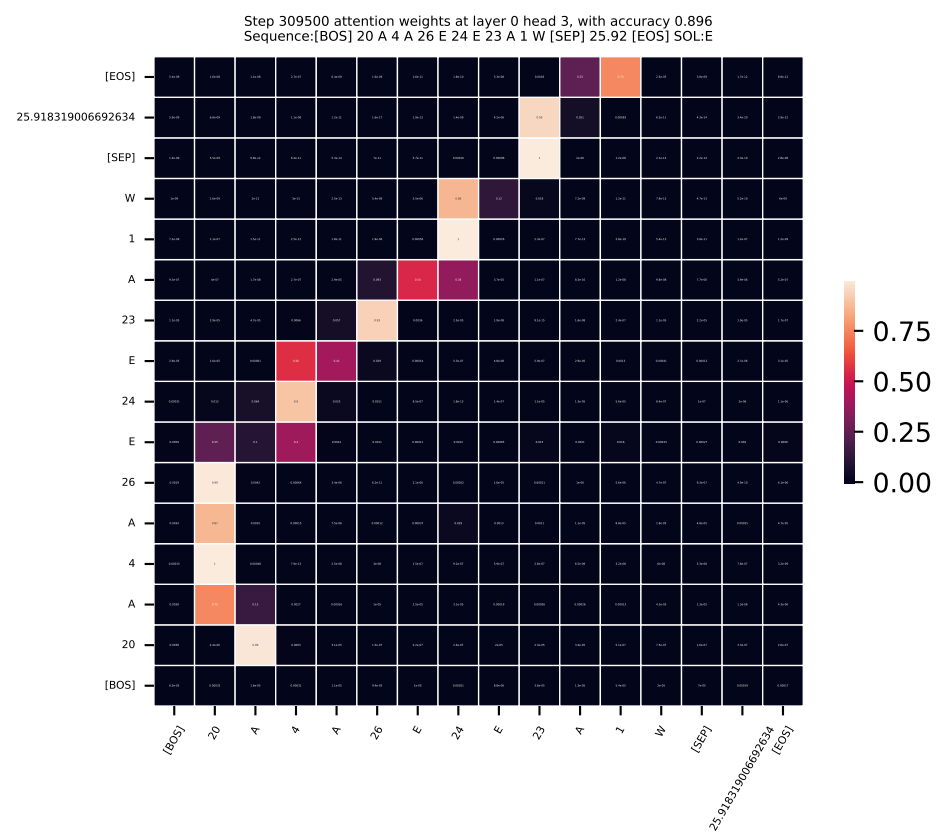

Step 300000 attention weights at layer 1 head 2, with accuracy 0.883
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

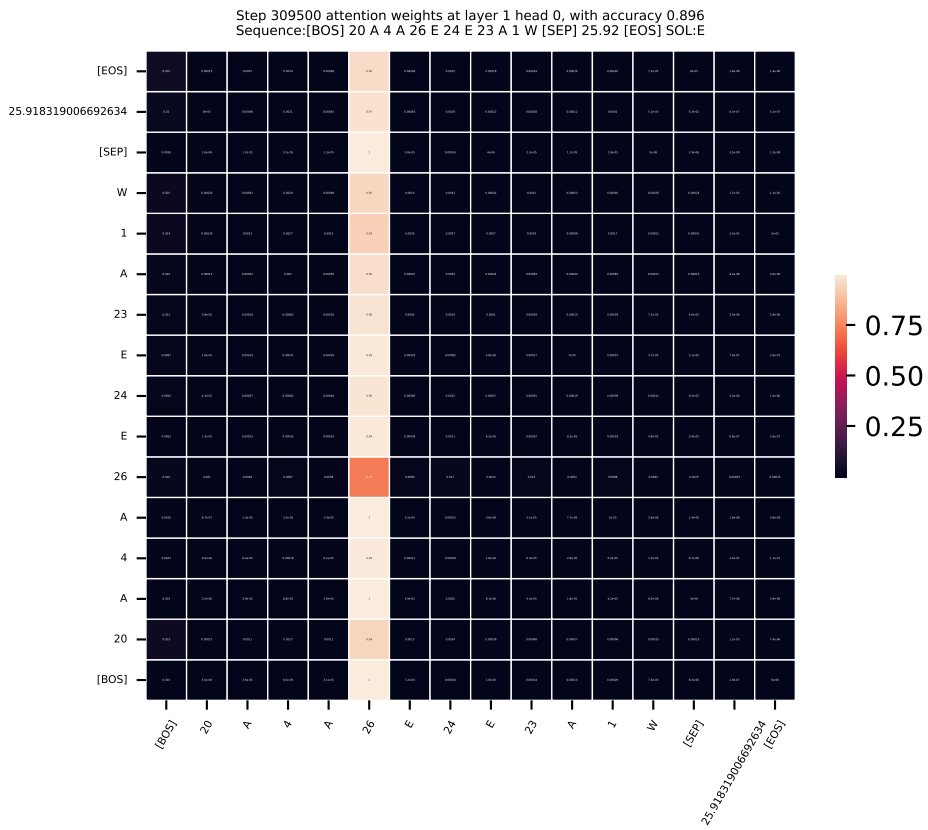

Step 300000 attention weights at layer 1 head 3, with accuracy 0.883
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
1.0
0.5
0.0

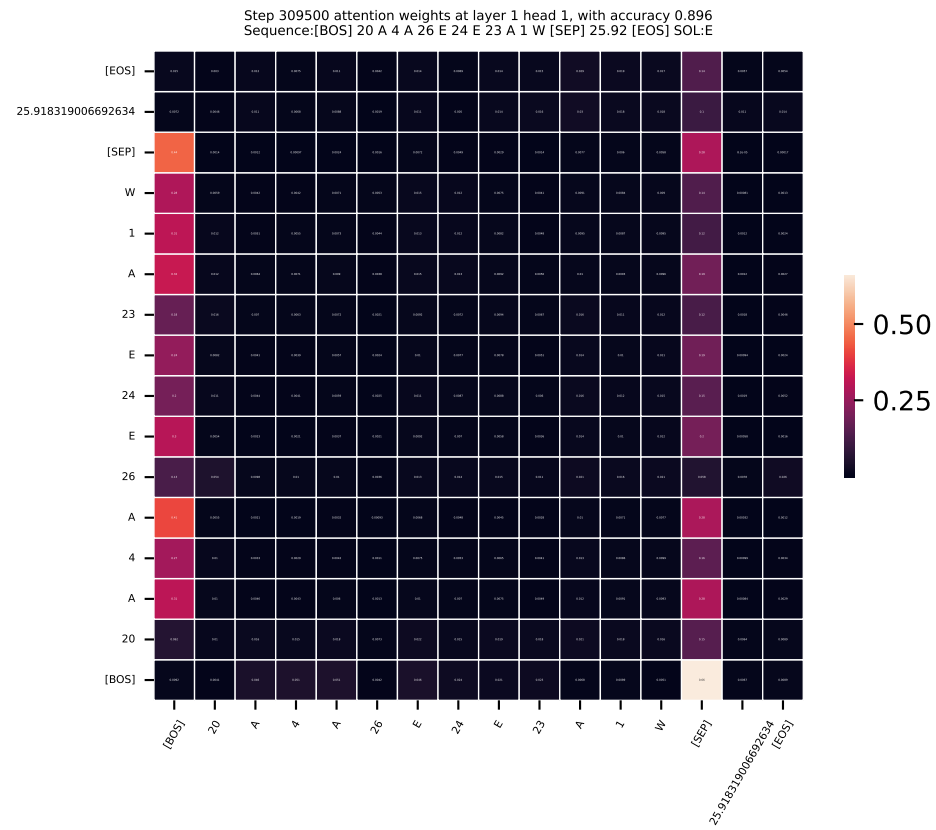

Step 309500 attention weights at layer 0 head 0, with accuracy 0.896
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

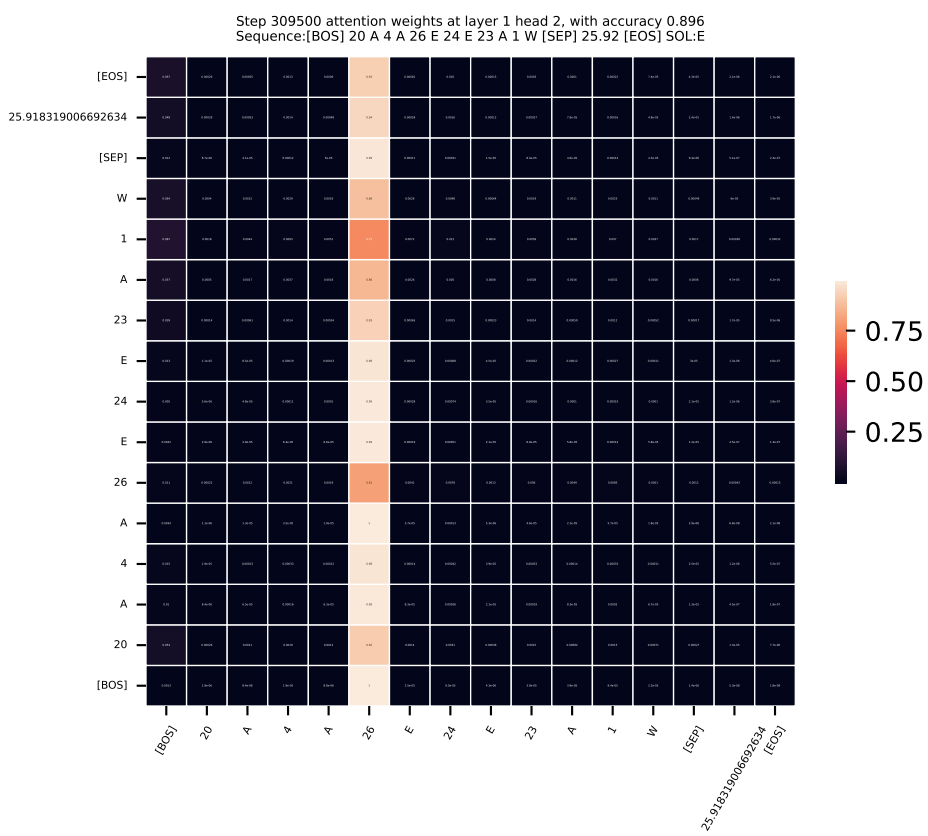

Step 309500 attention weights at layer 0 head 1, with accuracy 0.896
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

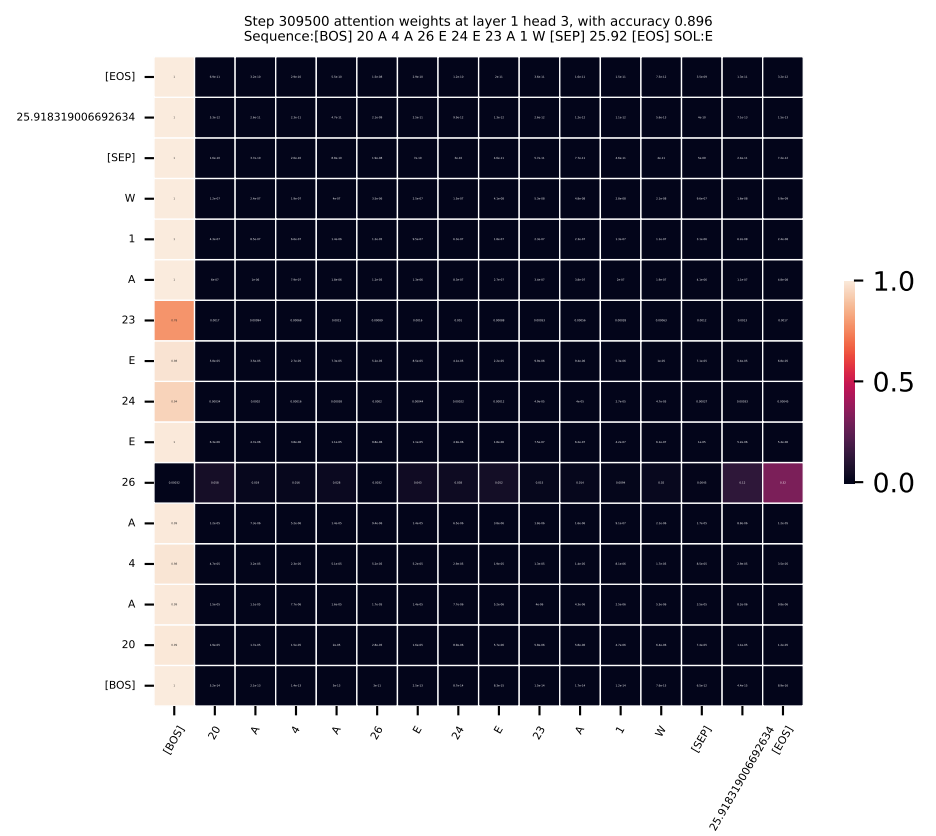

Step 309500 attention weights at layer 0 head 2, with accuracy 0.896
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

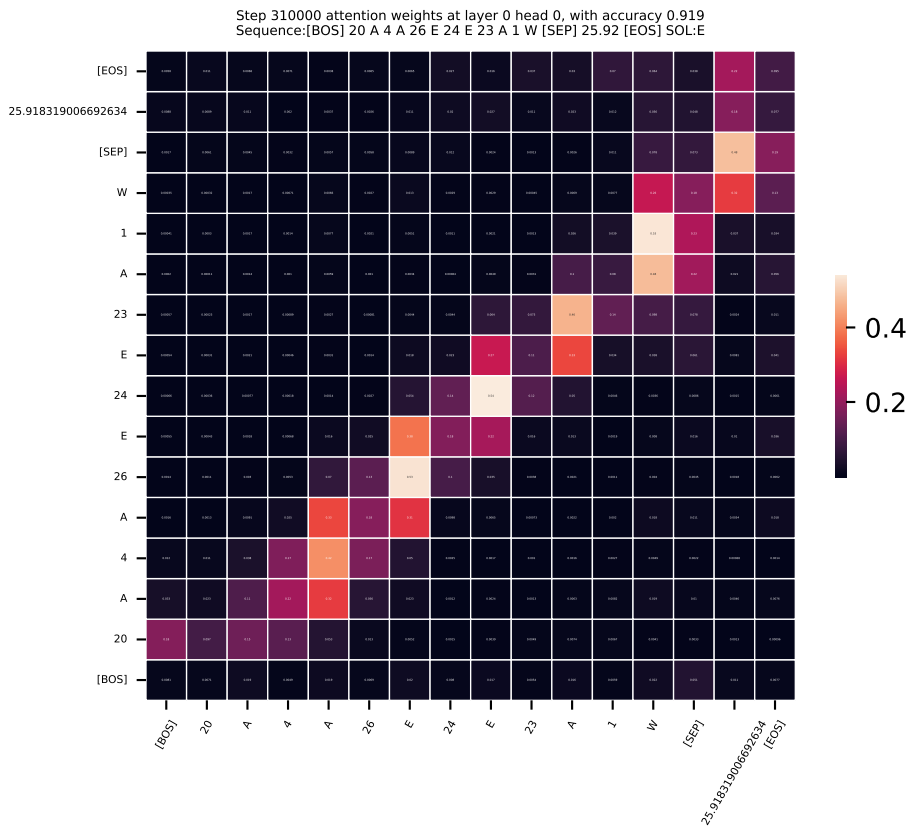

Step 309500 attention weights at layer 0 head 3, with accuracy 0.896
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

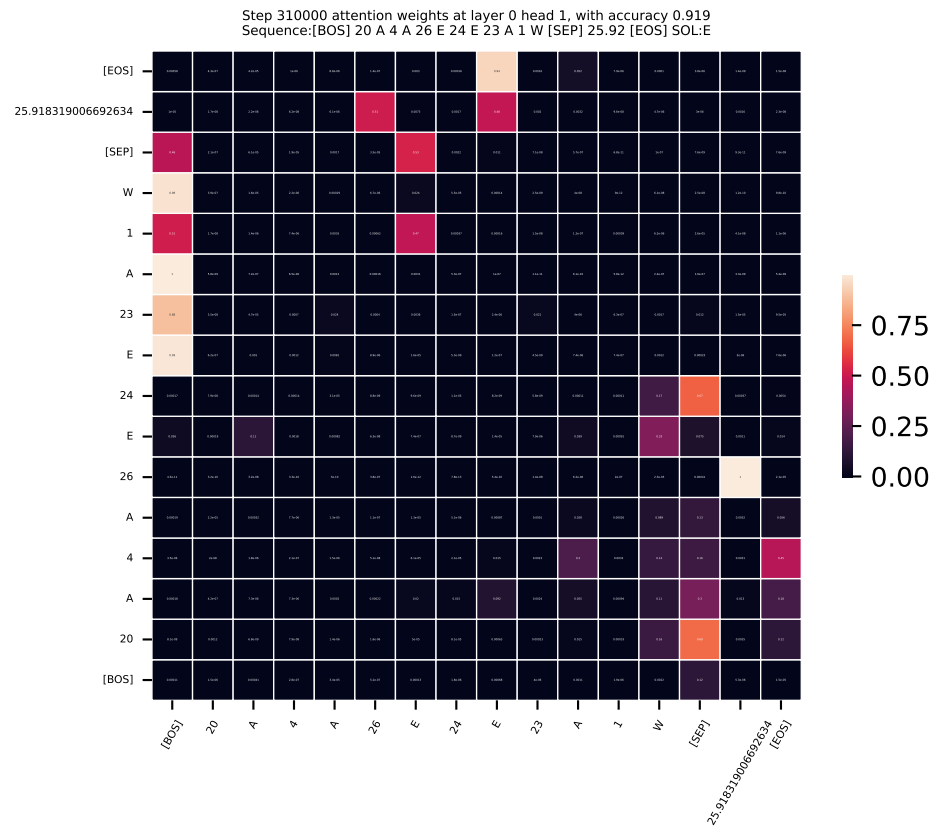

Step 309500 attention weights at layer 1 head 0, with accuracy 0.896
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

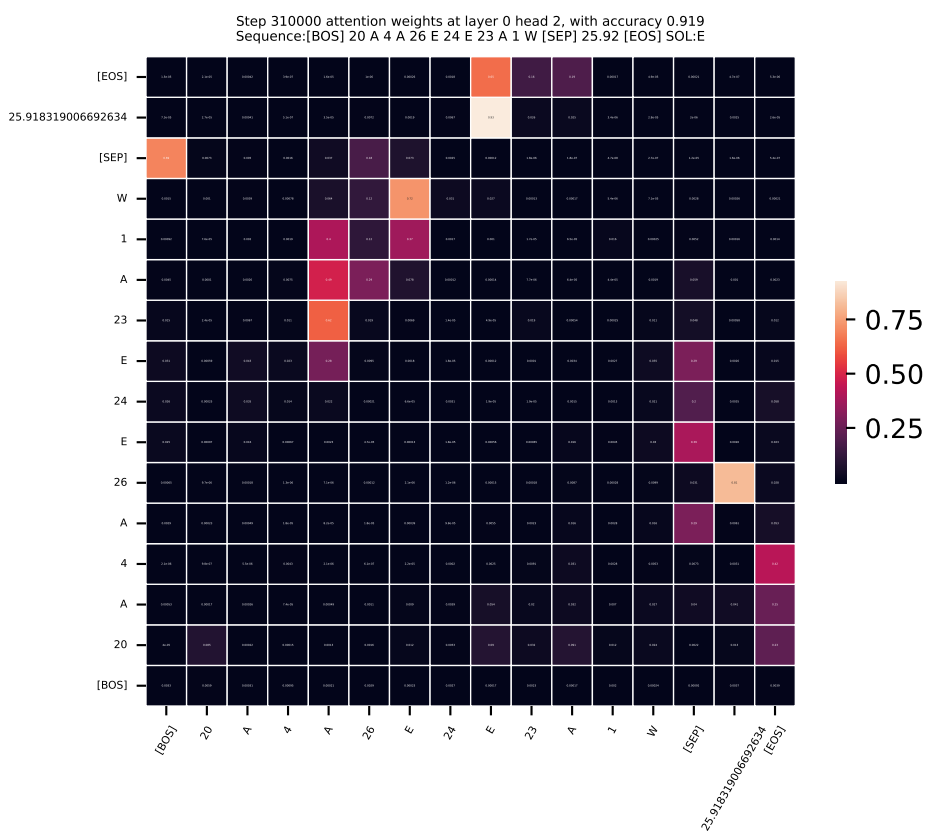

Step 309500 attention weights at layer 1 head 1, with accuracy 0.896
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

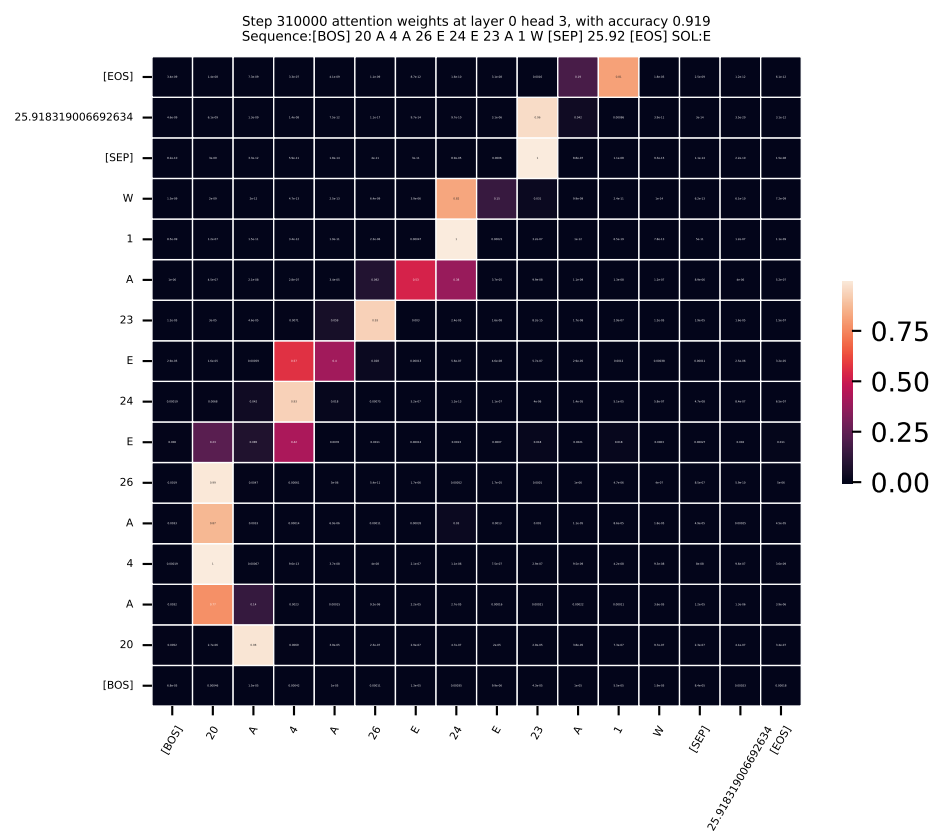

Step 309500 attention weights at layer 1 head 2, with accuracy 0.896
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

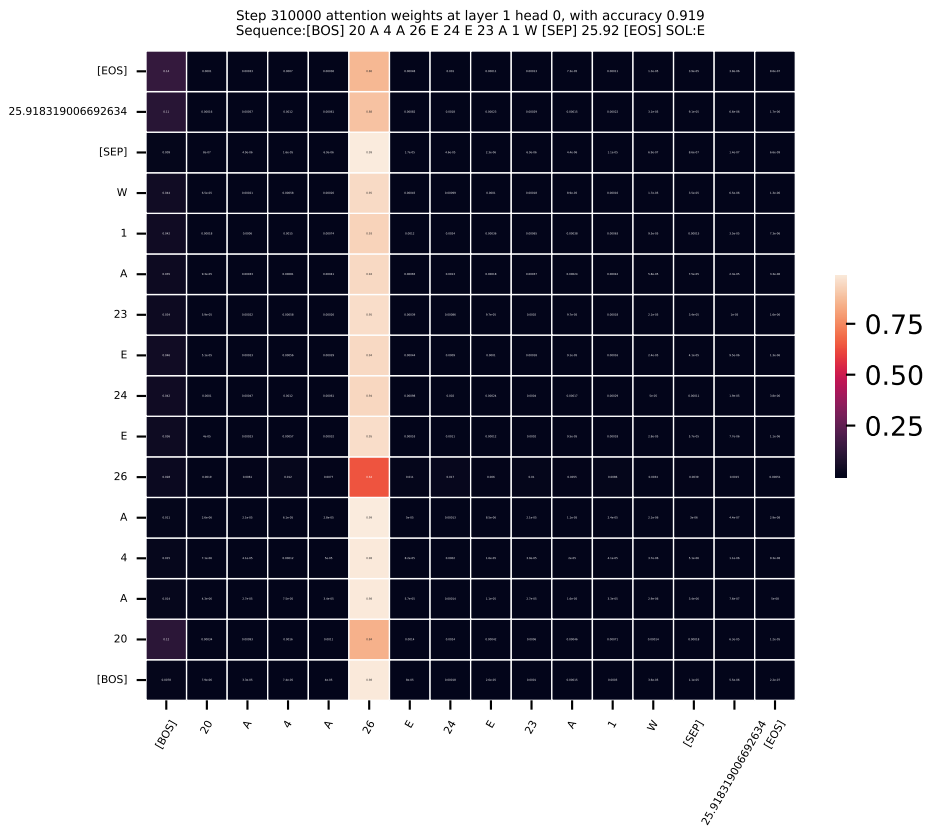

Step 309500 attention weights at layer 1 head 3, with accuracy 0.896
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
1.0
0.5
0.0

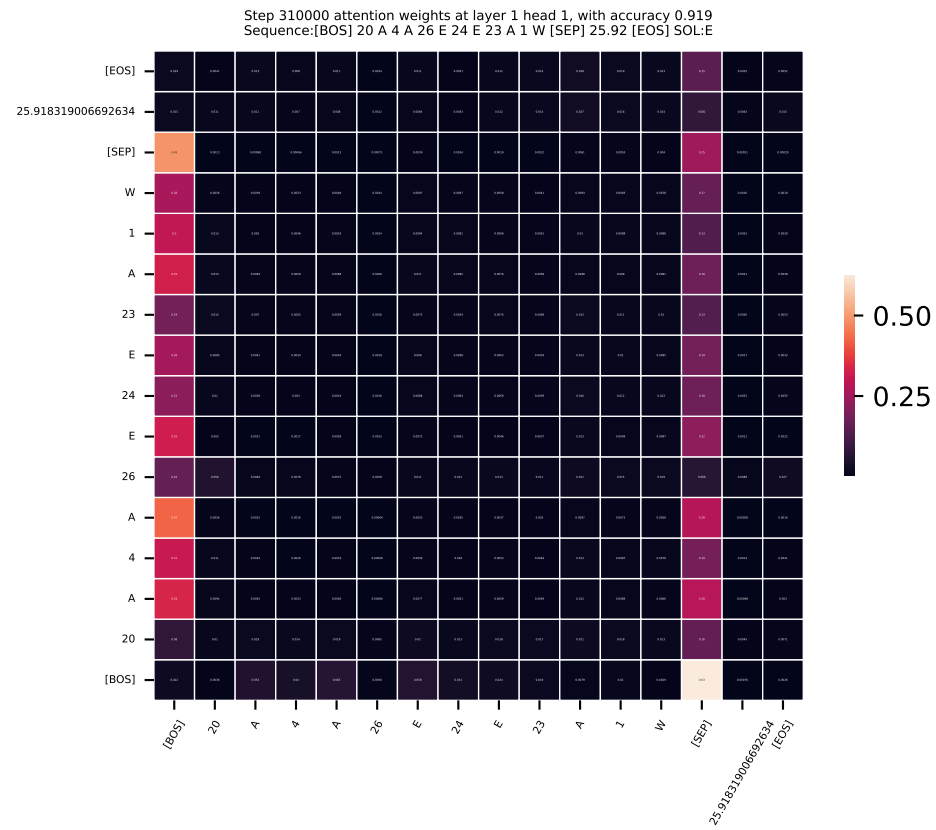

Step 310000 attention weights at layer 0 head 0, with accuracy 0.919
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

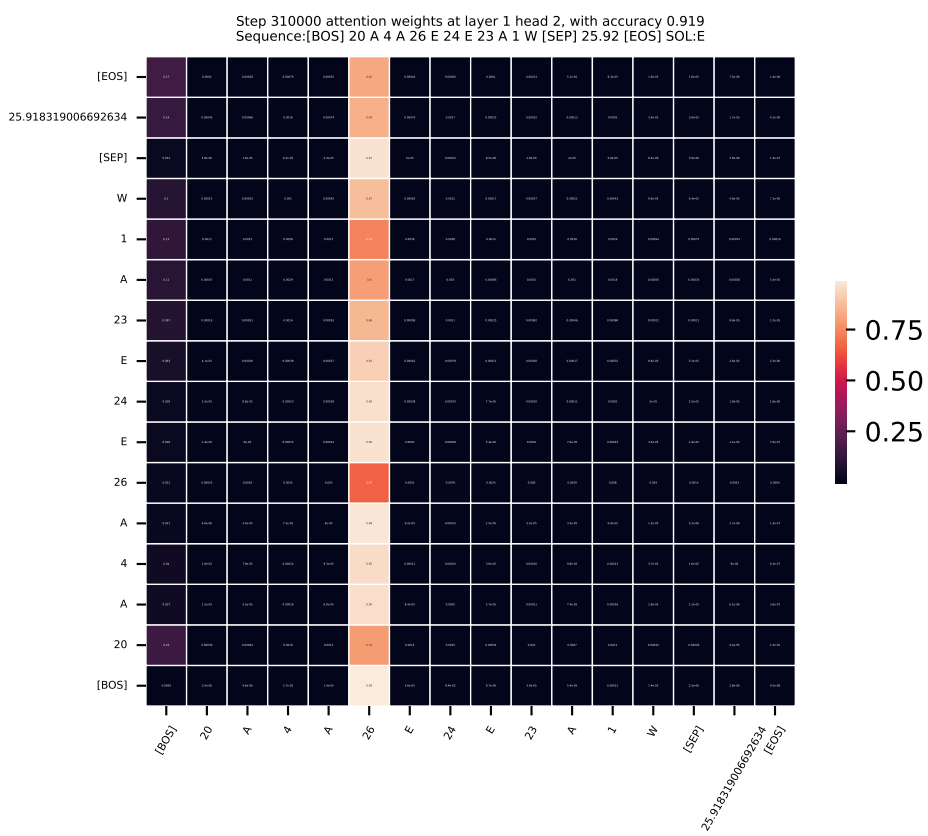

Step 310000 attention weights at layer 0 head 1, with accuracy 0.919
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

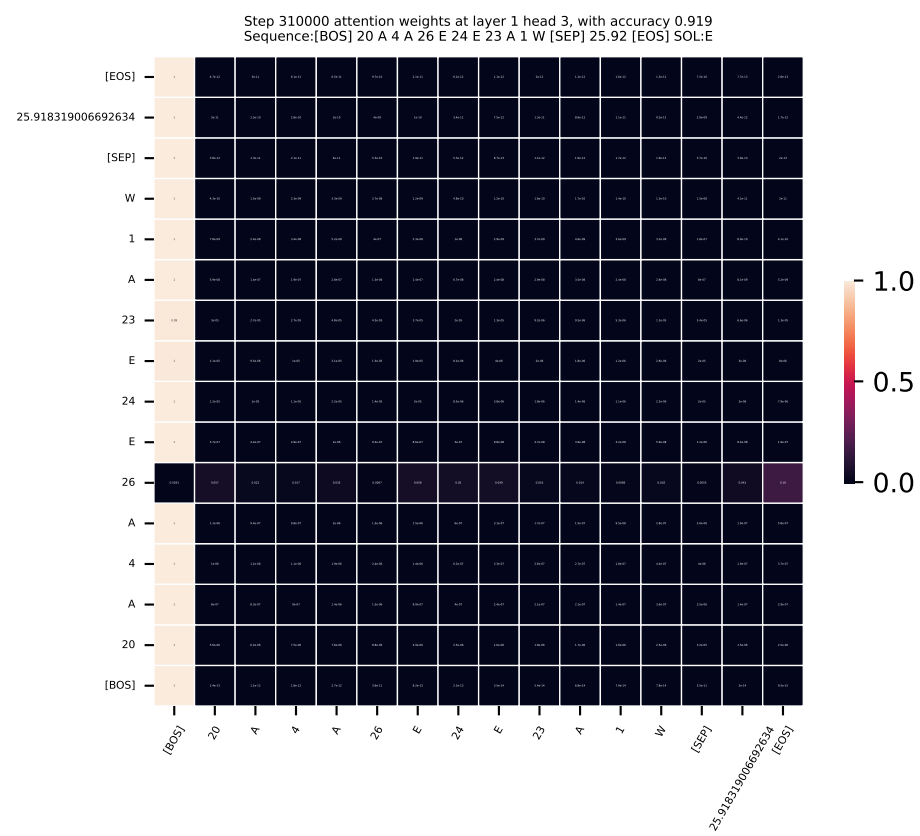

Step 310000 attention weights at layer 0 head 2, with accuracy 0.919
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

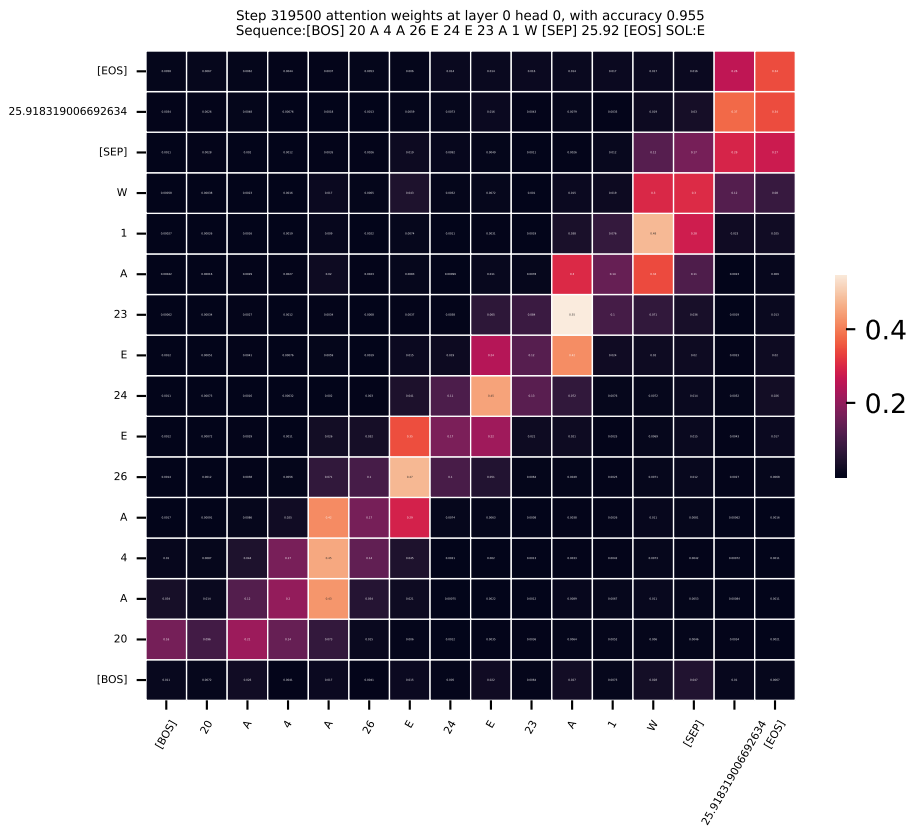

Step 310000 attention weights at layer 0 head 3, with accuracy 0.919
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

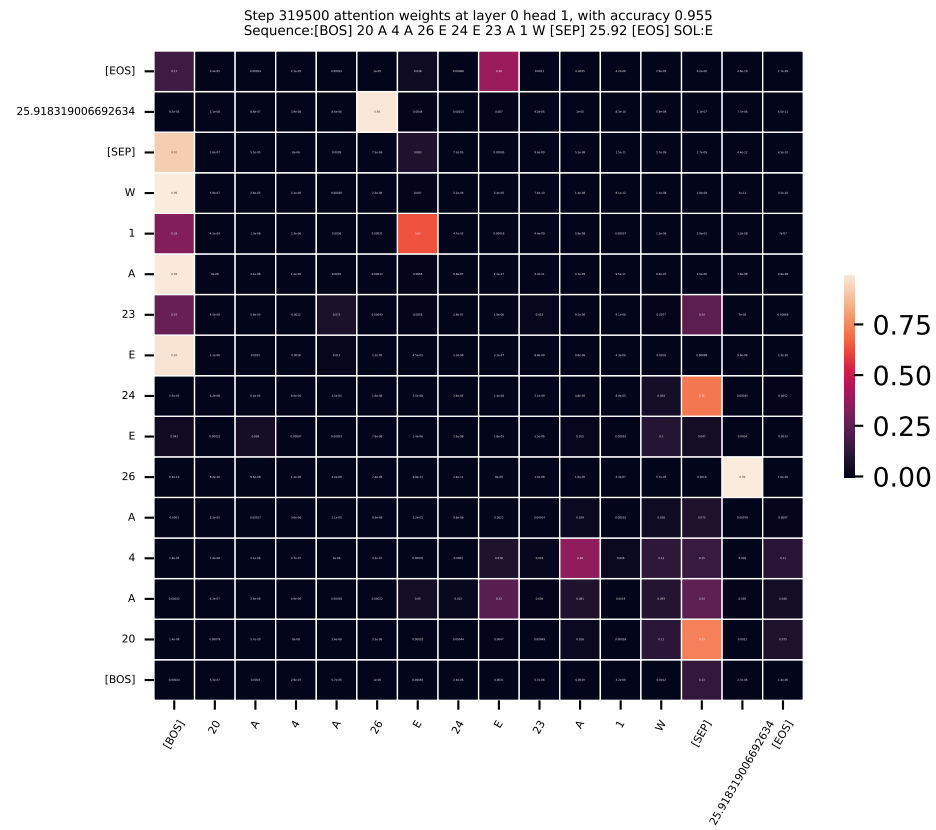

Step 310000 attention weights at layer 1 head 0, with accuracy 0.919
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

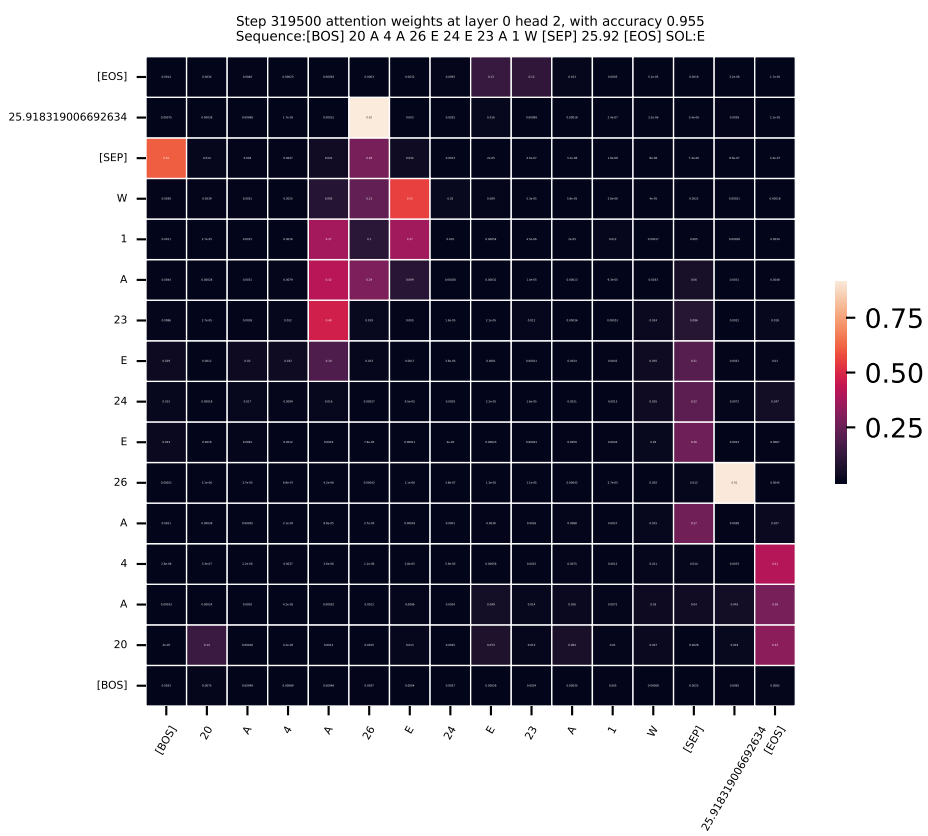

Step 310000 attention weights at layer 1 head 1, with accuracy 0.919
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

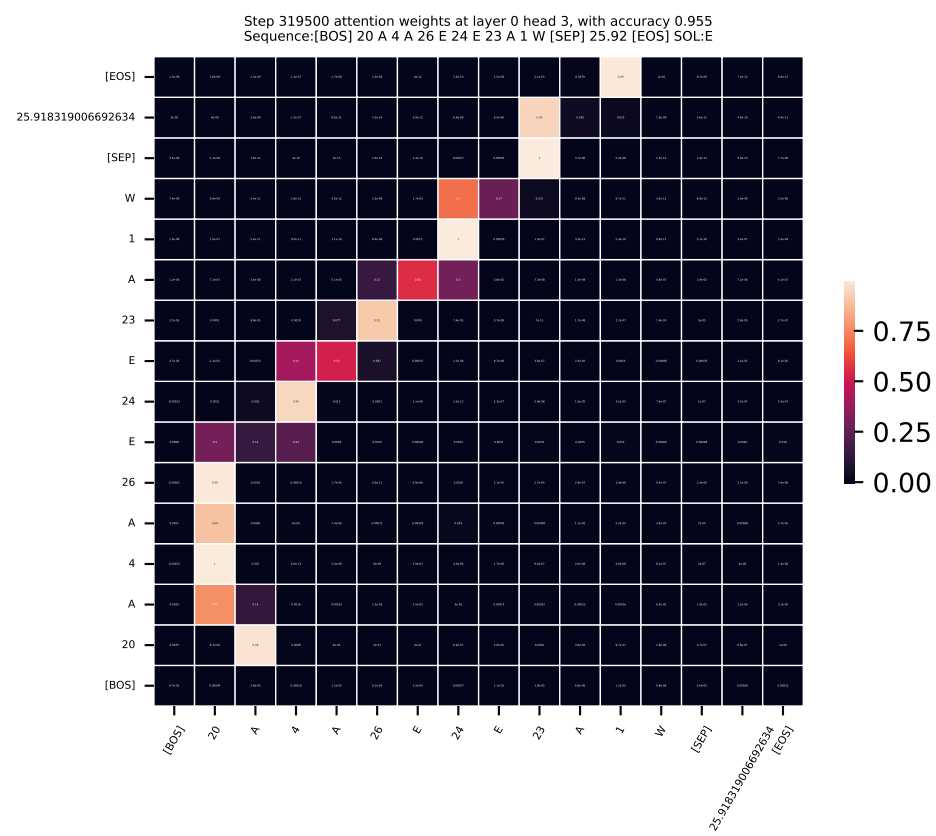

Step 310000 attention weights at layer 1 head 2, with accuracy 0.919
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

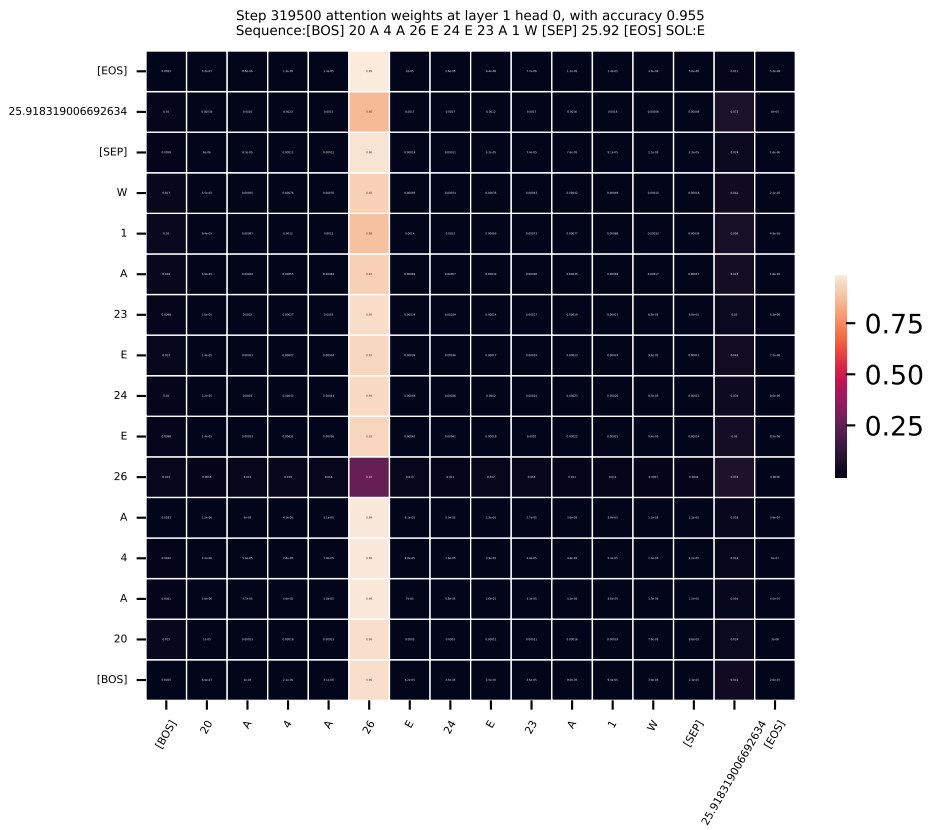

Step 310000 attention weights at layer 1 head 3, with accuracy 0.919
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
1.0
0.5
0.0

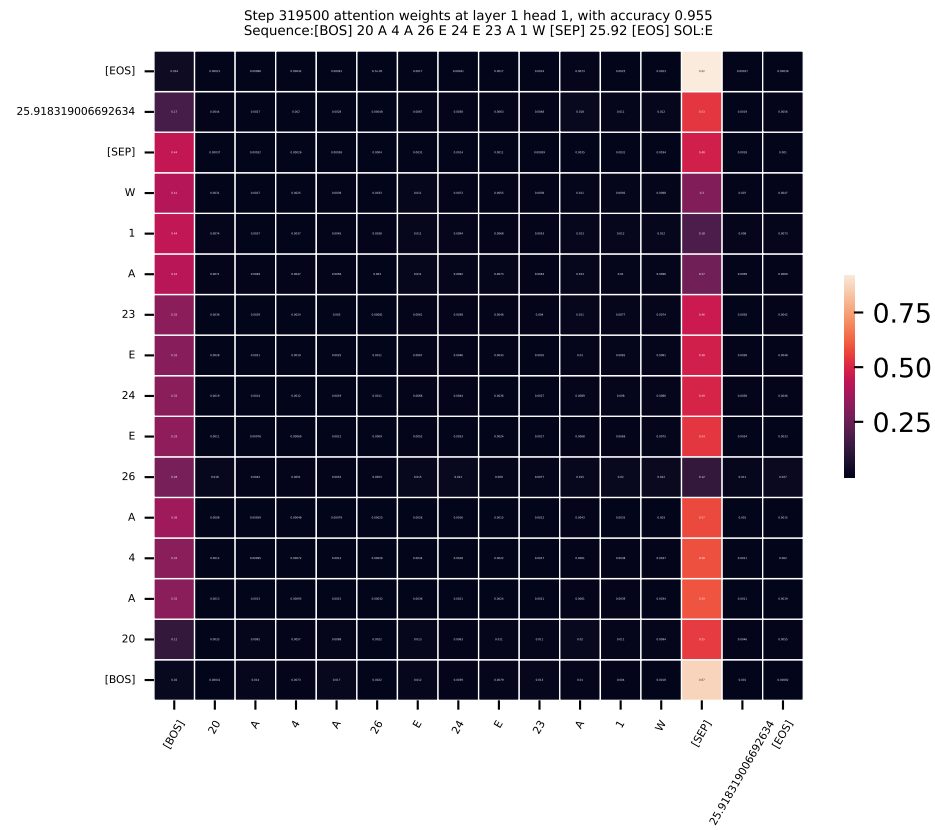

Step 319500 attention weights at layer 0 head 0, with accuracy 0.955
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

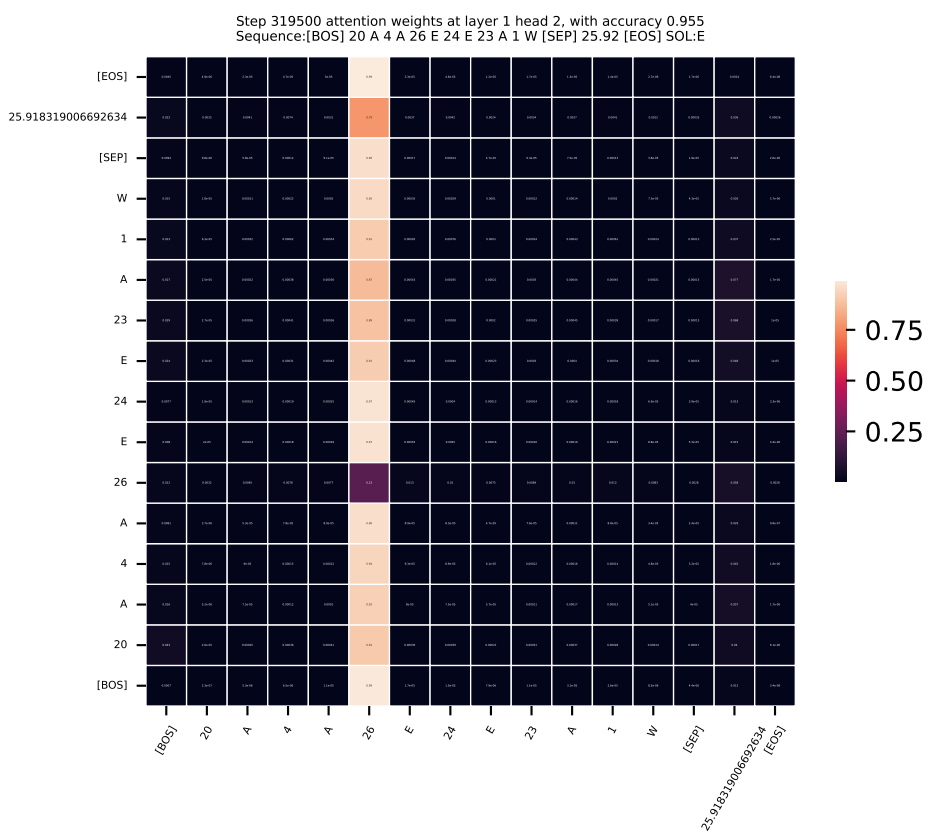

Step 319500 attention weights at layer 0 head 1, with accuracy 0.955
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

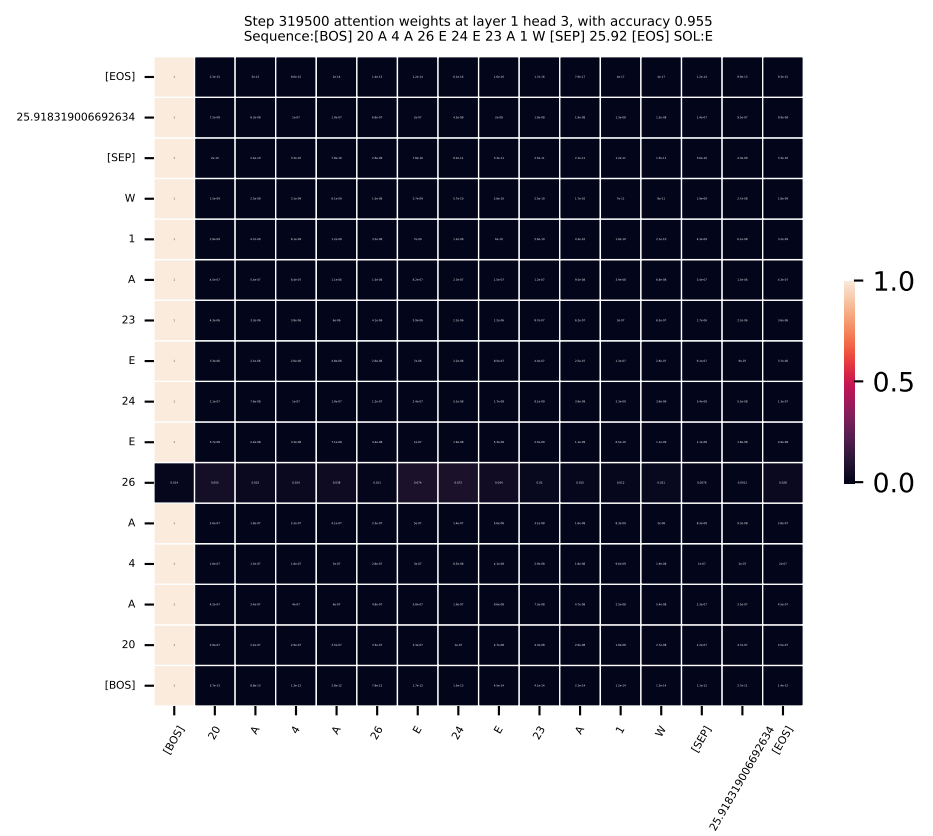

Step 319500 attention weights at layer 0 head 2, with accuracy 0.955
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

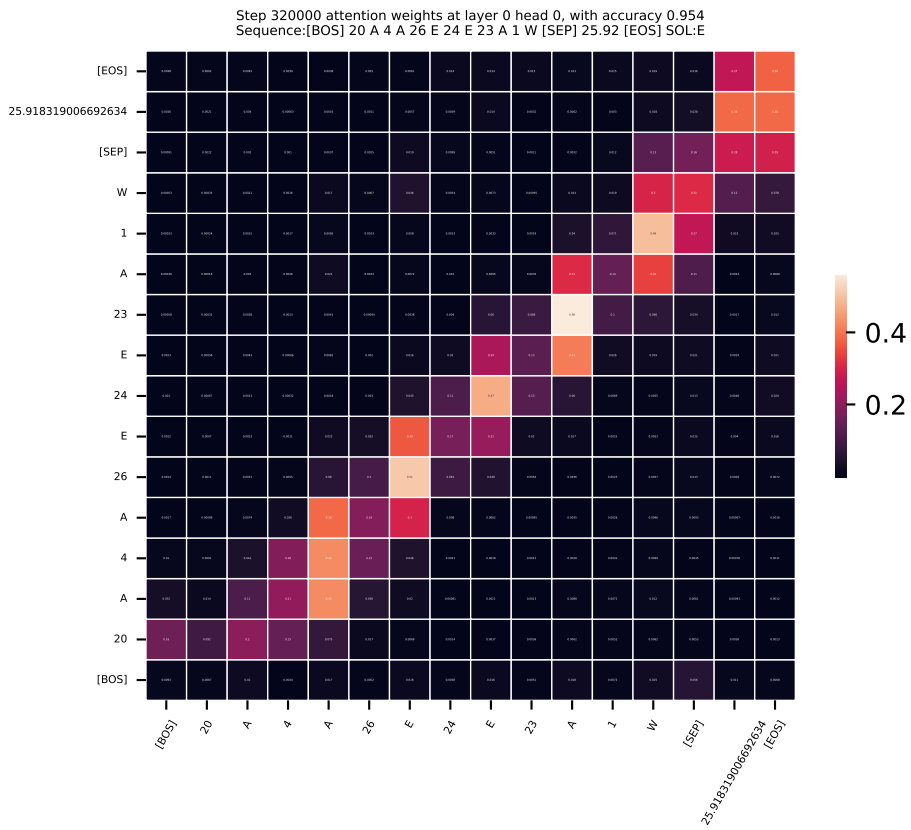

Step 319500 attention weights at layer 0 head 3, with accuracy 0.955
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

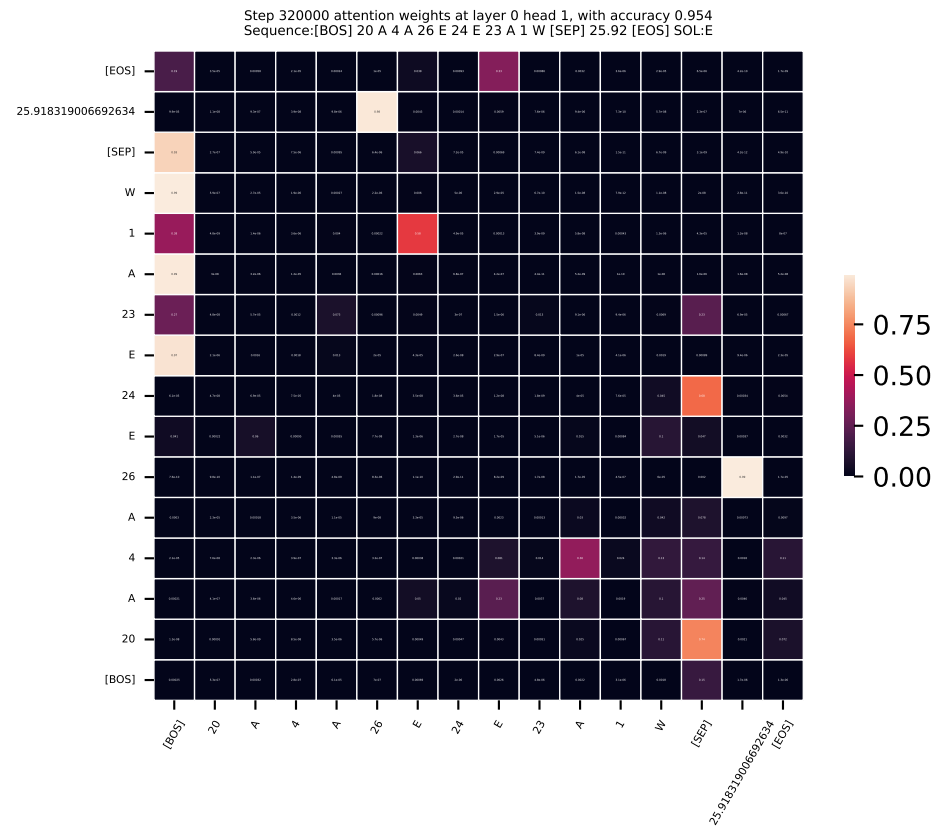

Step 319500 attention weights at layer 1 head 0, with accuracy 0.955
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

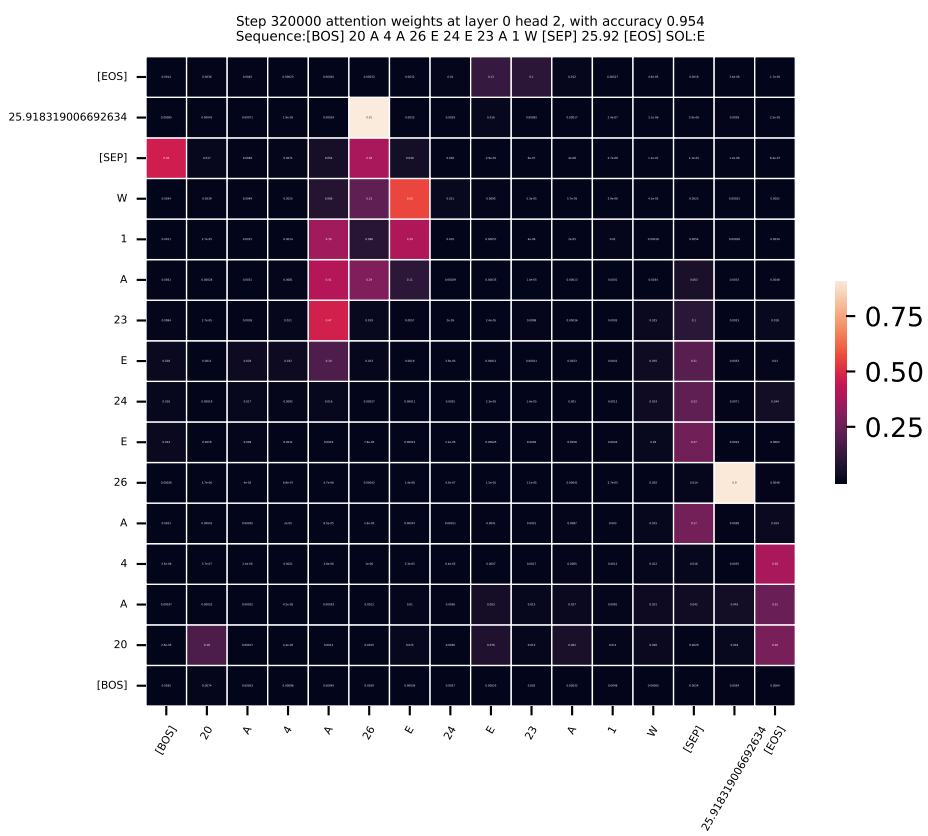

Step 319500 attention weights at layer 1 head 1, with accuracy 0.955
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

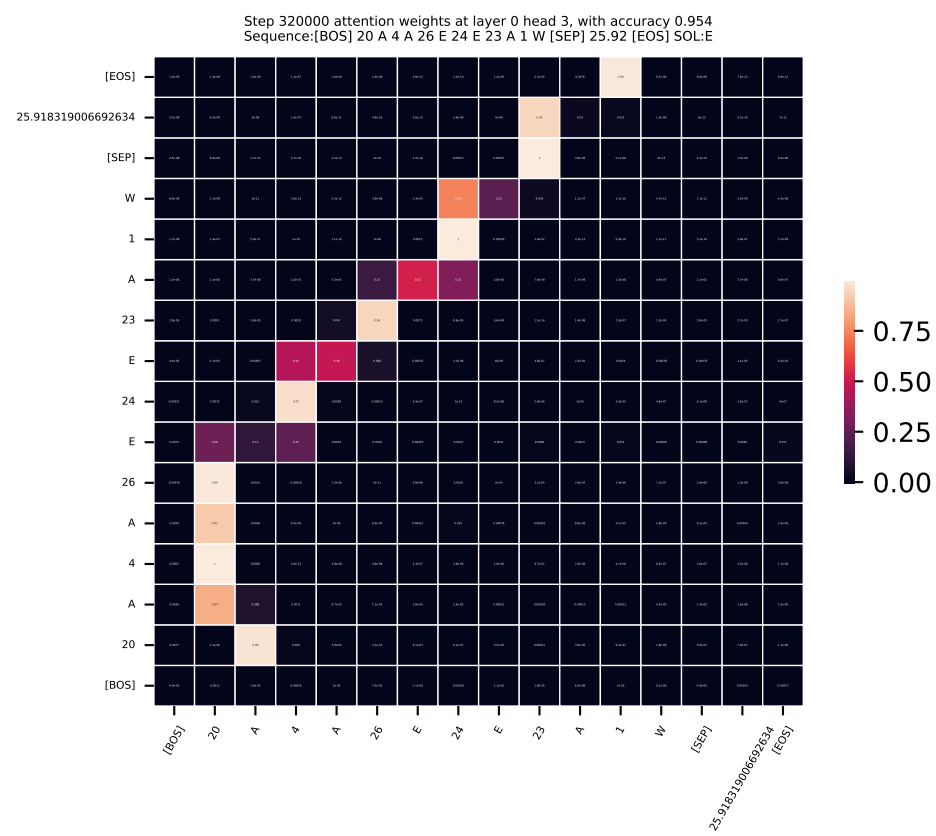

Step 319500 attention weights at layer 1 head 2, with accuracy 0.955
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

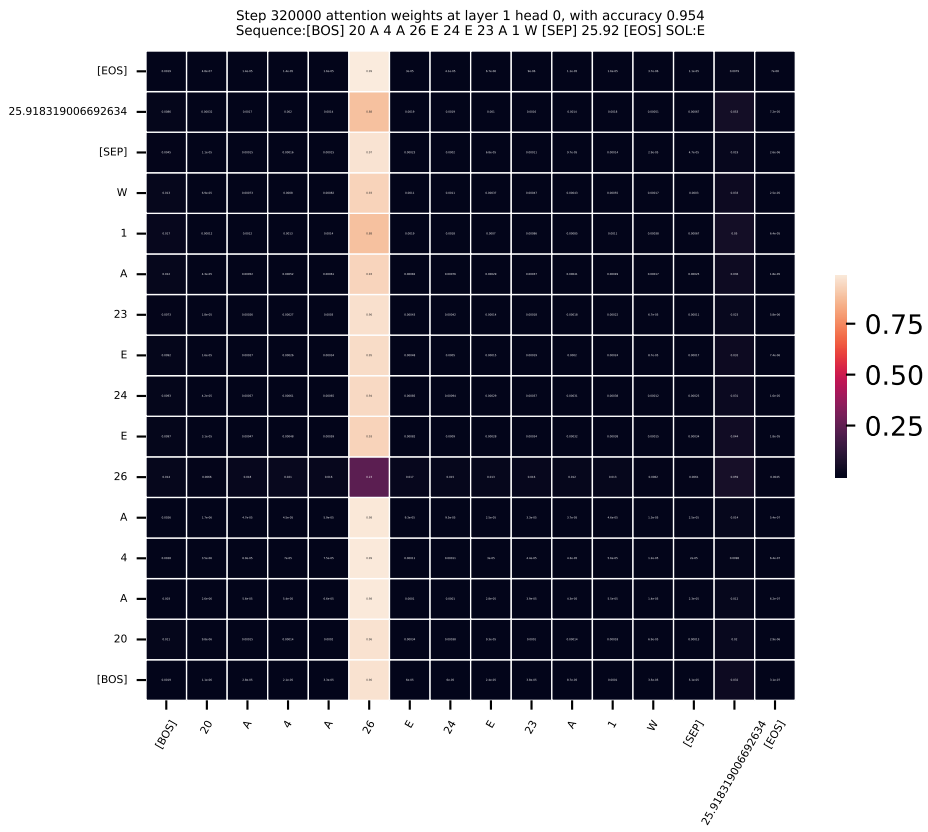

Step 319500 attention weights at layer 1 head 3, with accuracy 0.955
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
1.0
0.5
0.0

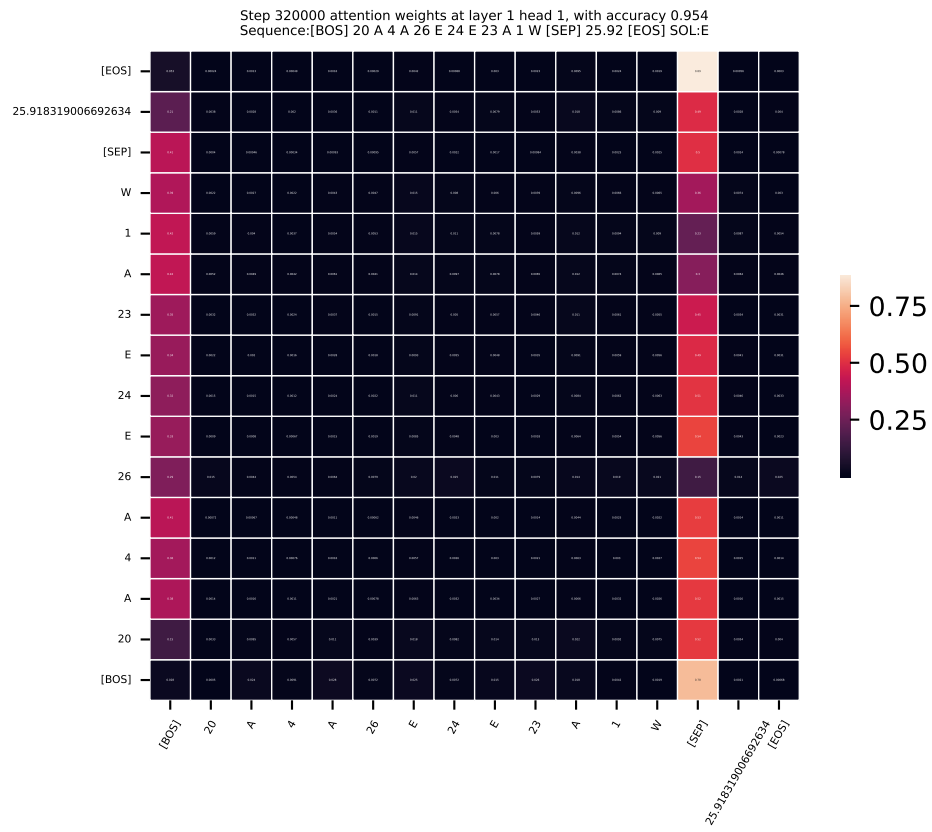

Step 320000 attention weights at layer 0 head 0, with accuracy 0.954
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

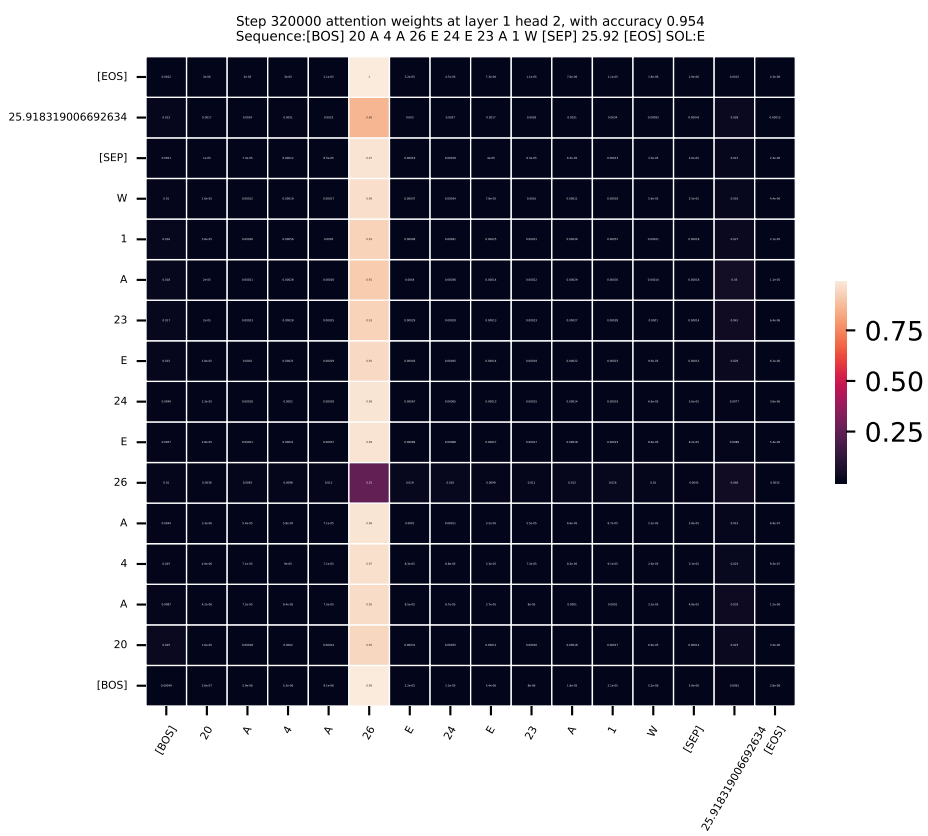

Step 320000 attention weights at layer 0 head 1, with accuracy 0.954
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

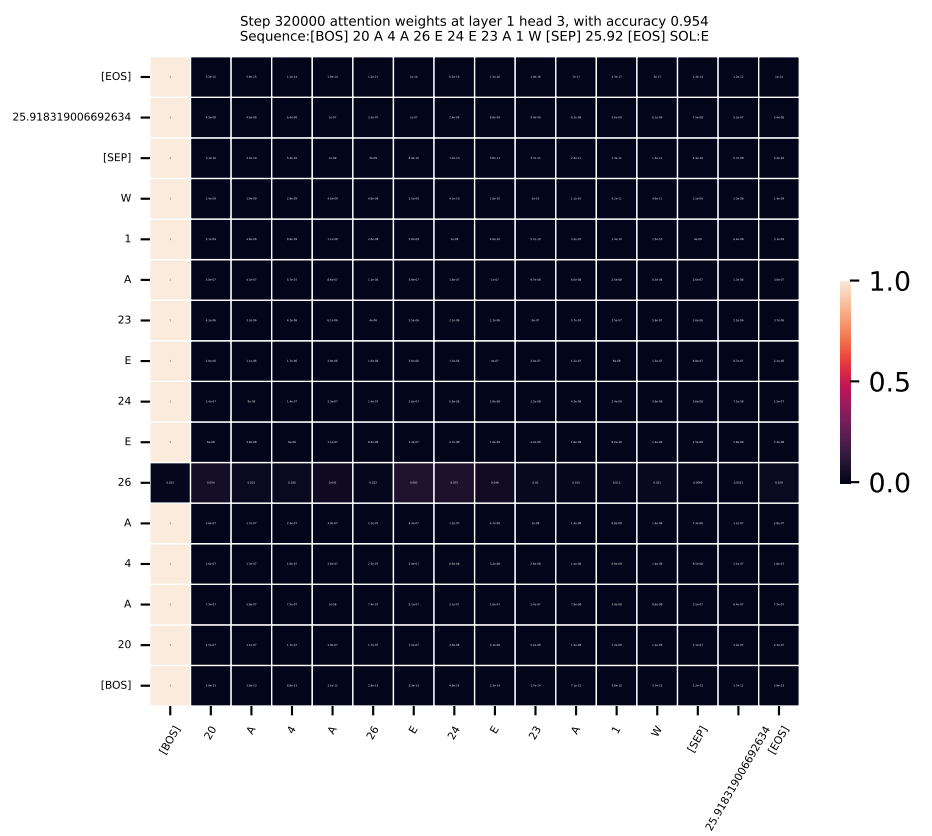

Step 320000 attention weights at layer 0 head 2, with accuracy 0.954
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

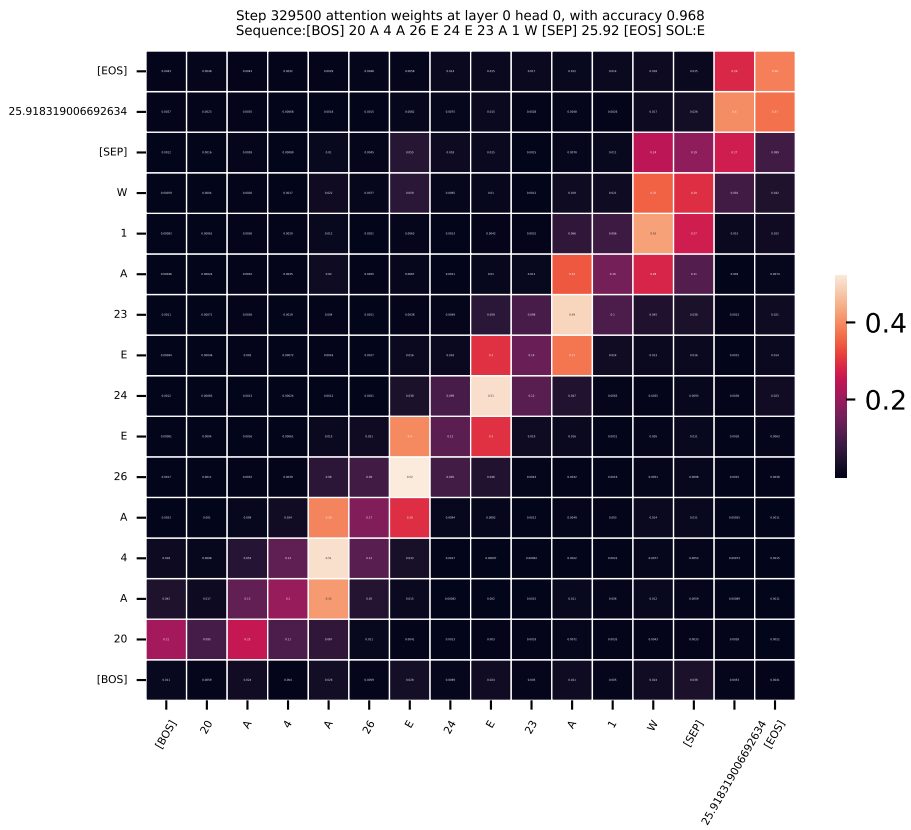

Step 320000 attention weights at layer 0 head 3, with accuracy 0.954
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

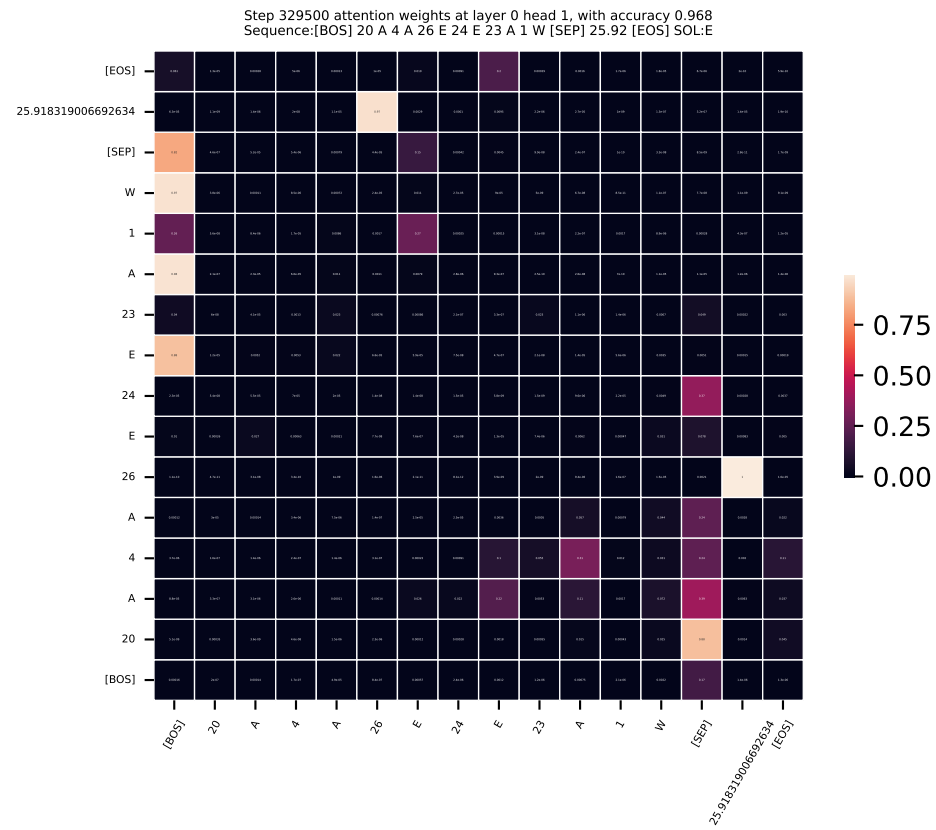

Step 320000 attention weights at layer 1 head 0, with accuracy 0.954
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

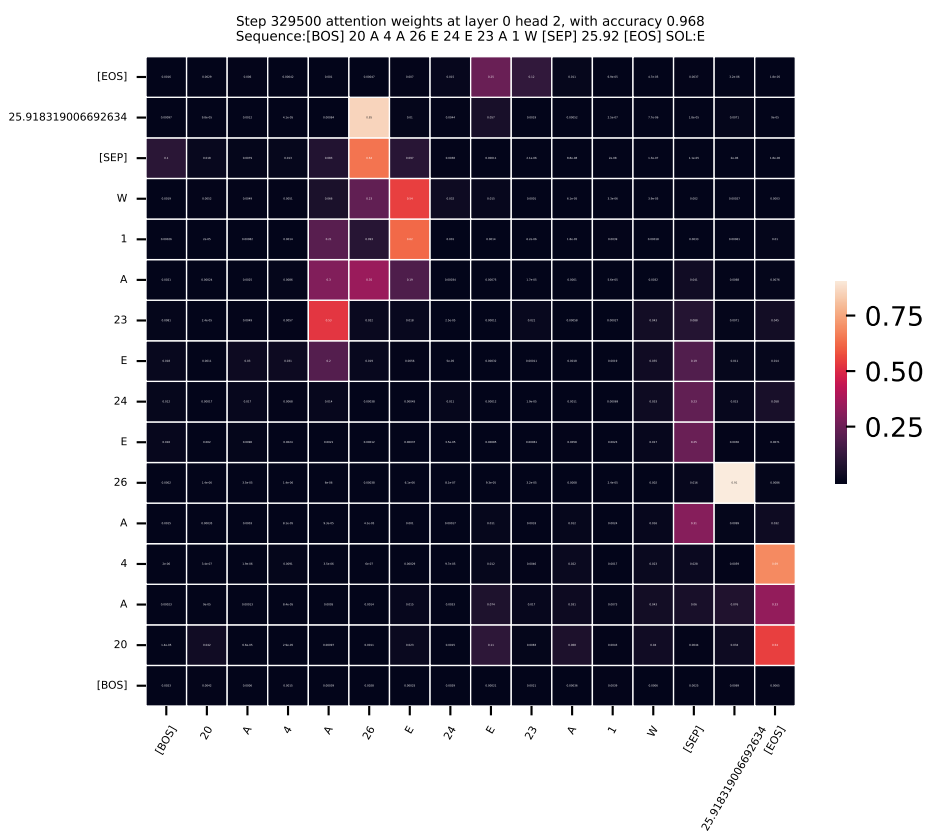

Step 320000 attention weights at layer 1 head 1, with accuracy 0.954
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

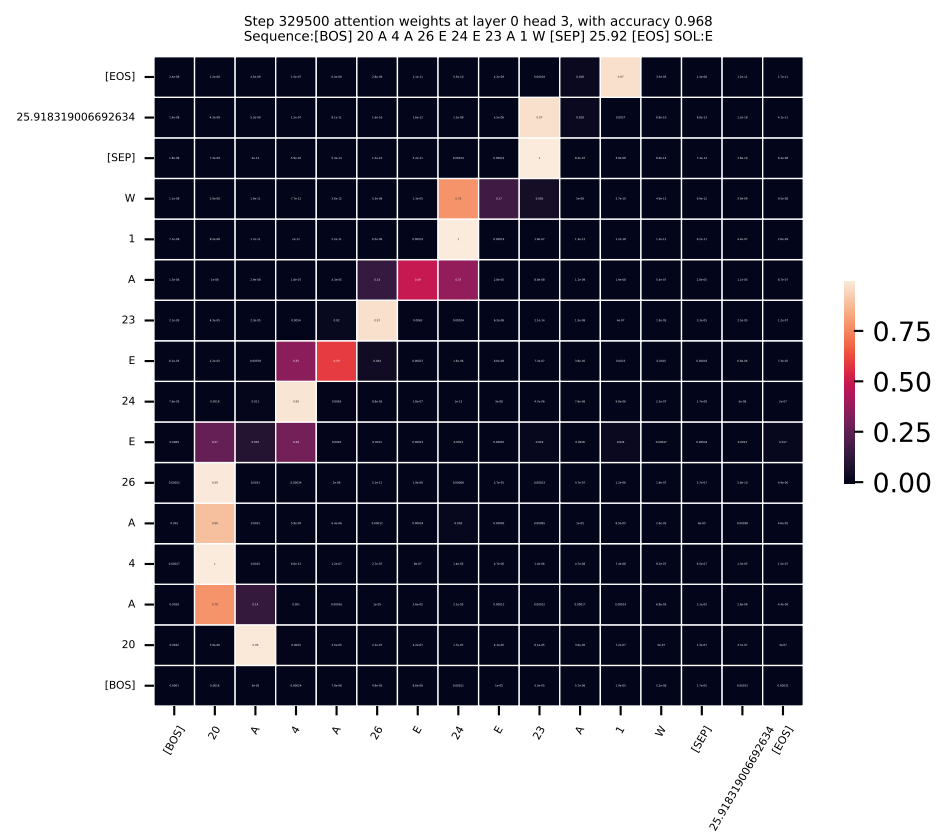

Step 320000 attention weights at layer 1 head 2, with accuracy 0.954
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

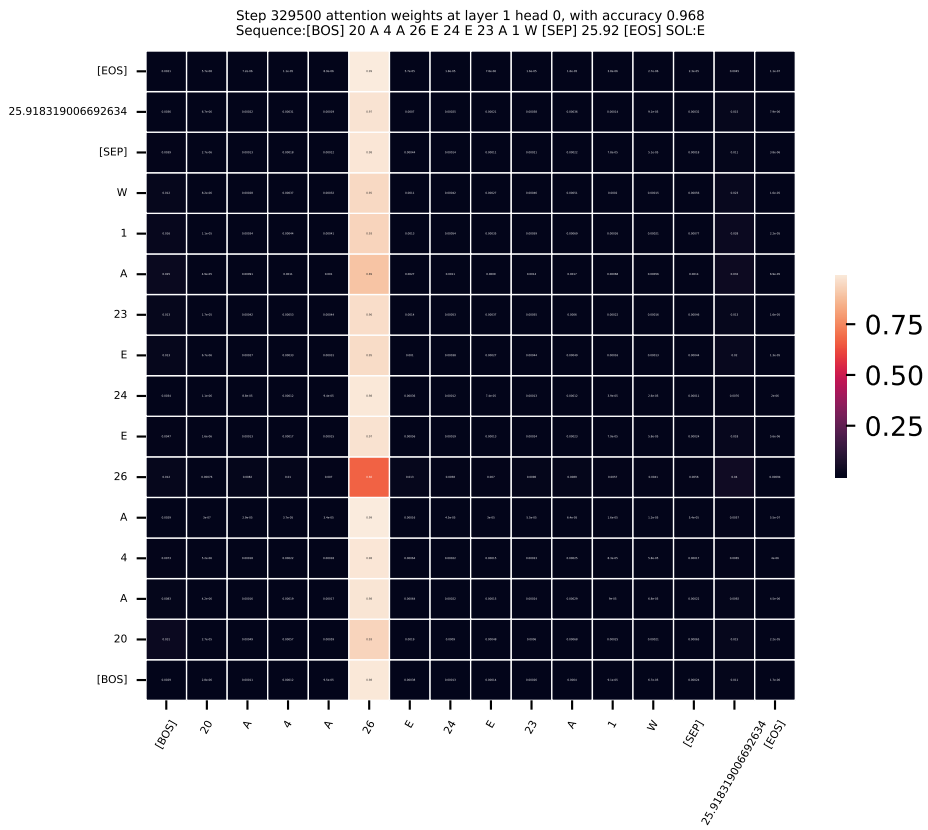

Step 320000 attention weights at layer 1 head 3, with accuracy 0.954
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
1.0
0.5
0.0

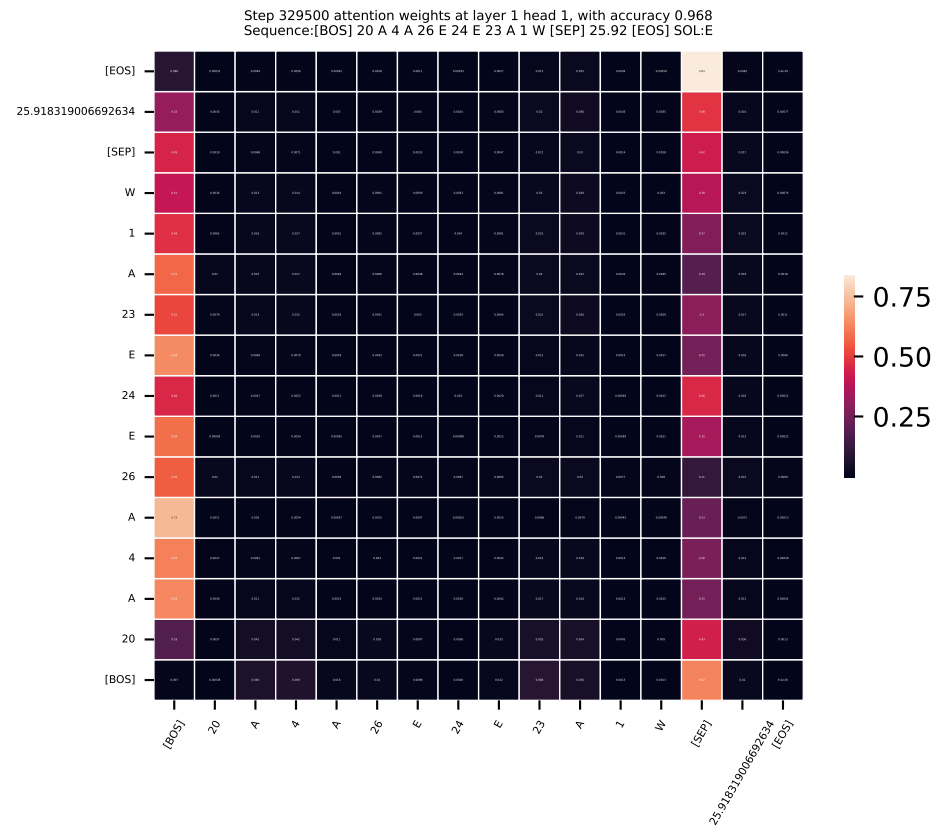

Step 329500 attention weights at layer 0 head 0, with accuracy 0.968
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

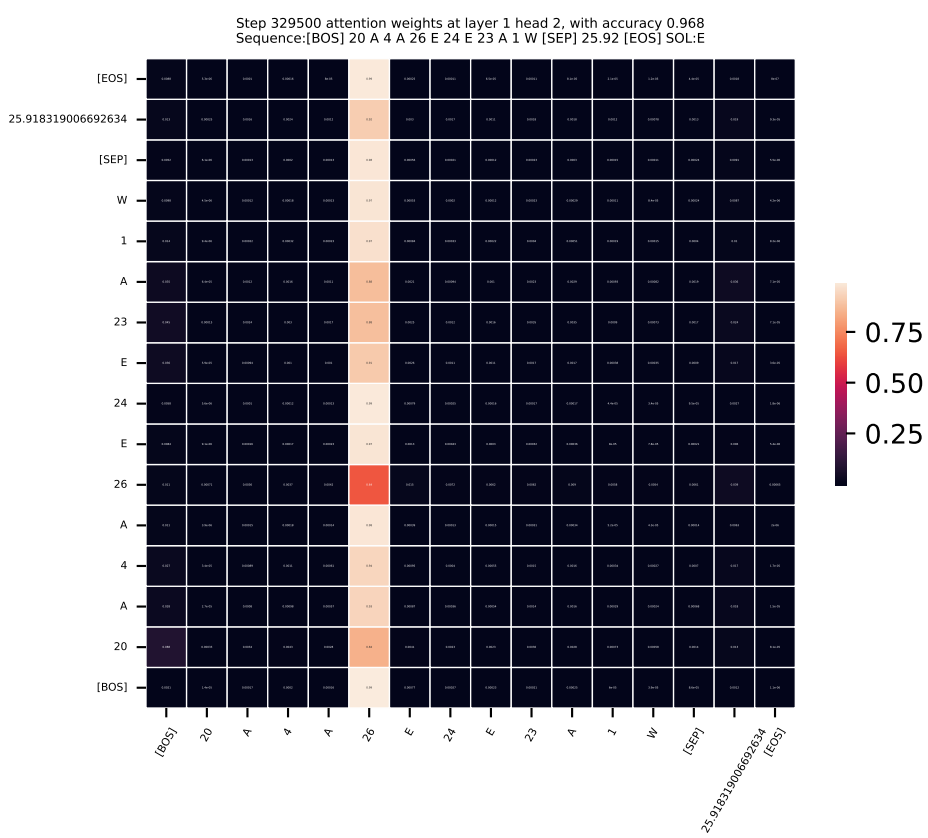

Step 329500 attention weights at layer 0 head 1, with accuracy 0.968
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

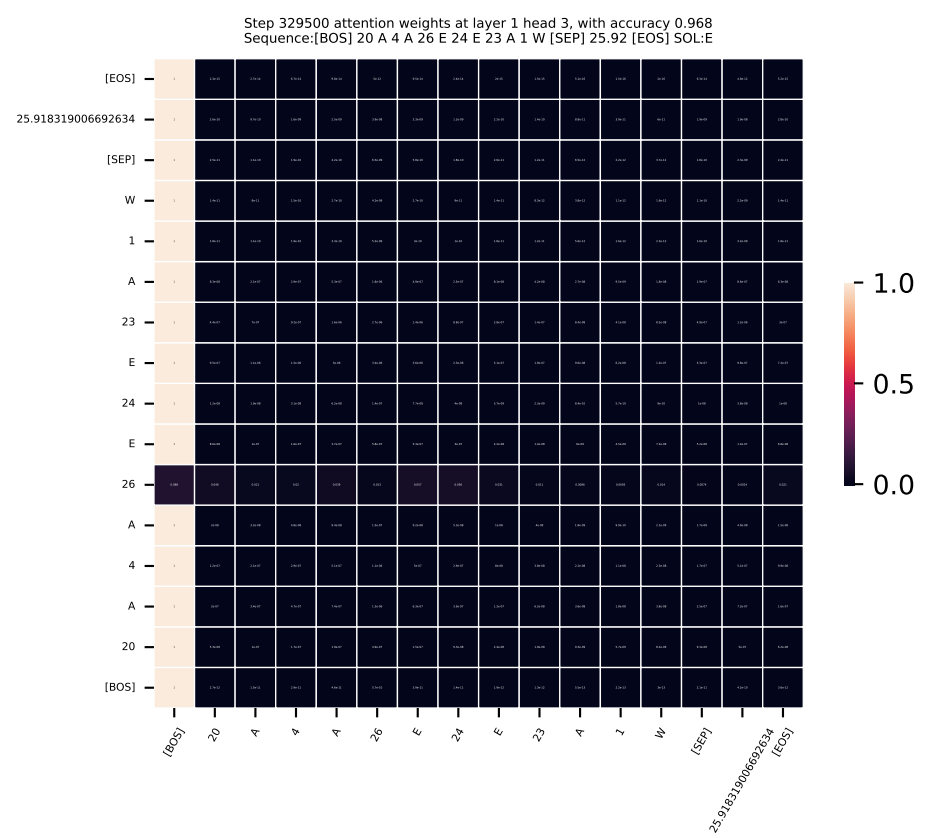

Step 329500 attention weights at layer 0 head 2, with accuracy 0.968
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

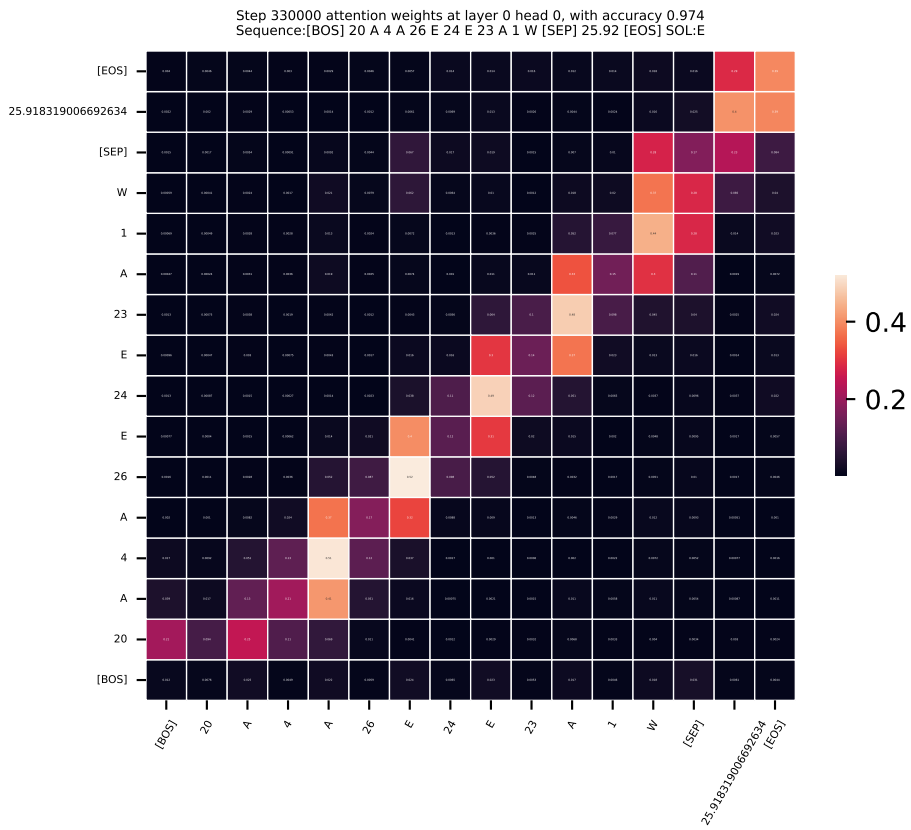

Step 329500 attention weights at layer 0 head 3, with accuracy 0.968
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

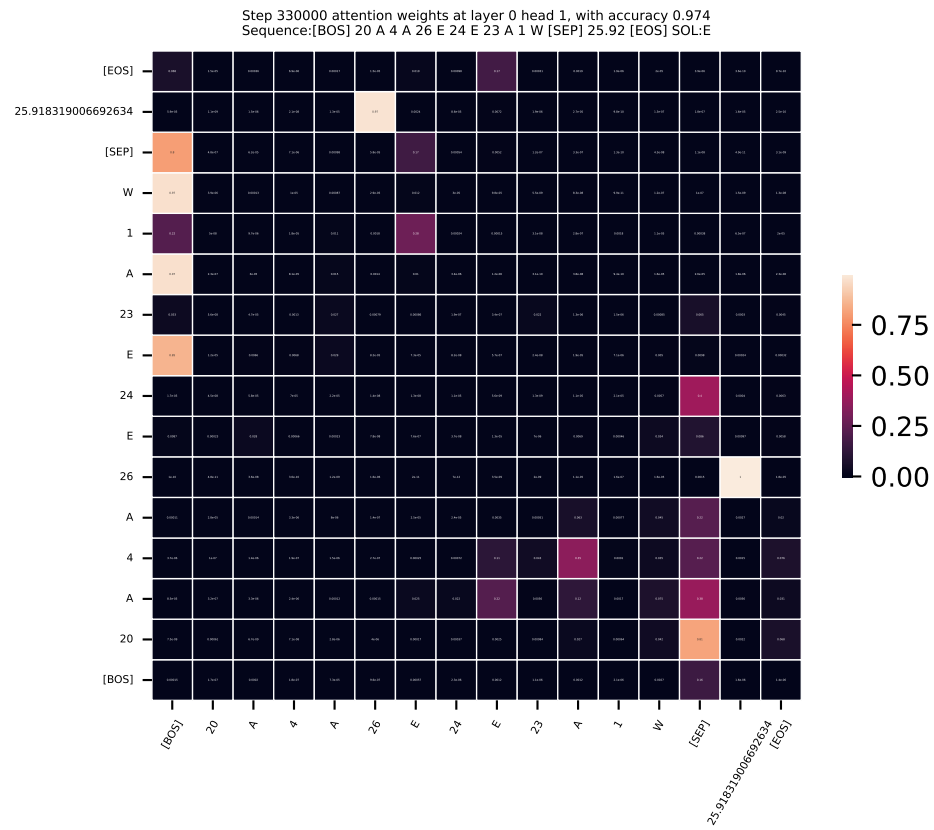

Step 329500 attention weights at layer 1 head 0, with accuracy 0.968
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

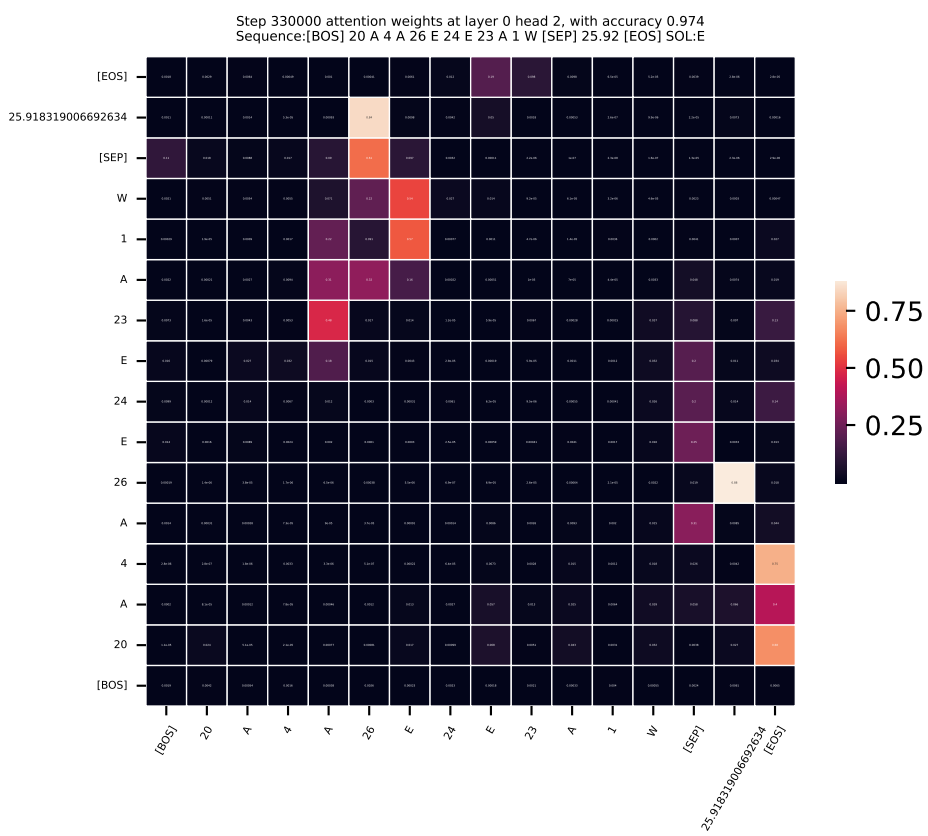

Step 329500 attention weights at layer 1 head 1, with accuracy 0.968
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

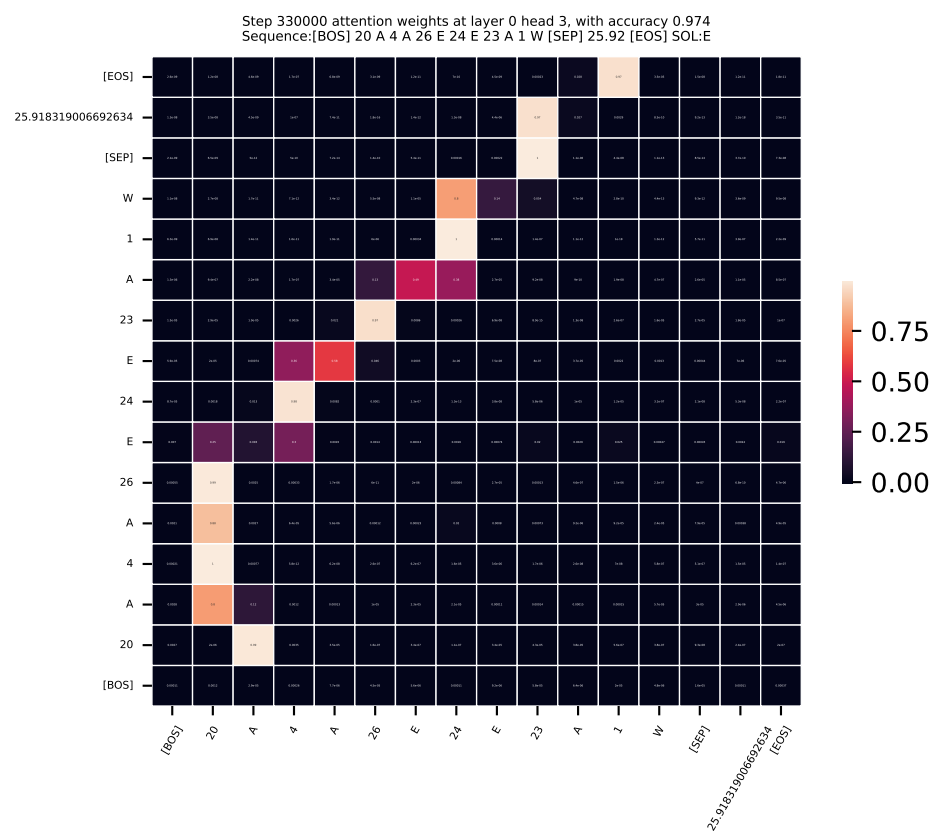

Step 329500 attention weights at layer 1 head 2, with accuracy 0.968
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

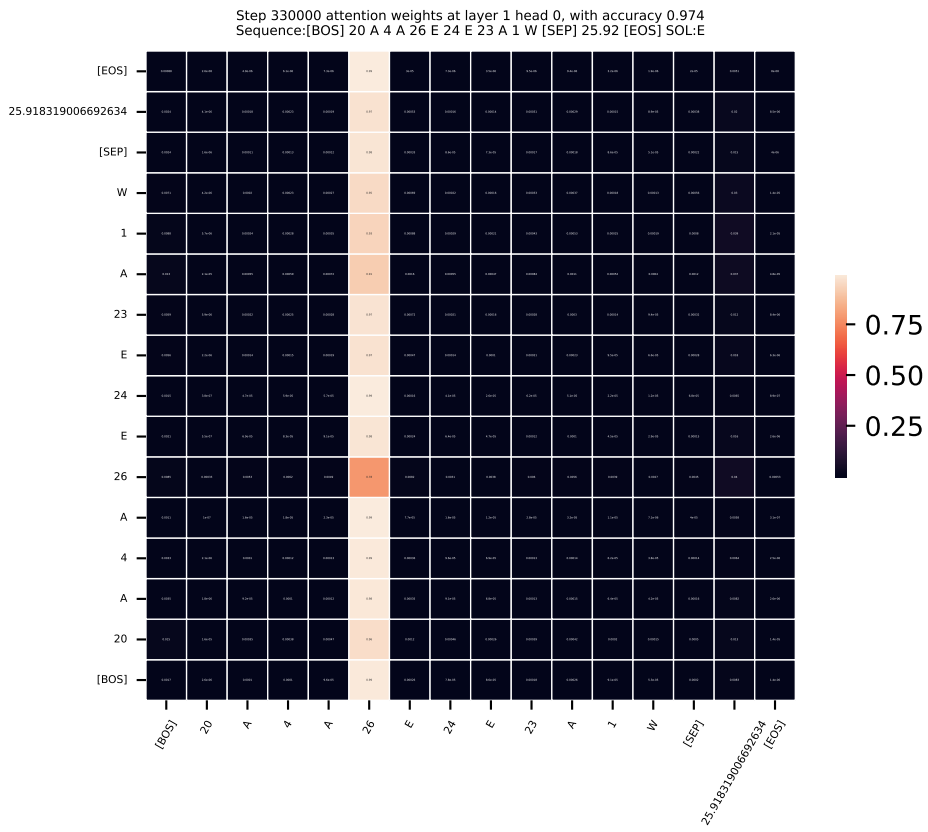

Step 329500 attention weights at layer 1 head 3, with accuracy 0.968
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
1.0
0.5
0.0

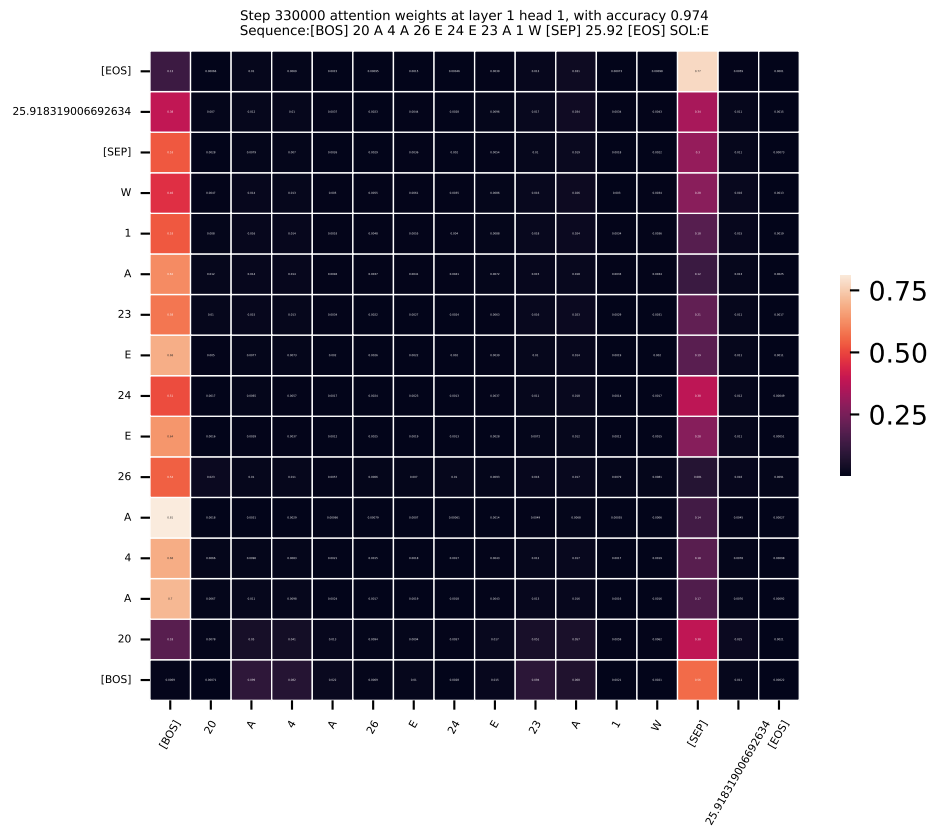

Step 330000 attention weights at layer 0 head 0, with accuracy 0.974
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

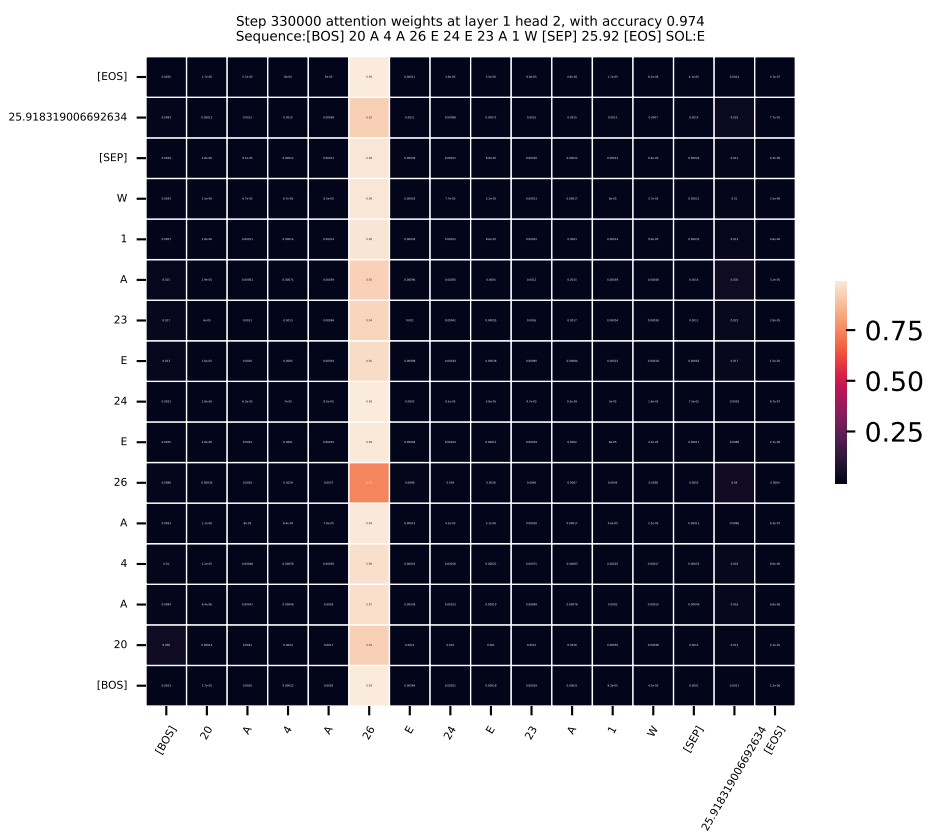

Step 330000 attention weights at layer 0 head 1, with accuracy 0.974
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

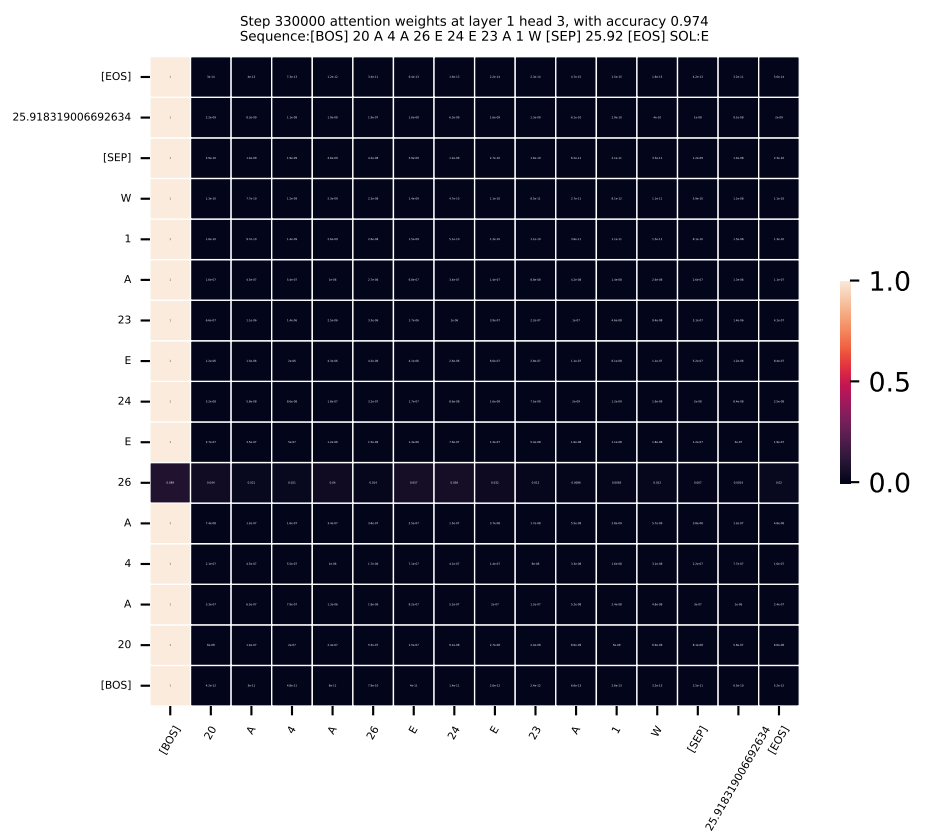

Step 330000 attention weights at layer 0 head 2, with accuracy 0.974
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

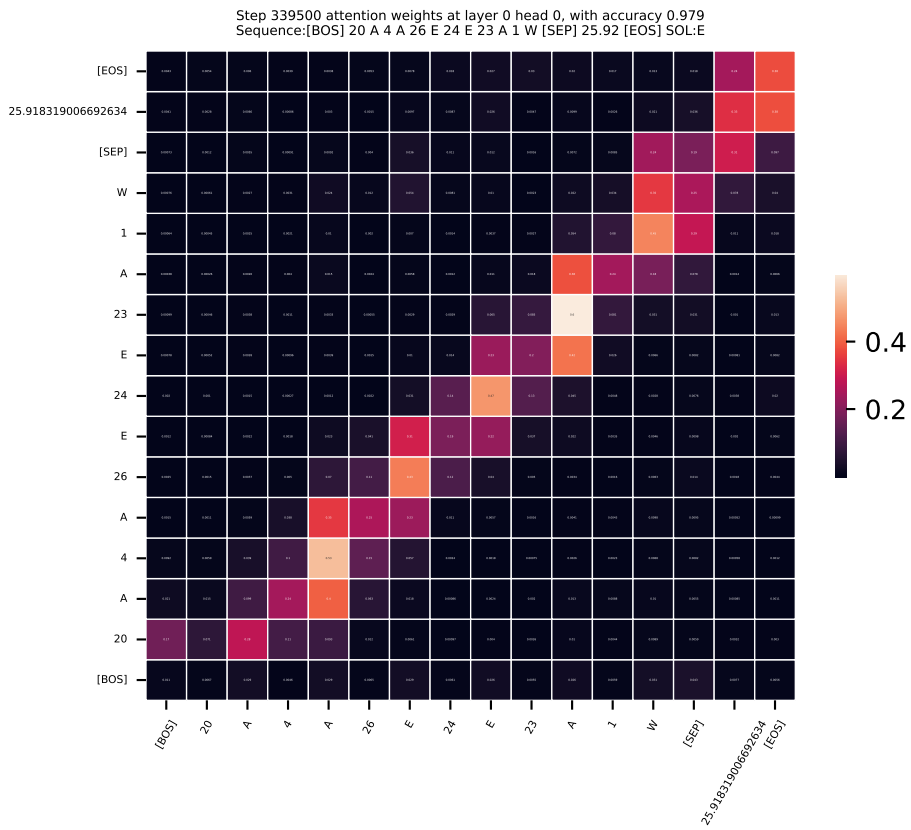

Step 330000 attention weights at layer 0 head 3, with accuracy 0.974
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

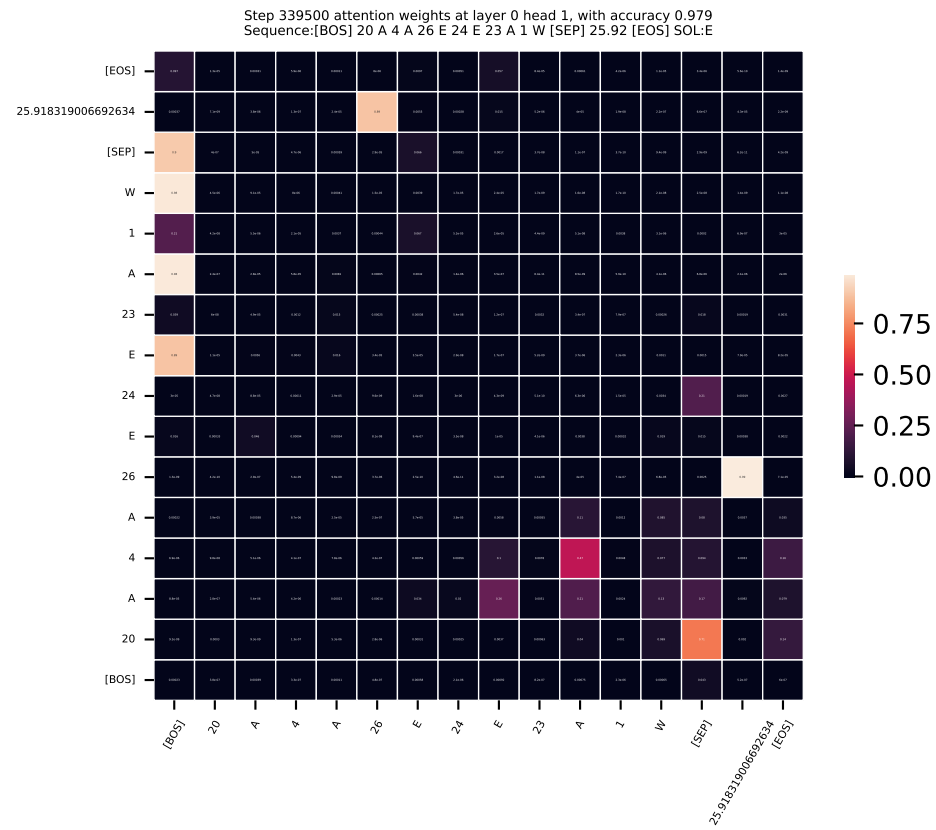

Step 330000 attention weights at layer 1 head 0, with accuracy 0.974
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

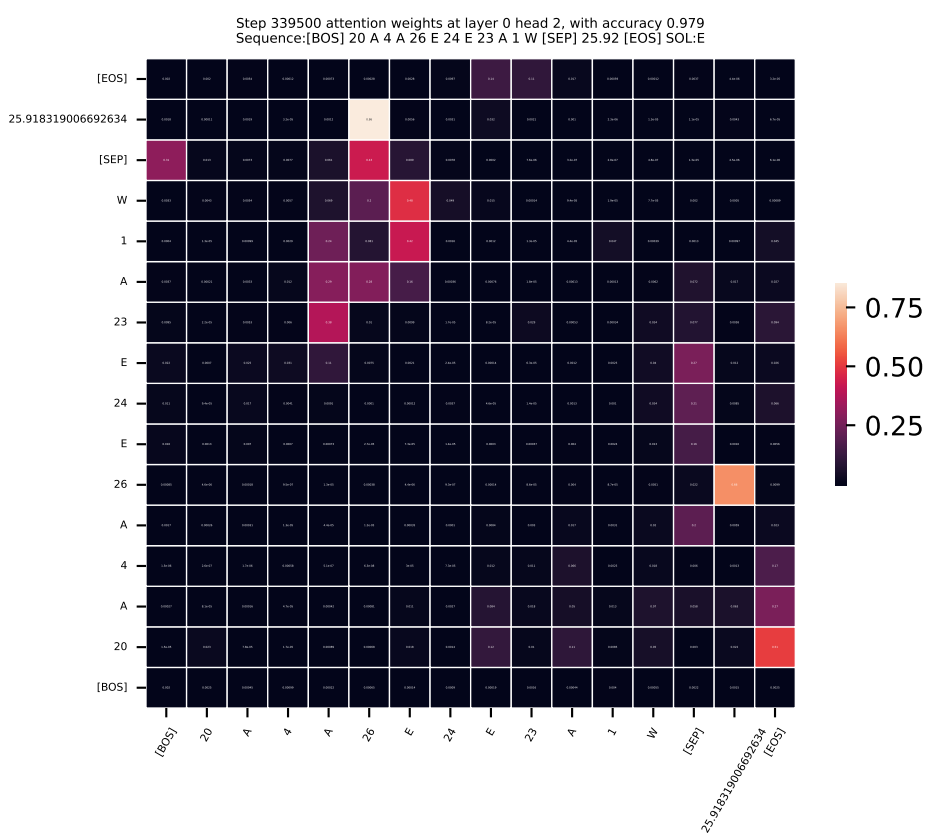

Step 330000 attention weights at layer 1 head 1, with accuracy 0.974
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

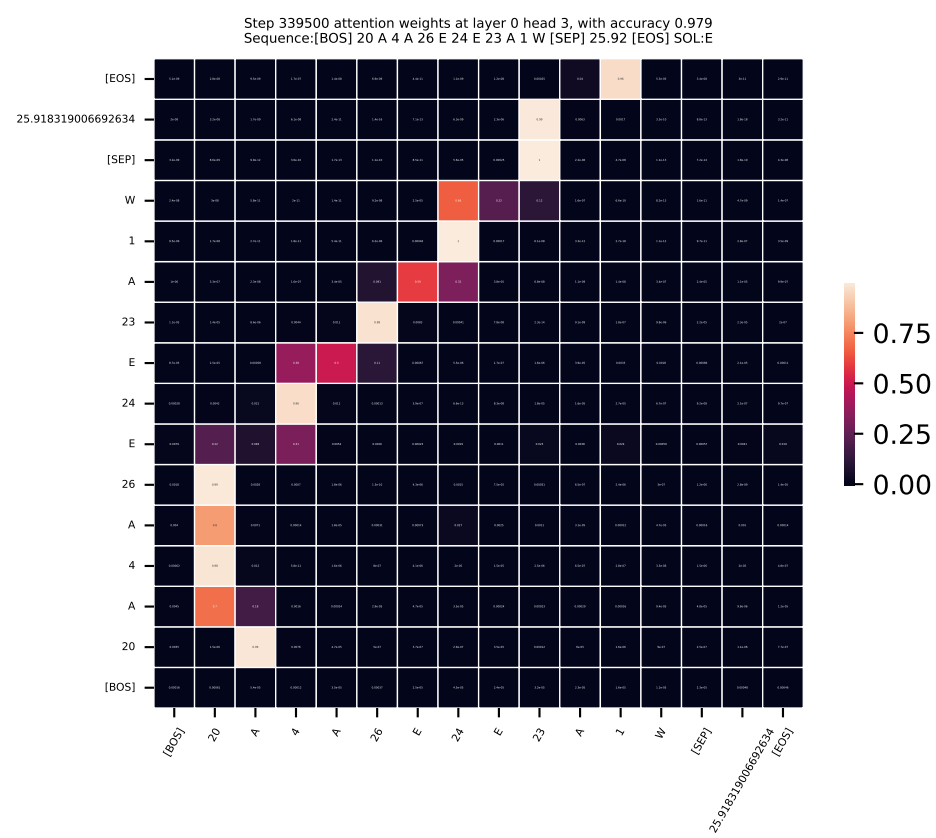

Step 330000 attention weights at layer 1 head 2, with accuracy 0.974
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

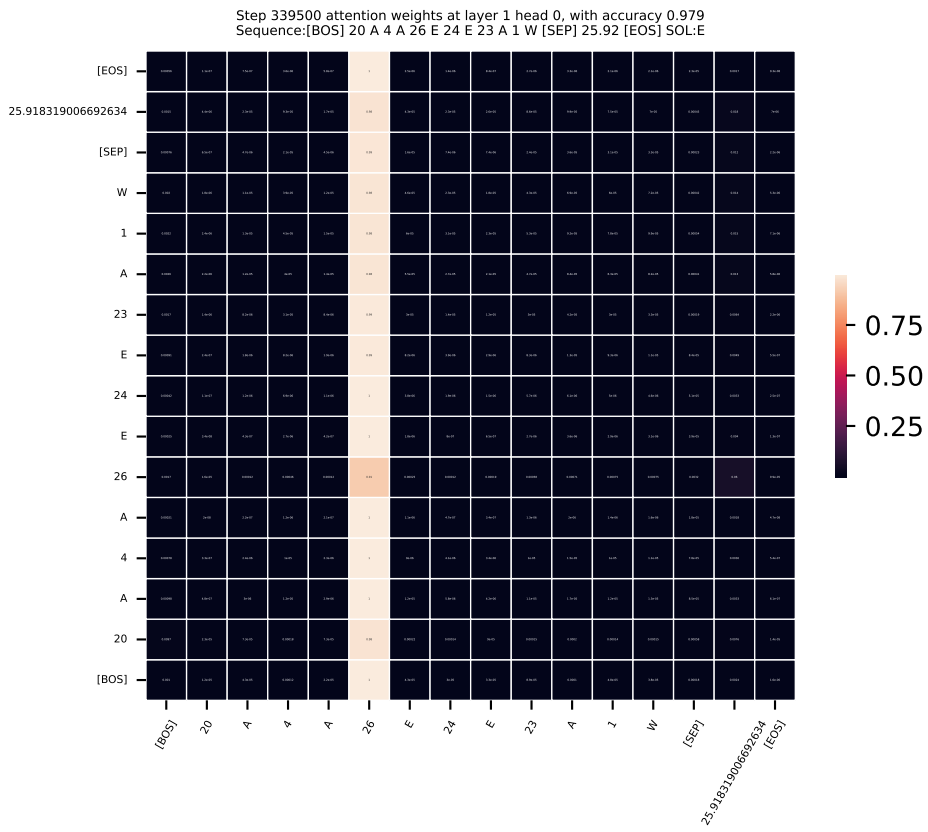

Step 330000 attention weights at layer 1 head 3, with accuracy 0.974
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
1.0
0.5
0.0

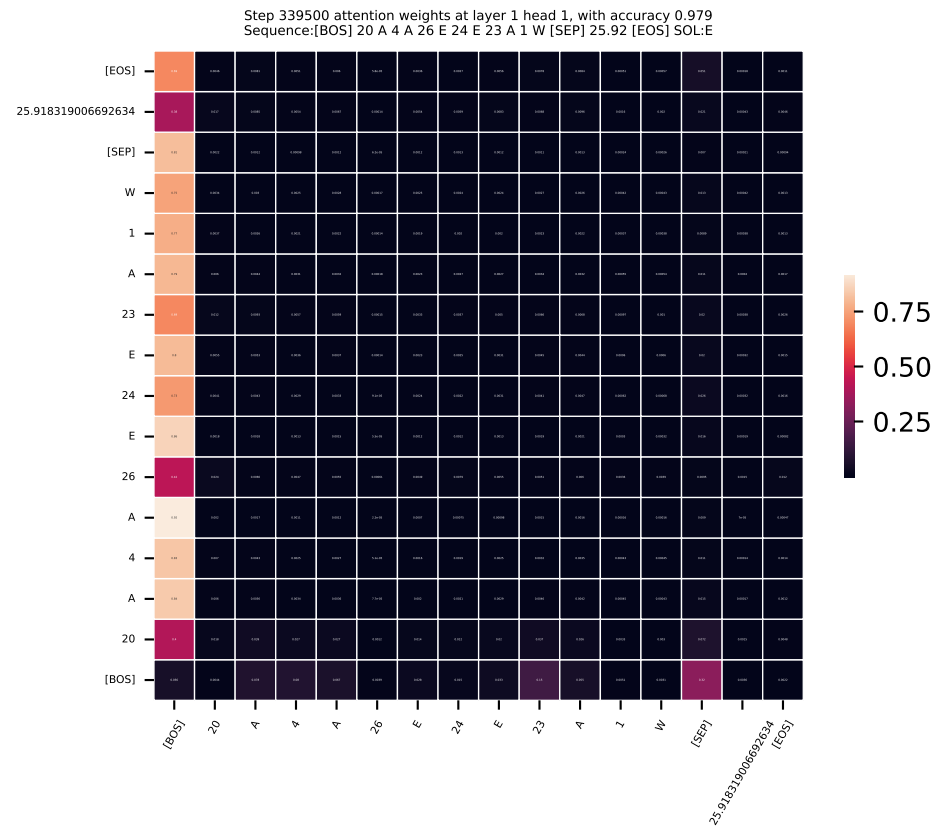

Step 339500 attention weights at layer 0 head 0, with accuracy 0.979
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

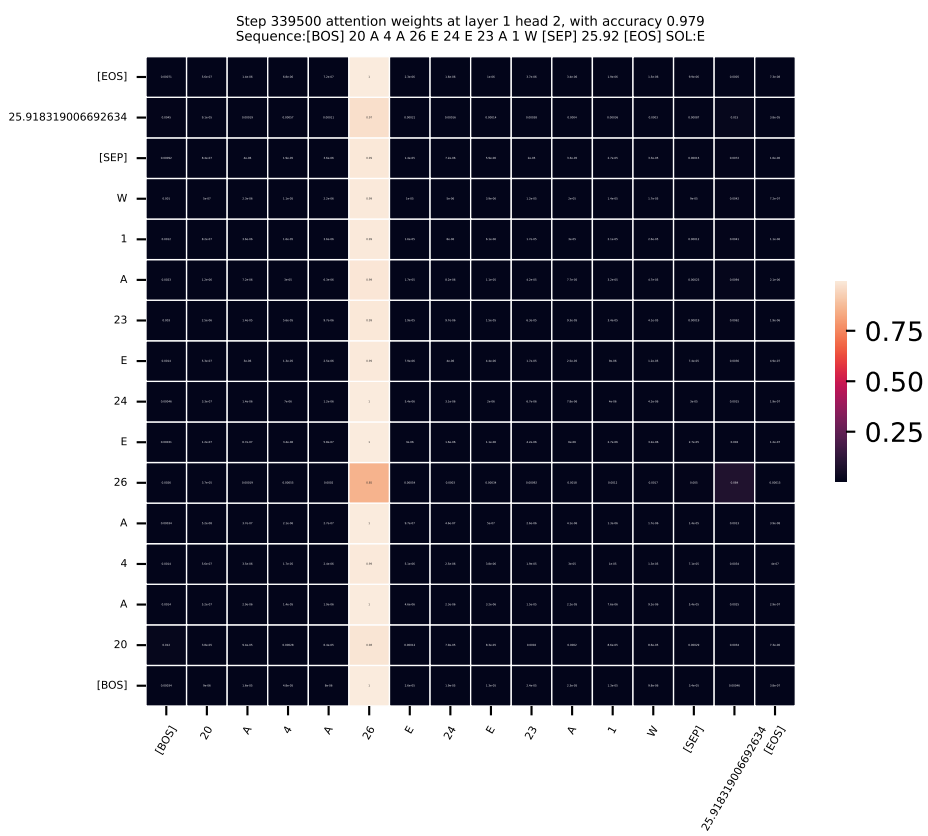

Step 339500 attention weights at layer 0 head 1, with accuracy 0.979
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

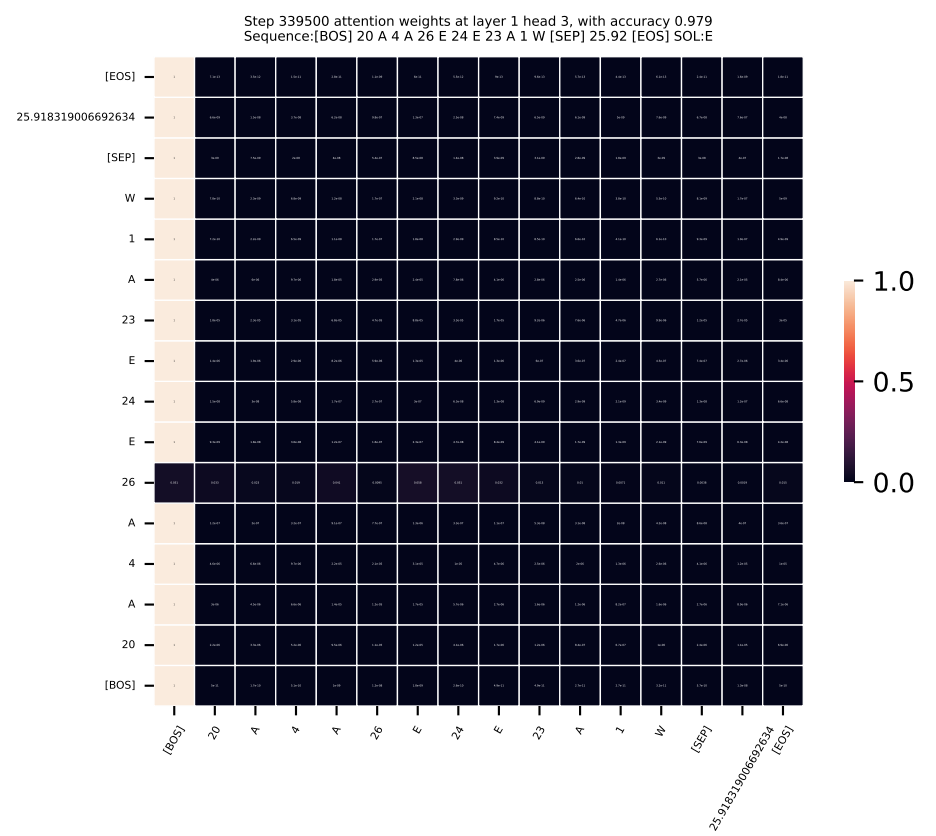

Step 339500 attention weights at layer 0 head 2, with accuracy 0.979
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

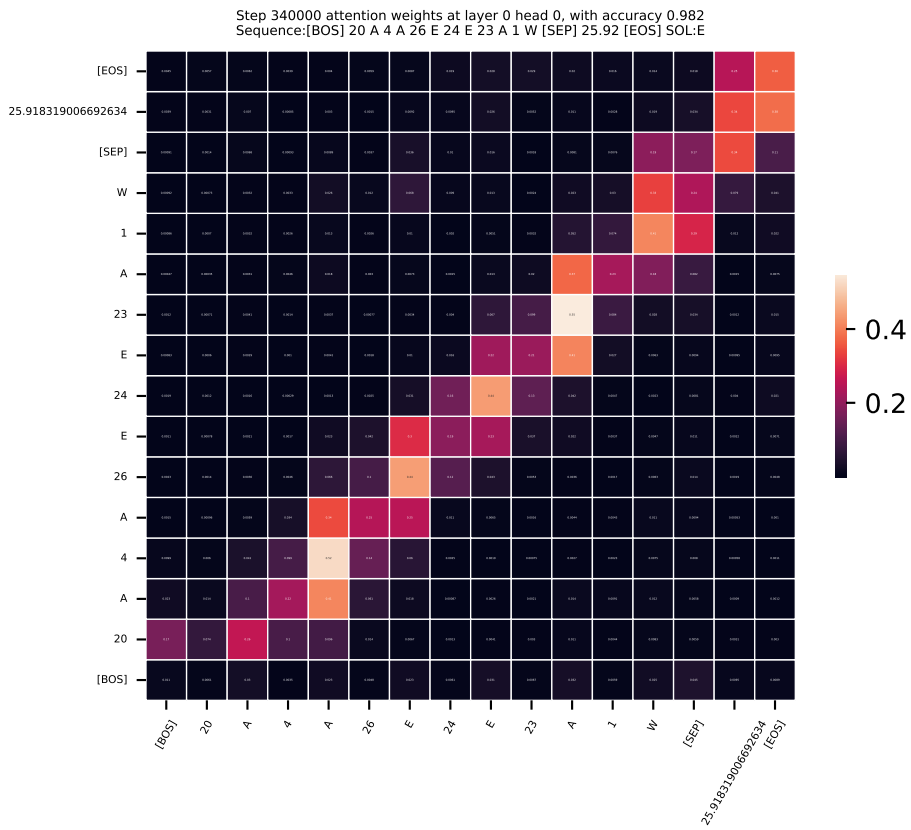

Step 339500 attention weights at layer 0 head 3, with accuracy 0.979
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

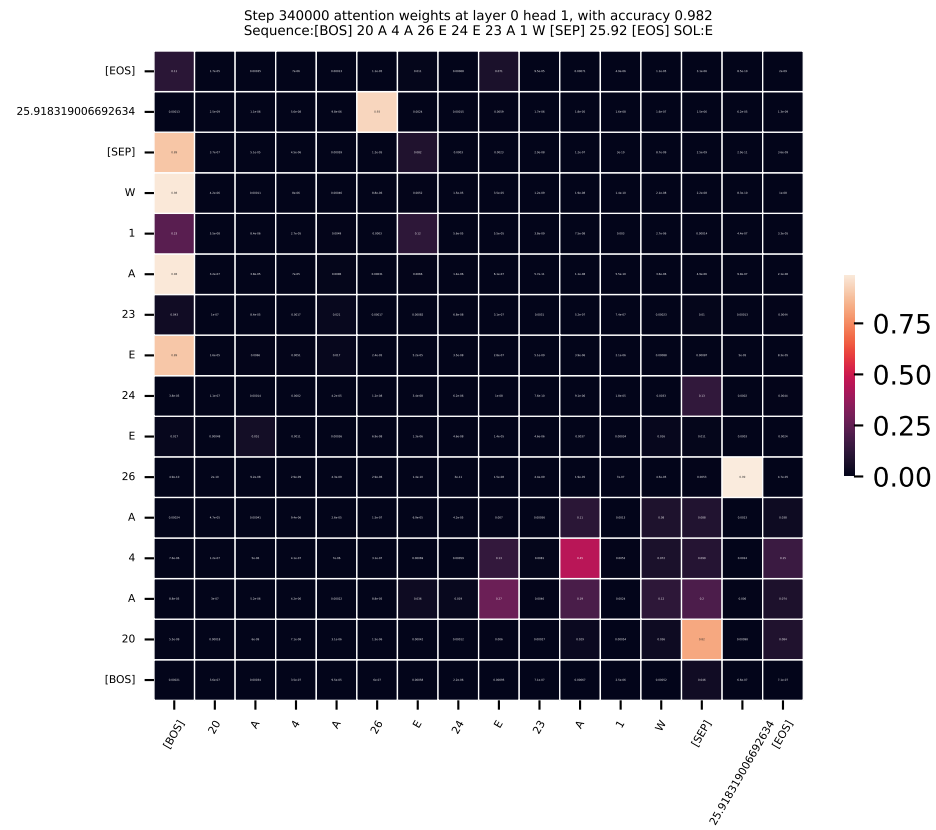

Step 339500 attention weights at layer 1 head 0, with accuracy 0.979
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

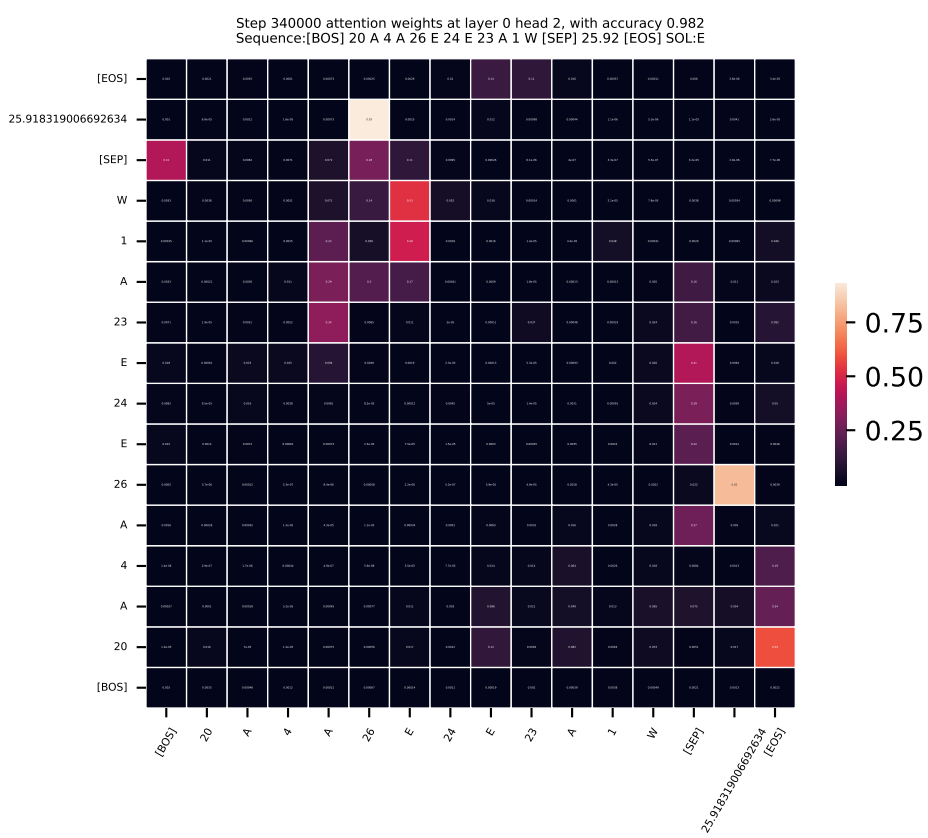

Step 339500 attention weights at layer 1 head 1, with accuracy 0.979
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

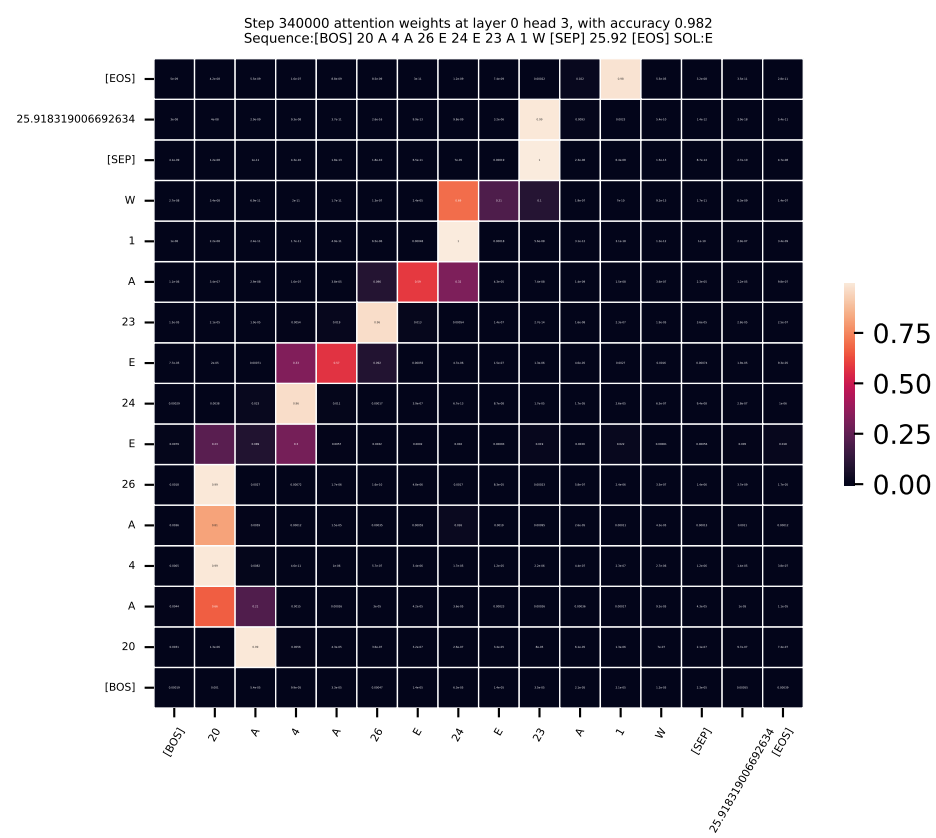

Step 339500 attention weights at layer 1 head 2, with accuracy 0.979
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

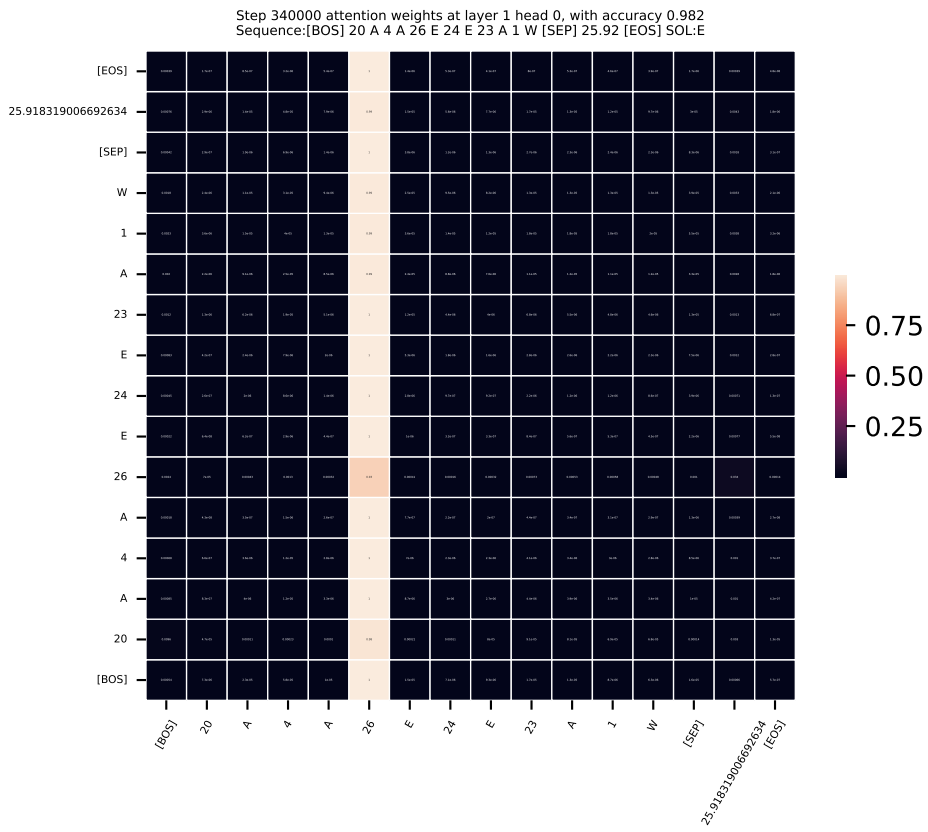

Step 339500 attention weights at layer 1 head 3, with accuracy 0.979
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
1.0
0.5
0.0

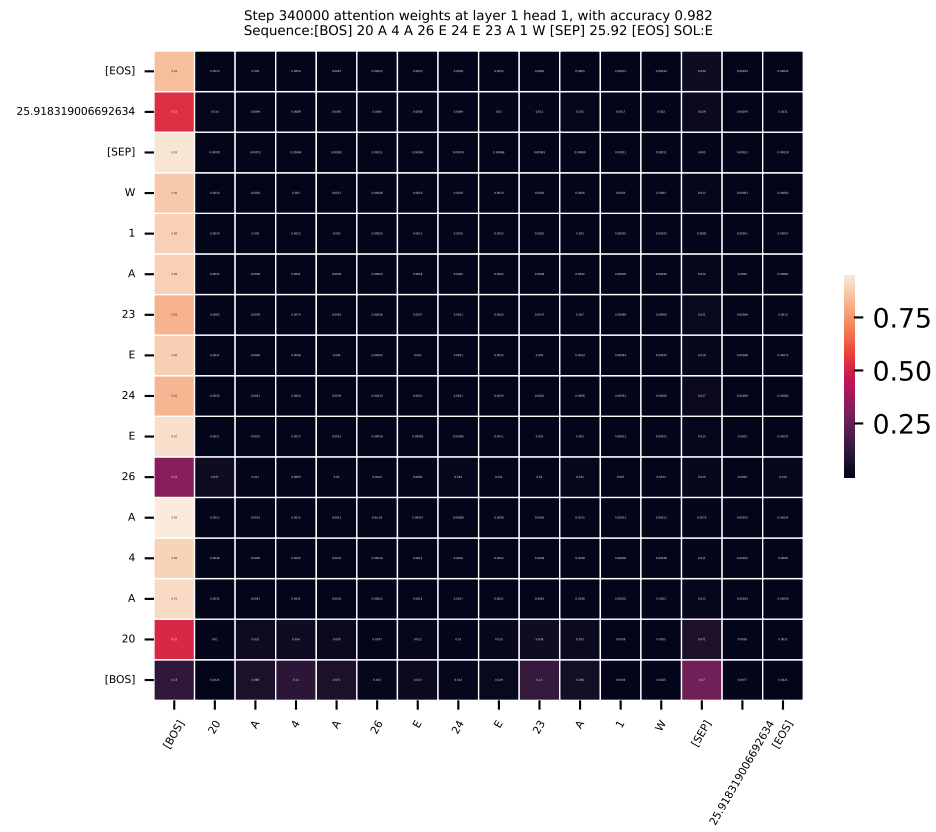

Step 340000 attention weights at layer 0 head 0, with accuracy 0.982
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

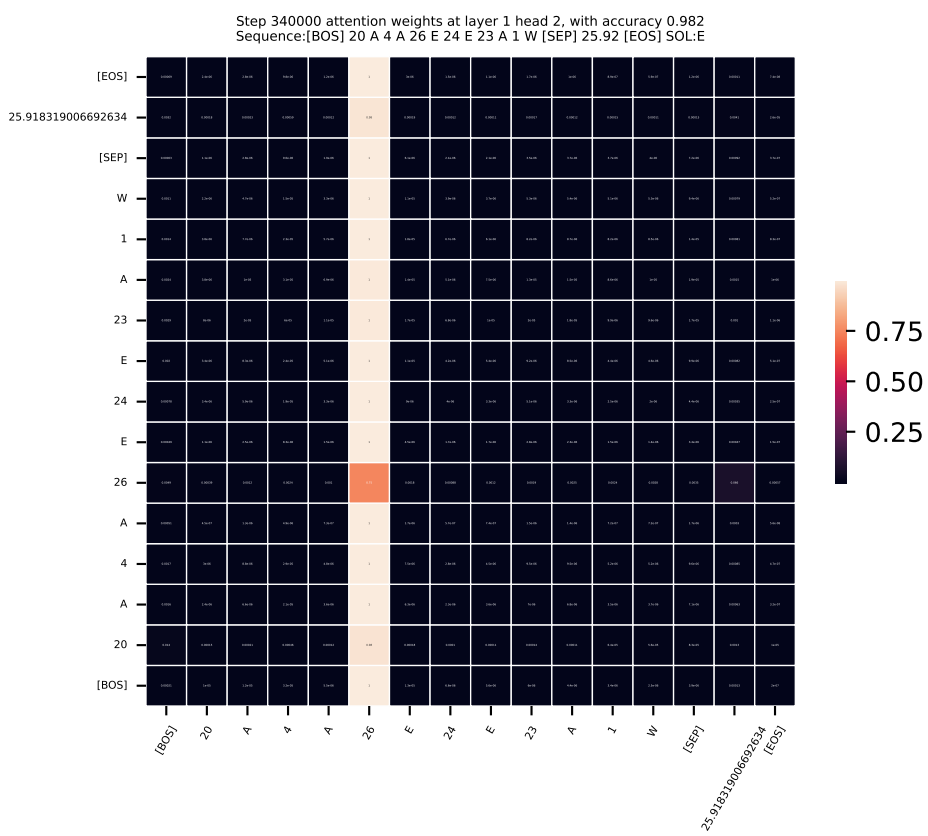

Step 340000 attention weights at layer 0 head 1, with accuracy 0.982
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

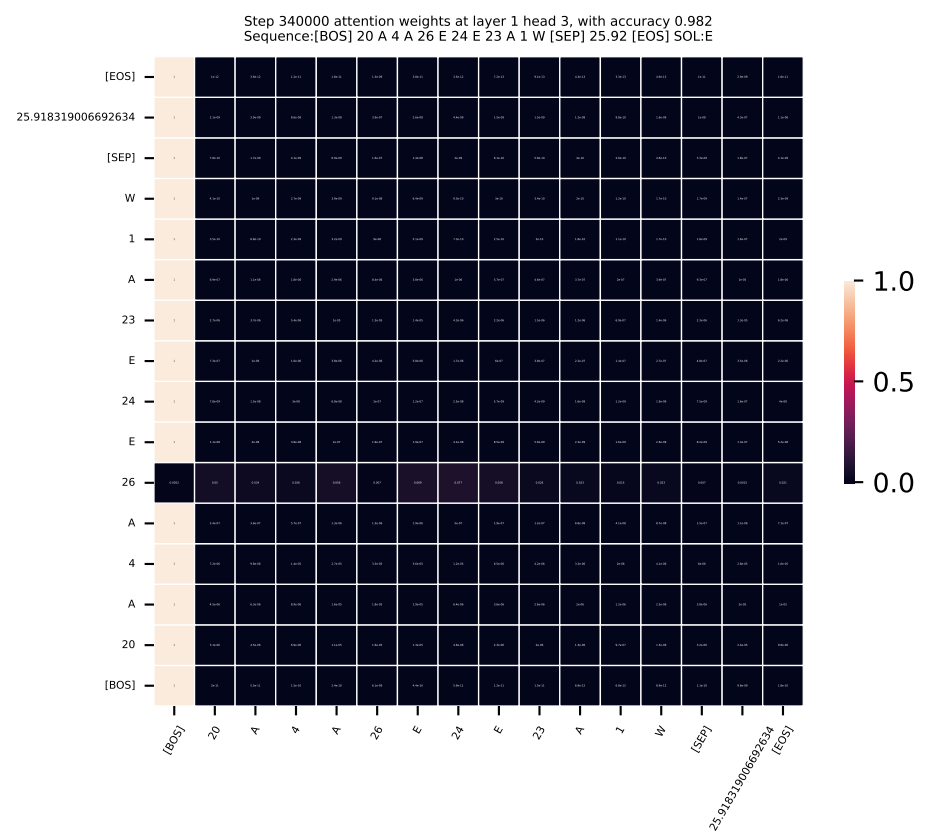

Step 340000 attention weights at layer 0 head 2, with accuracy 0.982
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

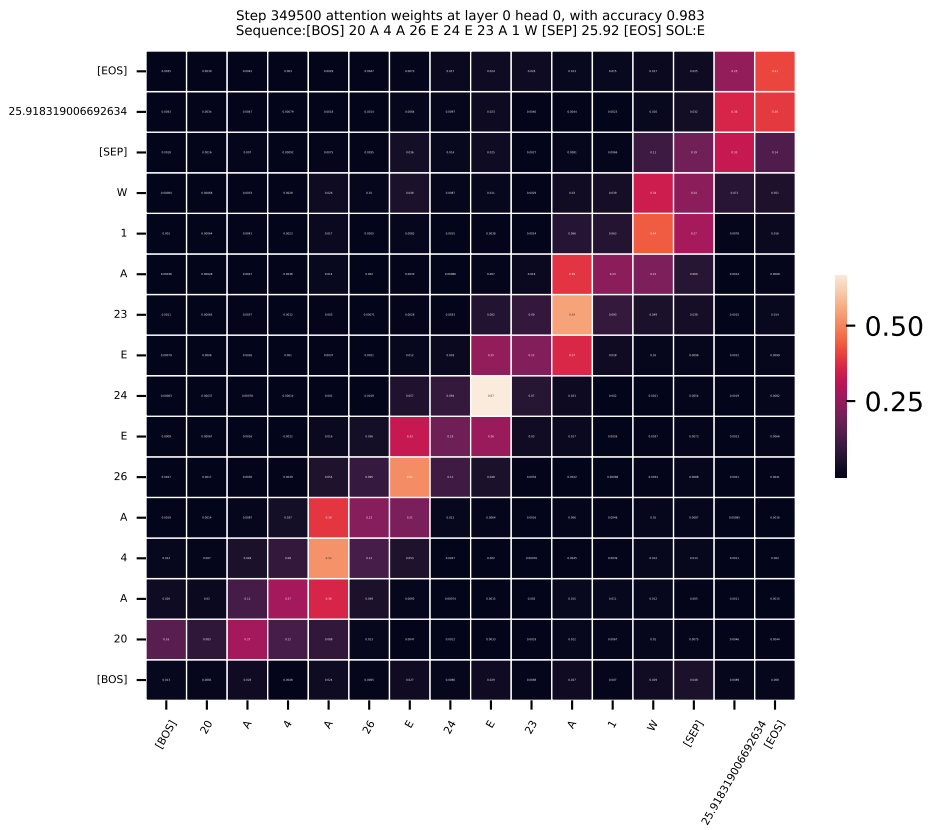

Step 340000 attention weights at layer 0 head 3, with accuracy 0.982
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

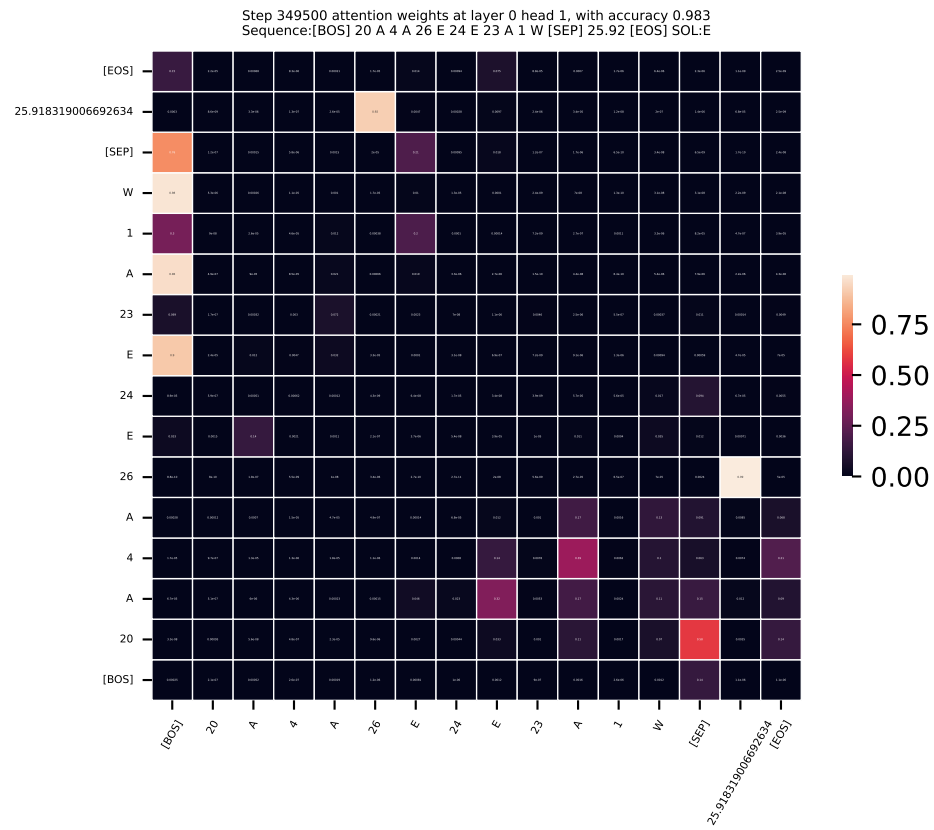

Step 340000 attention weights at layer 1 head 0, with accuracy 0.982
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

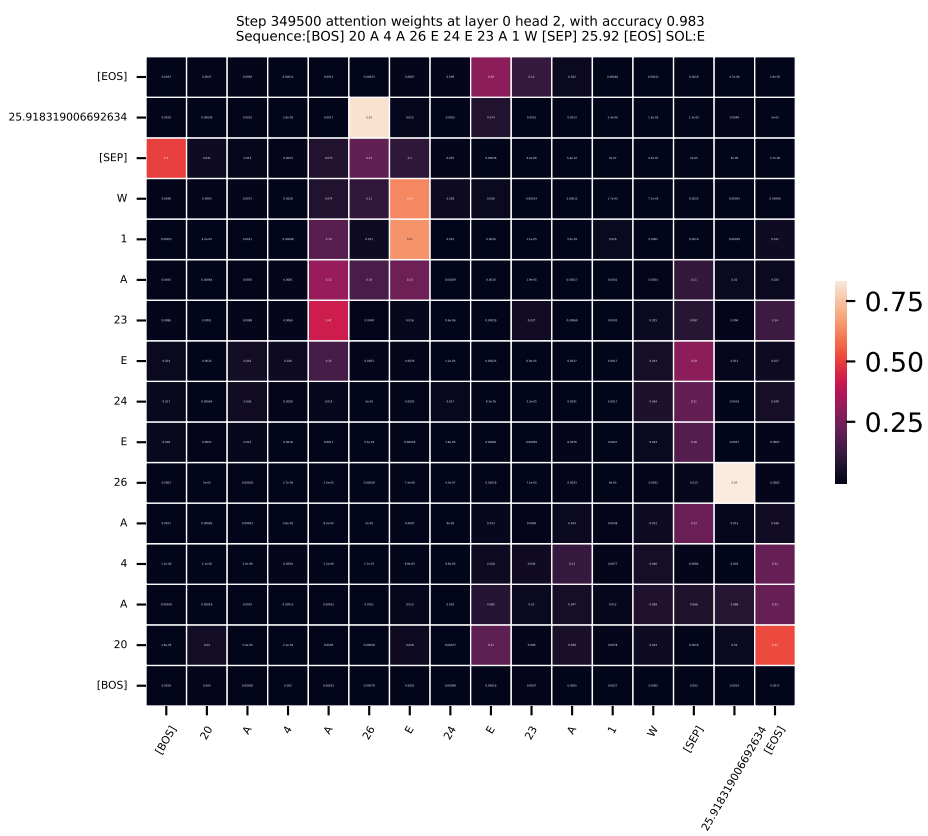

Step 340000 attention weights at layer 1 head 1, with accuracy 0.982
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

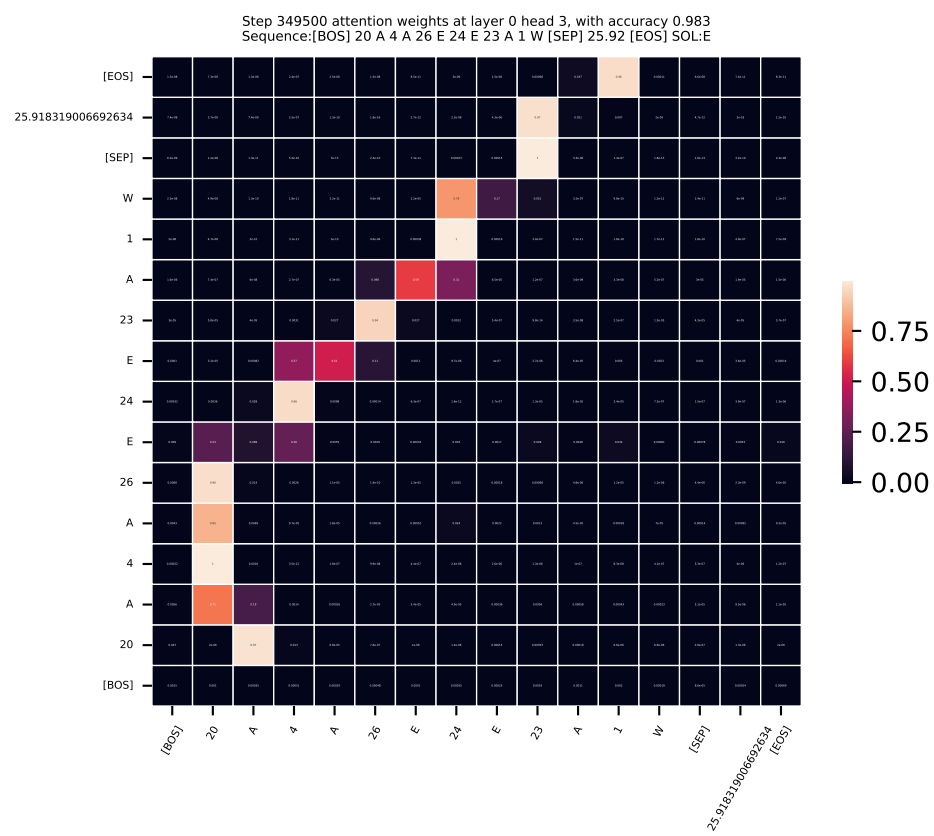

Step 340000 attention weights at layer 1 head 2, with accuracy 0.982
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

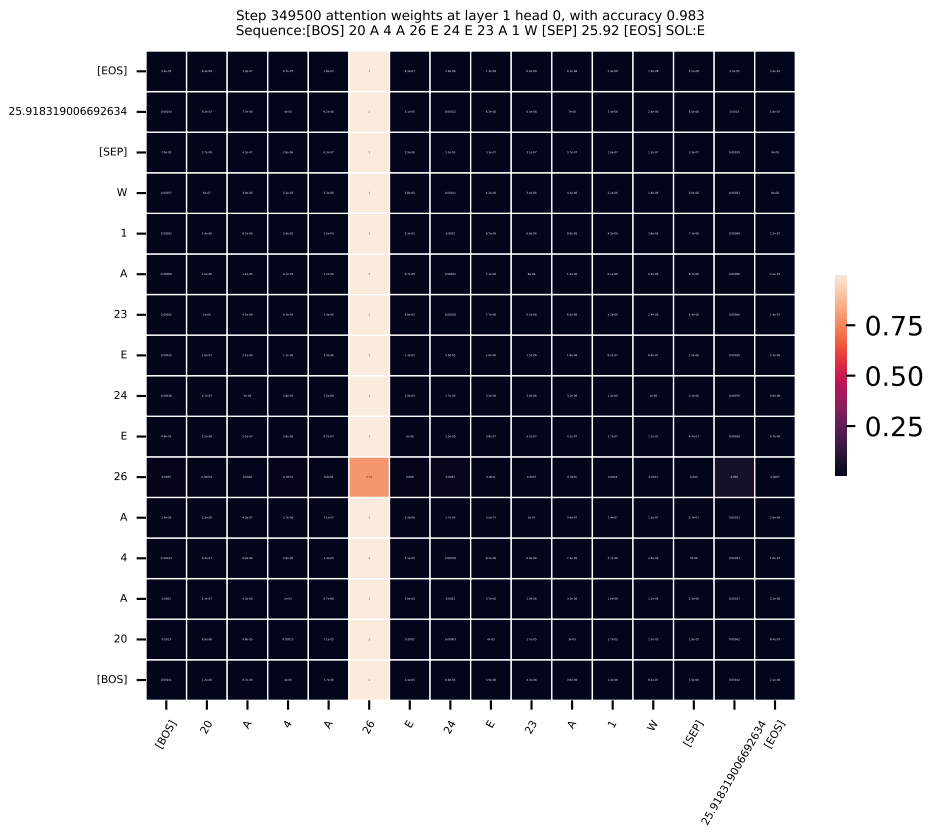

Step 340000 attention weights at layer 1 head 3, with accuracy 0.982
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
1.0
0.5
0.0

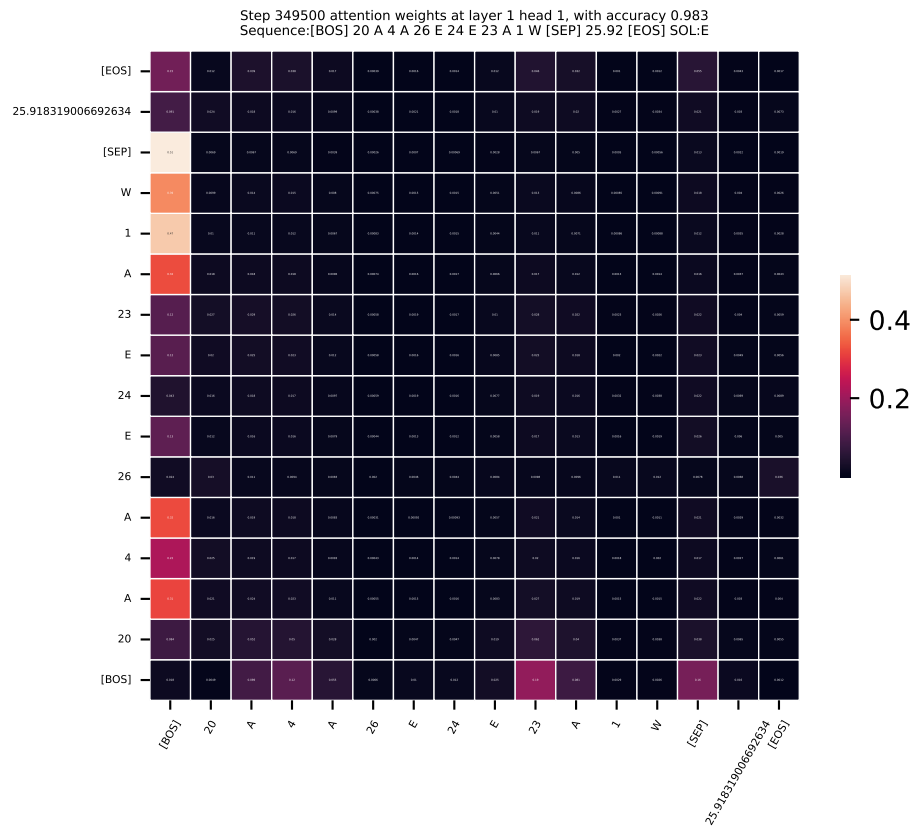

Step 349500 attention weights at layer 0 head 0, with accuracy 0.983
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

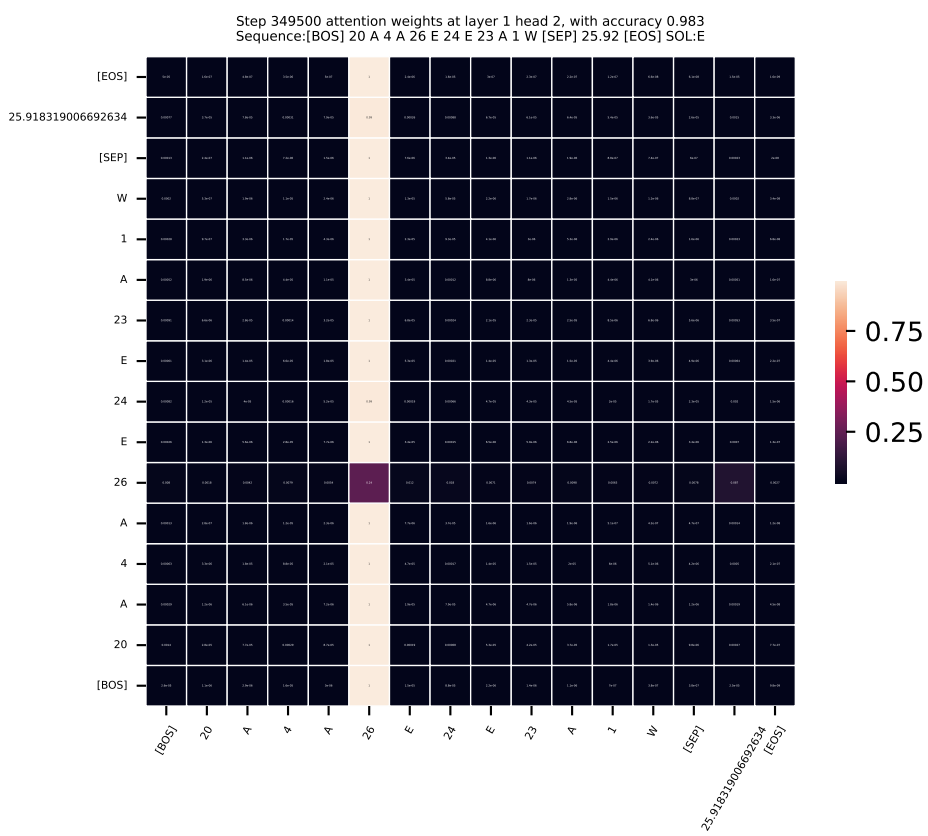

Step 349500 attention weights at layer 0 head 1, with accuracy 0.983
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

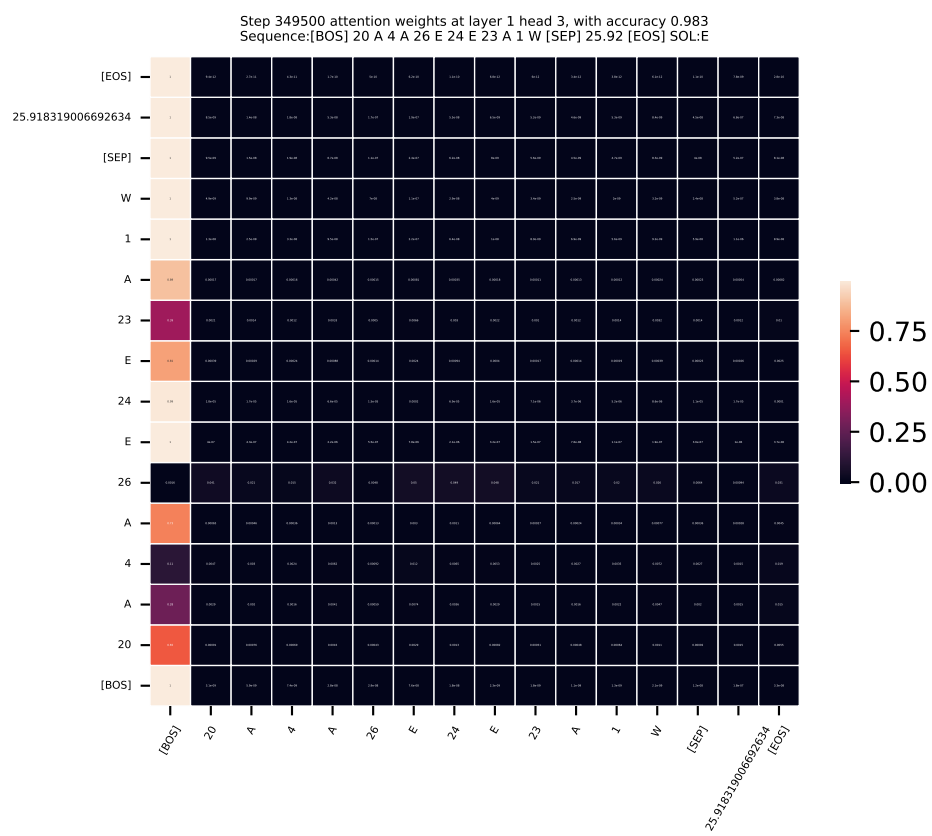

Step 349500 attention weights at layer 0 head 2, with accuracy 0.983
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

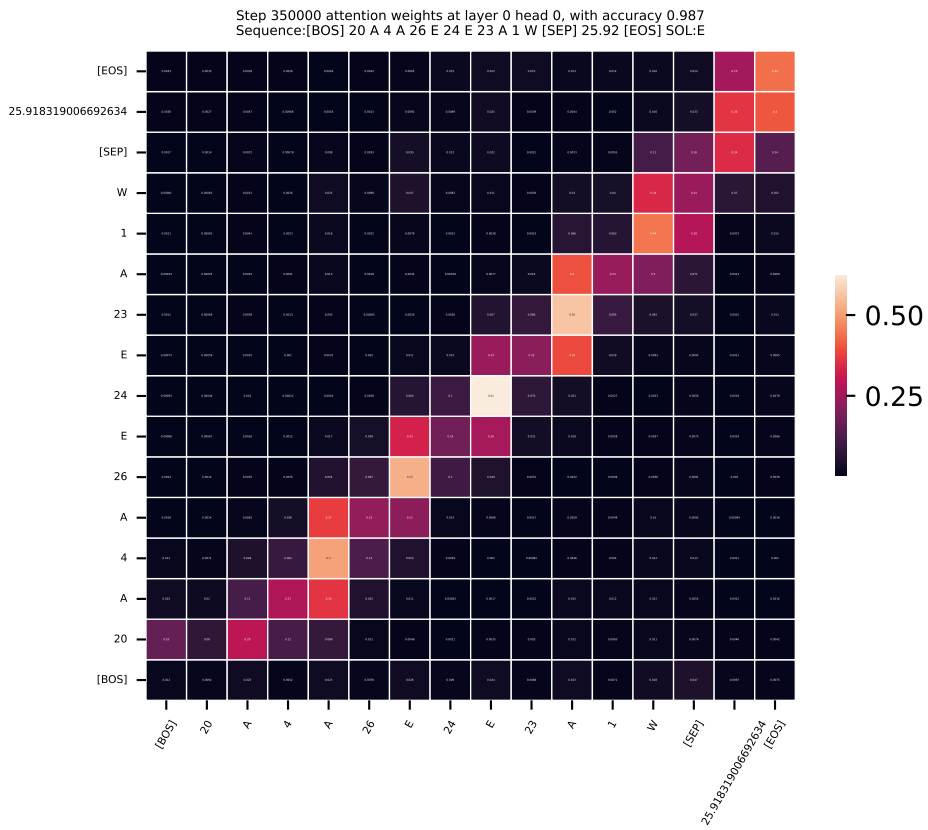

Step 349500 attention weights at layer 0 head 3, with accuracy 0.983
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

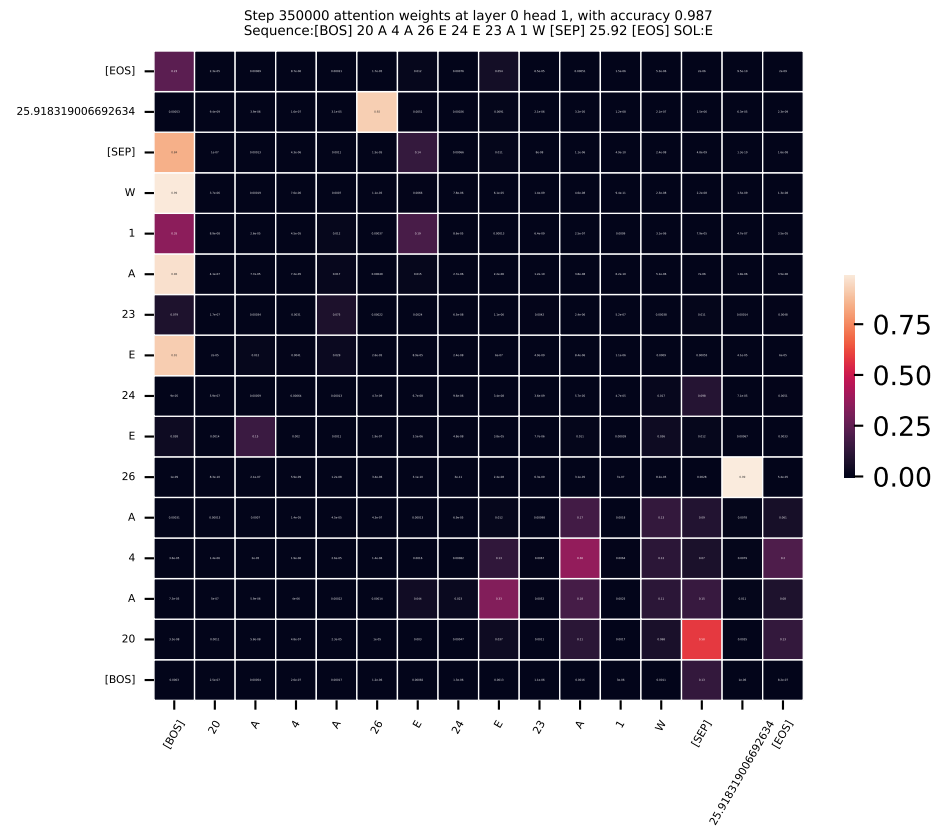

Step 349500 attention weights at layer 1 head 0, with accuracy 0.983
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

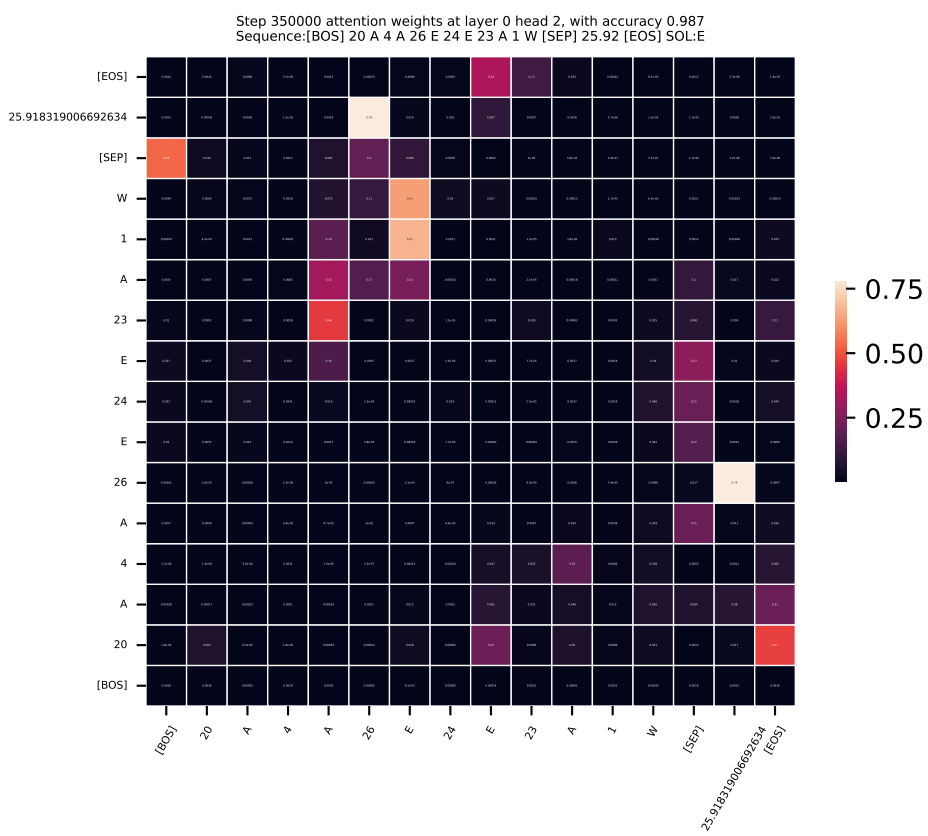

Step 349500 attention weights at layer 1 head 1, with accuracy 0.983
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.4
0.2

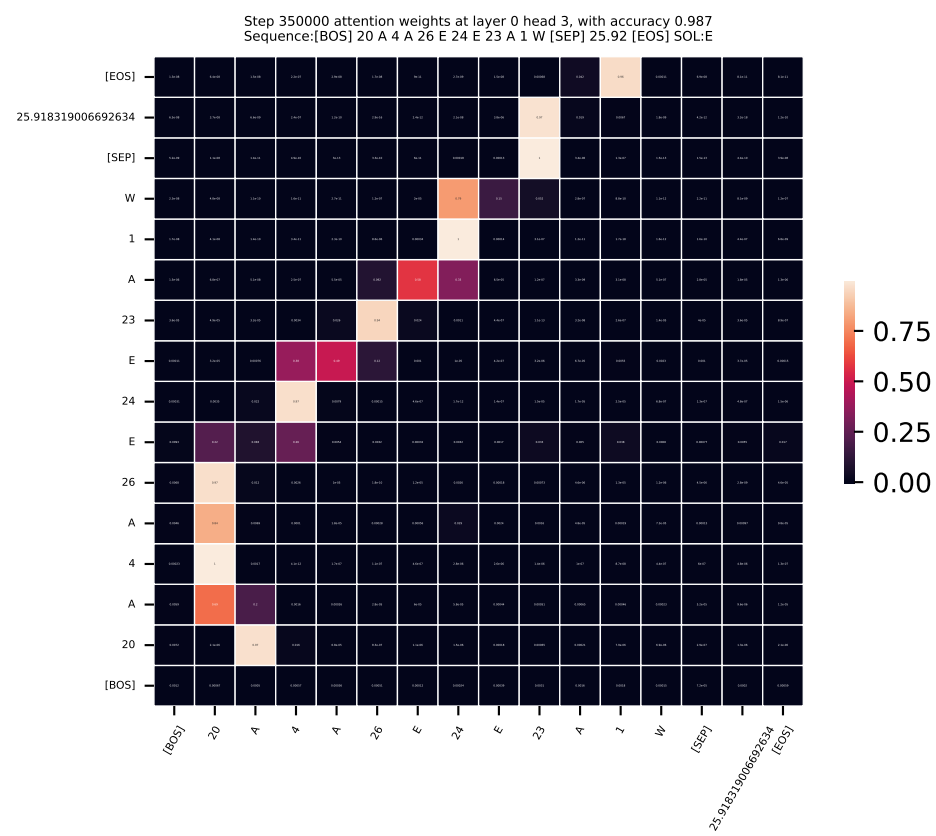

Step 349500 attention weights at layer 1 head 2, with accuracy 0.983
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

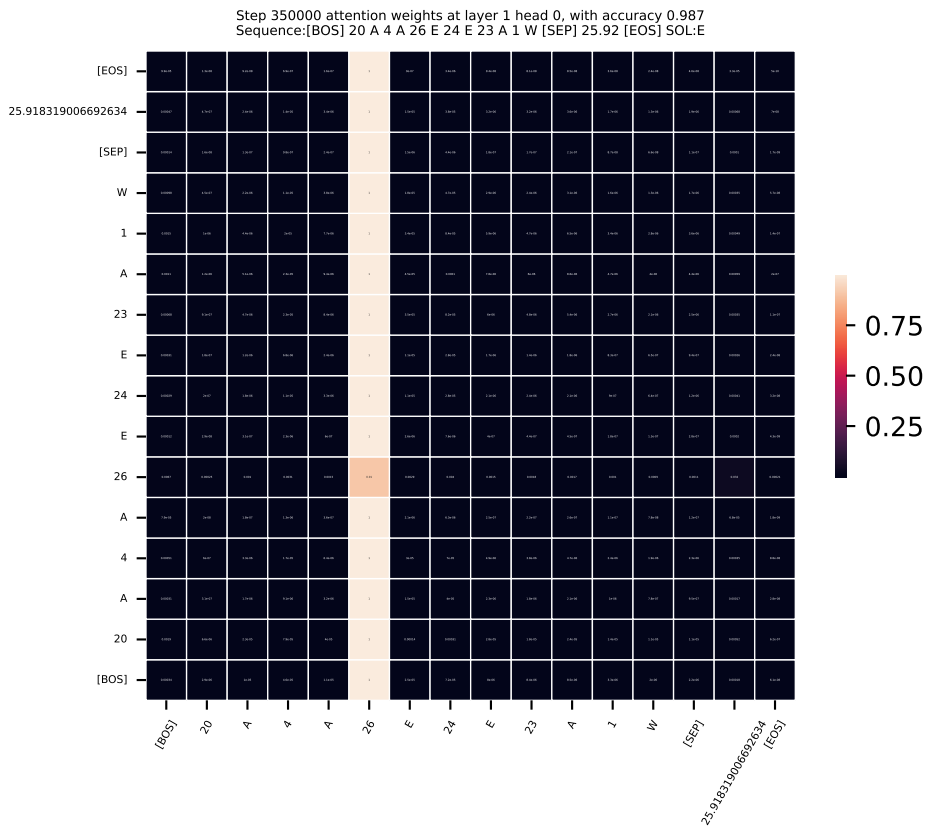

Step 349500 attention weights at layer 1 head 3, with accuracy 0.983
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25
0.00

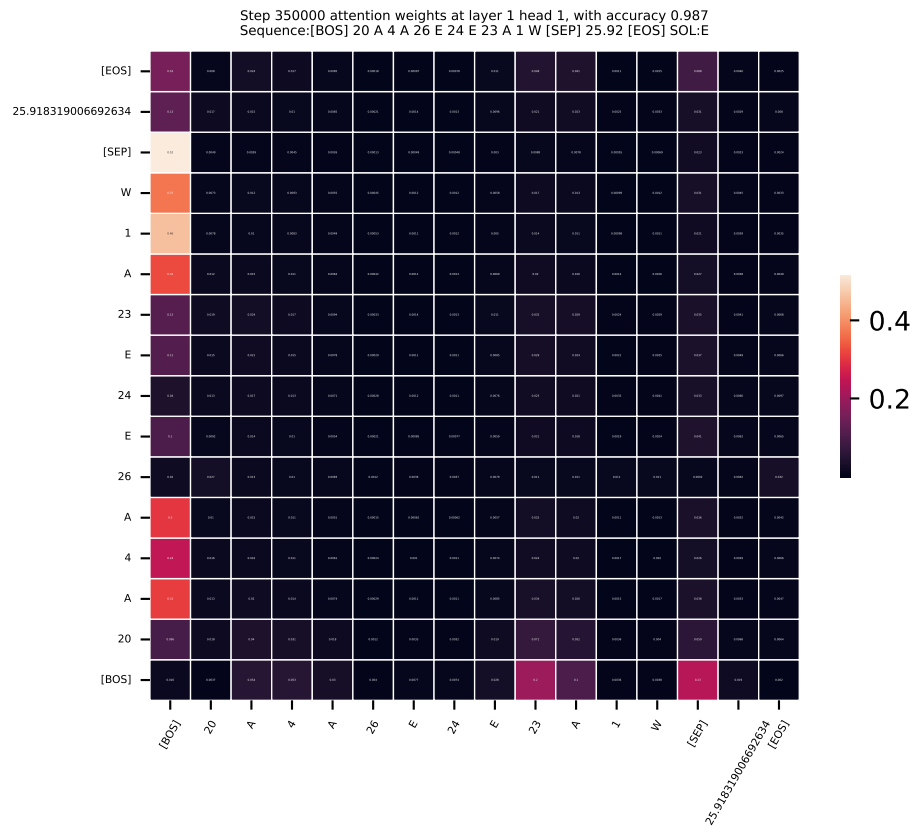

Step 350000 attention weights at layer 0 head 0, with accuracy 0.987
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

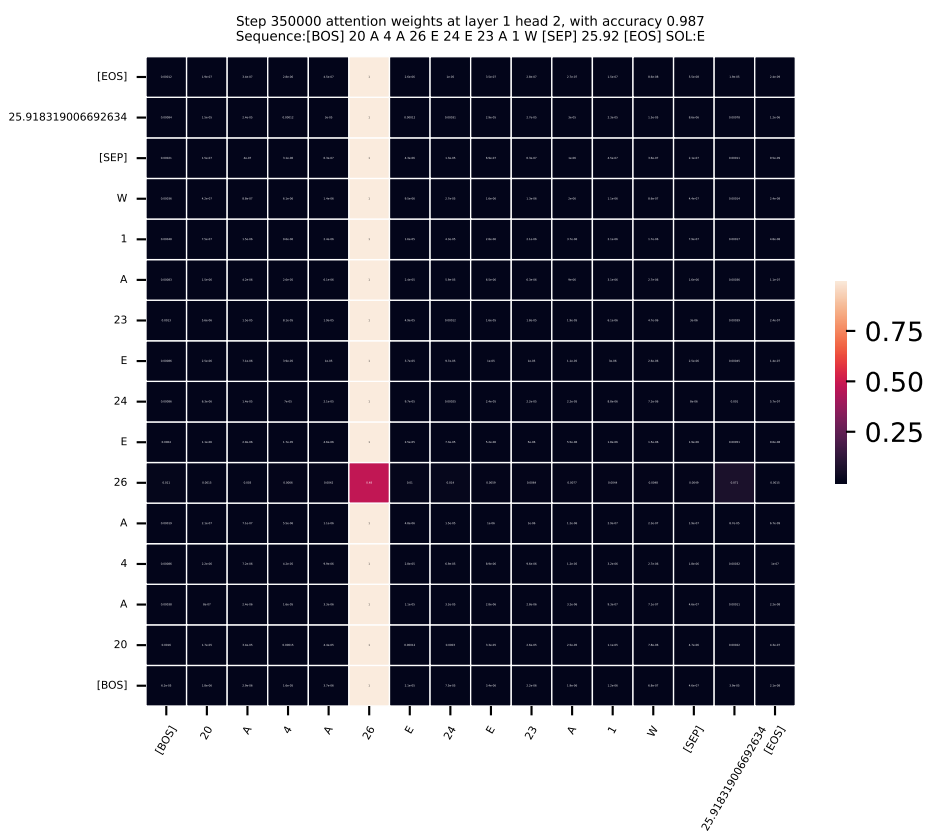

Step 350000 attention weights at layer 0 head 1, with accuracy 0.987
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

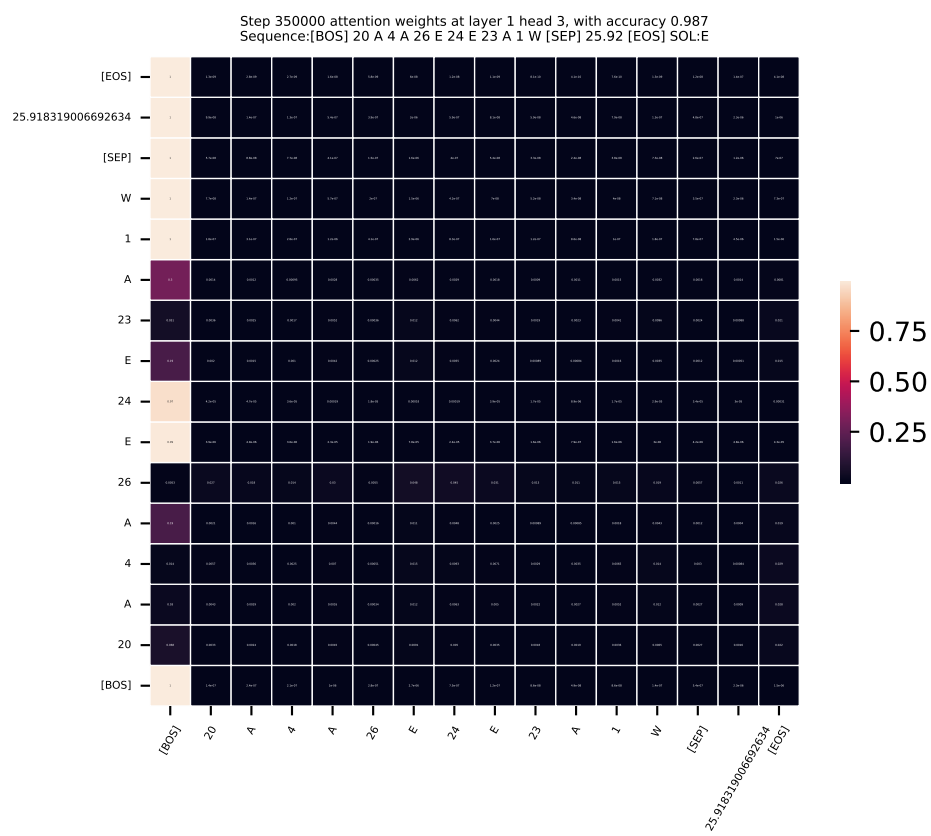

Step 350000 attention weights at layer 0 head 2, with accuracy 0.987
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

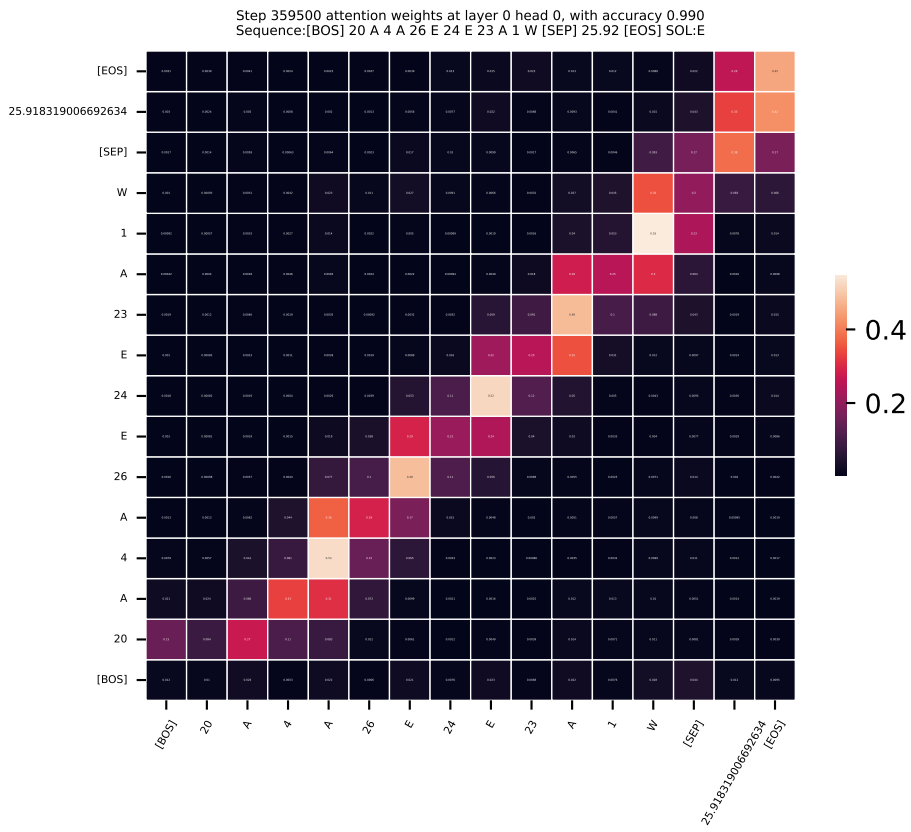

Step 350000 attention weights at layer 0 head 3, with accuracy 0.987
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

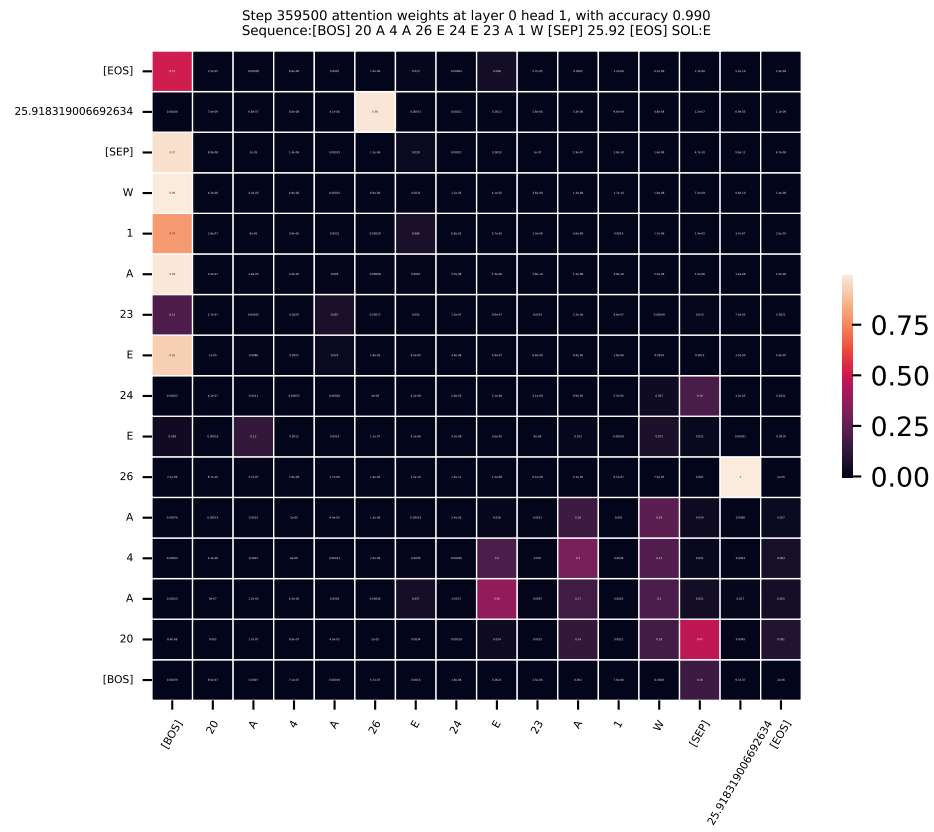

Step 350000 attention weights at layer 1 head 0, with accuracy 0.987
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

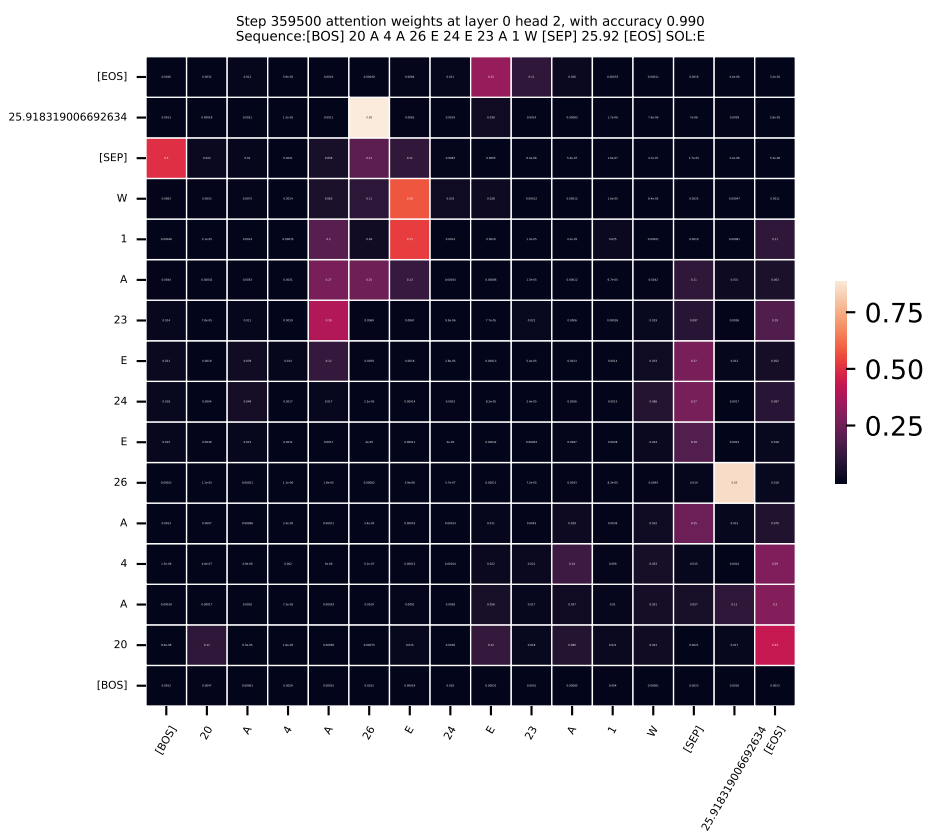

Step 350000 attention weights at layer 1 head 1, with accuracy 0.987
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

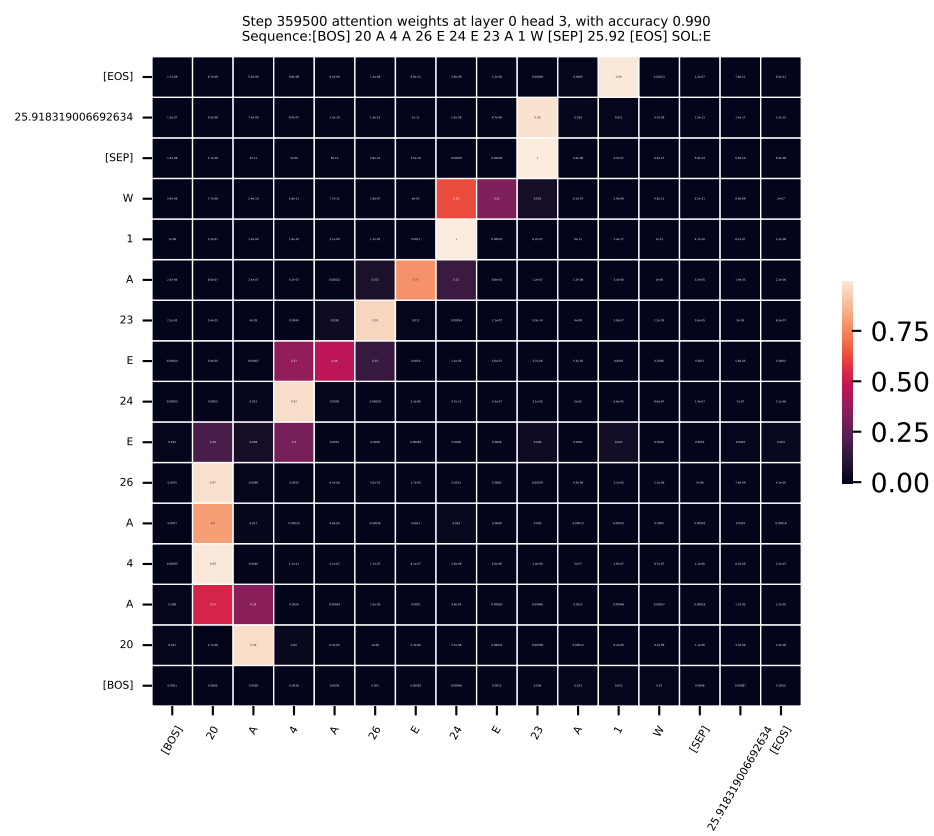

Step 350000 attention weights at layer 1 head 2, with accuracy 0.987
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

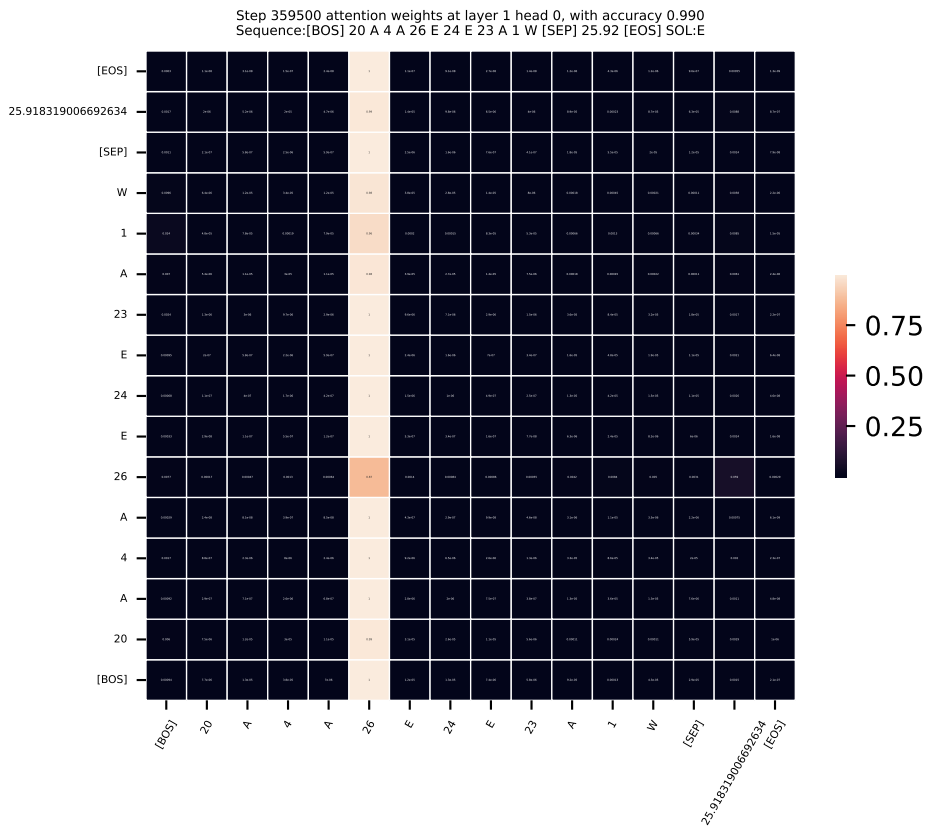

Step 350000 attention weights at layer 1 head 3, with accuracy 0.987
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

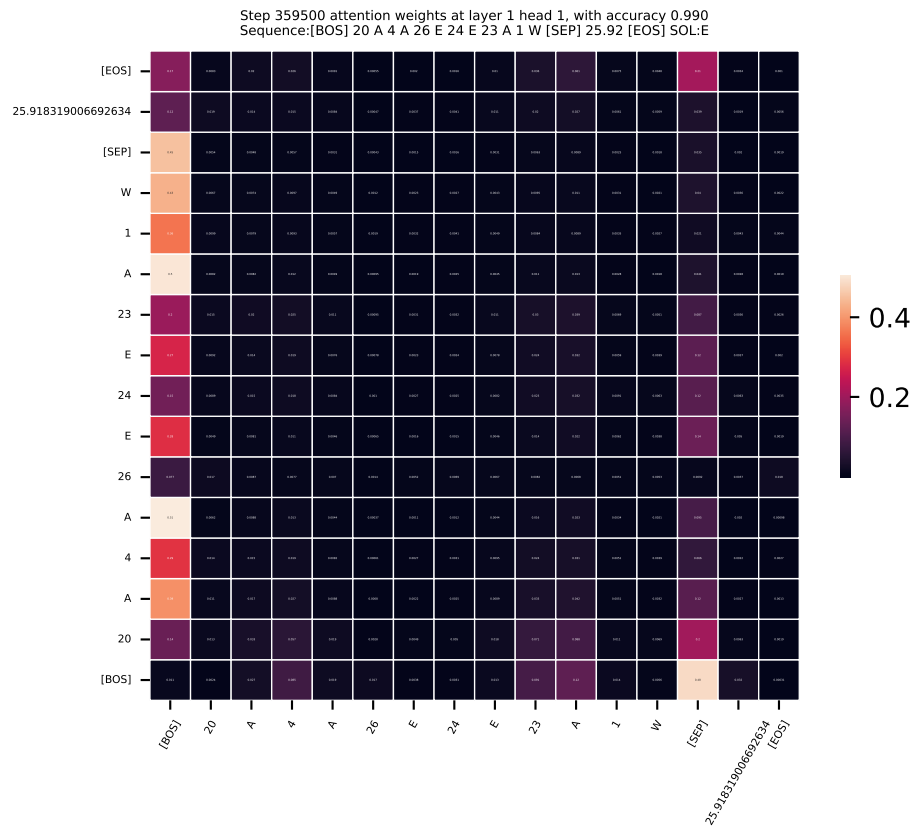

Step 359500 attention weights at layer 0 head 0, with accuracy 0.990
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

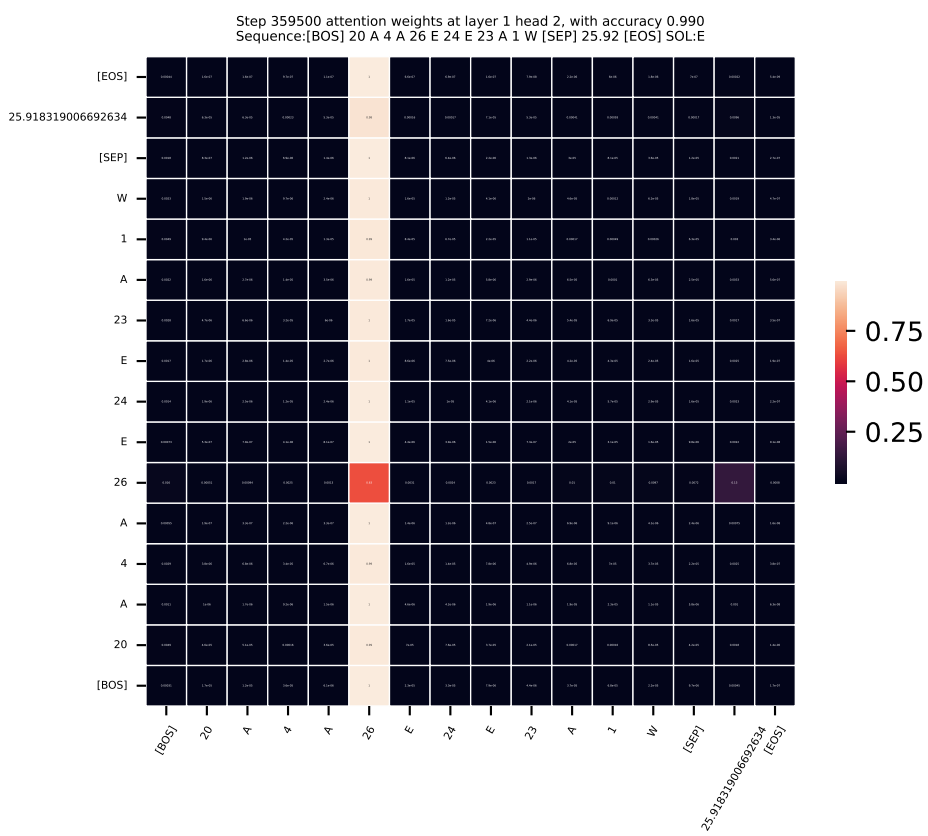

Step 359500 attention weights at layer 0 head 1, with accuracy 0.990
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

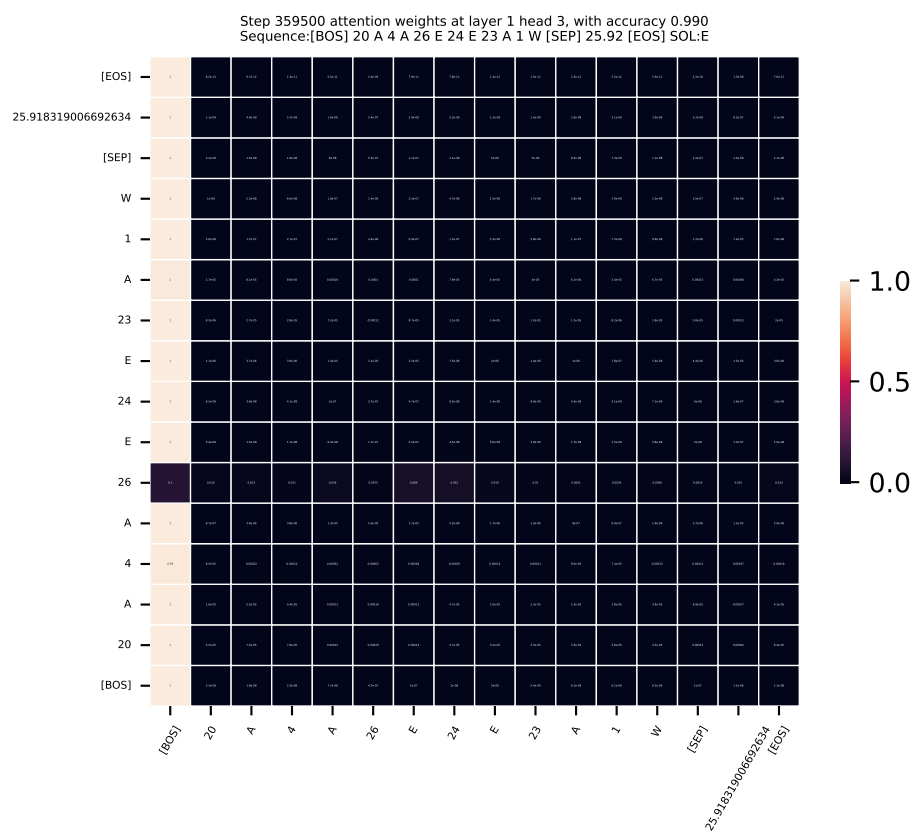

Step 359500 attention weights at layer 0 head 2, with accuracy 0.990
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

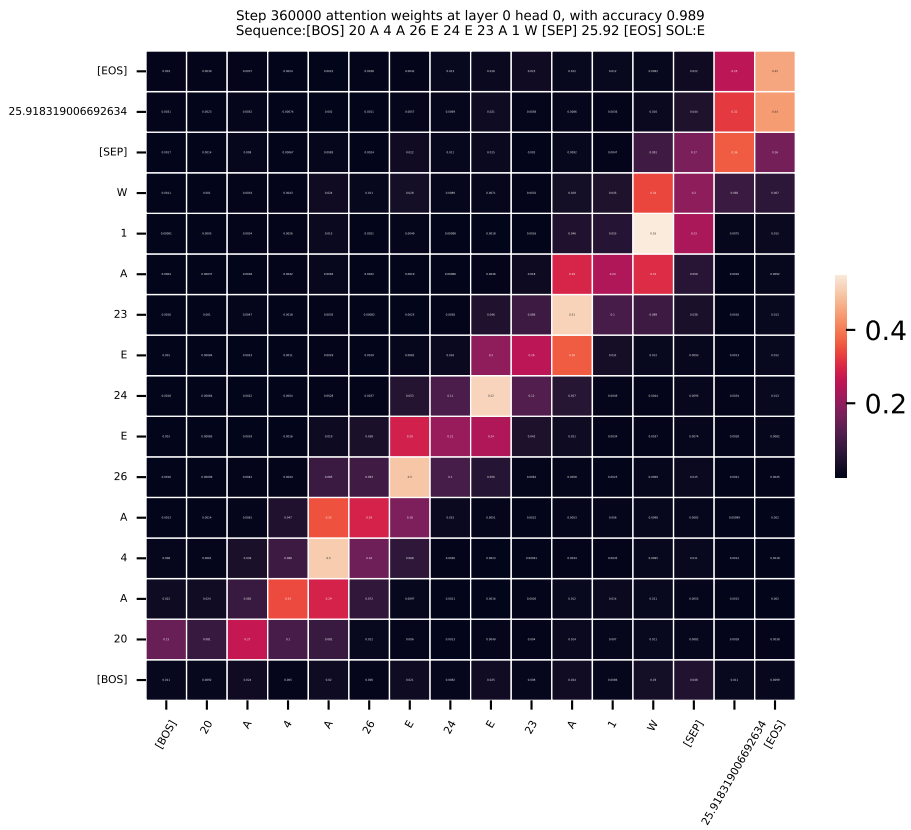

Step 359500 attention weights at layer 0 head 3, with accuracy 0.990
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

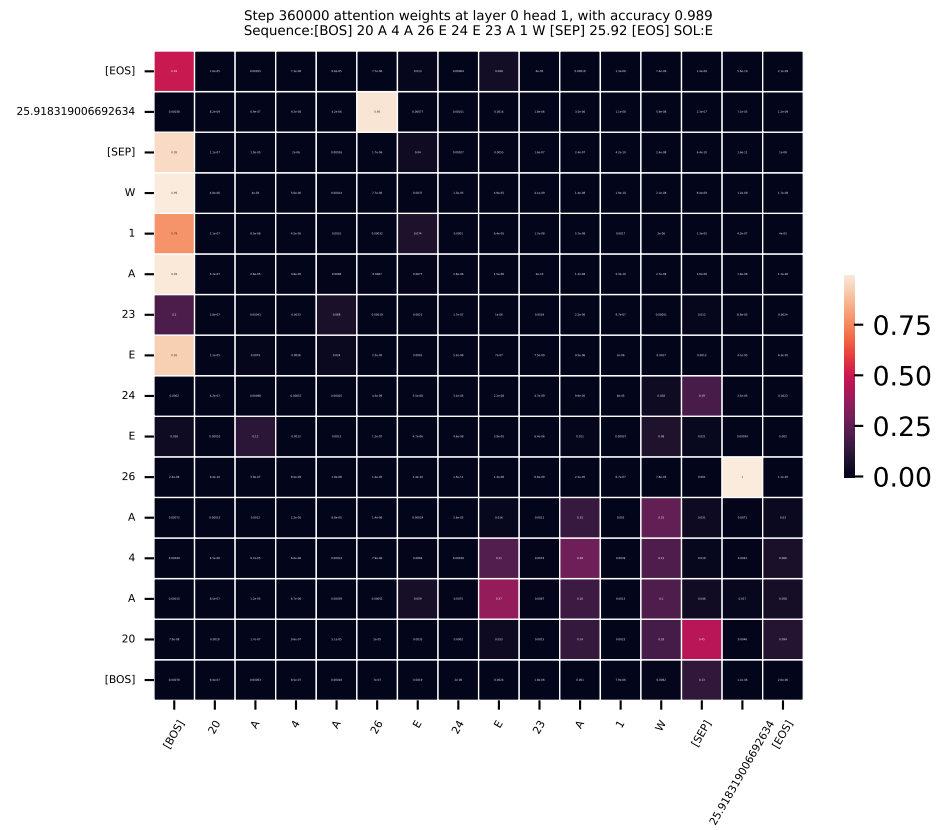

Step 359500 attention weights at layer 1 head 0, with accuracy 0.990
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

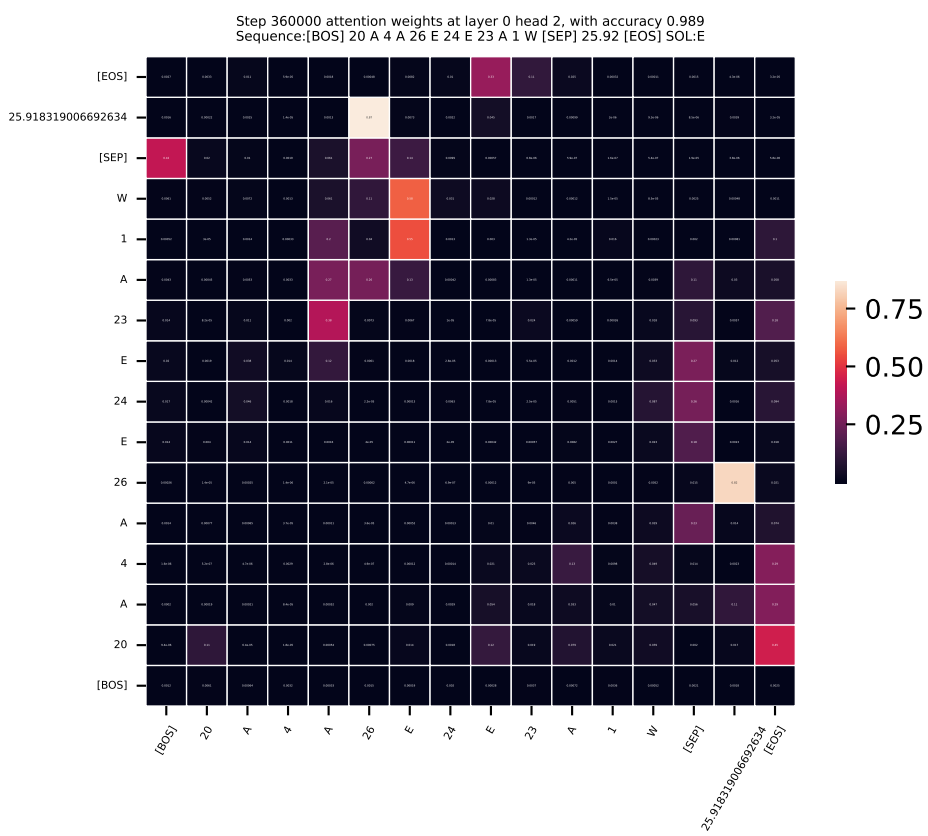

Step 359500 attention weights at layer 1 head 1, with accuracy 0.990
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

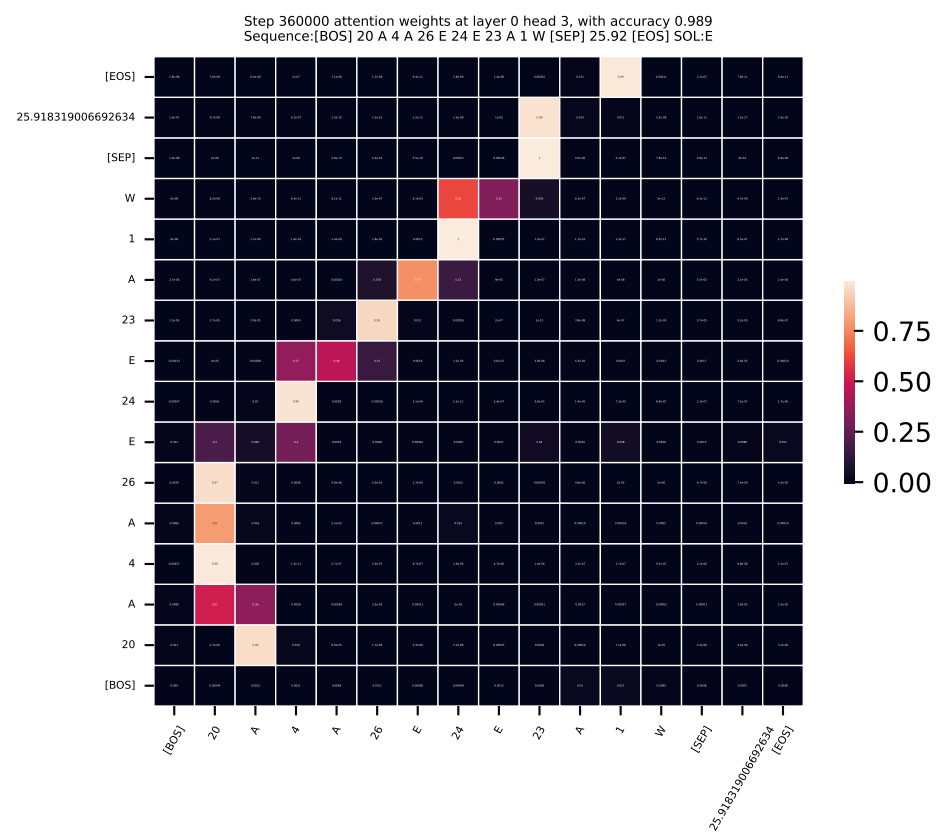

Step 359500 attention weights at layer 1 head 2, with accuracy 0.990
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

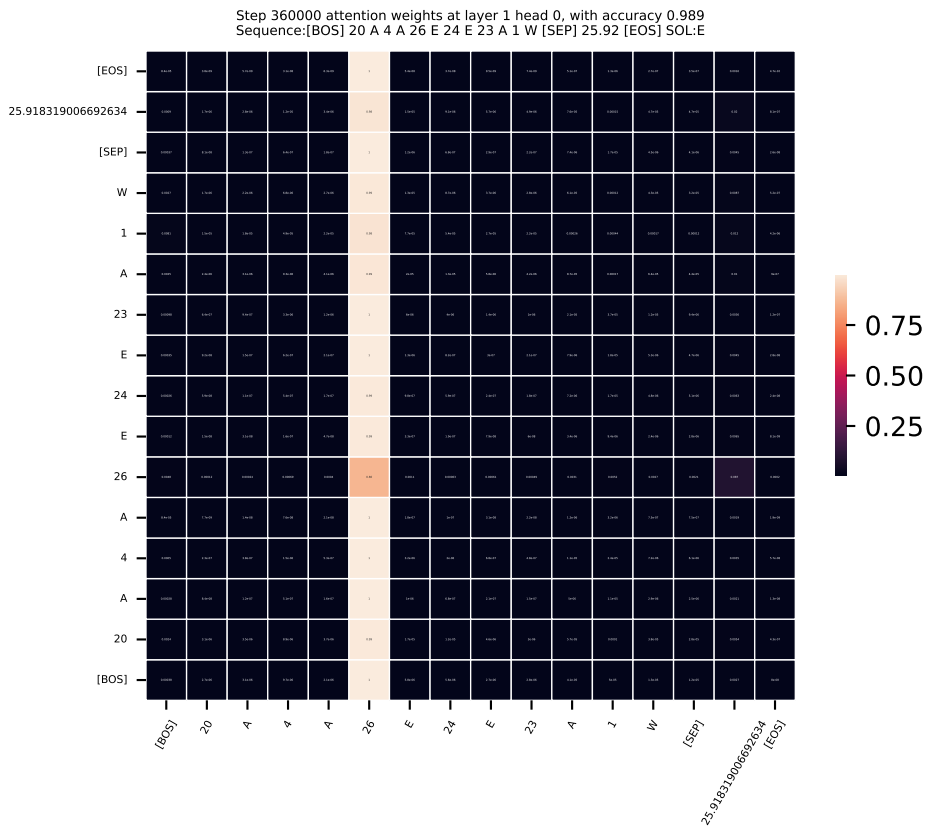

Step 359500 attention weights at layer 1 head 3, with accuracy 0.990
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
1.0
0.5
0.0

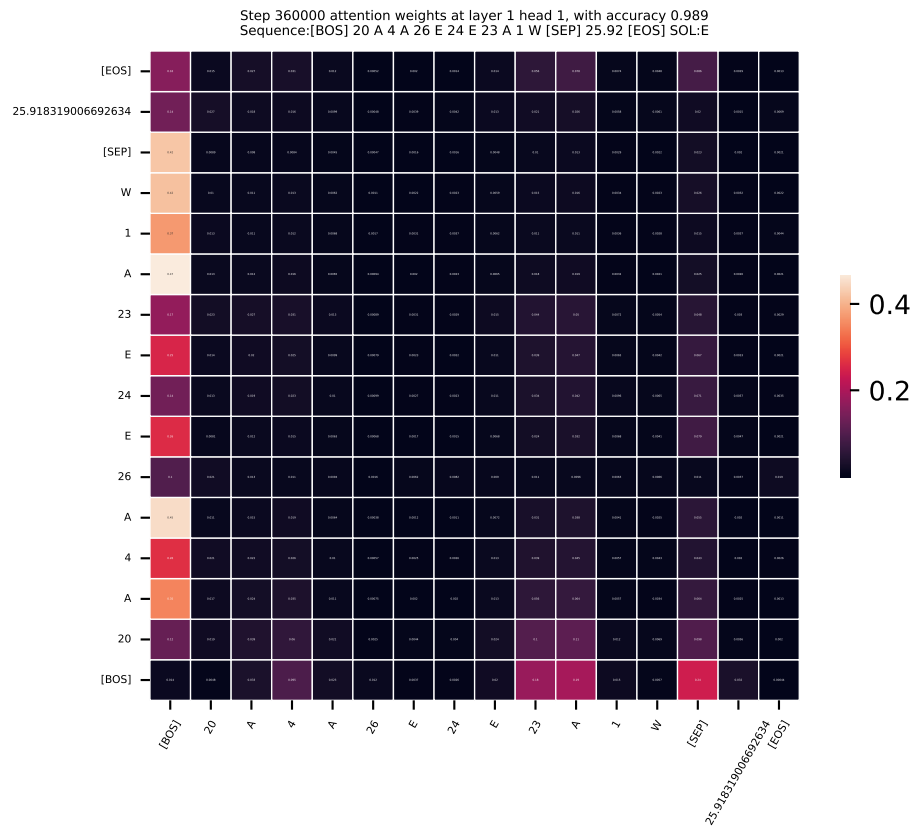

Step 360000 attention weights at layer 0 head 0, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

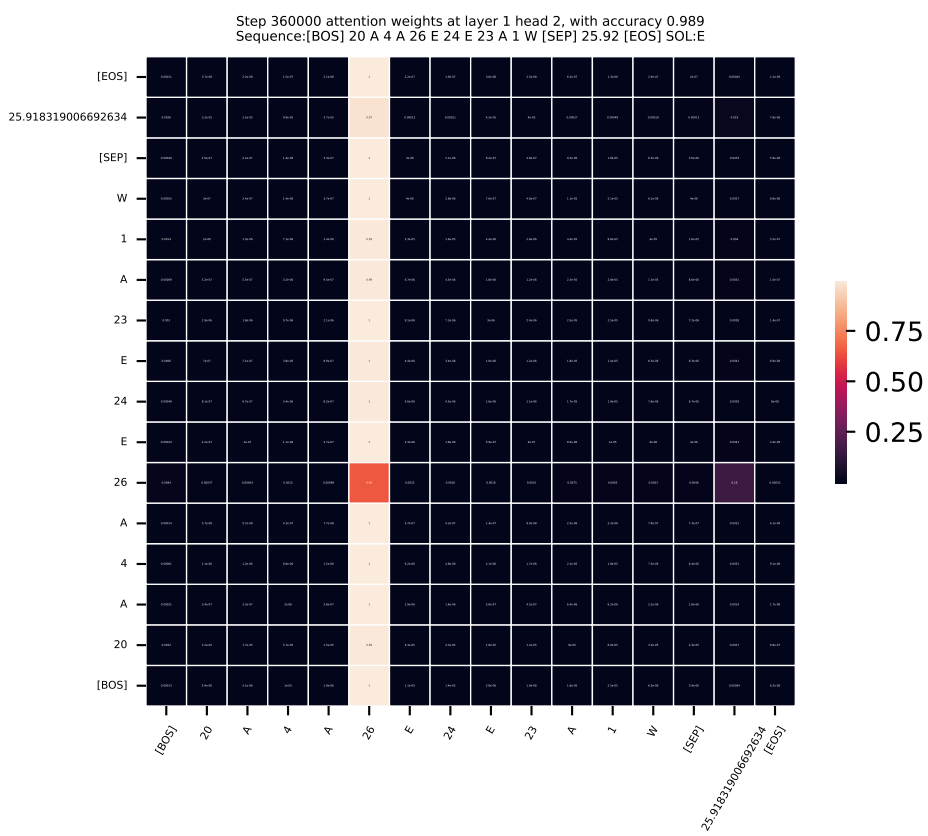

Step 360000 attention weights at layer 0 head 1, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

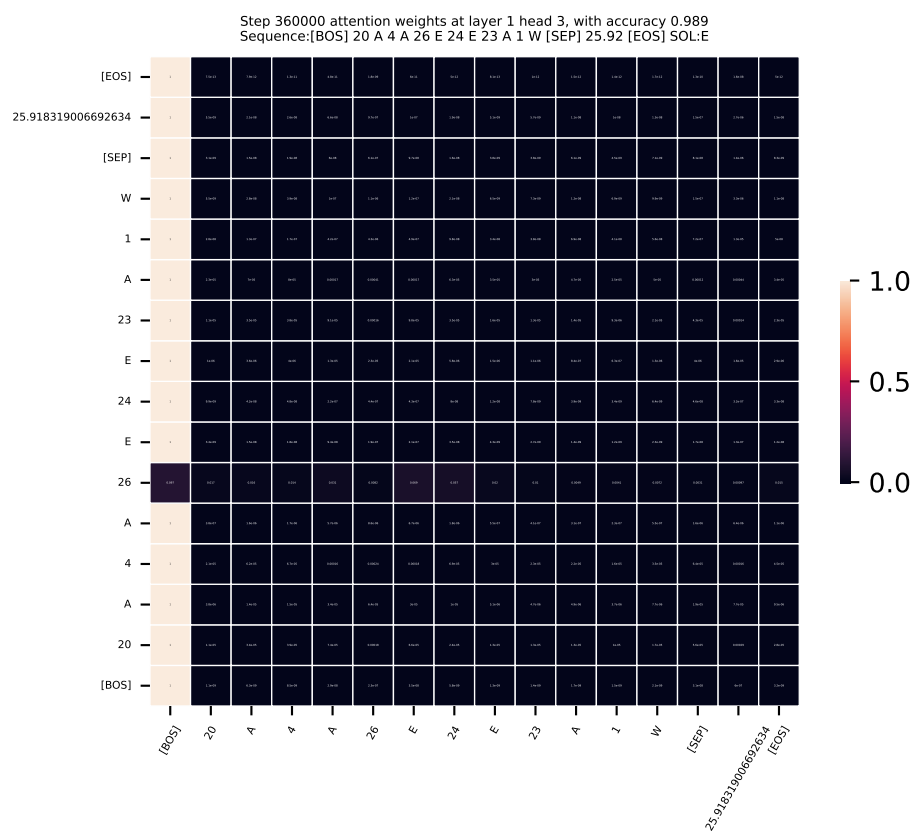

Step 360000 attention weights at layer 0 head 2, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

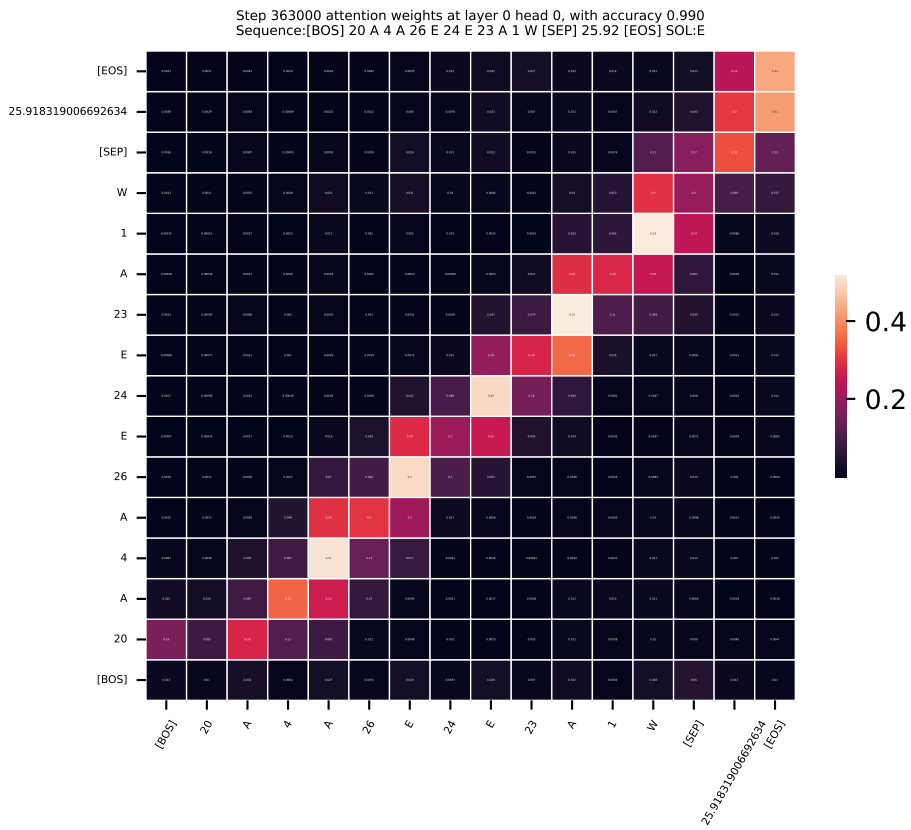

Step 360000 attention weights at layer 0 head 3, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

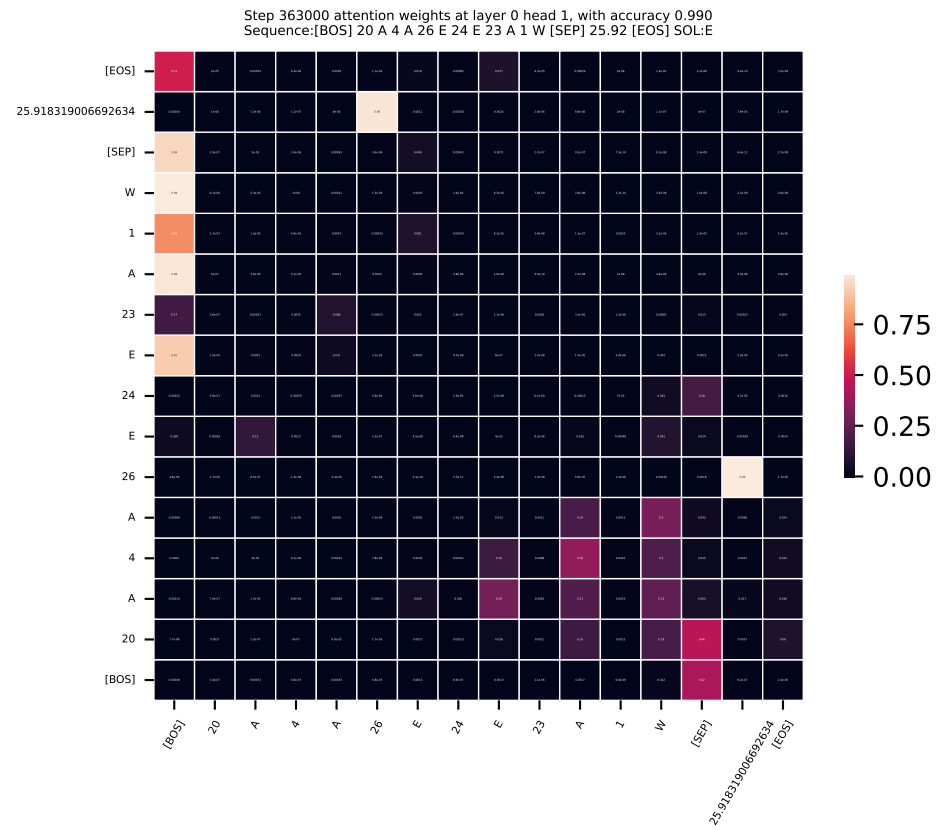

Step 360000 attention weights at layer 1 head 0, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

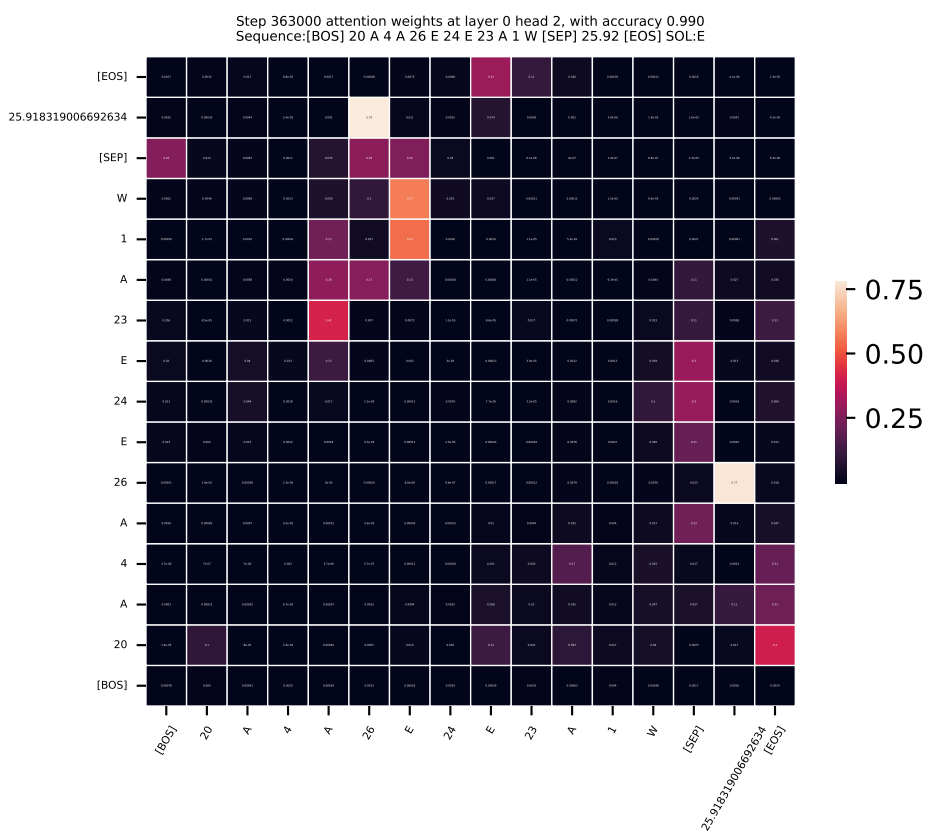

Step 360000 attention weights at layer 1 head 1, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

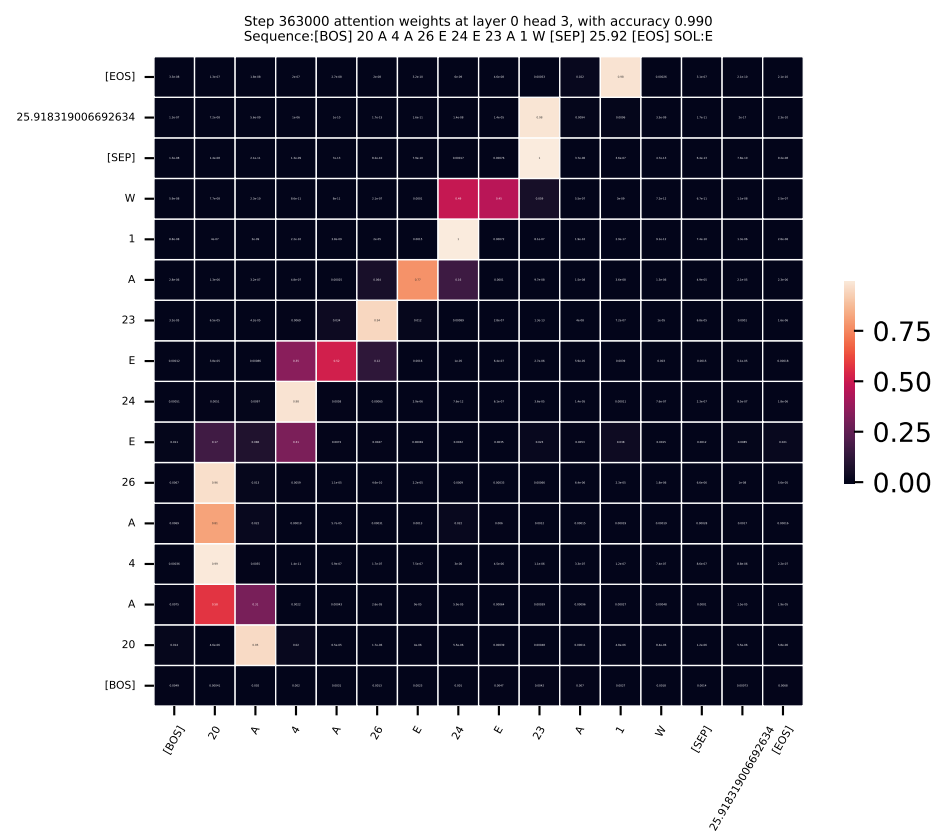

Step 360000 attention weights at layer 1 head 2, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

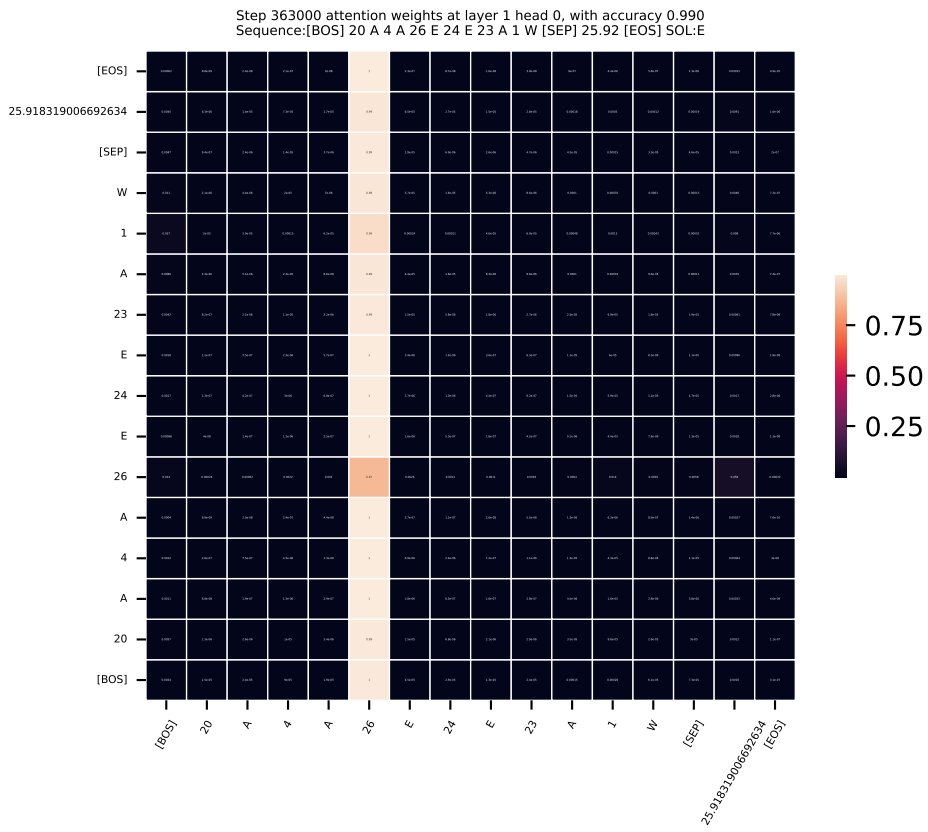

Step 360000 attention weights at layer 1 head 3, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
1.0
0.5
0.0

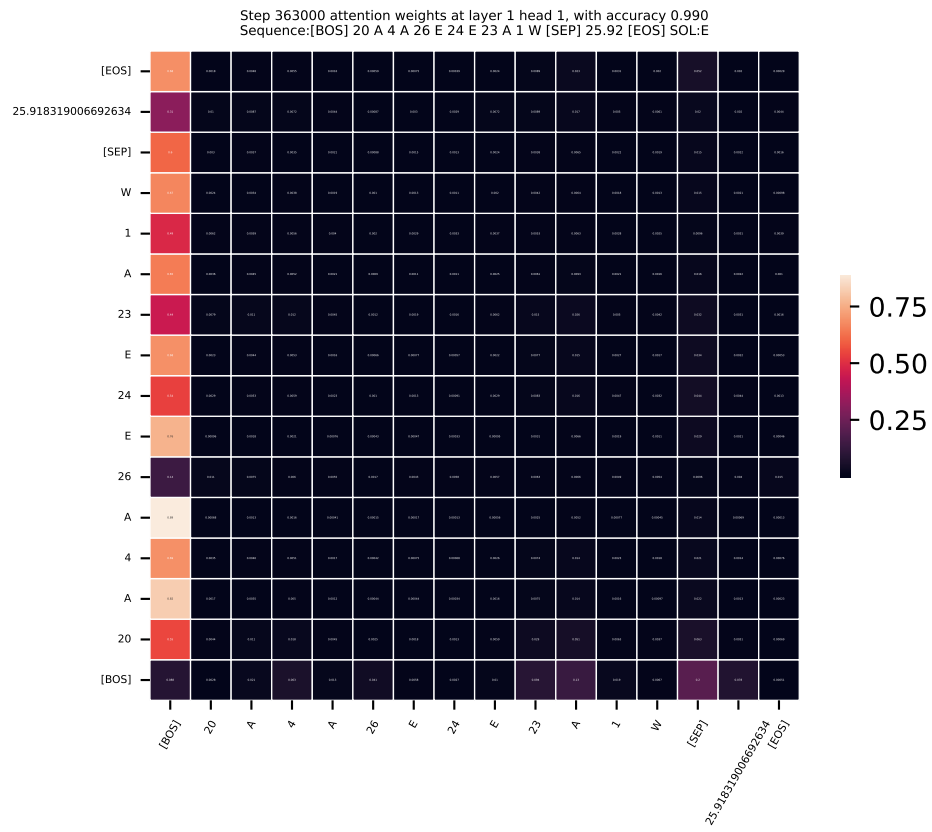

Step 363000 attention weights at layer 0 head 0, with accuracy 0.990
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

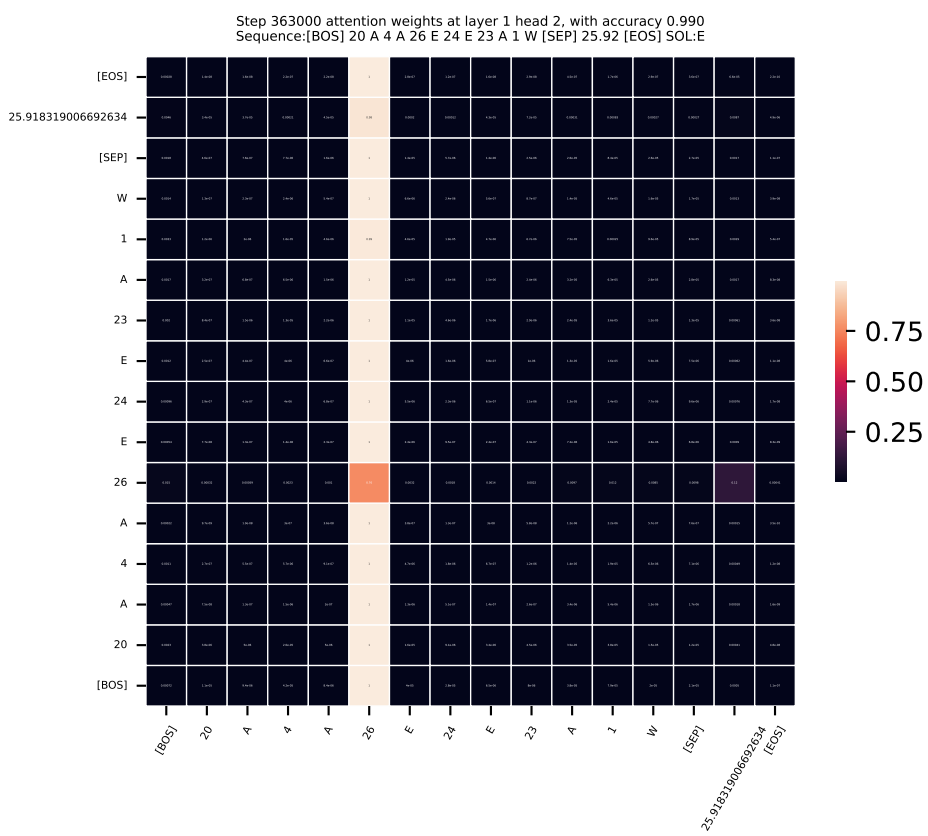

Step 363000 attention weights at layer 0 head 1, with accuracy 0.990
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

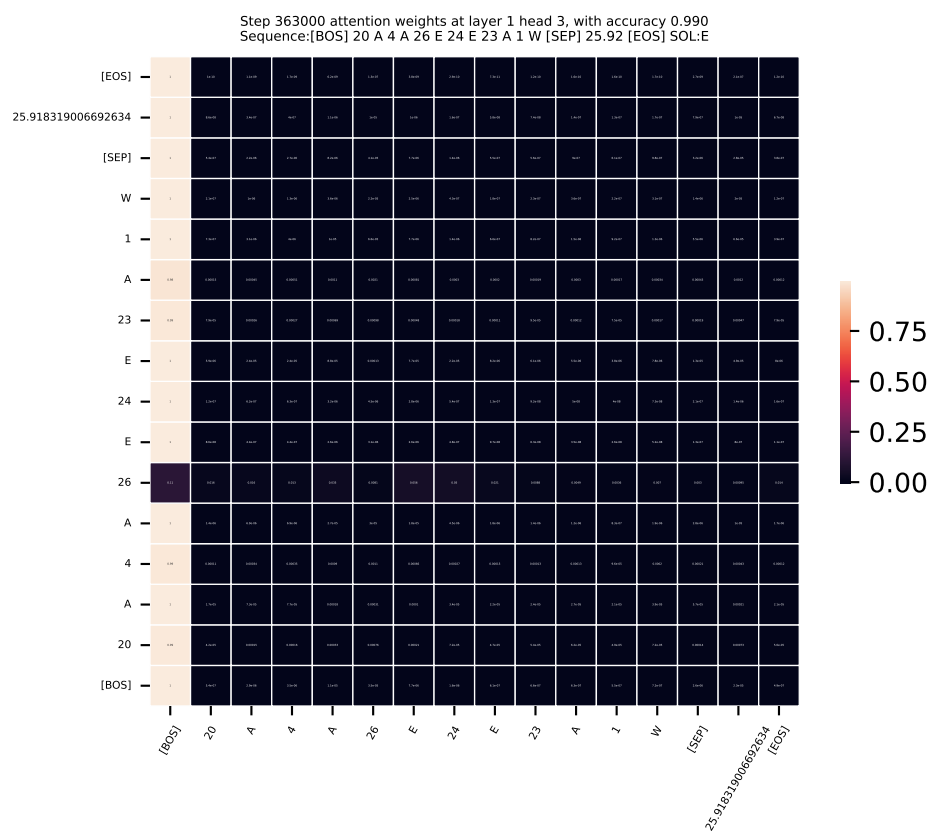

Step 363000 attention weights at layer 0 head 2, with accuracy 0.990
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

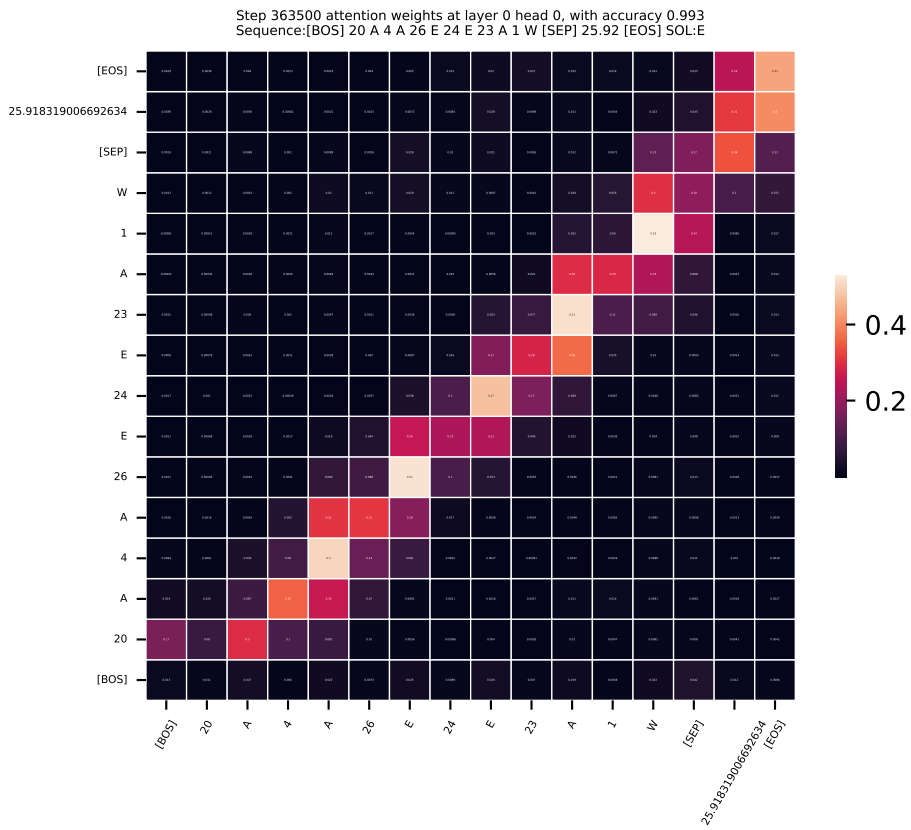

Step 363000 attention weights at layer 0 head 3, with accuracy 0.990
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

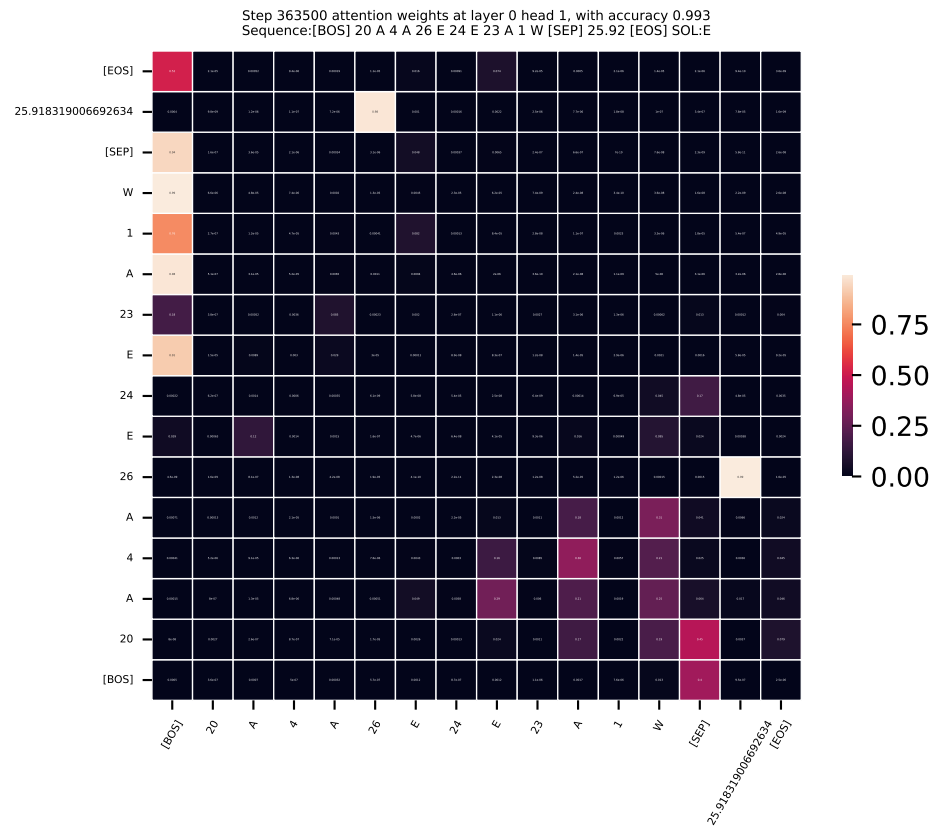

Step 363000 attention weights at layer 1 head 0, with accuracy 0.990
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

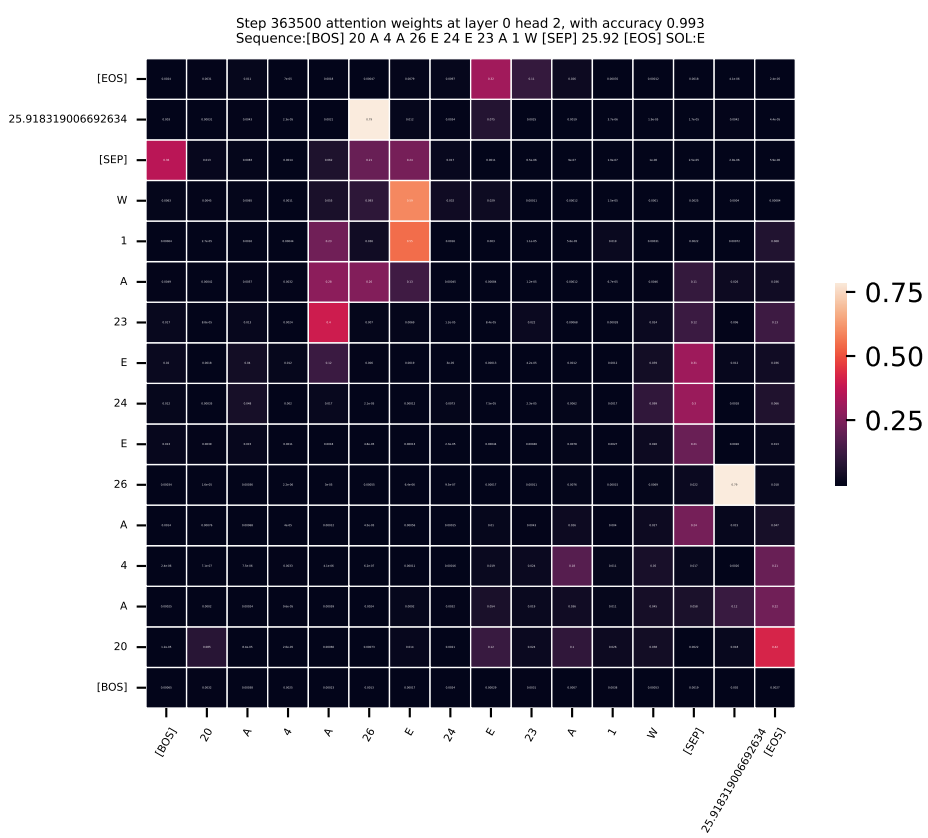

Step 363000 attention weights at layer 1 head 1, with accuracy 0.990
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

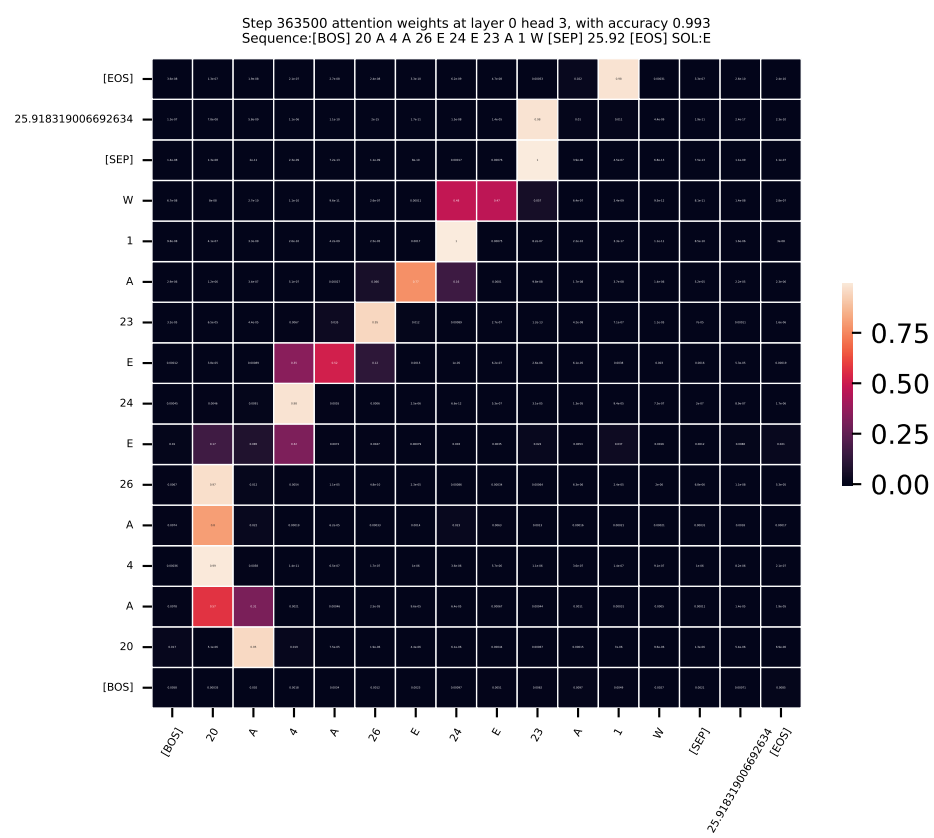

Step 363000 attention weights at layer 1 head 2, with accuracy 0.990
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

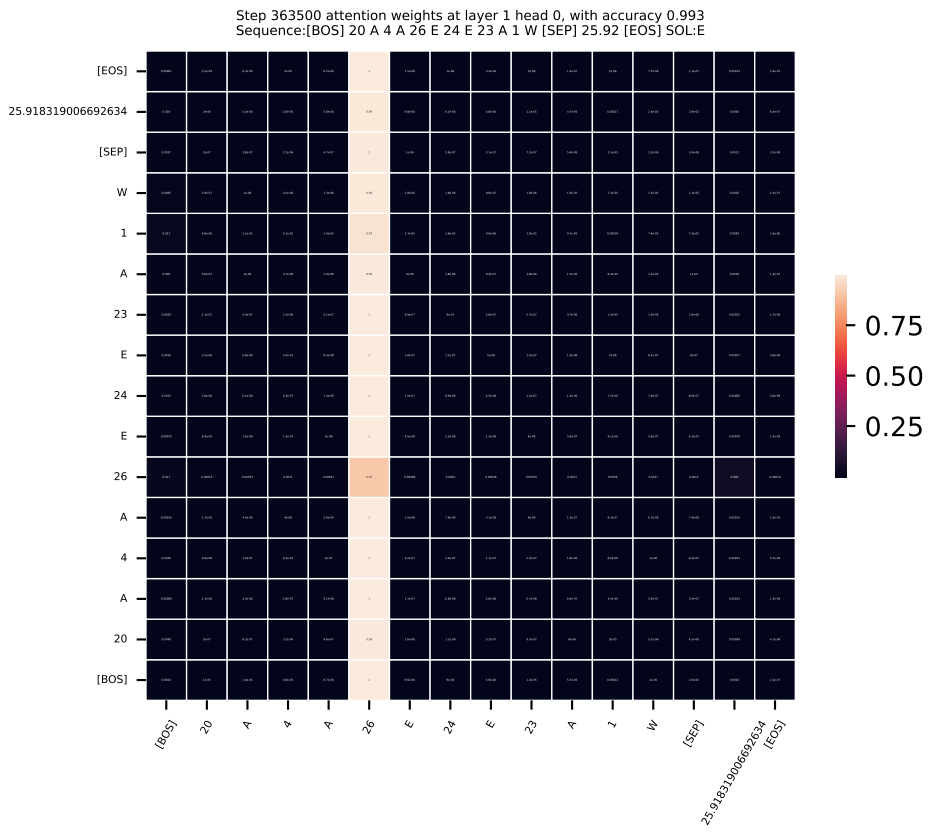

Step 363000 attention weights at layer 1 head 3, with accuracy 0.990
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

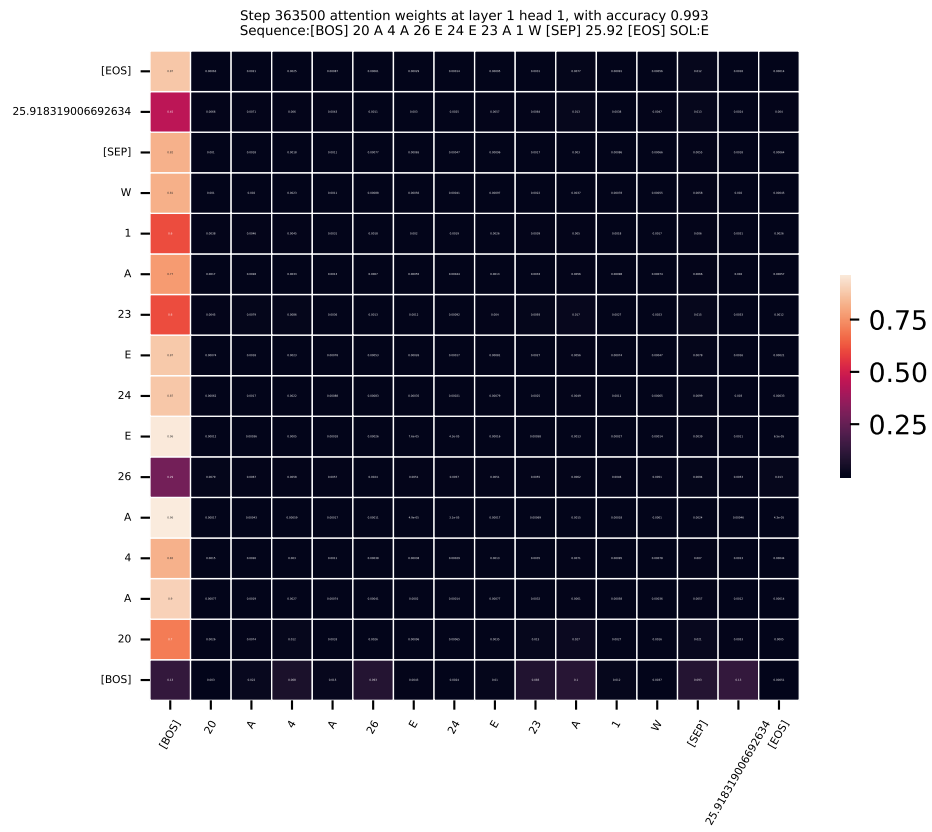

Step 363500 attention weights at layer 0 head 0, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

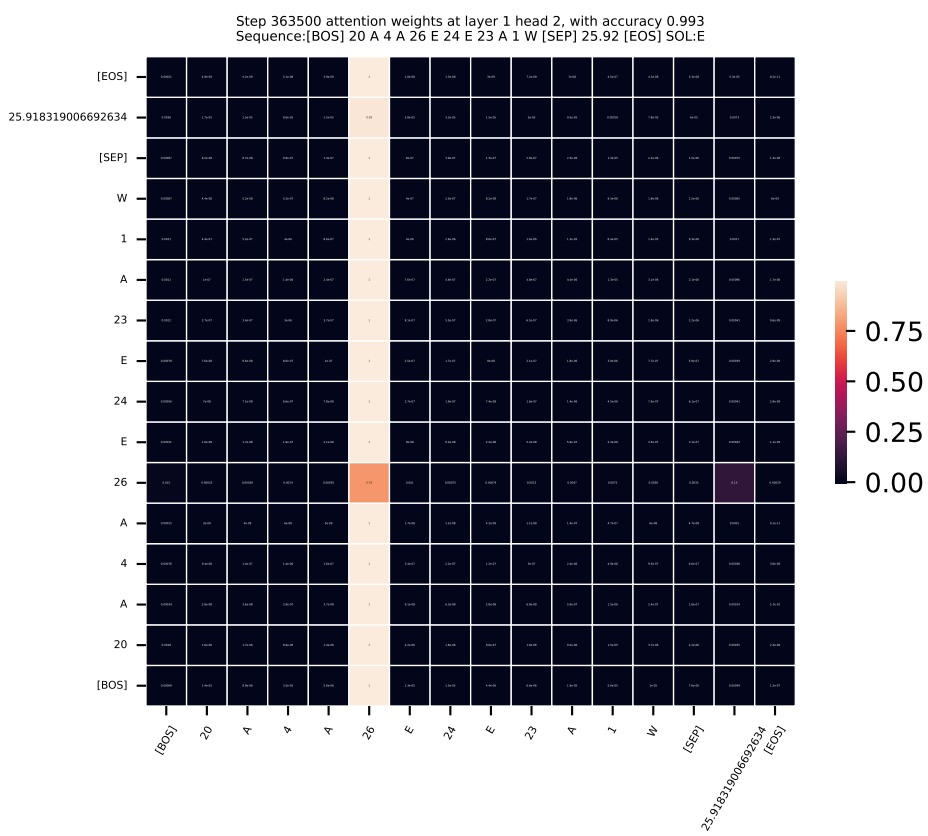

Step 363500 attention weights at layer 0 head 1, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

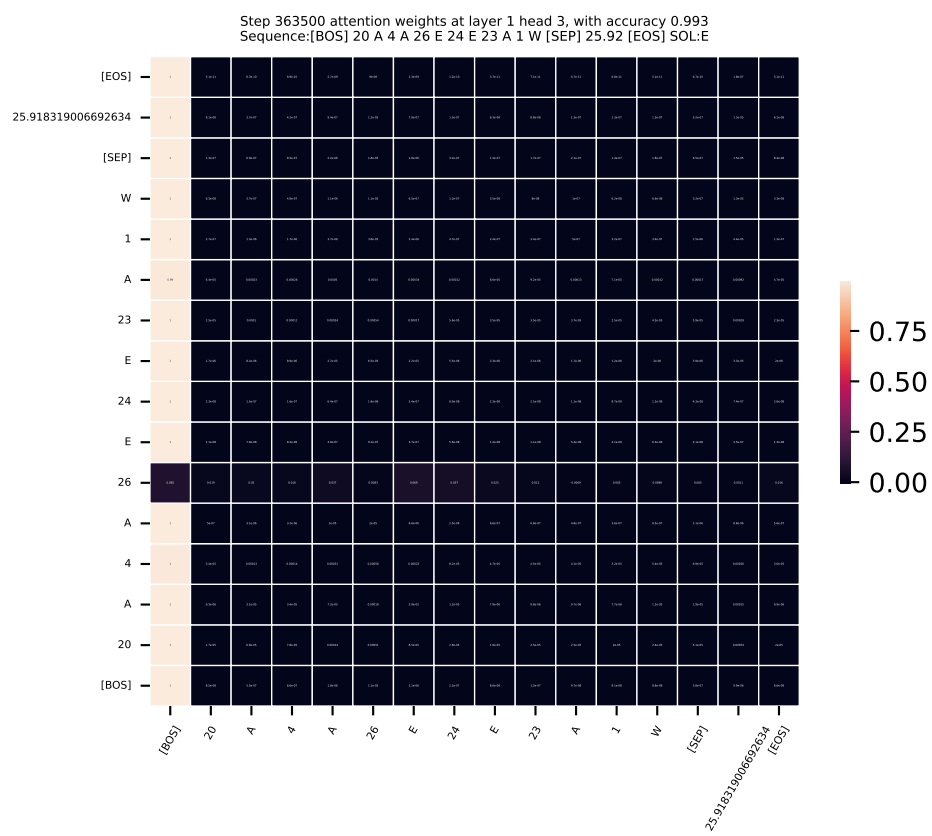

Step 363500 attention weights at layer 0 head 2, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

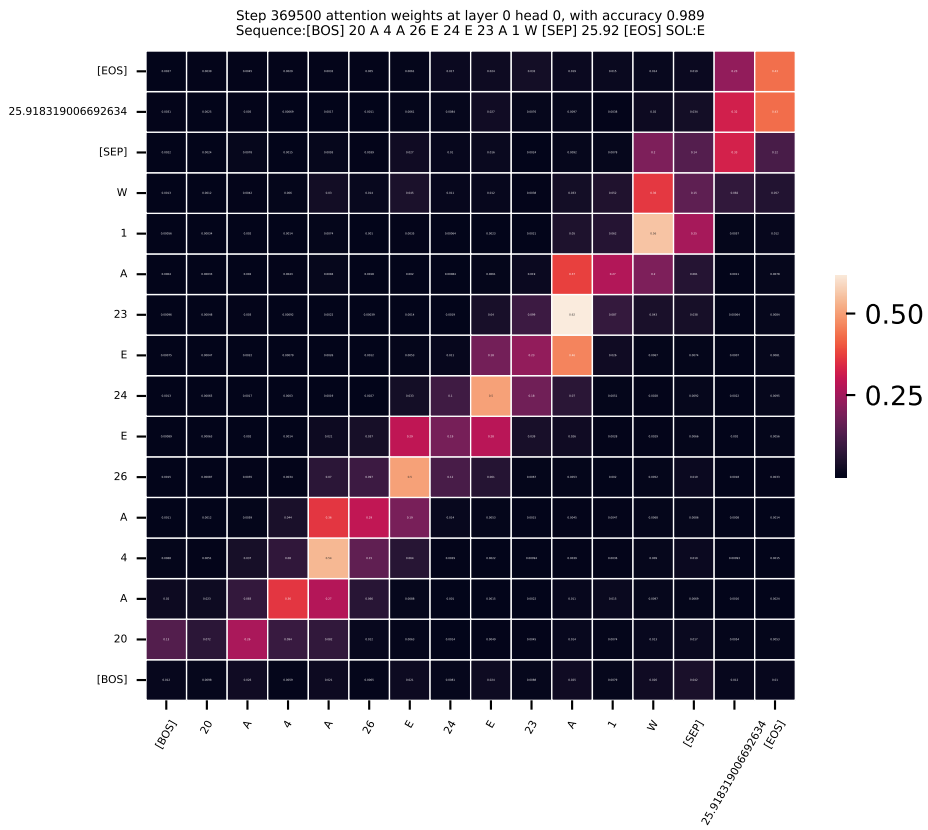

Step 363500 attention weights at layer 0 head 3, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

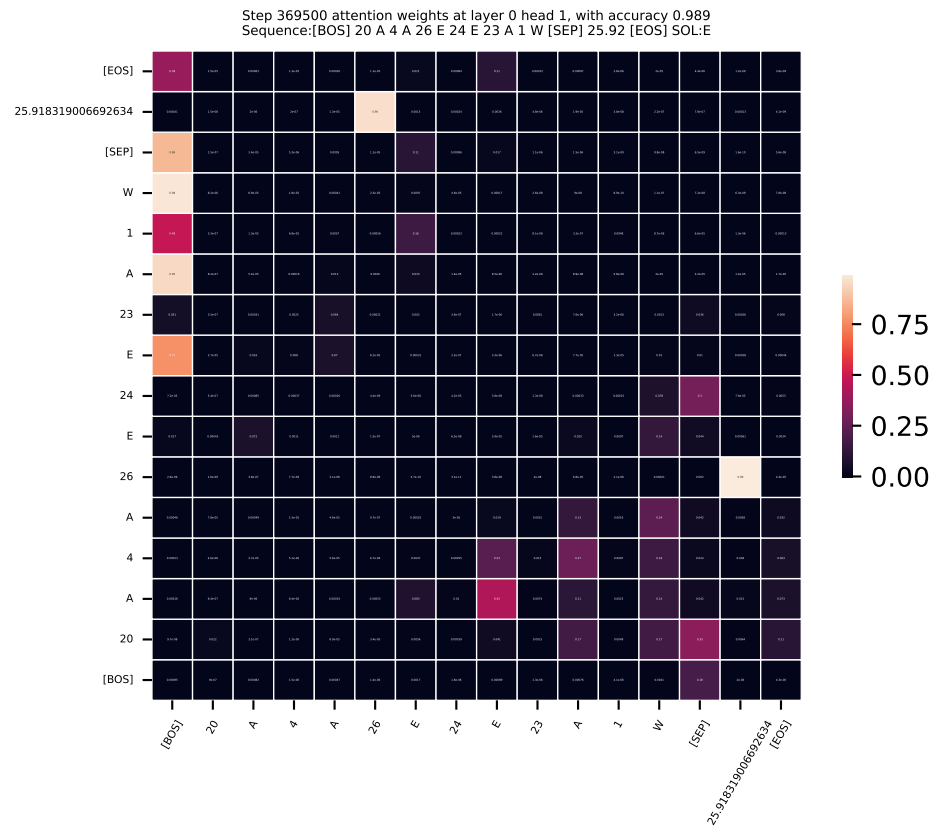

Step 363500 attention weights at layer 1 head 0, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

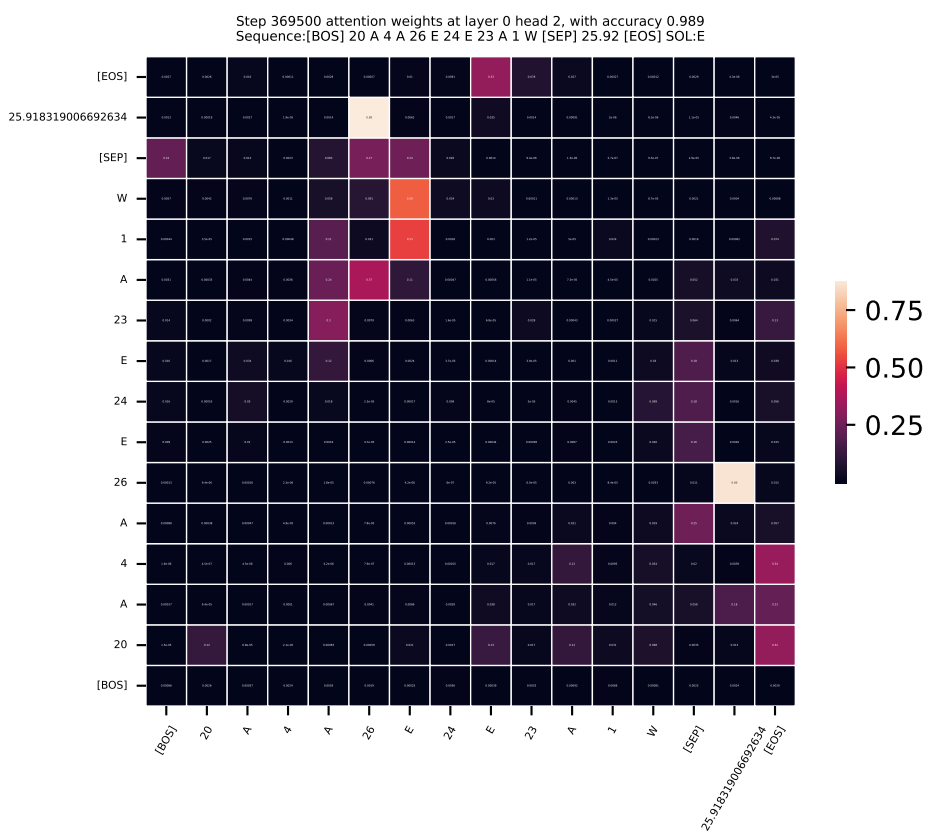

Step 363500 attention weights at layer 1 head 1, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

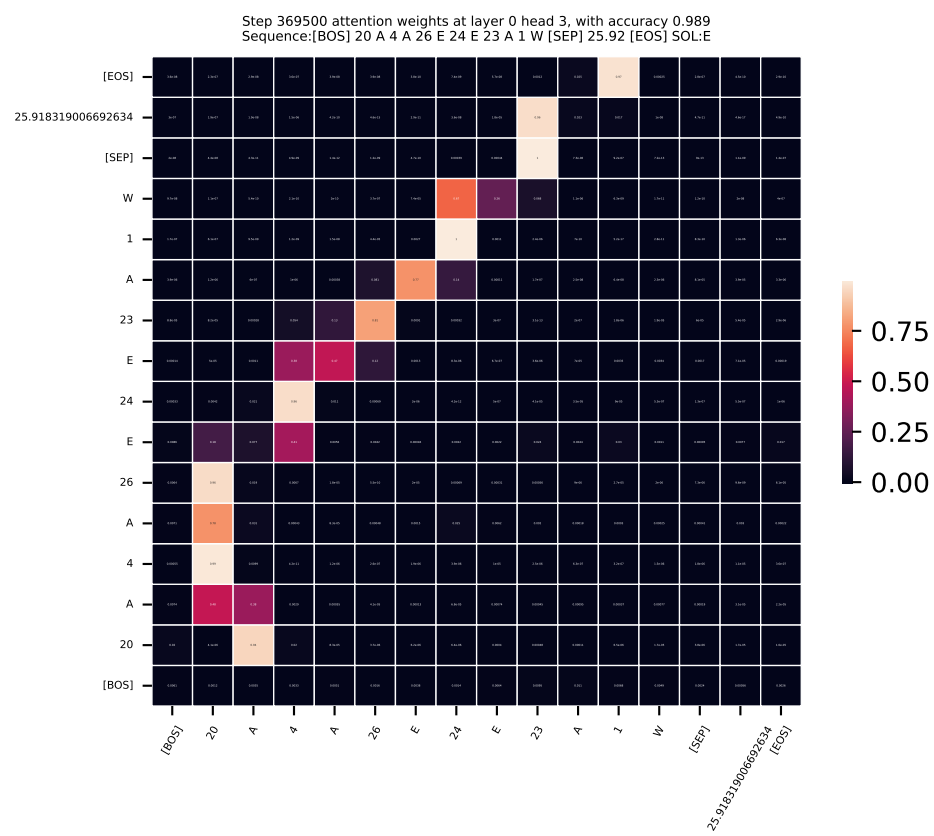

Step 363500 attention weights at layer 1 head 2, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

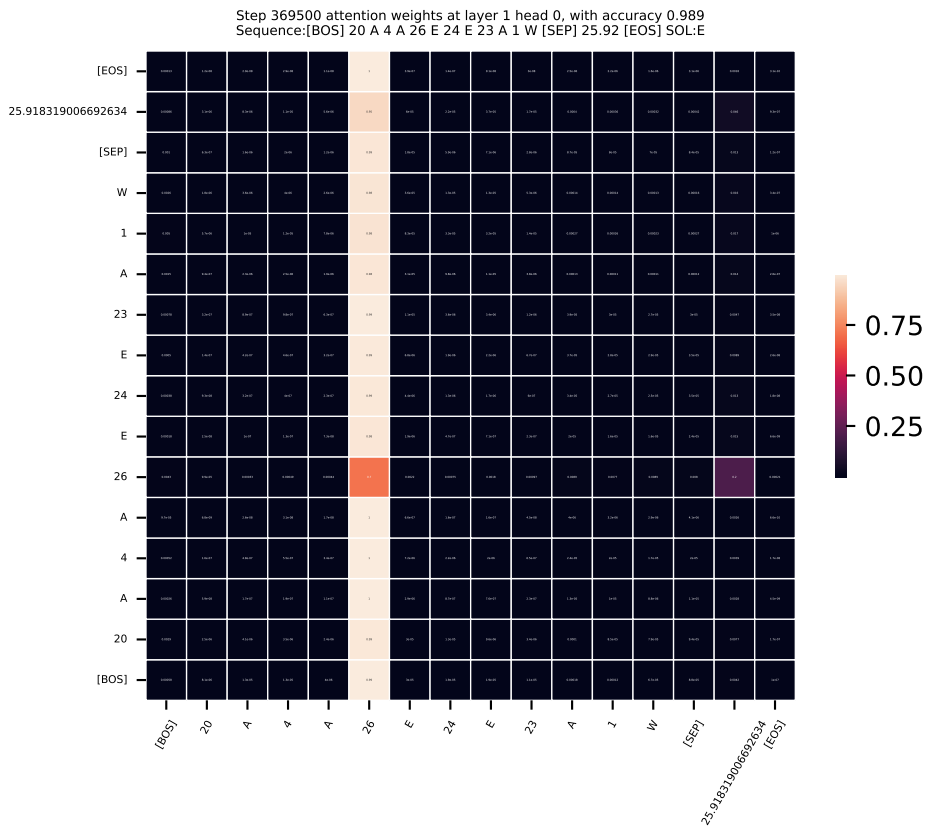

Step 363500 attention weights at layer 1 head 3, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

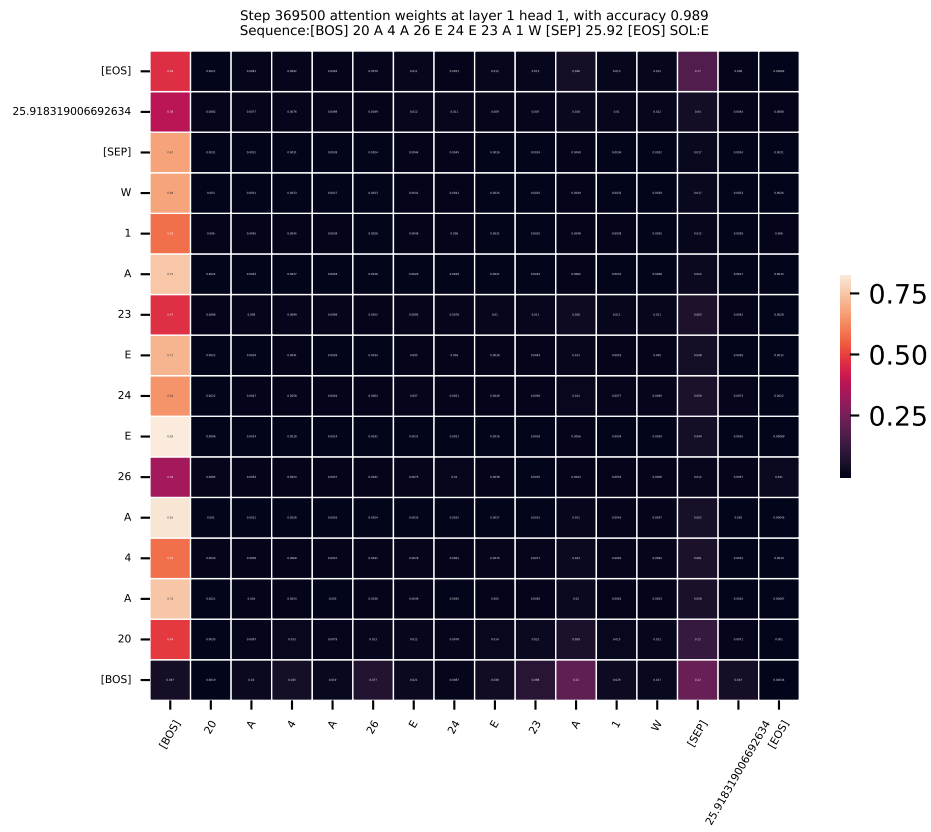

Step 369500 attention weights at layer 0 head 0, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

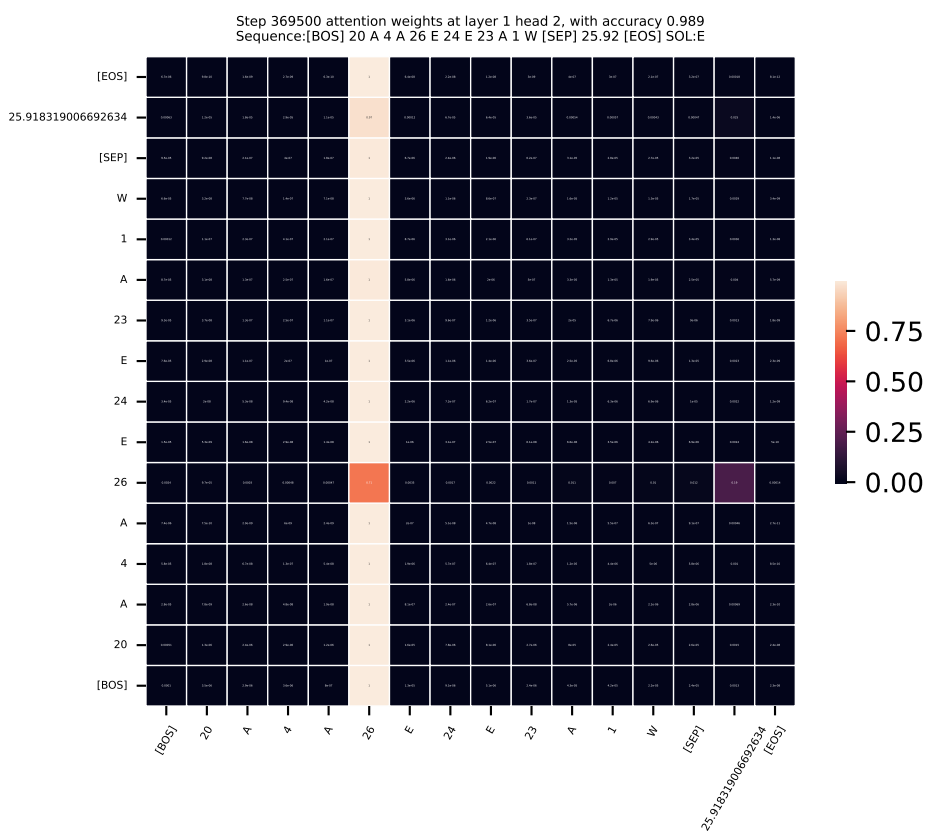

Step 369500 attention weights at layer 0 head 1, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

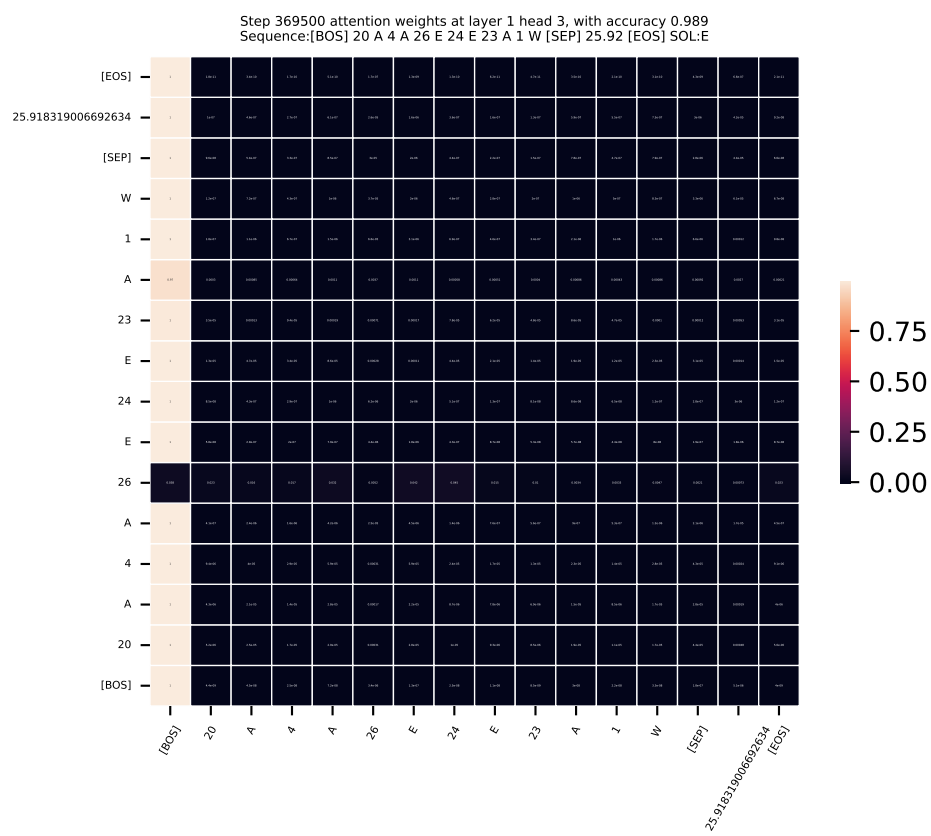

Step 369500 attention weights at layer 0 head 2, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

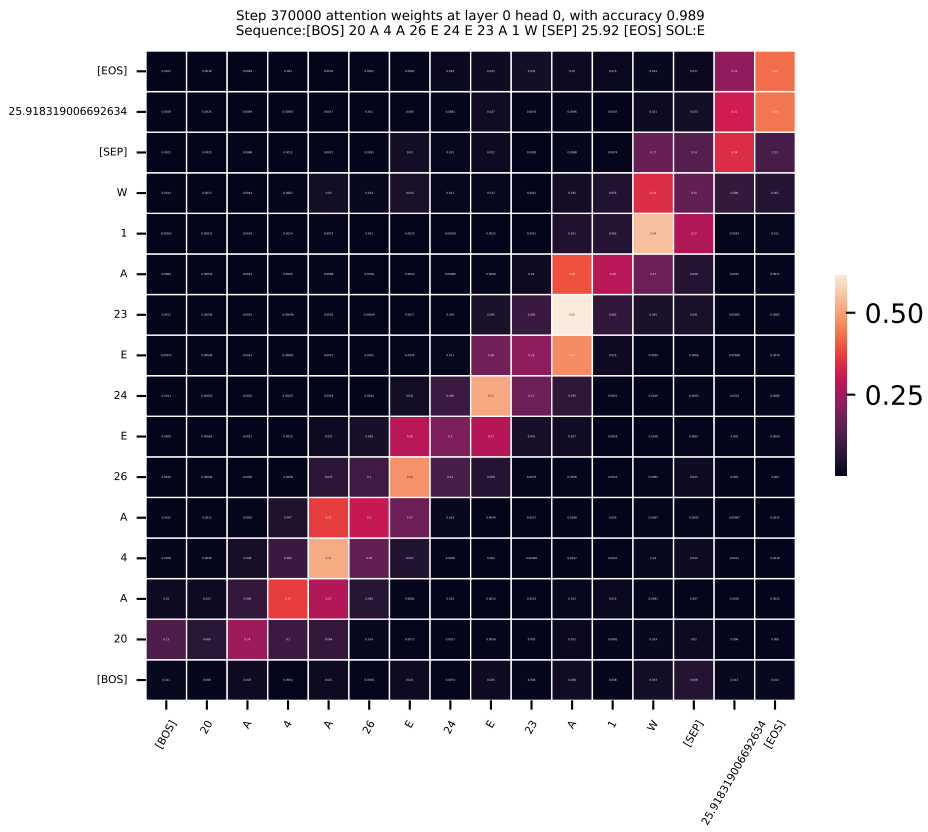

Step 369500 attention weights at layer 0 head 3, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

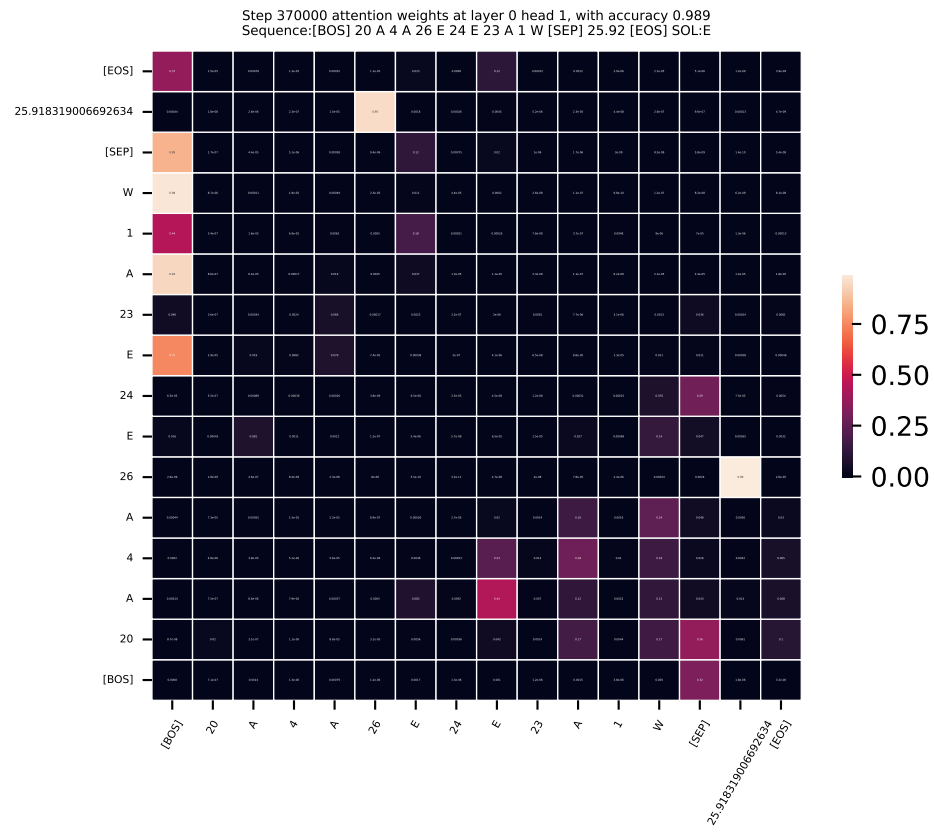

Step 369500 attention weights at layer 1 head 0, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

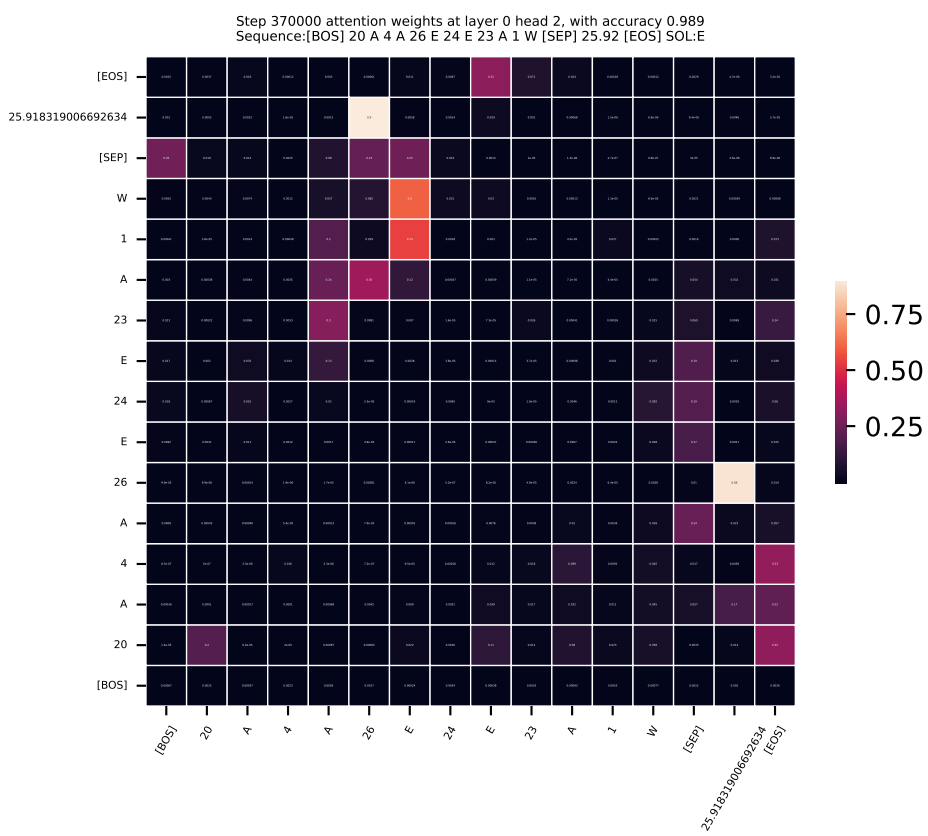

Step 369500 attention weights at layer 1 head 1, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

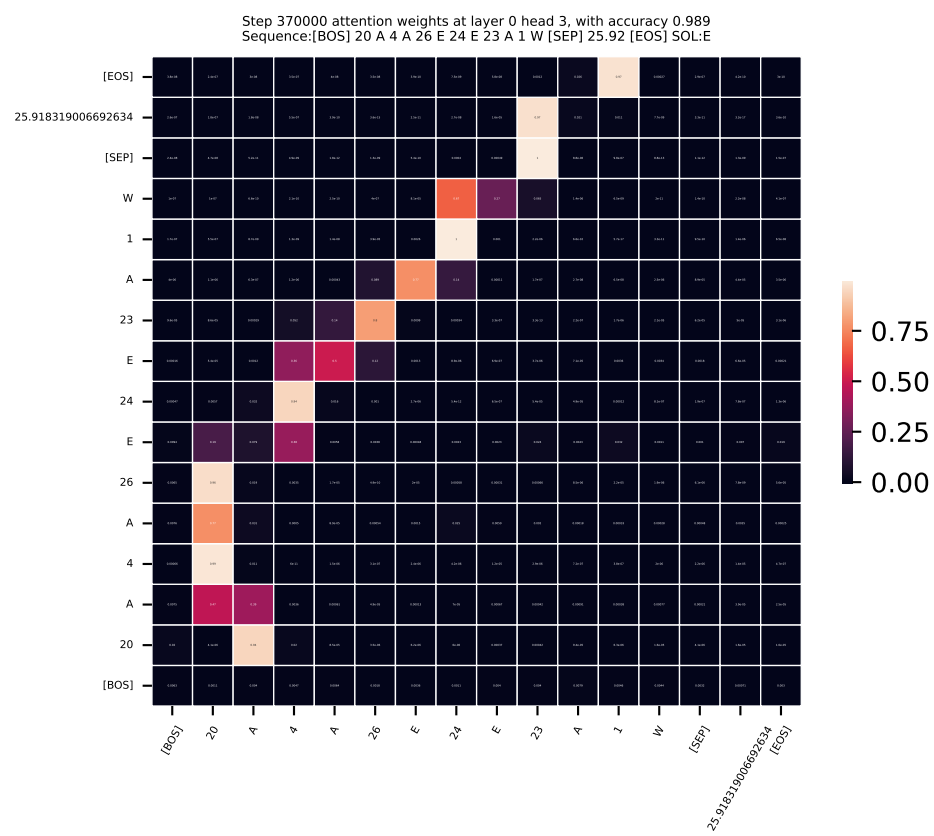

Step 369500 attention weights at layer 1 head 2, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

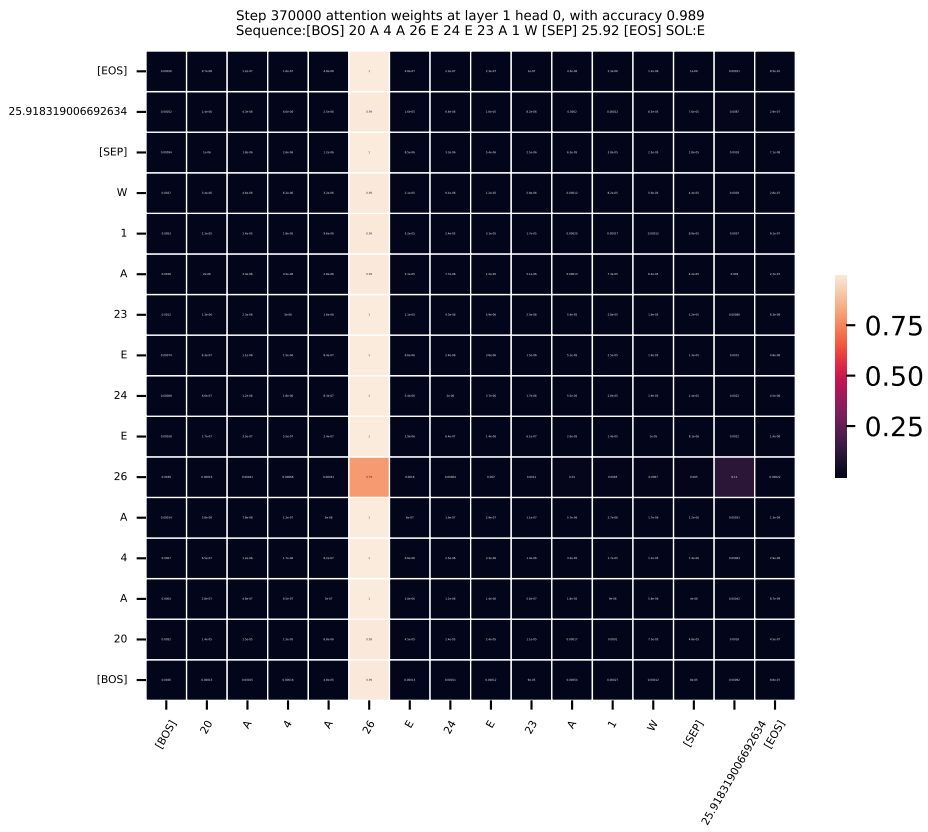

Step 369500 attention weights at layer 1 head 3, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

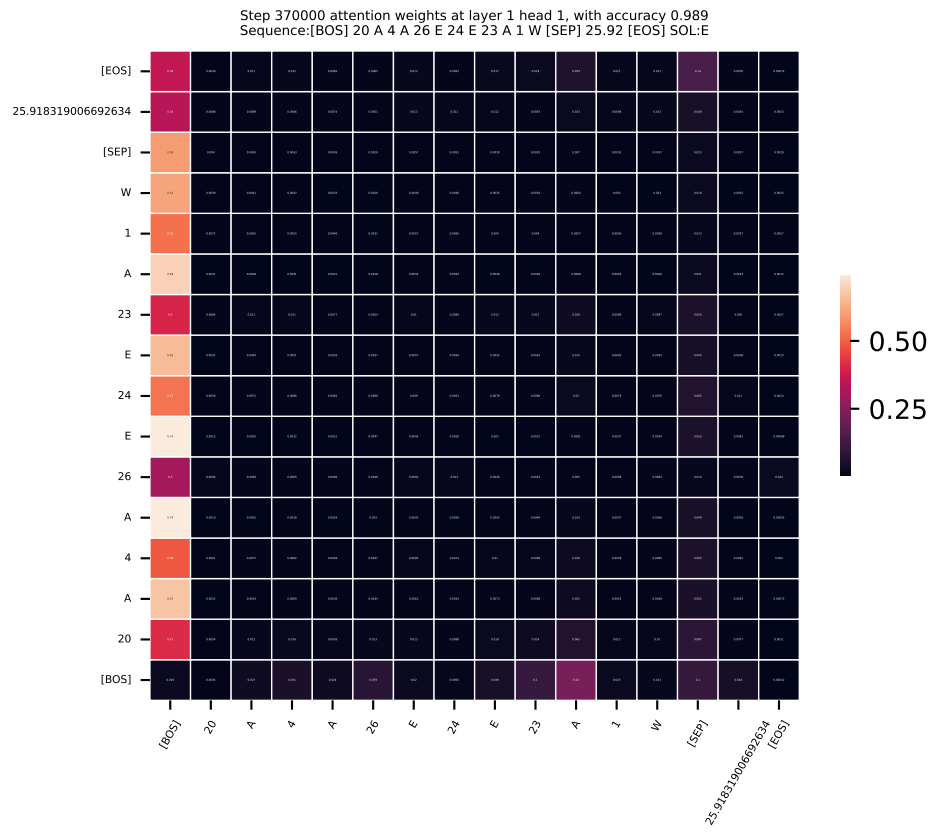

Step 370000 attention weights at layer 0 head 0, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

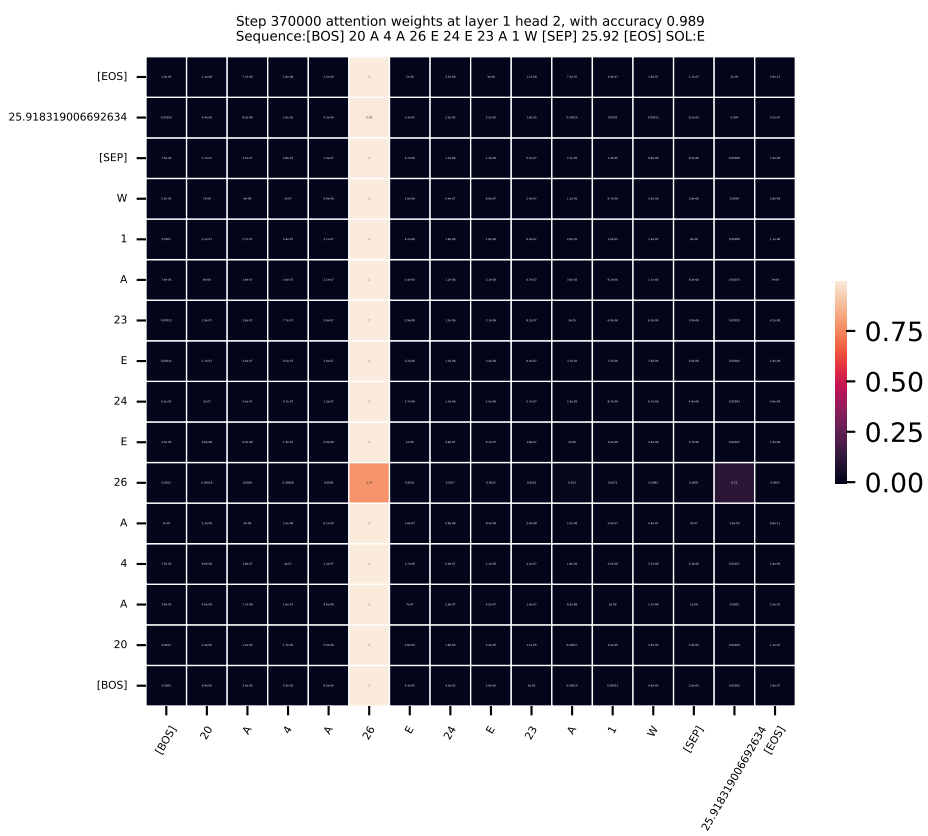

Step 370000 attention weights at layer 0 head 1, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

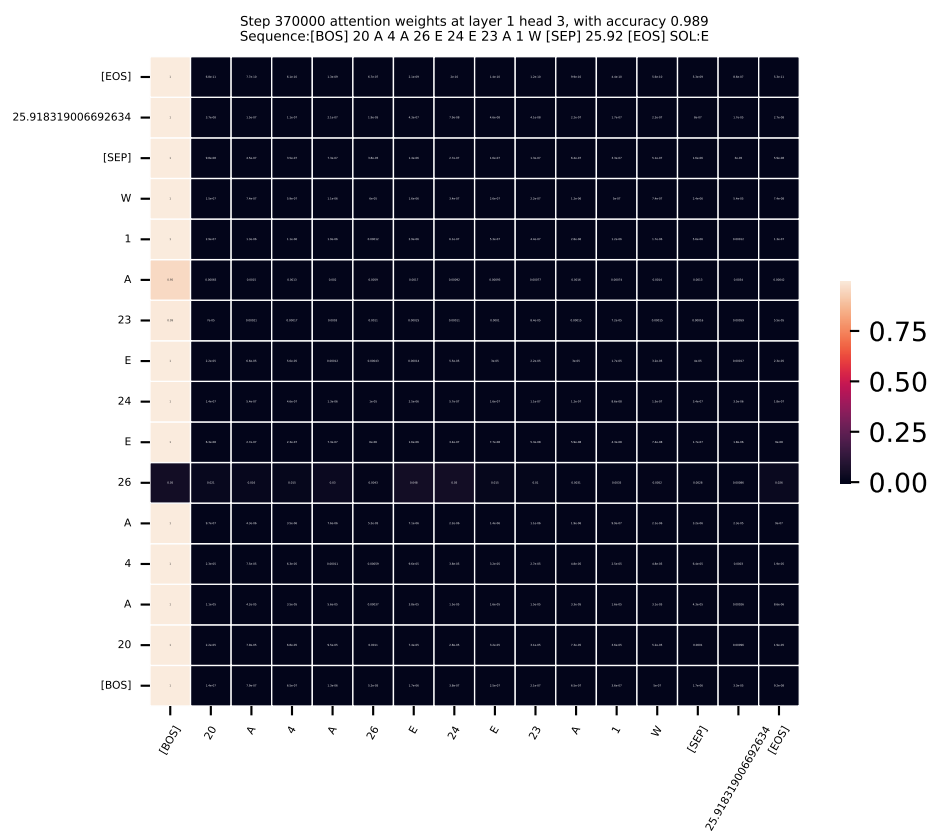

Step 370000 attention weights at layer 0 head 2, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

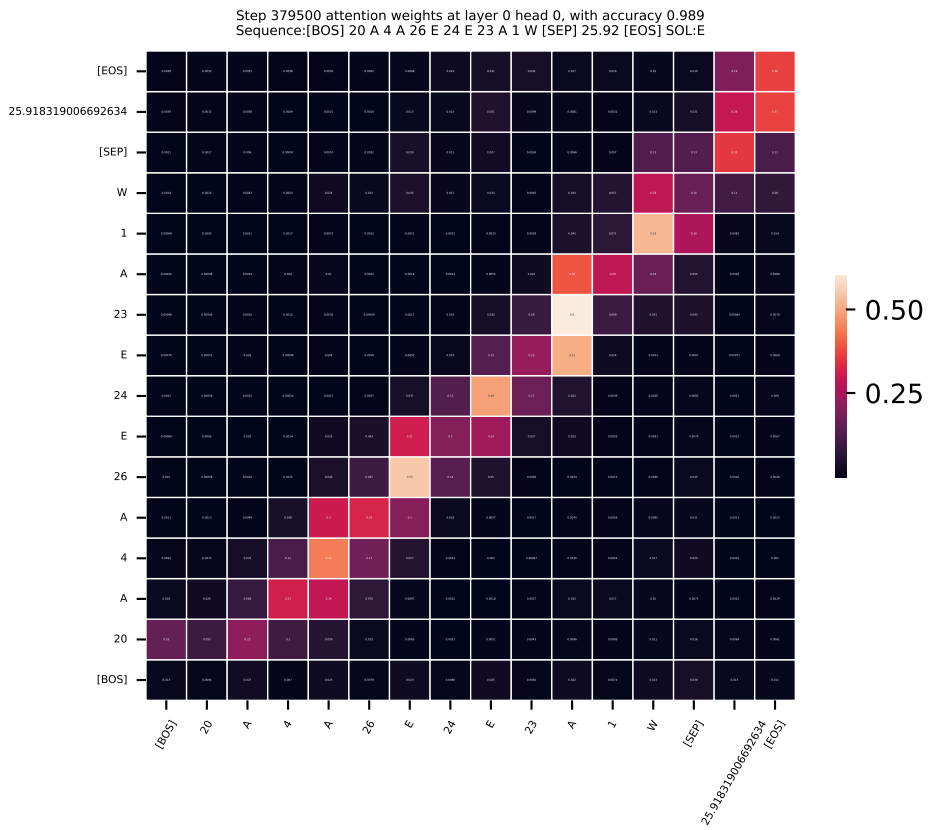

Step 370000 attention weights at layer 0 head 3, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

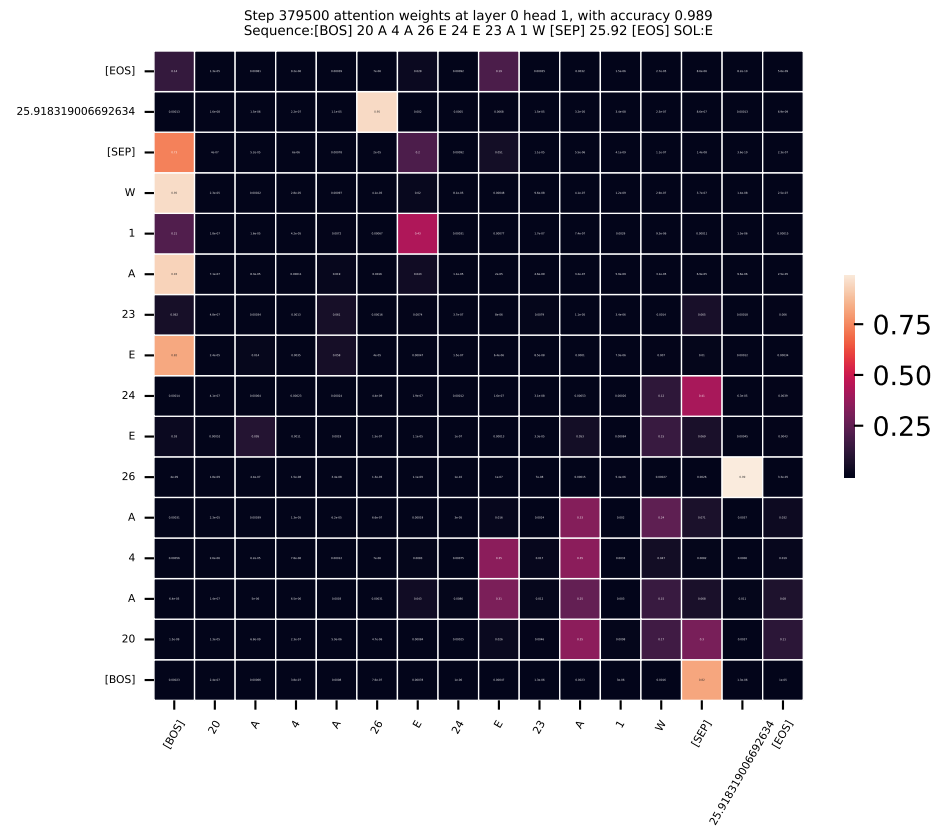

Step 370000 attention weights at layer 1 head 0, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

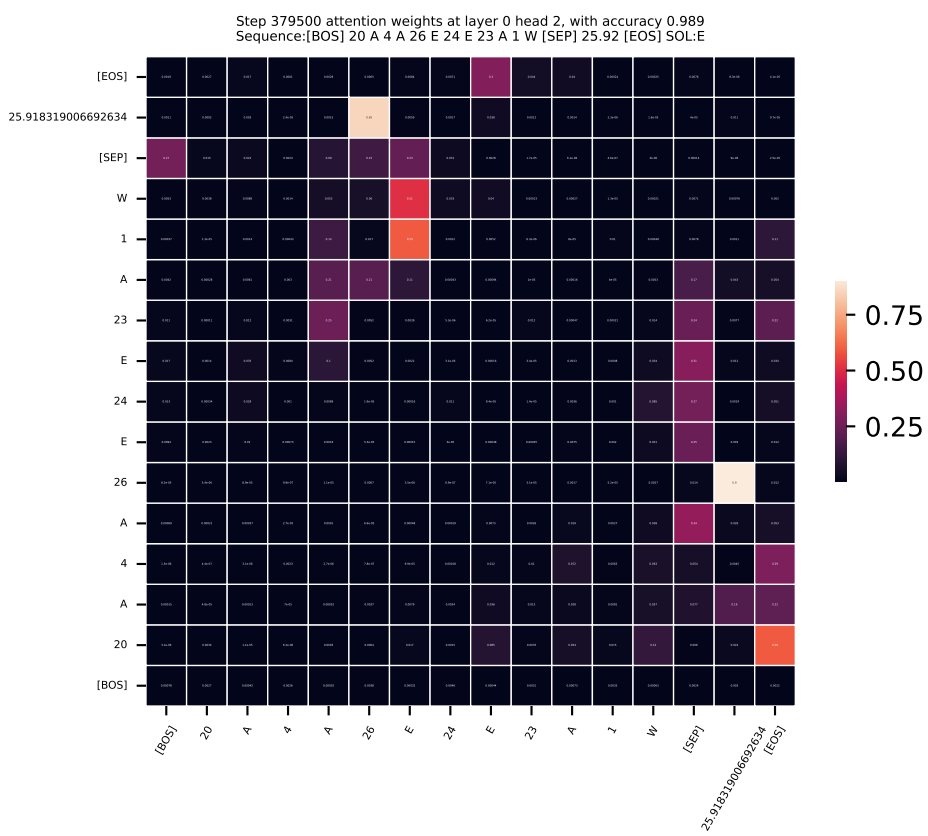

Step 370000 attention weights at layer 1 head 1, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

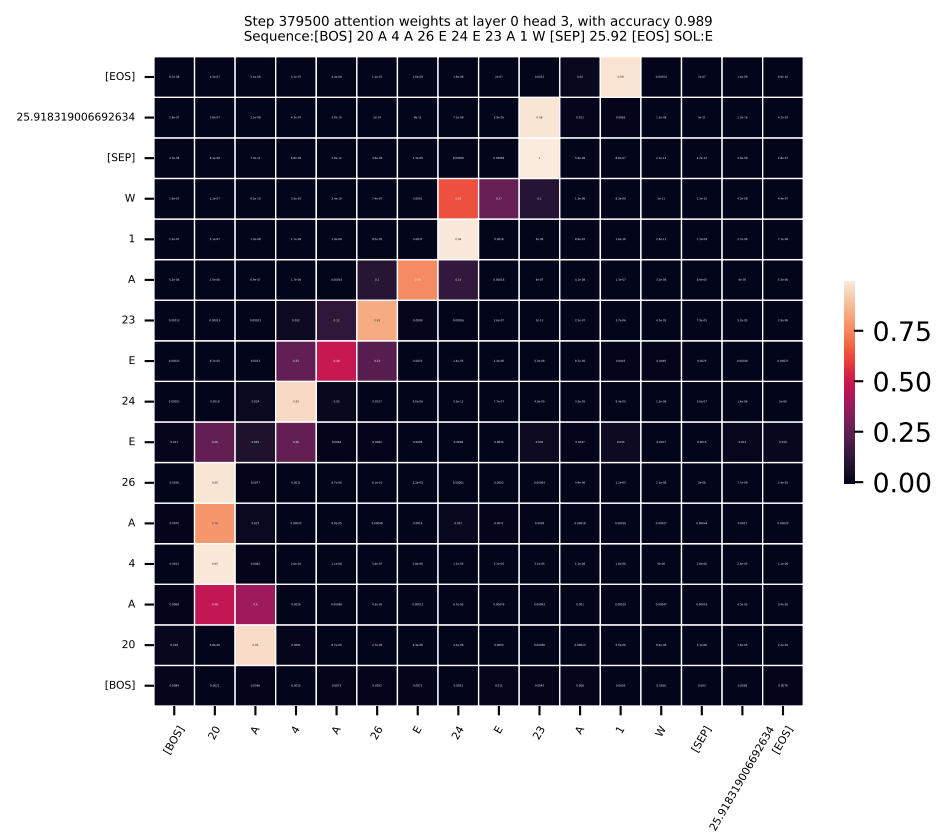

Step 370000 attention weights at layer 1 head 2, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

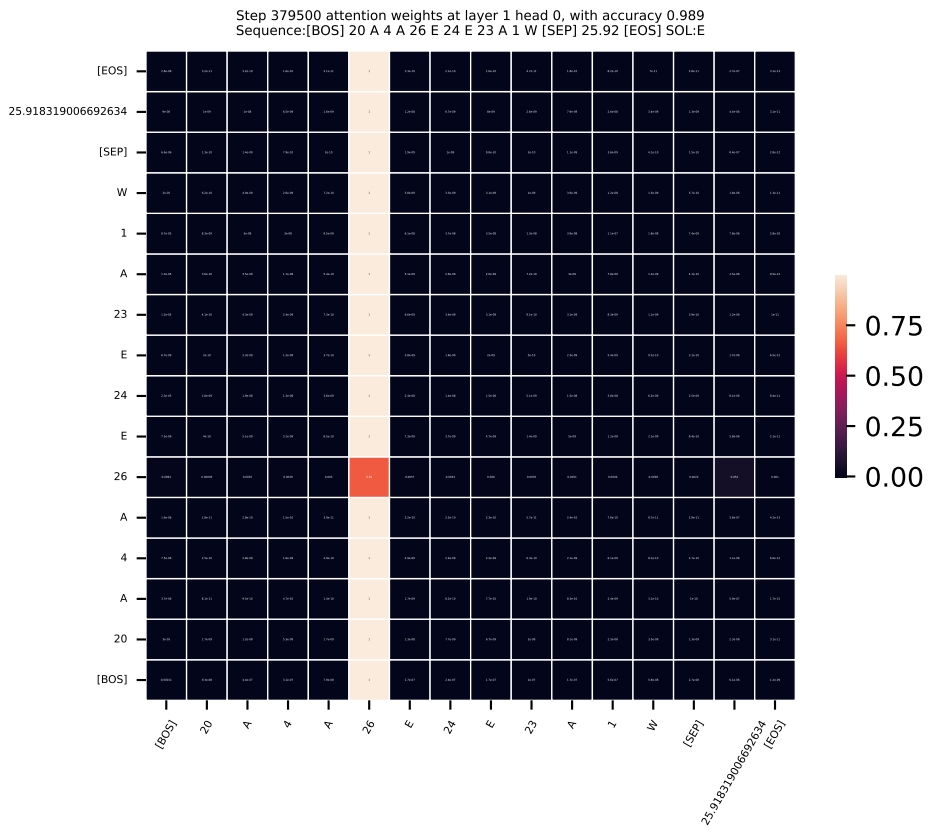

Step 370000 attention weights at layer 1 head 3, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

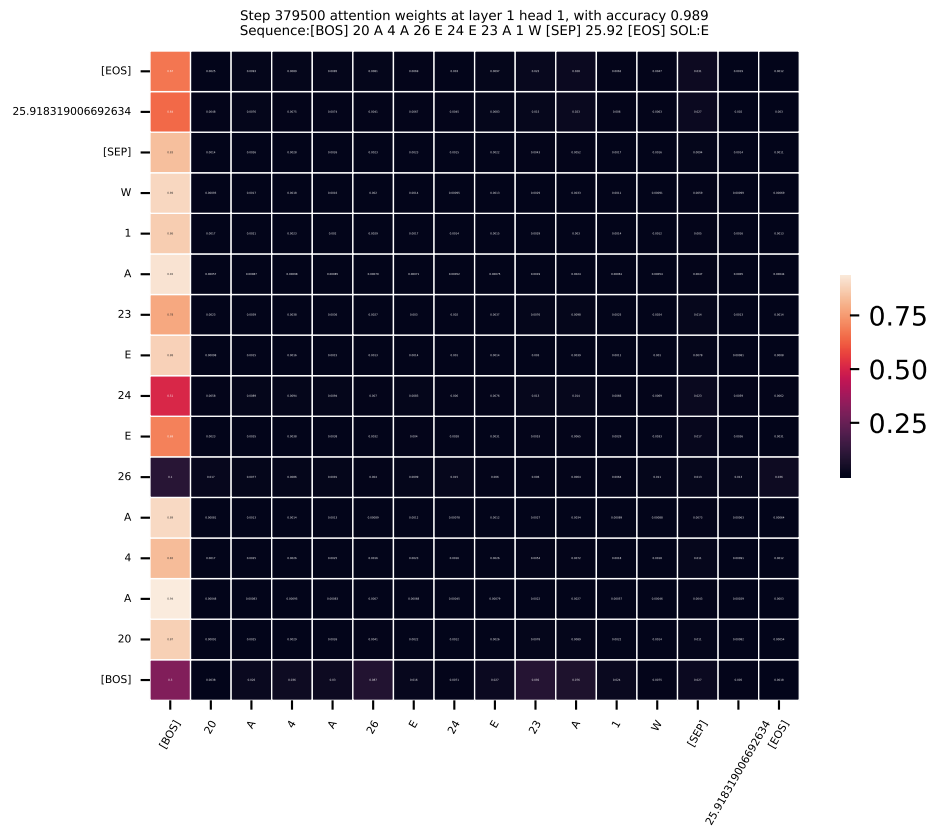

Step 379500 attention weights at layer 0 head 0, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

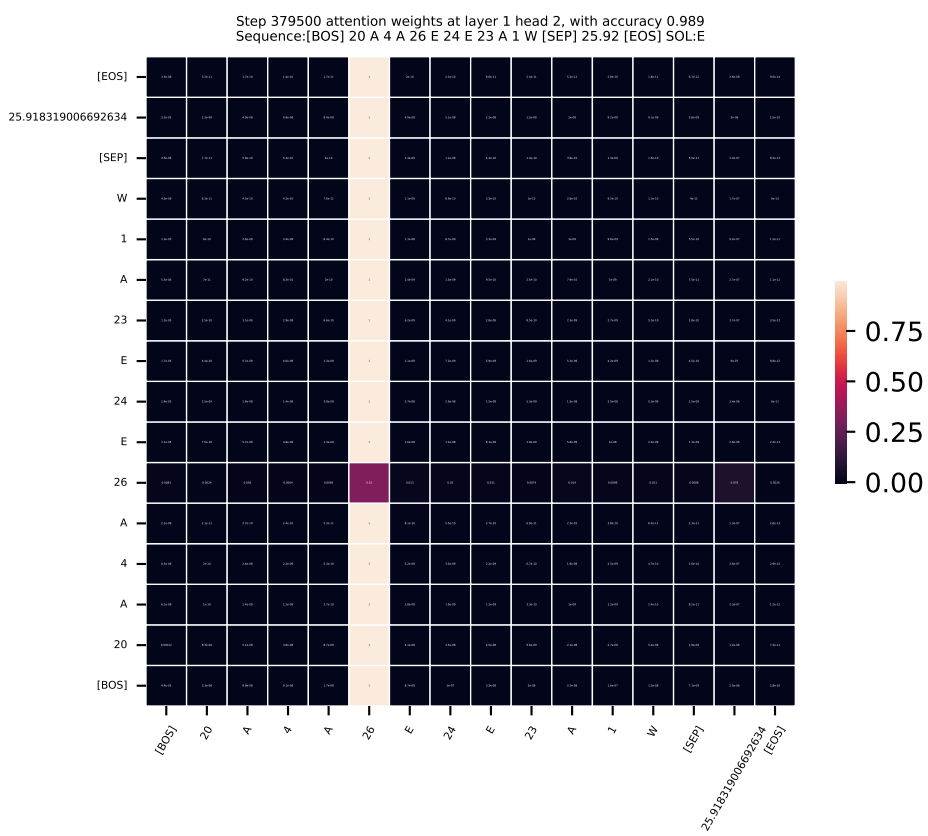

Step 379500 attention weights at layer 0 head 1, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

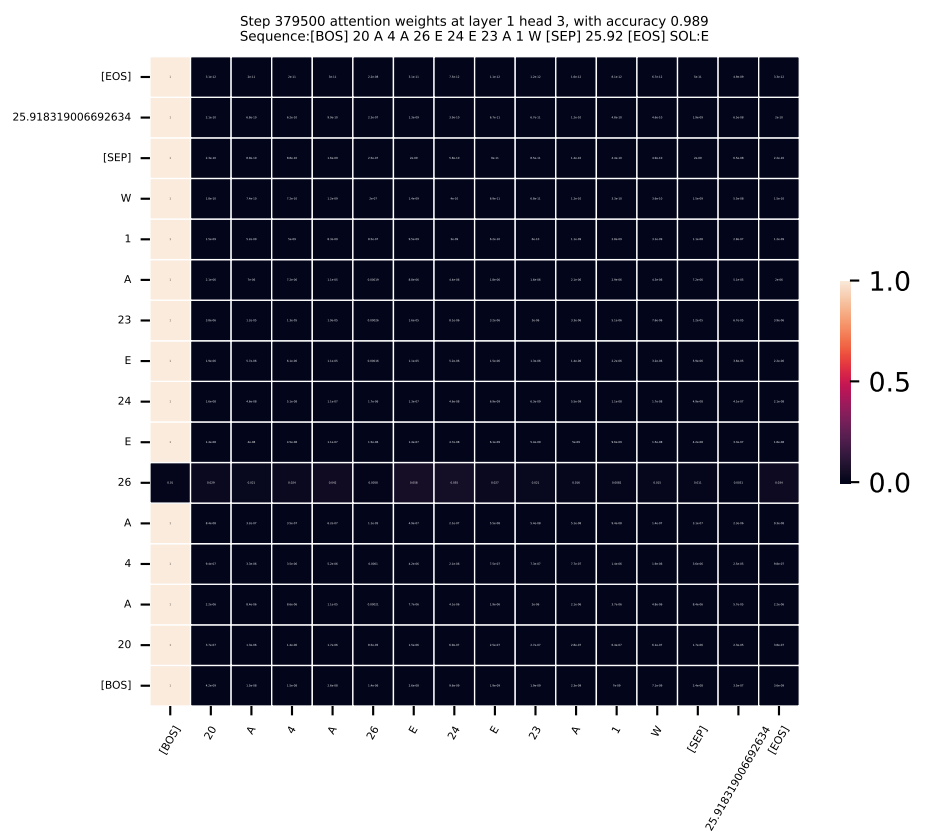

Step 379500 attention weights at layer 0 head 2, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

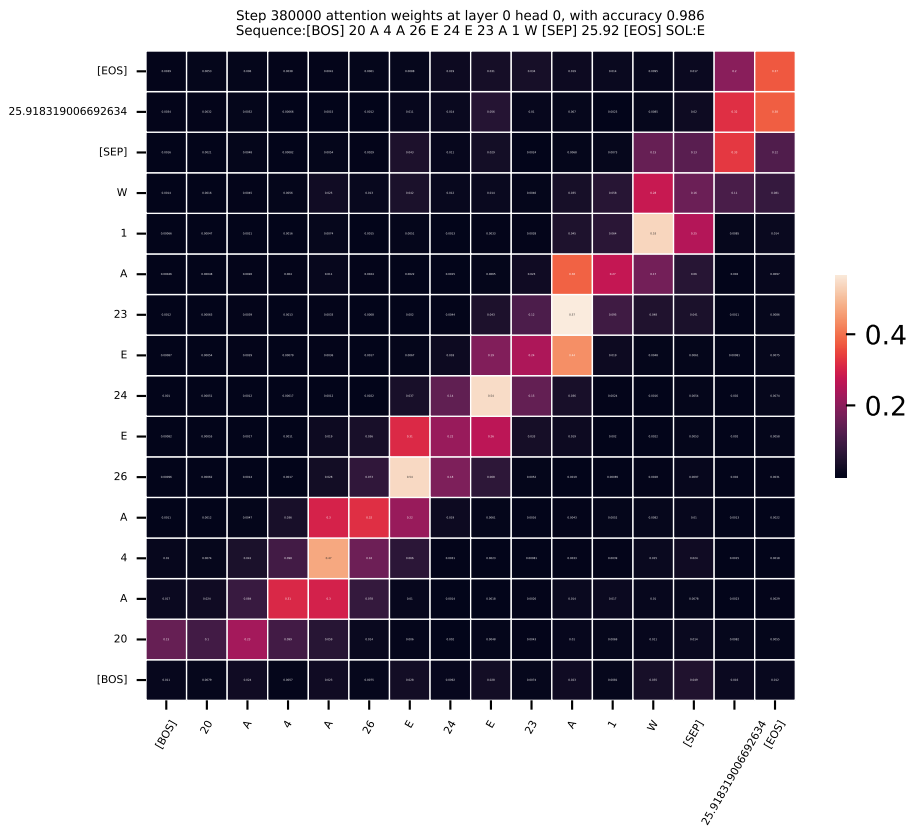

Step 379500 attention weights at layer 0 head 3, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

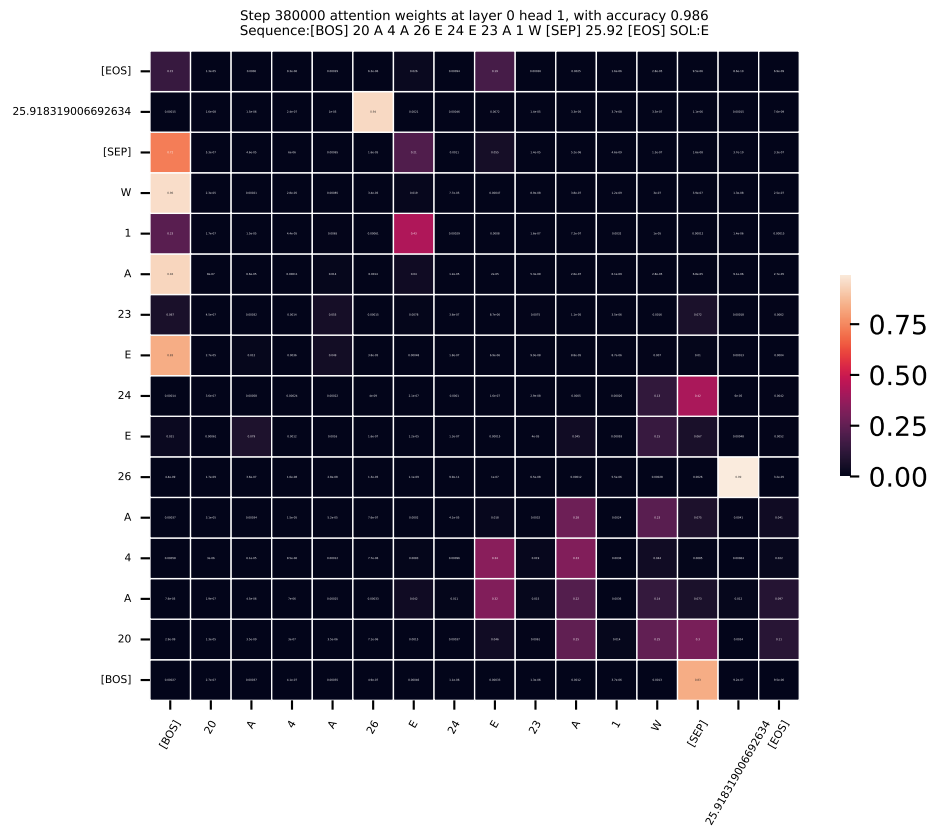

Step 379500 attention weights at layer 1 head 0, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

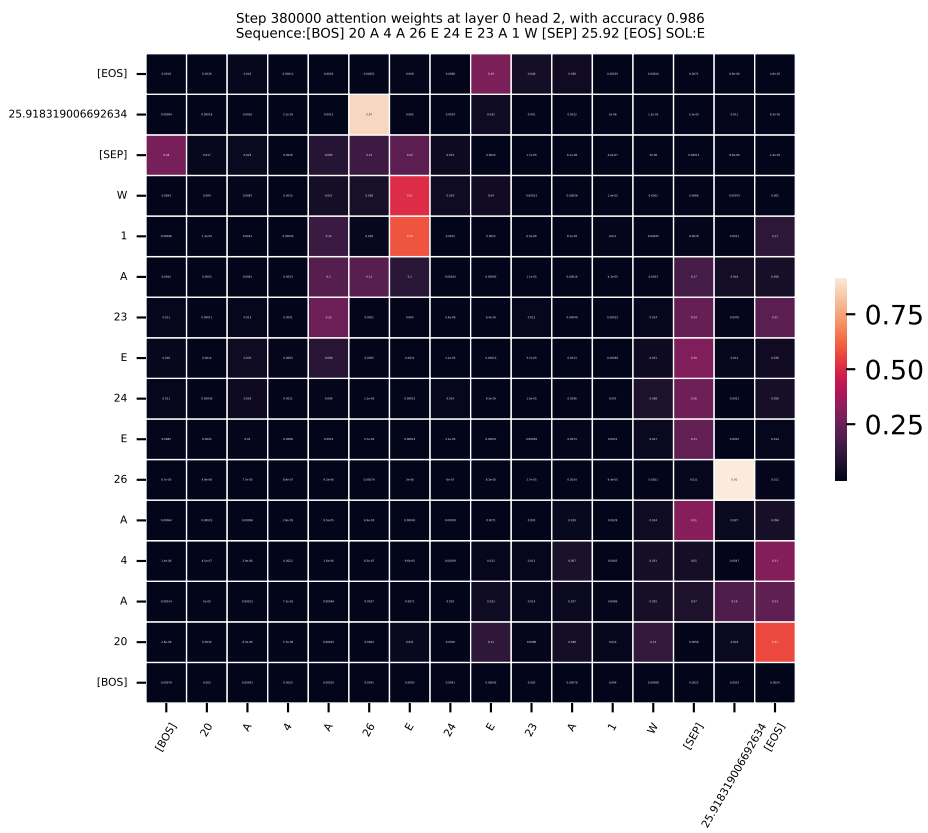

Step 379500 attention weights at layer 1 head 1, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

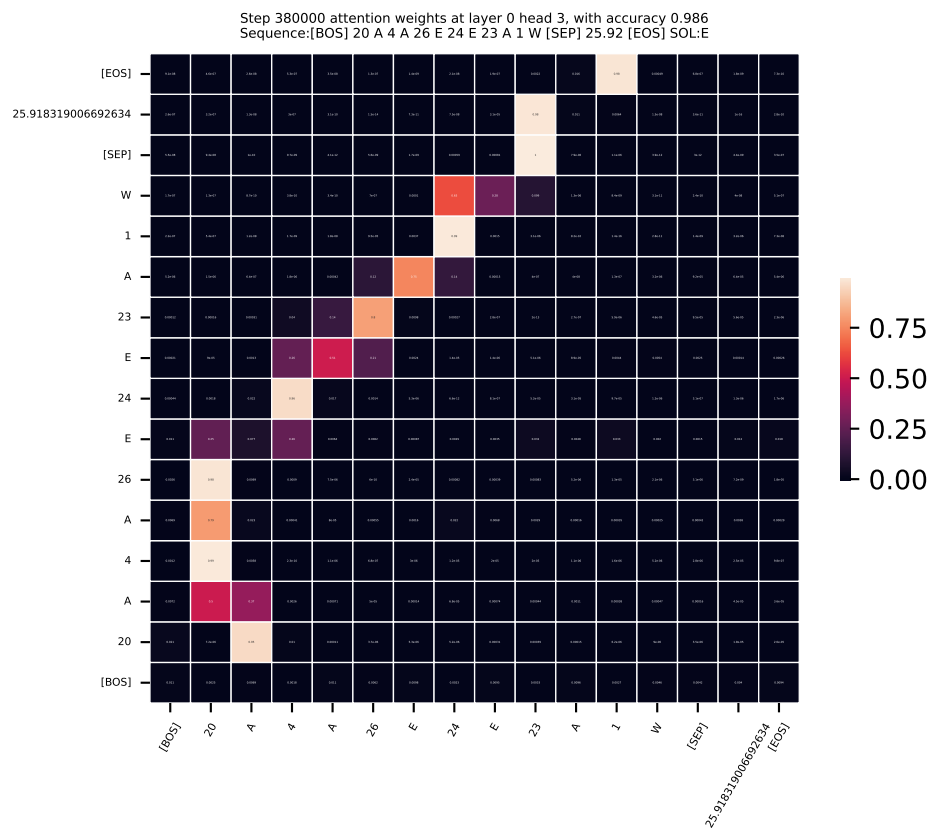

Step 379500 attention weights at layer 1 head 2, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

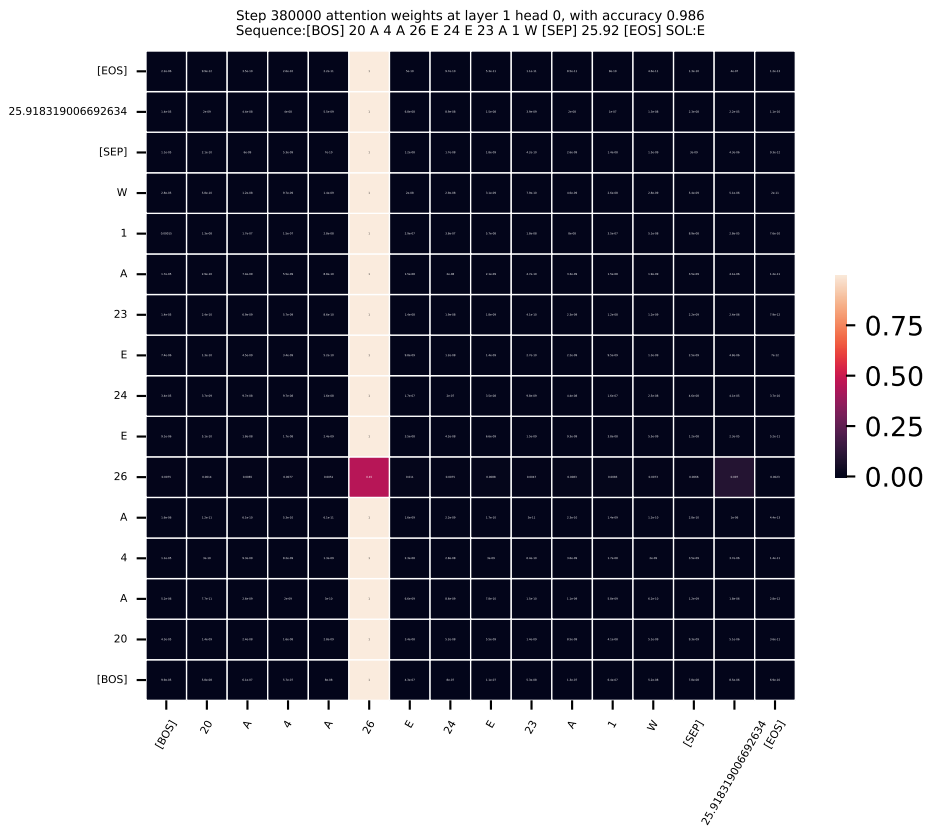

Step 379500 attention weights at layer 1 head 3, with accuracy 0.989
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
1.0
0.5
0.0

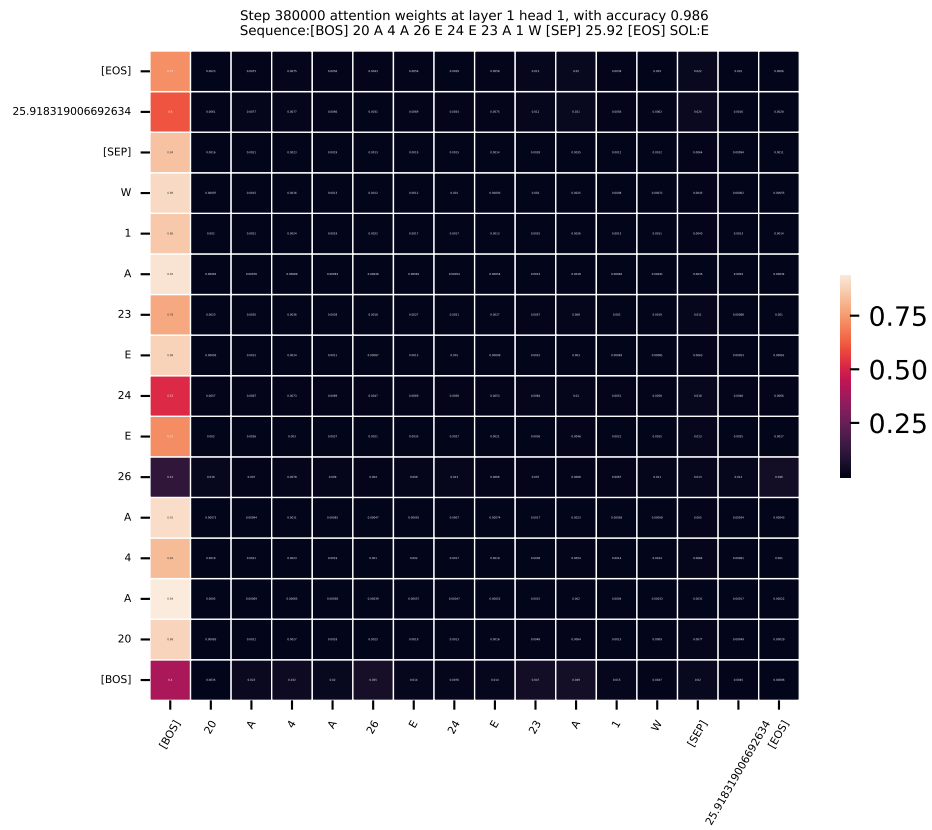

Step 380000 attention weights at layer 0 head 0, with accuracy 0.986
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

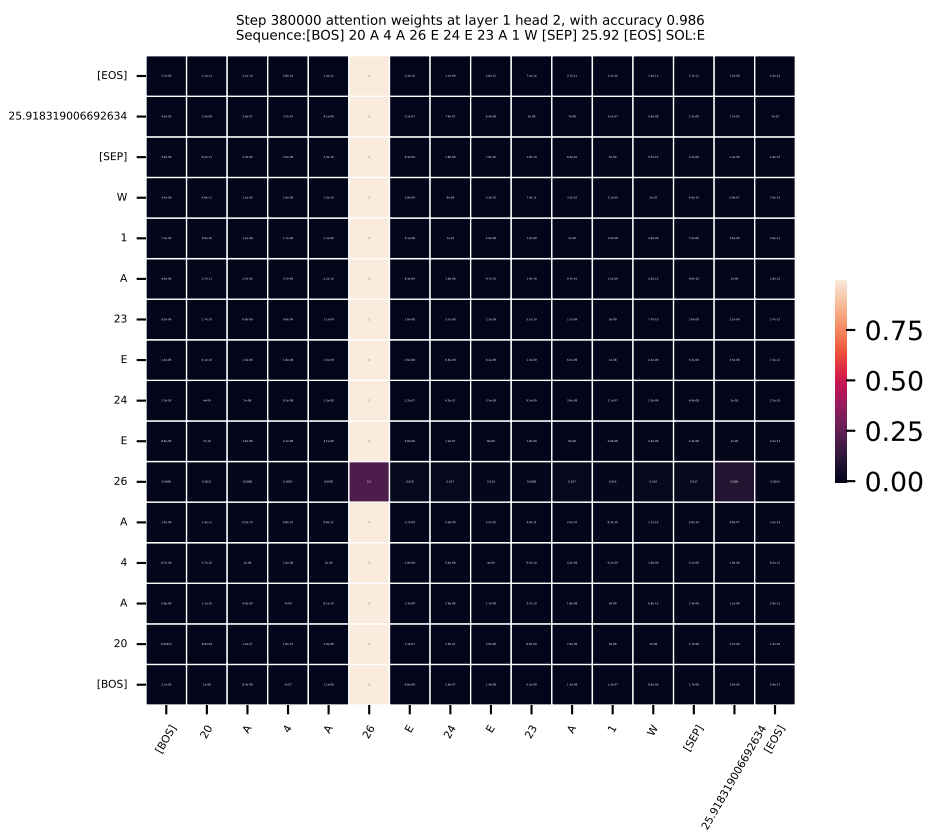

Step 380000 attention weights at layer 0 head 1, with accuracy 0.986
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

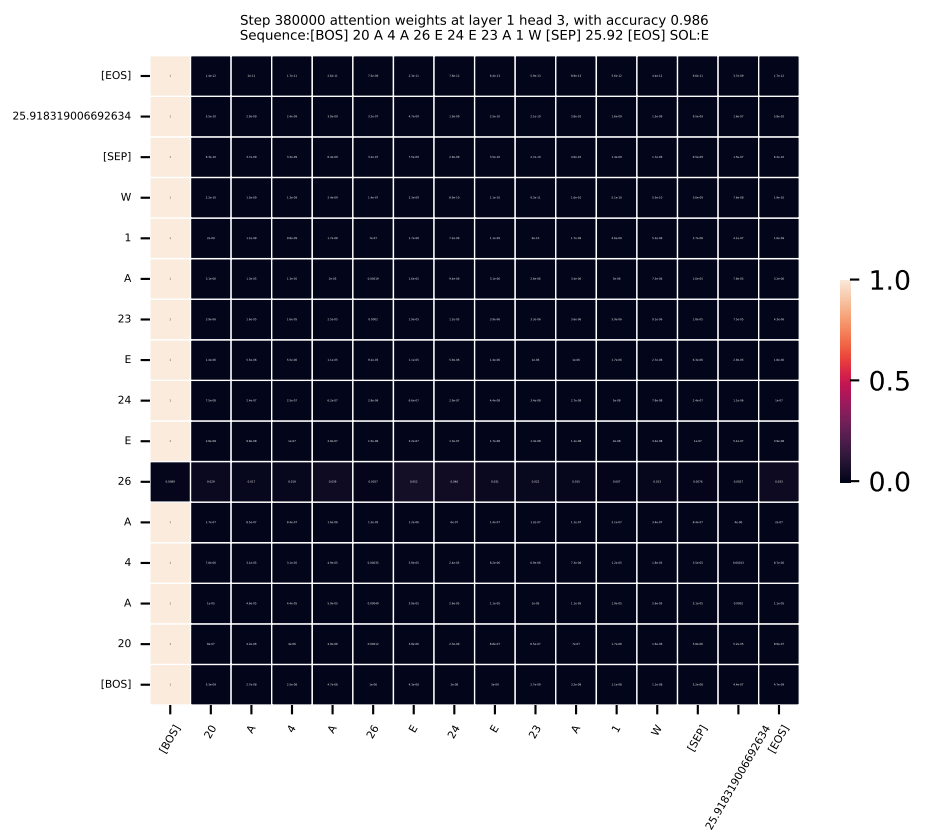

Step 380000 attention weights at layer 0 head 2, with accuracy 0.986
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

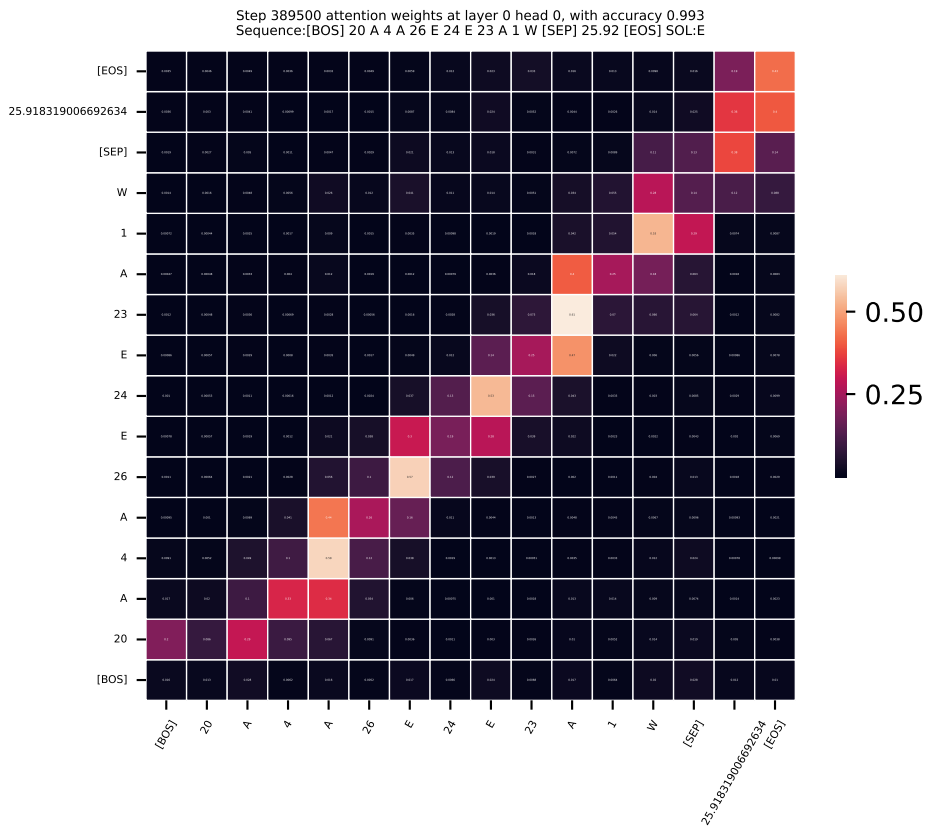

Step 380000 attention weights at layer 0 head 3, with accuracy 0.986
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

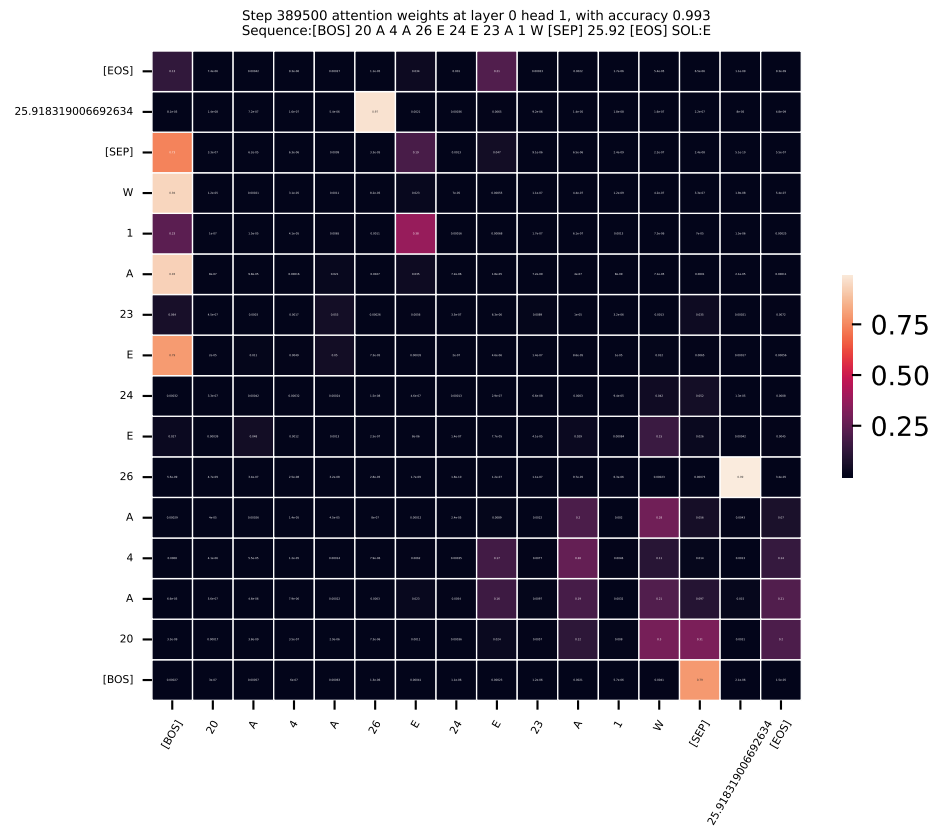

Step 380000 attention weights at layer 1 head 0, with accuracy 0.986
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

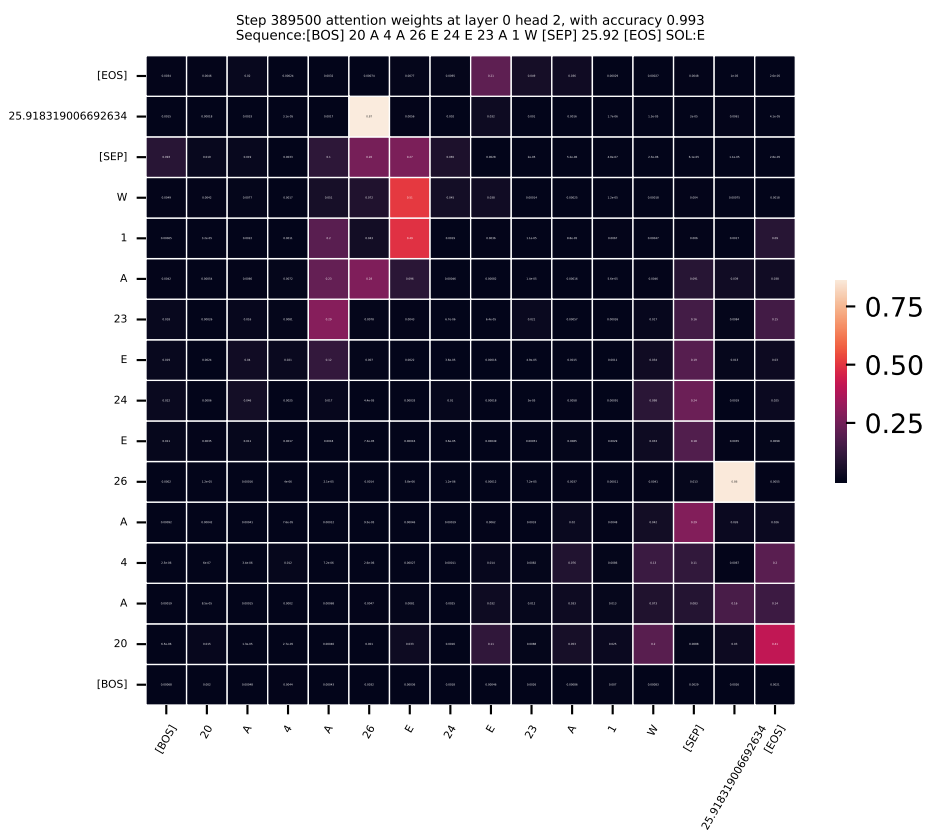

Step 380000 attention weights at layer 1 head 1, with accuracy 0.986
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

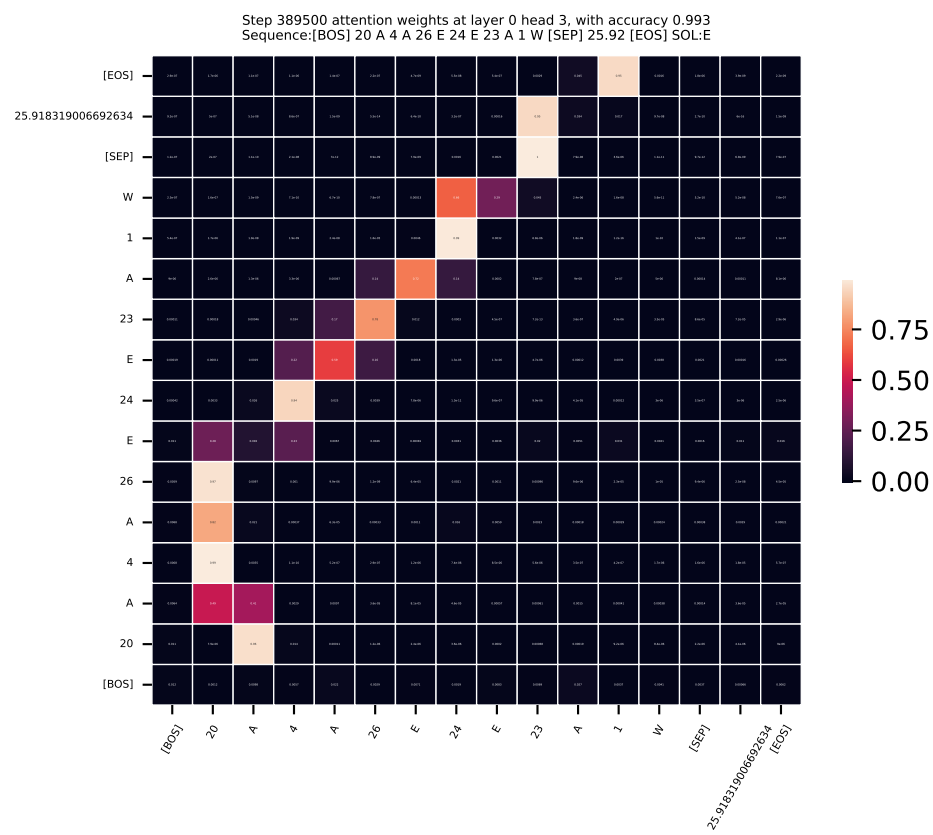

Step 380000 attention weights at layer 1 head 2, with accuracy 0.986
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

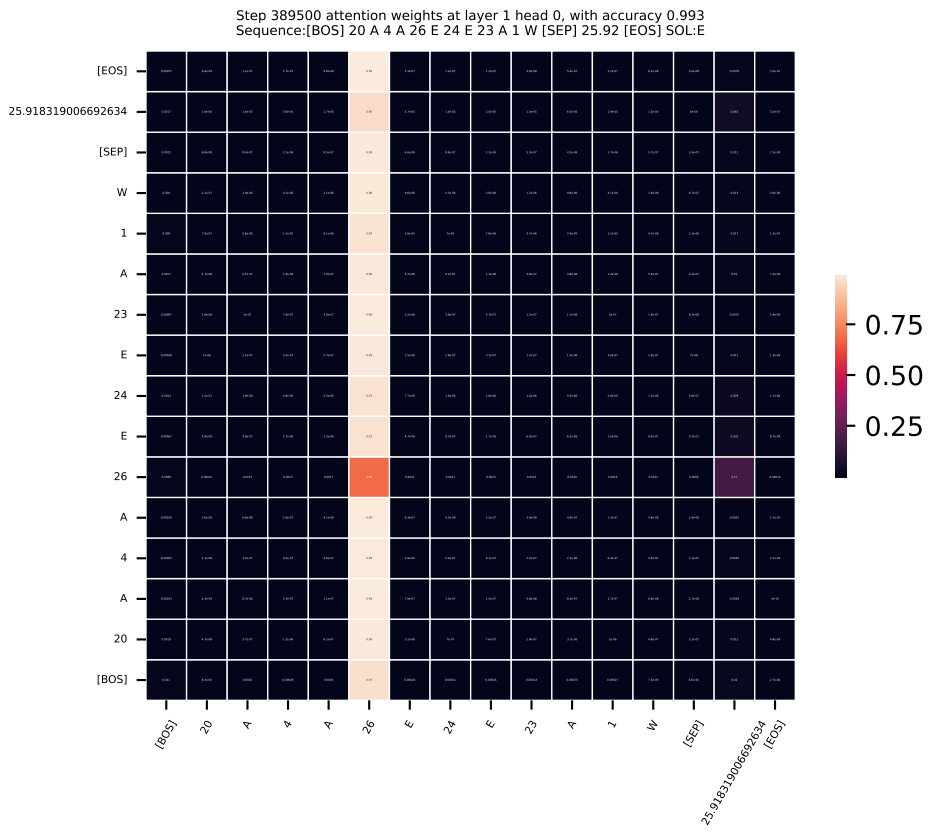

Step 380000 attention weights at layer 1 head 3, with accuracy 0.986
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
1.0
0.5
0.0

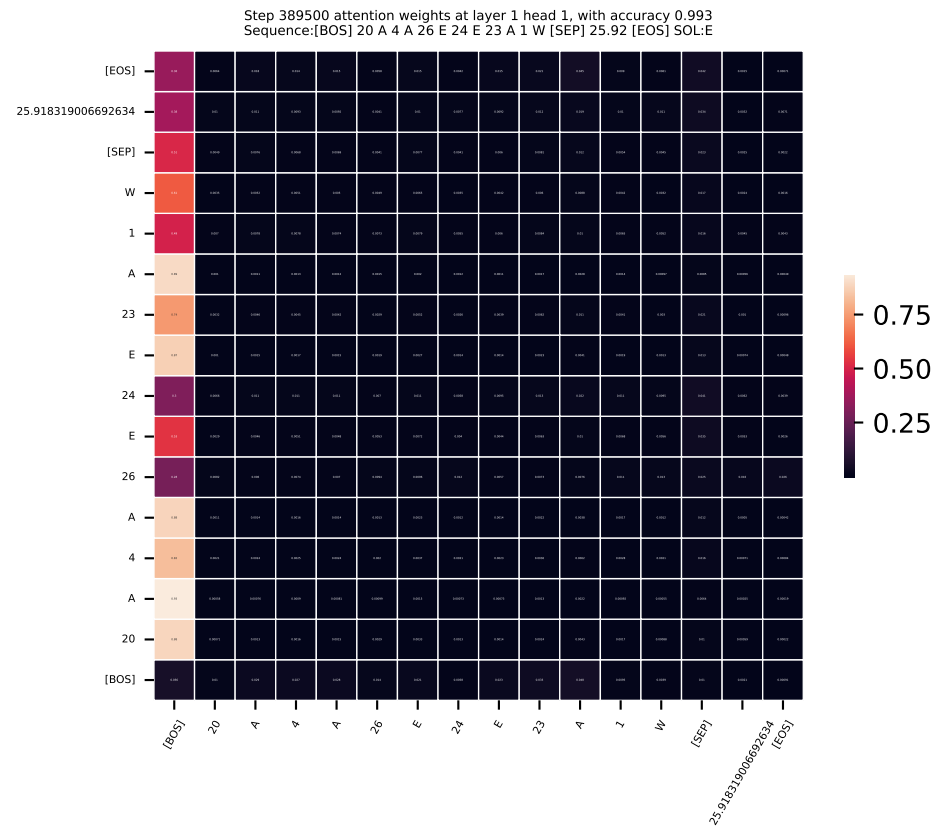

Step 389500 attention weights at layer 0 head 0, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

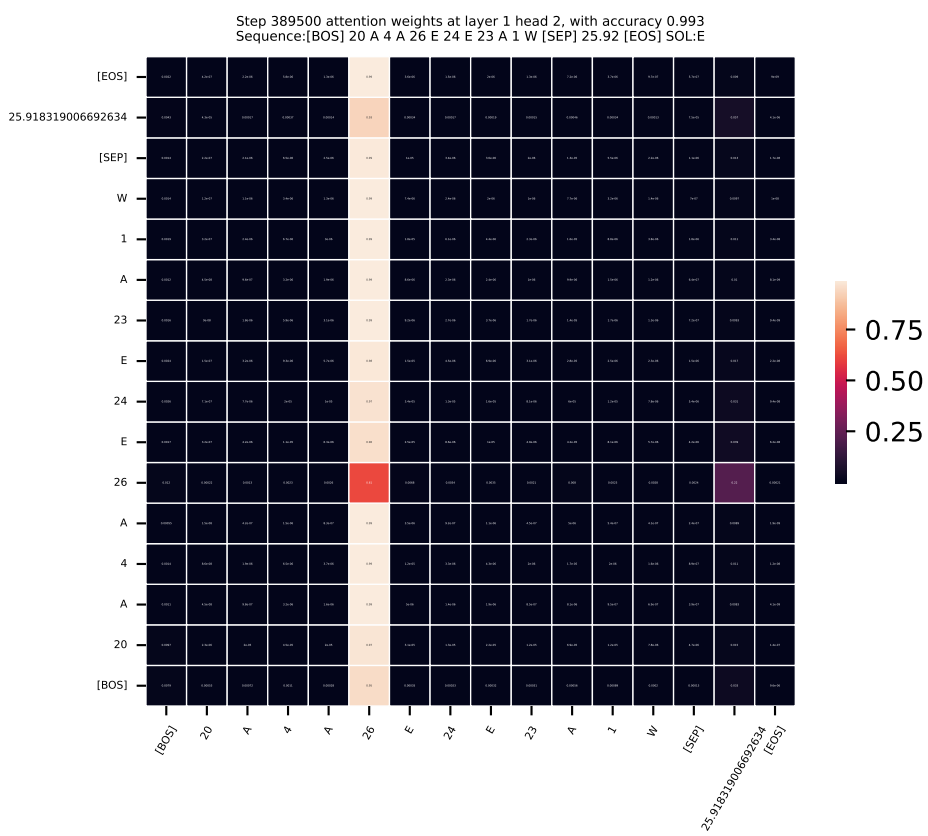

Step 389500 attention weights at layer 0 head 1, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

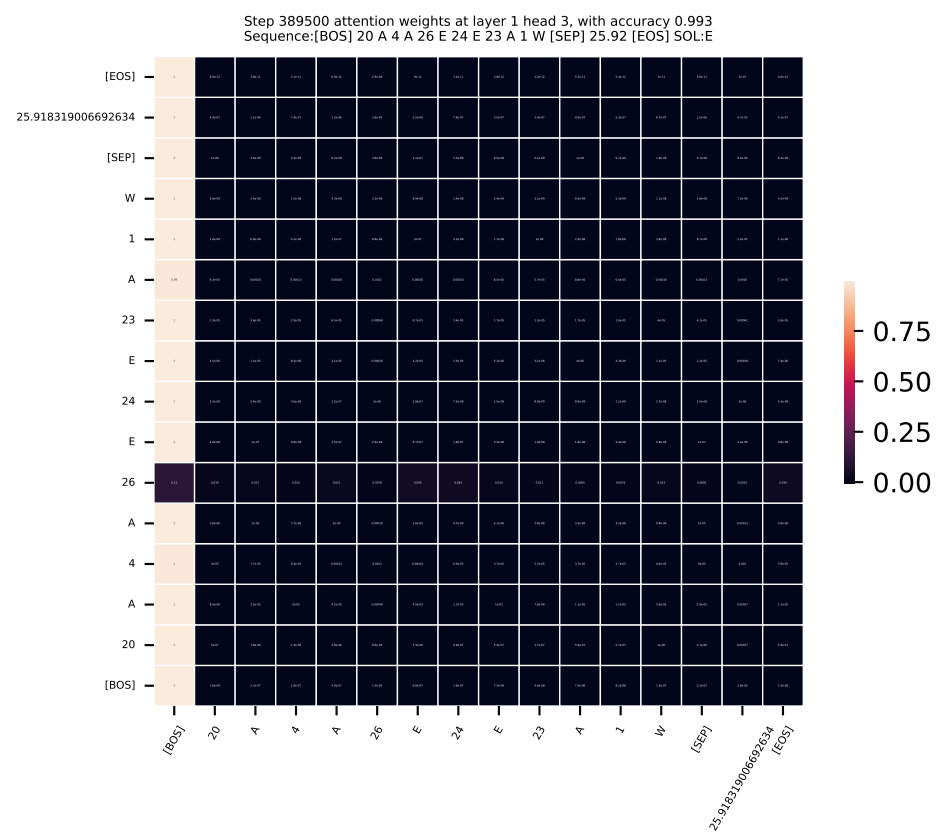

Step 389500 attention weights at layer 0 head 2, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

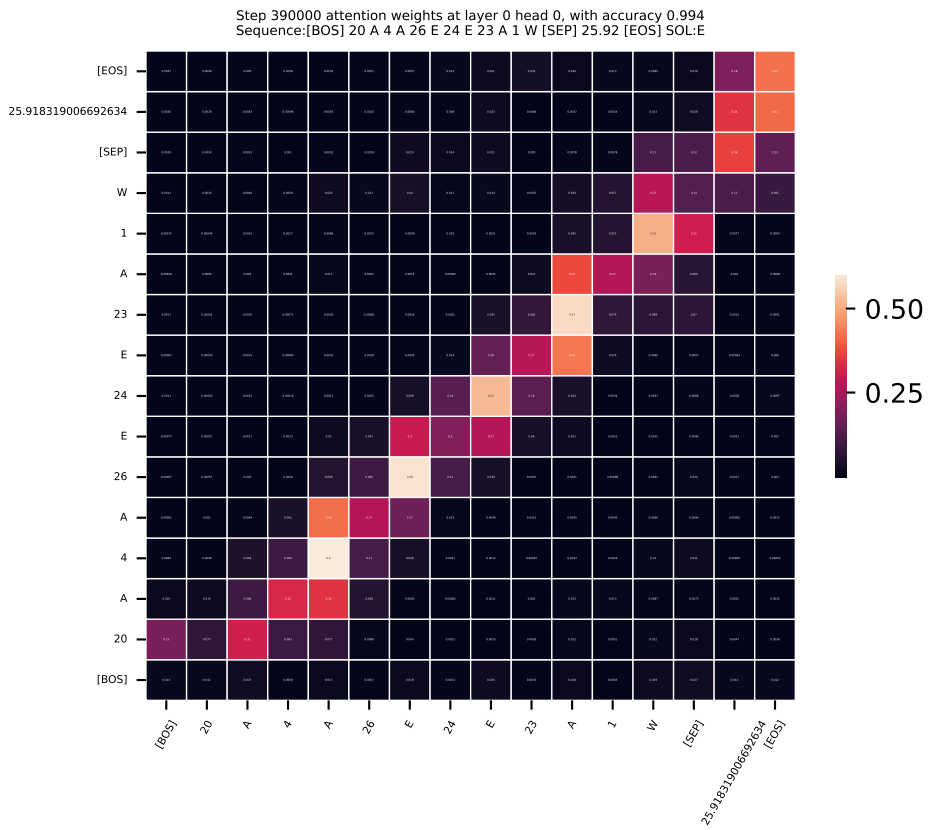

Step 389500 attention weights at layer 0 head 3, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

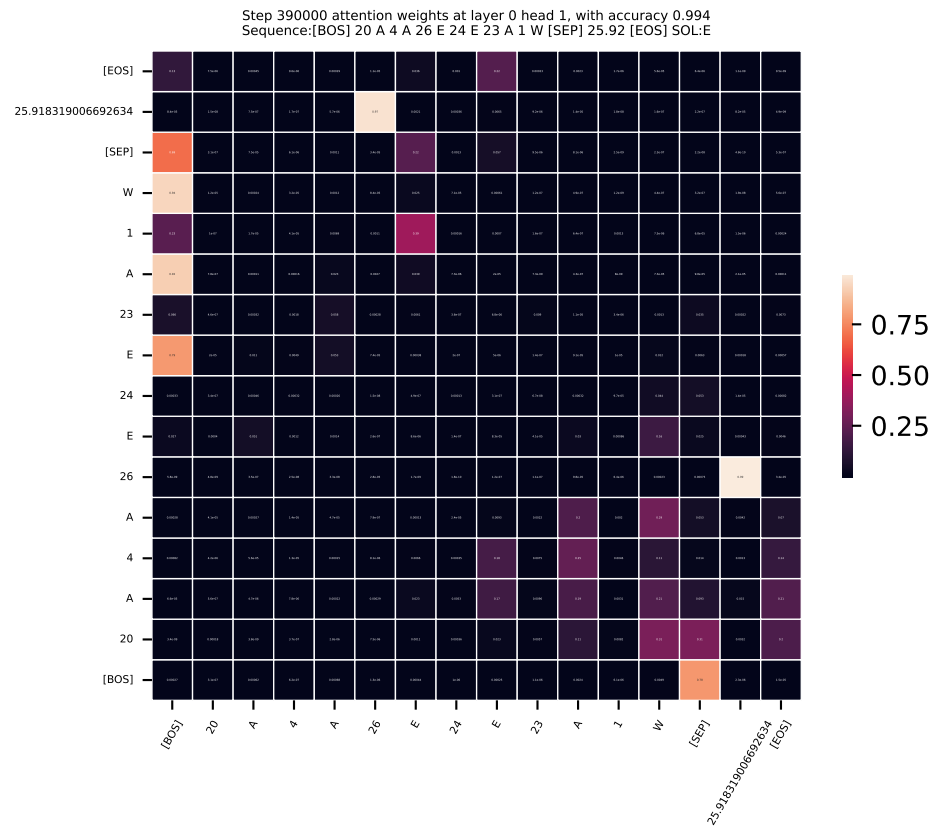

Step 389500 attention weights at layer 1 head 0, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

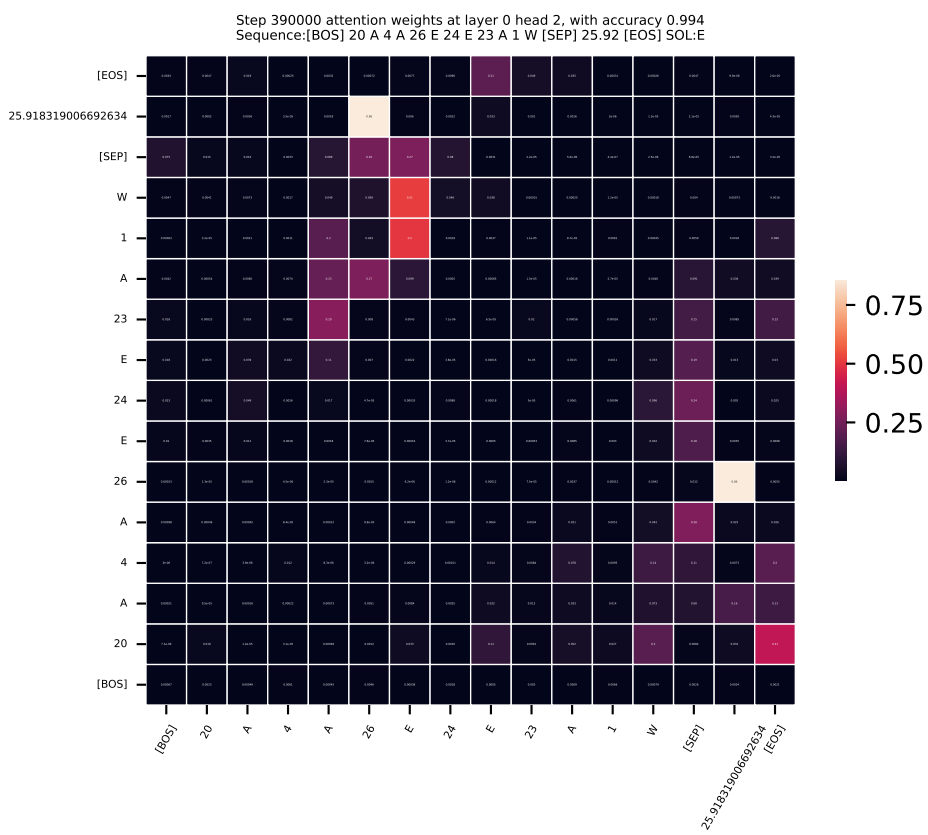

Step 389500 attention weights at layer 1 head 1, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

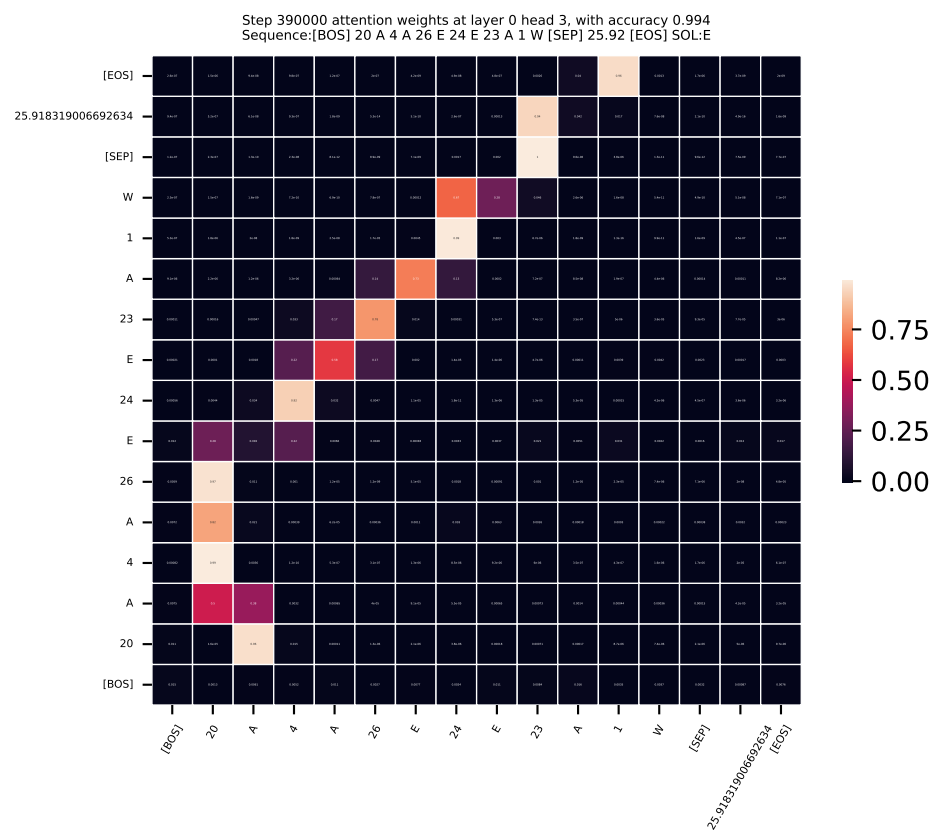

Step 389500 attention weights at layer 1 head 2, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

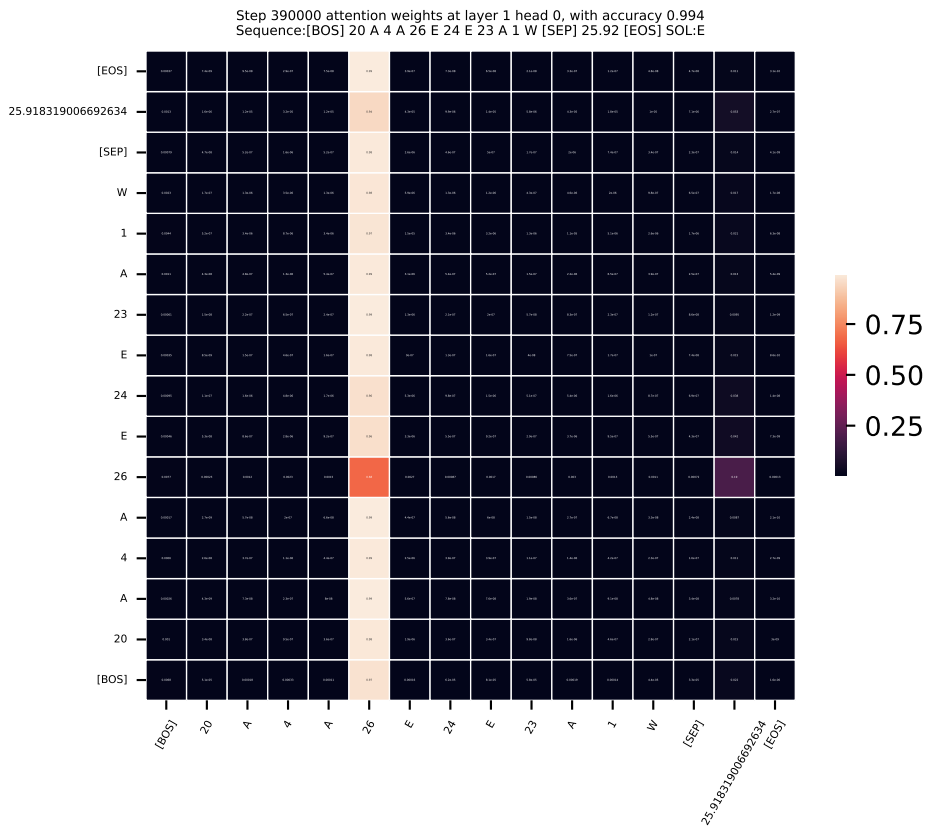

Step 389500 attention weights at layer 1 head 3, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

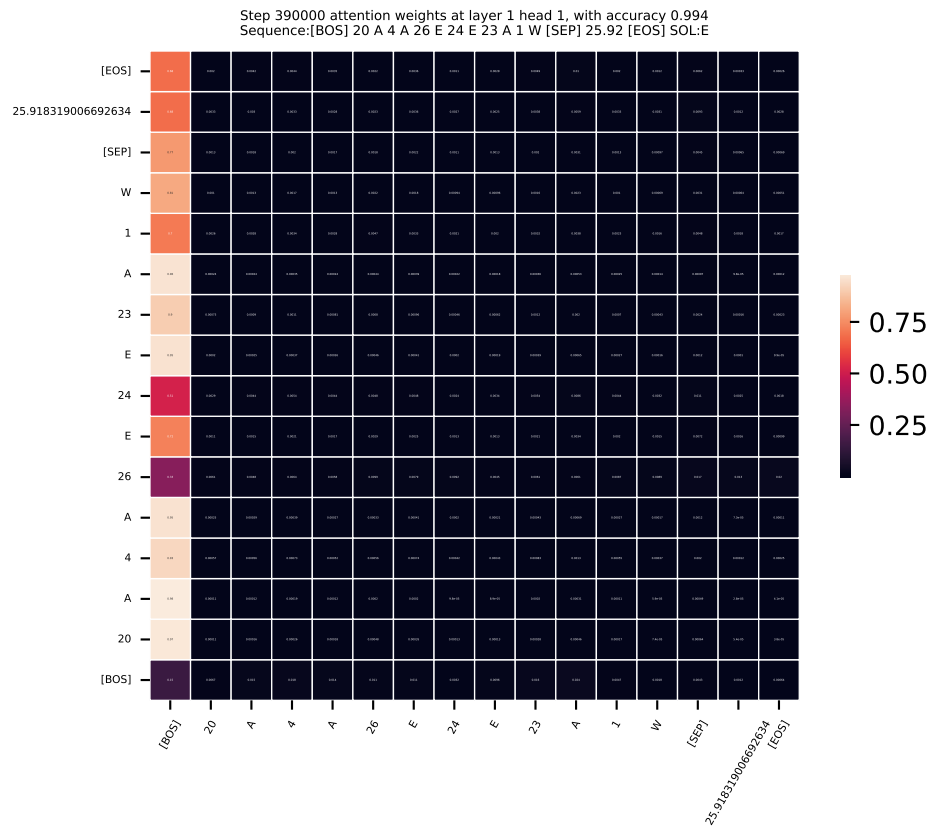

Step 390000 attention weights at layer 0 head 0, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

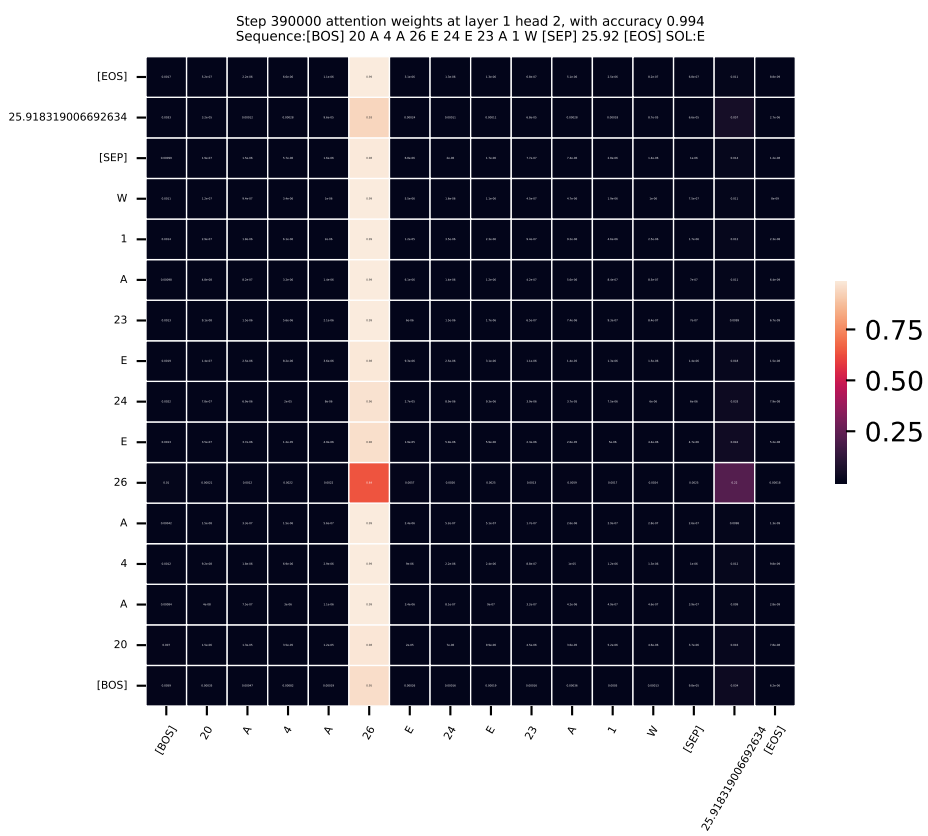

Step 390000 attention weights at layer 0 head 1, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

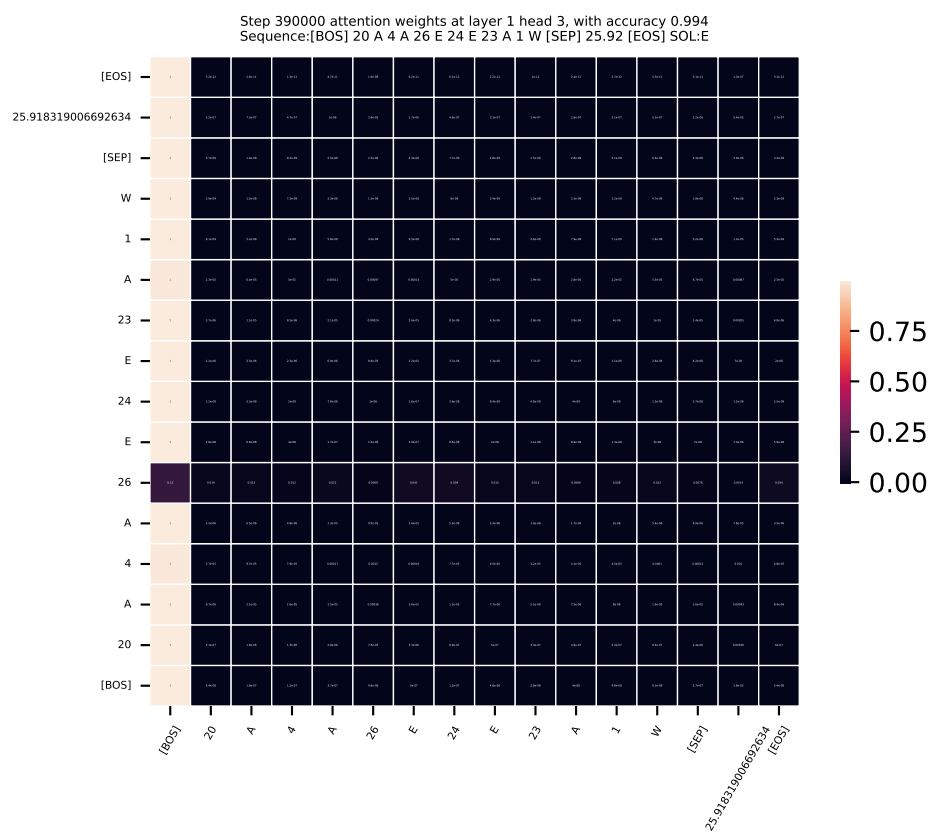

Step 390000 attention weights at layer 0 head 2, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

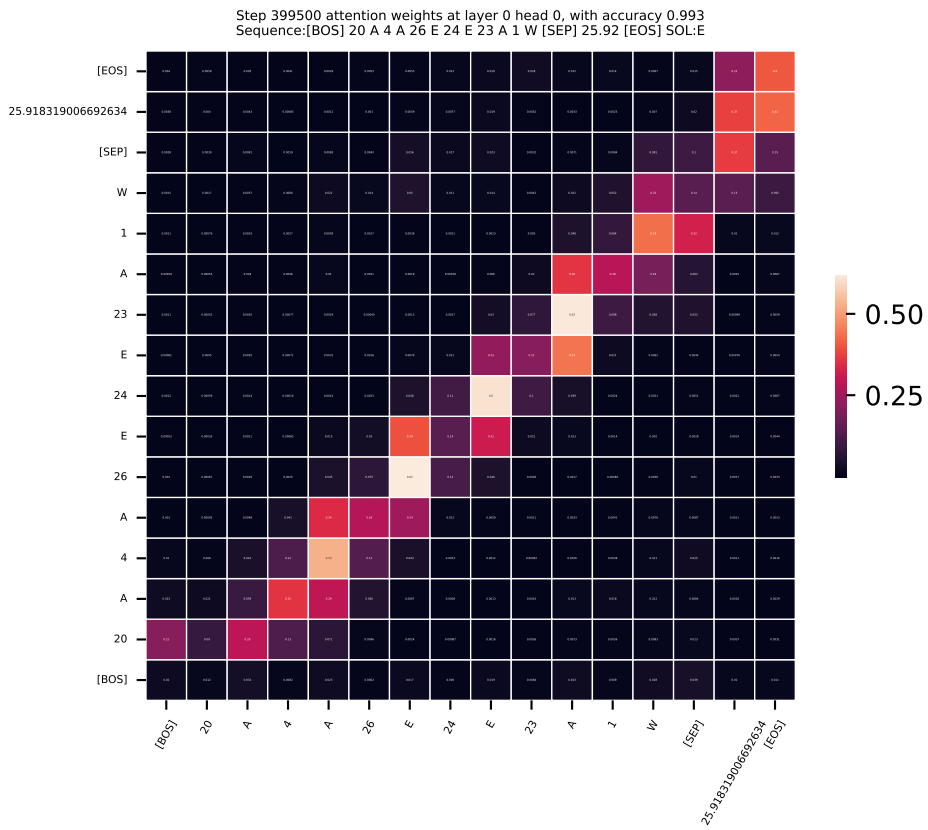

Step 390000 attention weights at layer 0 head 3, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

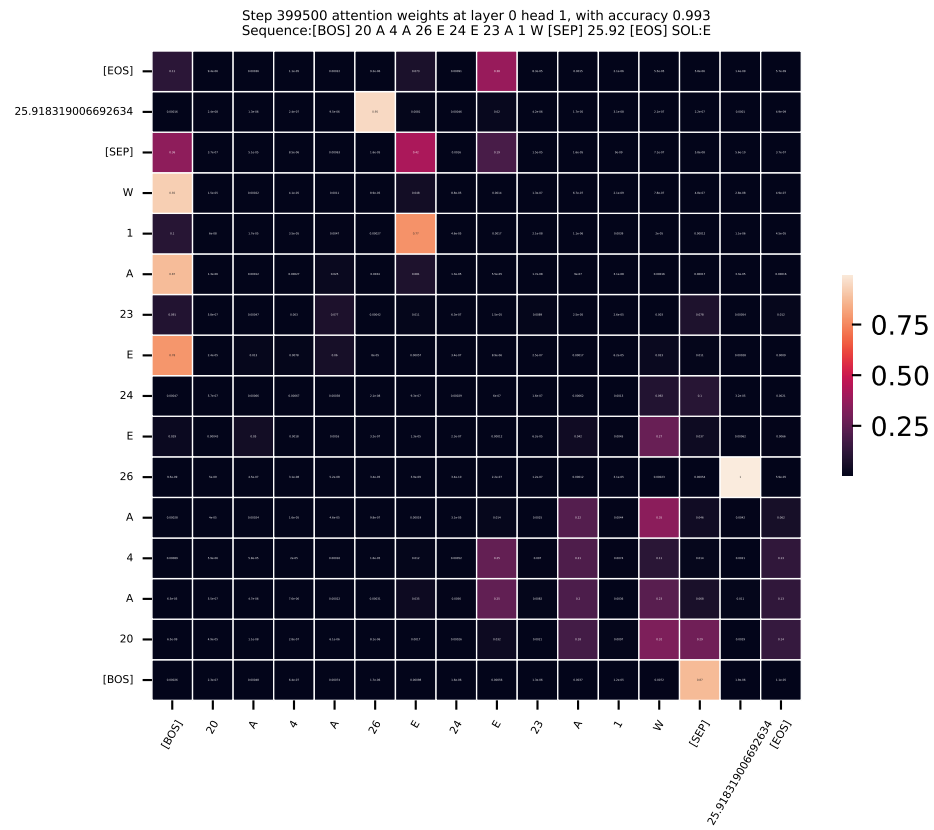

Step 390000 attention weights at layer 1 head 0, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

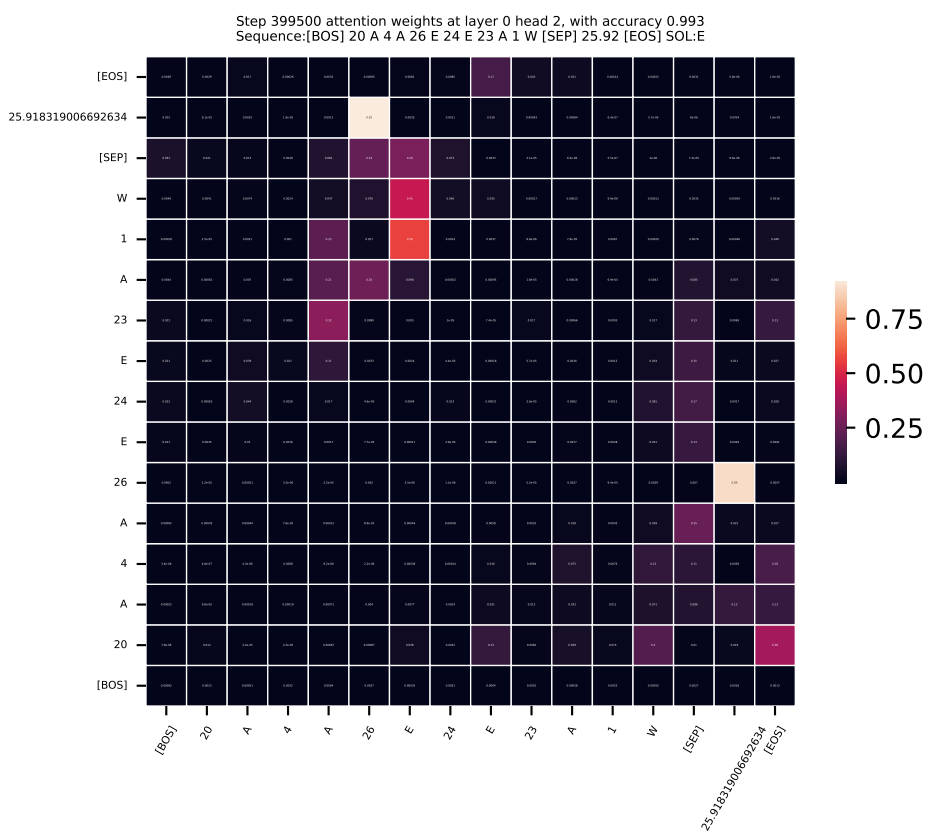

Step 390000 attention weights at layer 1 head 1, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

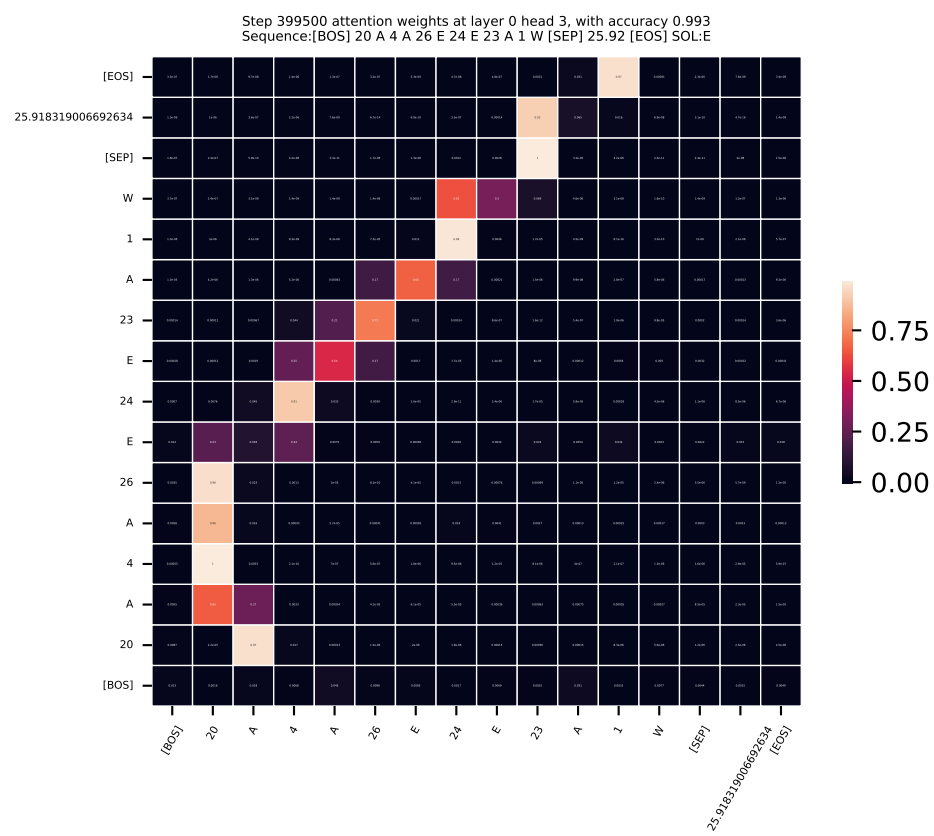

Step 390000 attention weights at layer 1 head 2, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

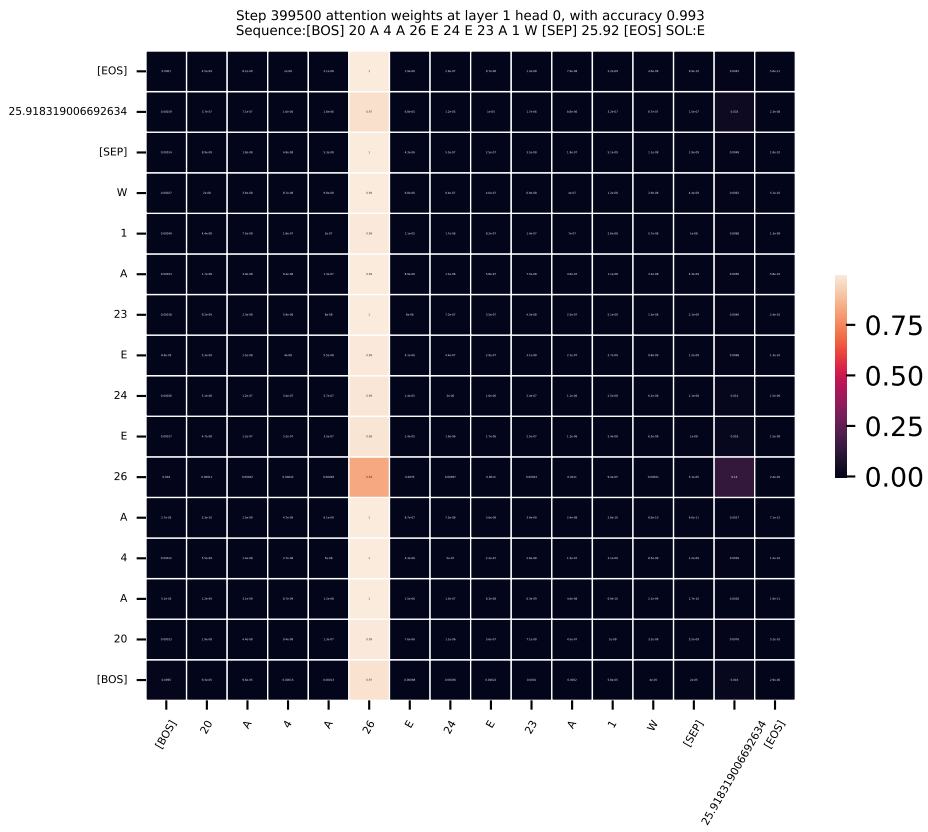

Step 390000 attention weights at layer 1 head 3, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

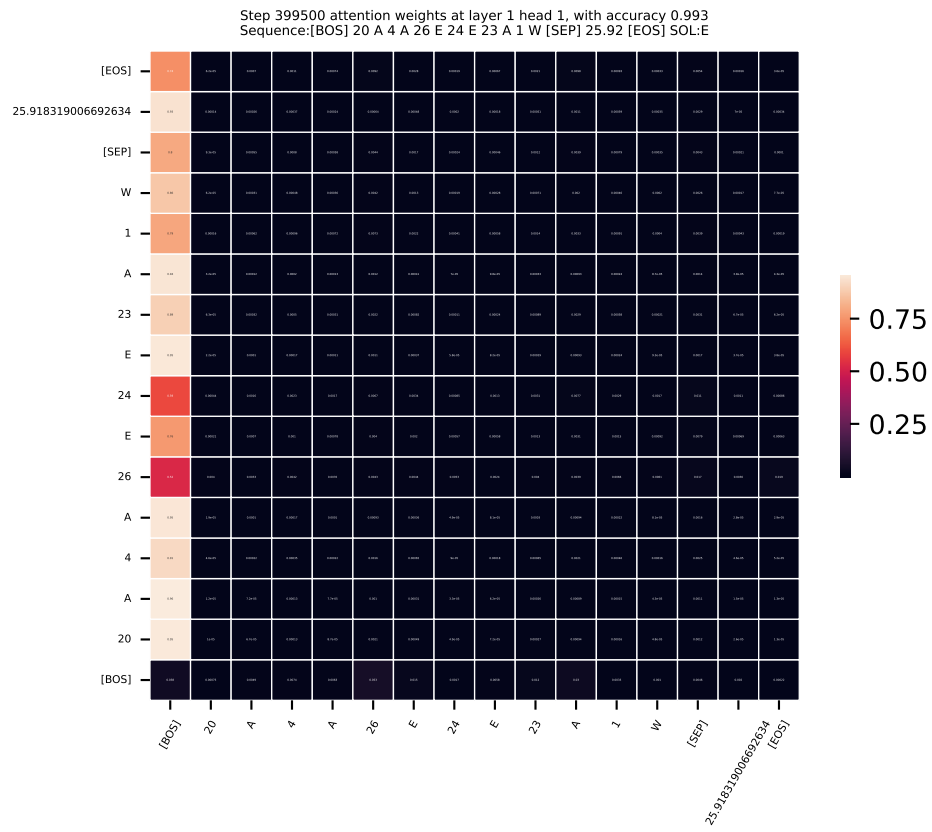

Step 399500 attention weights at layer 0 head 0, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

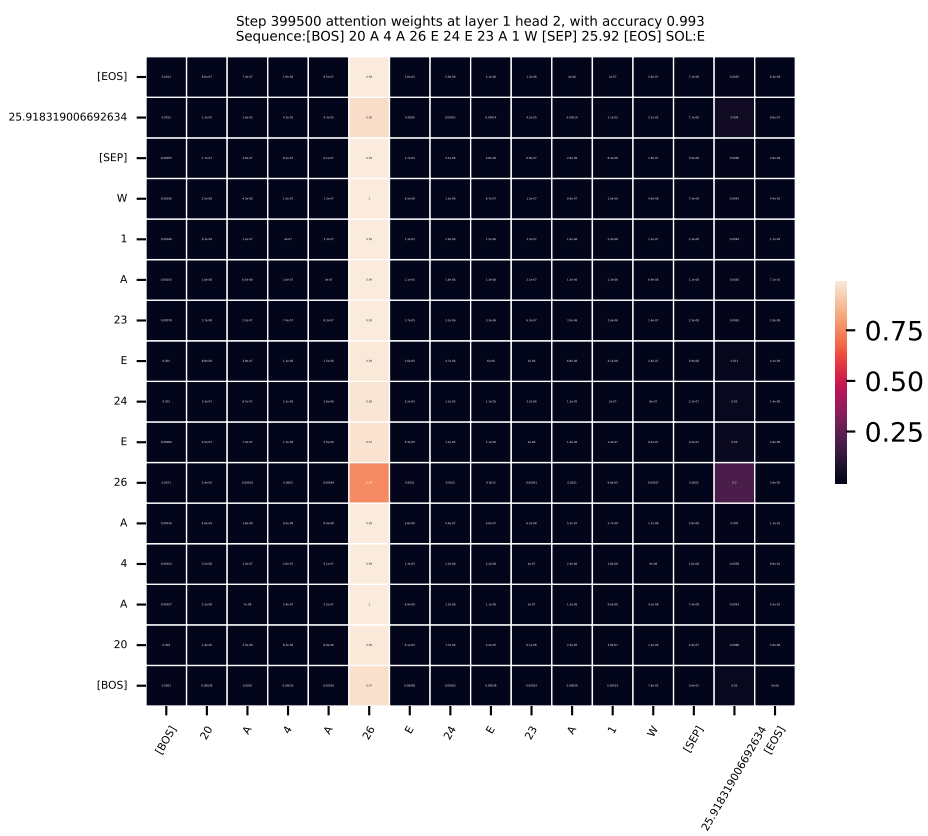

Step 399500 attention weights at layer 0 head 1, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

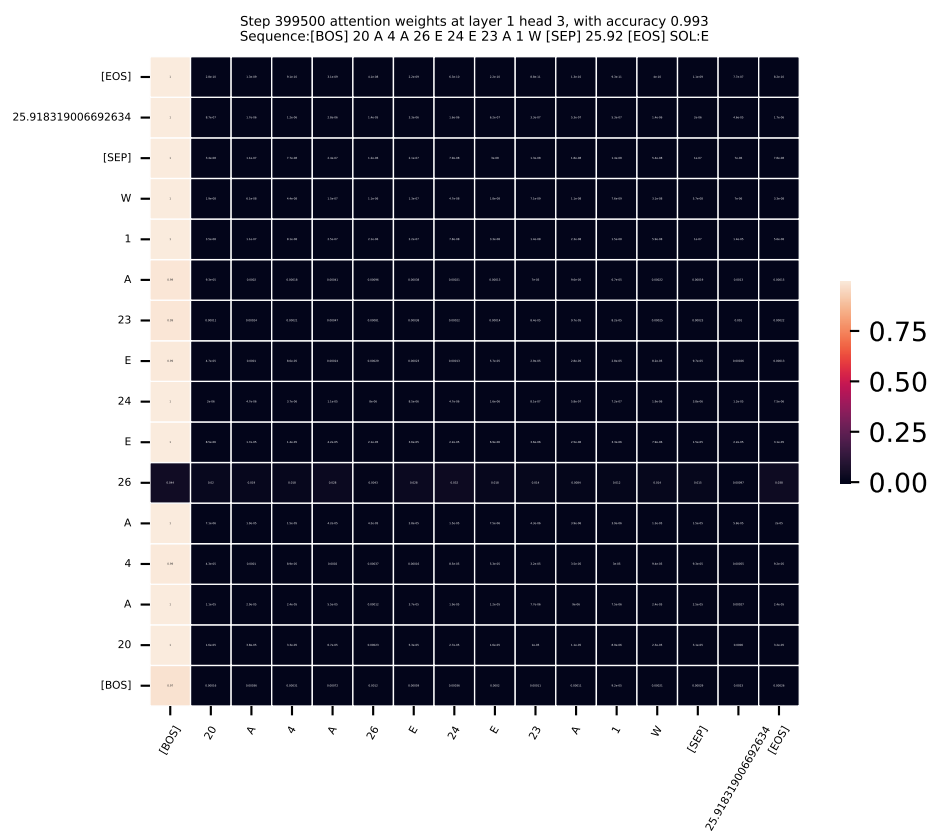

Step 399500 attention weights at layer 0 head 2, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

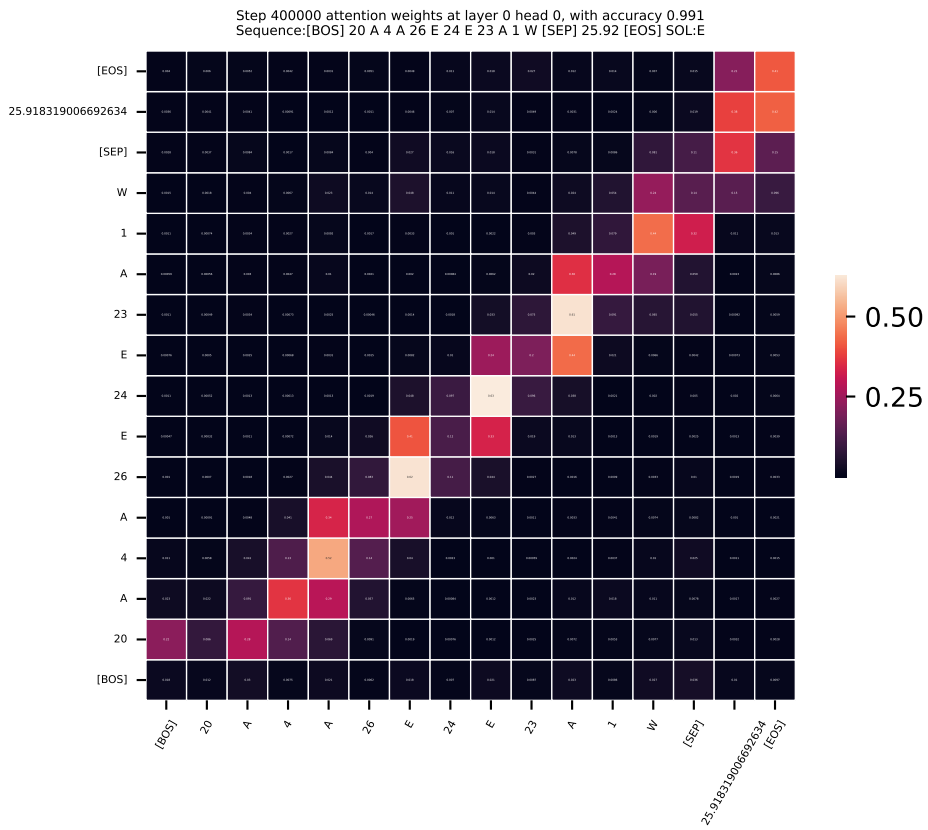

Step 399500 attention weights at layer 0 head 3, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

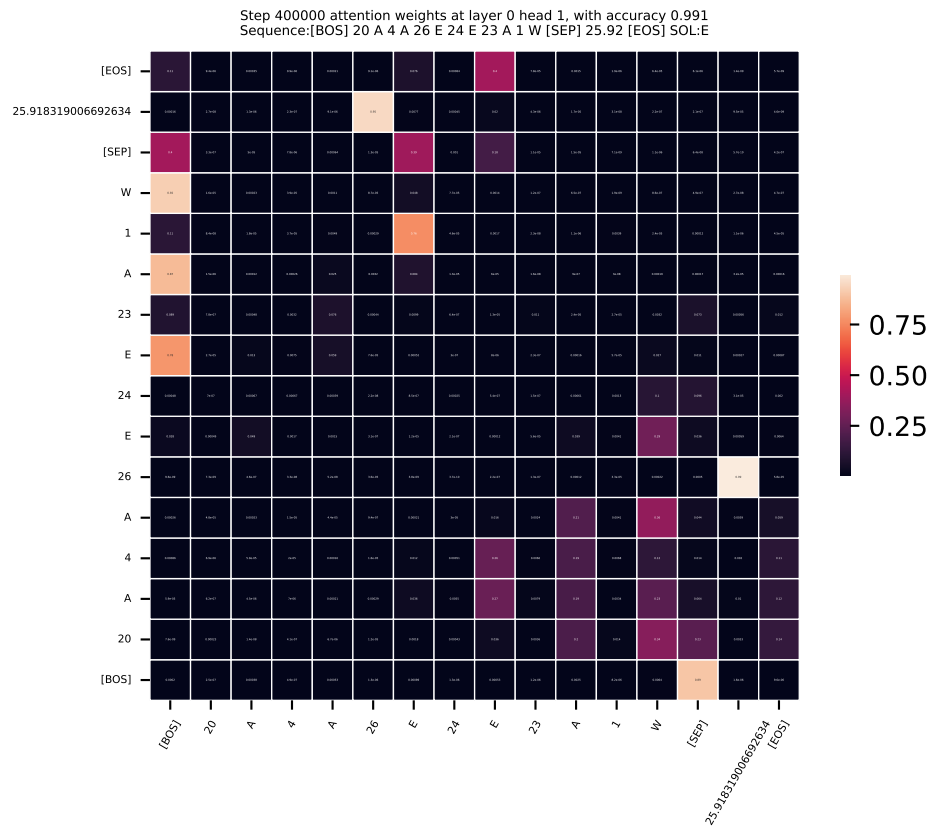

Step 399500 attention weights at layer 1 head 0, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

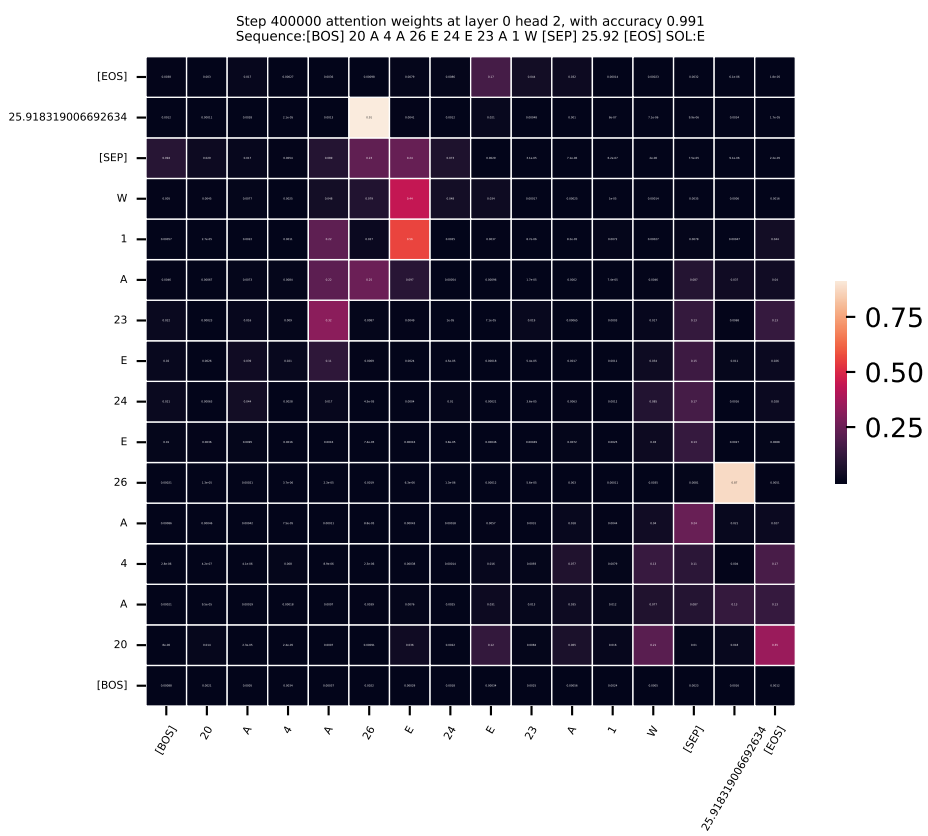

Step 399500 attention weights at layer 1 head 1, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

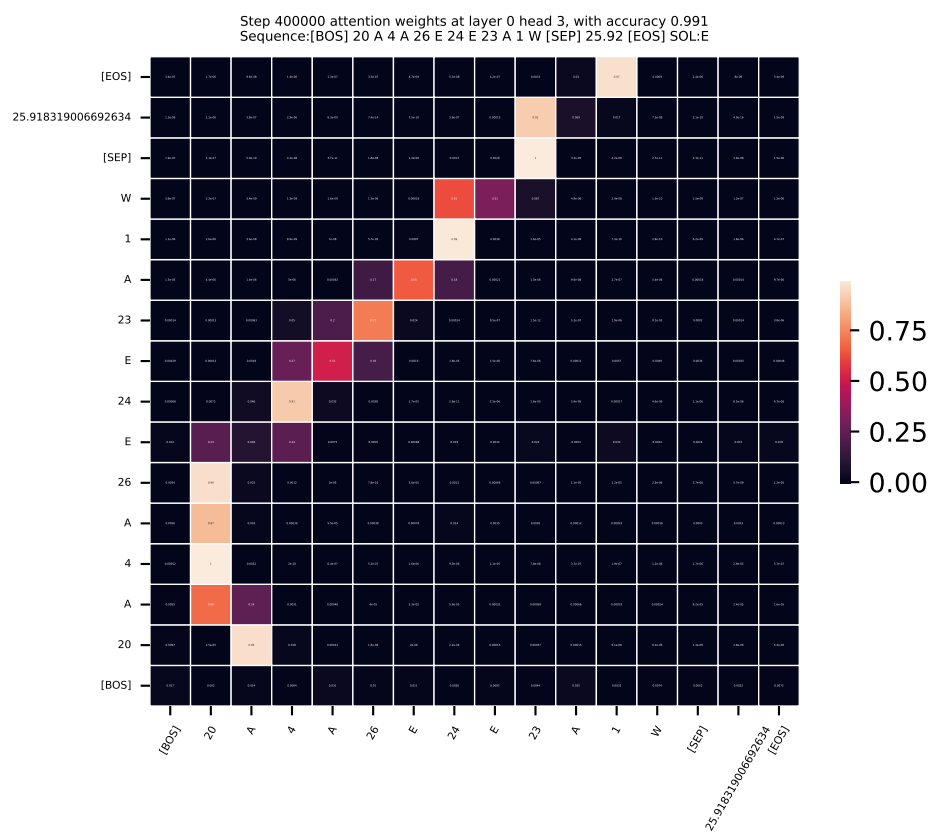

Step 399500 attention weights at layer 1 head 2, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

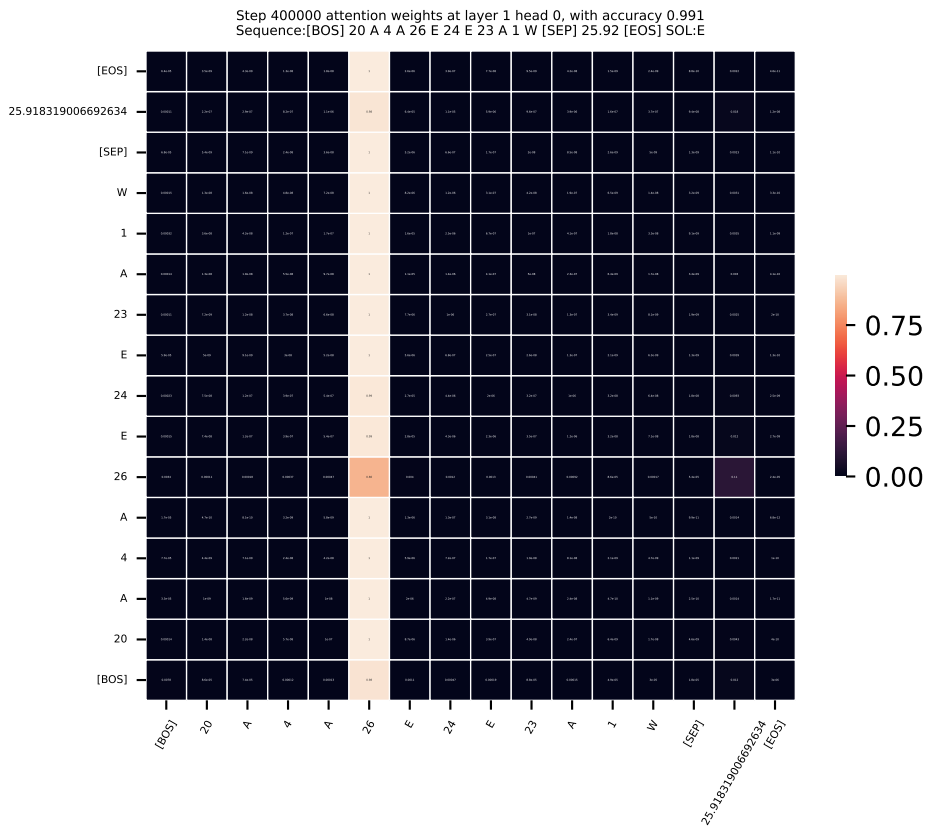

Step 399500 attention weights at layer 1 head 3, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

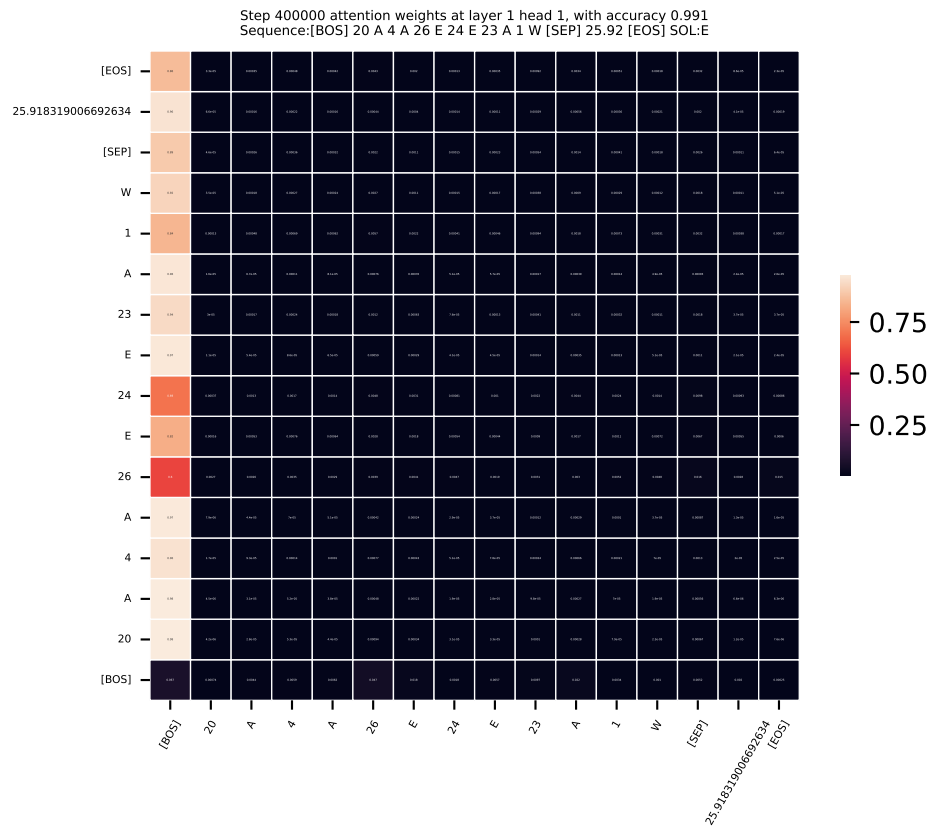

Step 400000 attention weights at layer 0 head 0, with accuracy 0.991
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.50
0.25

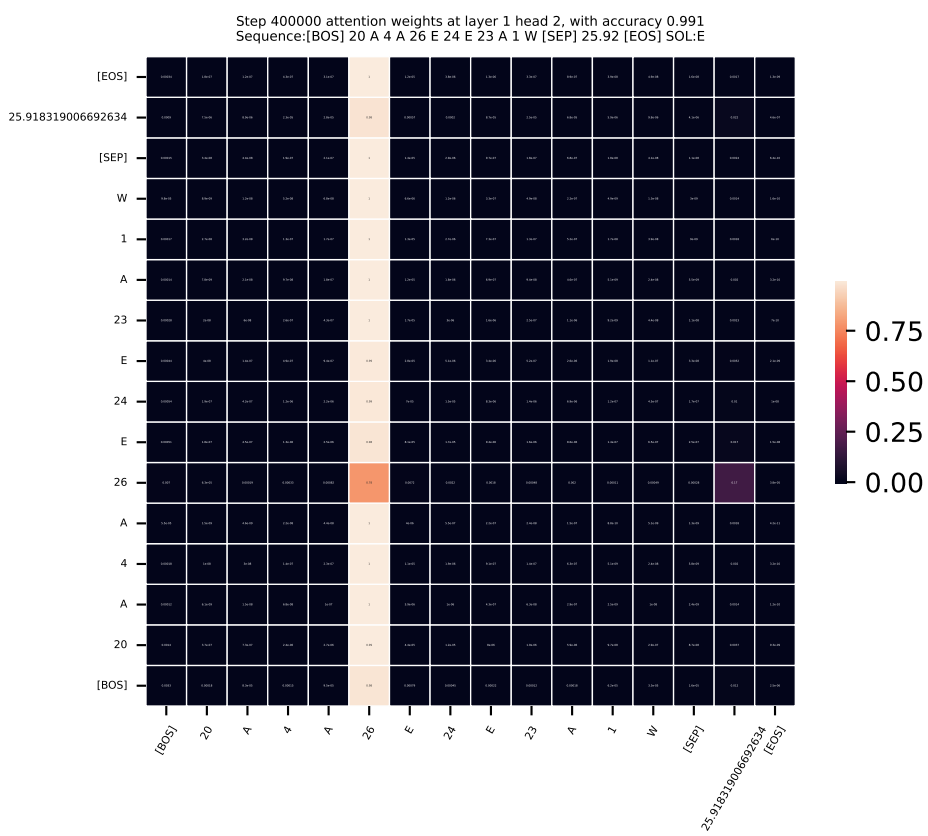

Step 400000 attention weights at layer 0 head 1, with accuracy 0.991
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

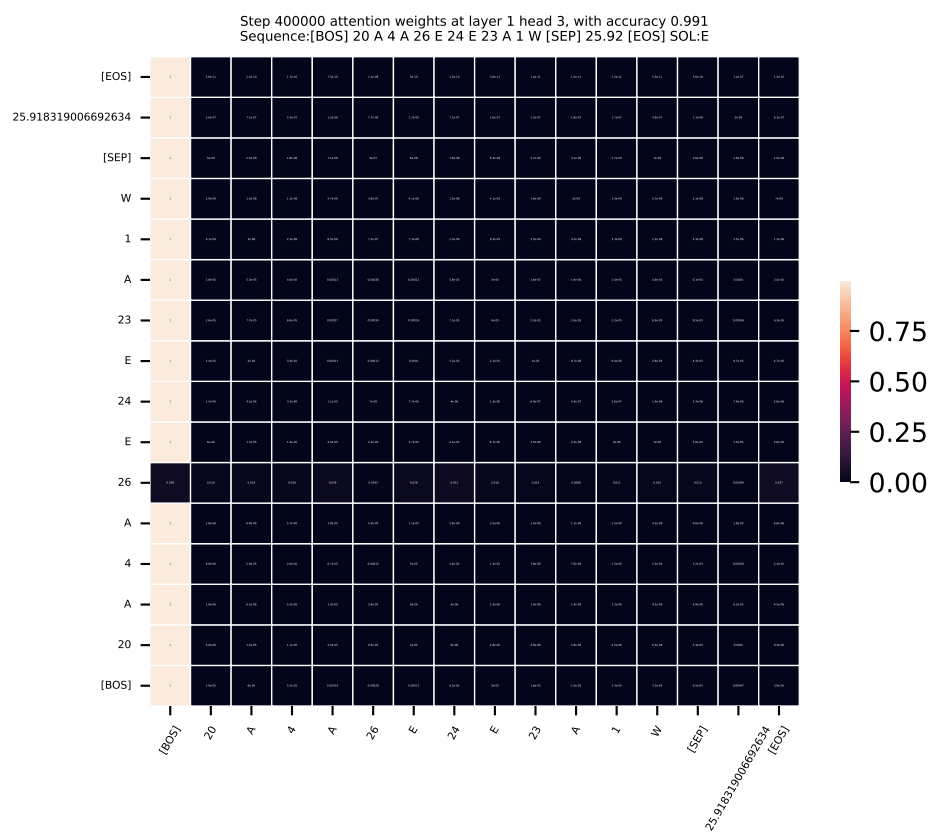

Step 400000 attention weights at layer 0 head 2, with accuracy 0.991
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

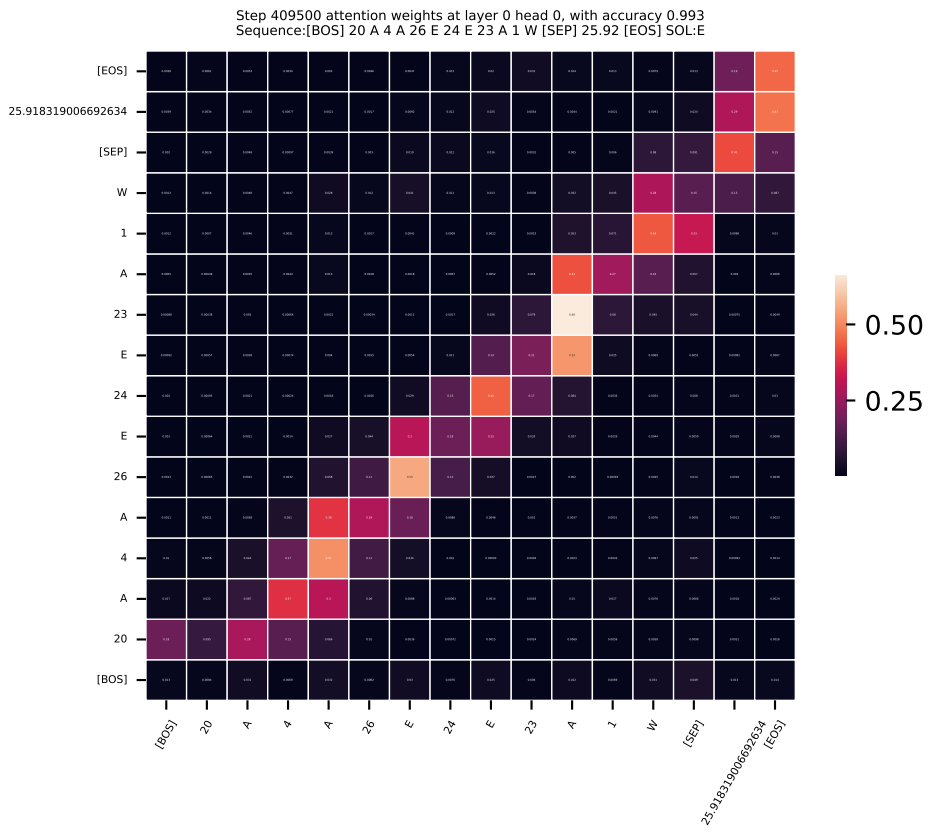

Step 400000 attention weights at layer 0 head 3, with accuracy 0.991
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25
0.00

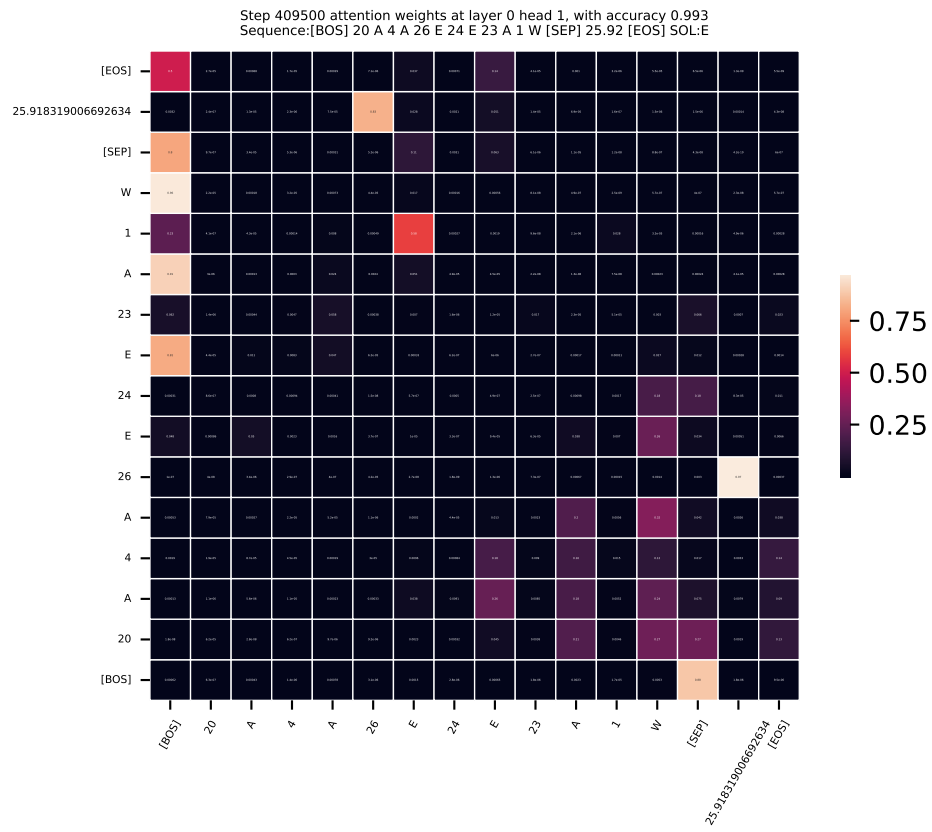

Step 400000 attention weights at layer 1 head 0, with accuracy 0.991
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

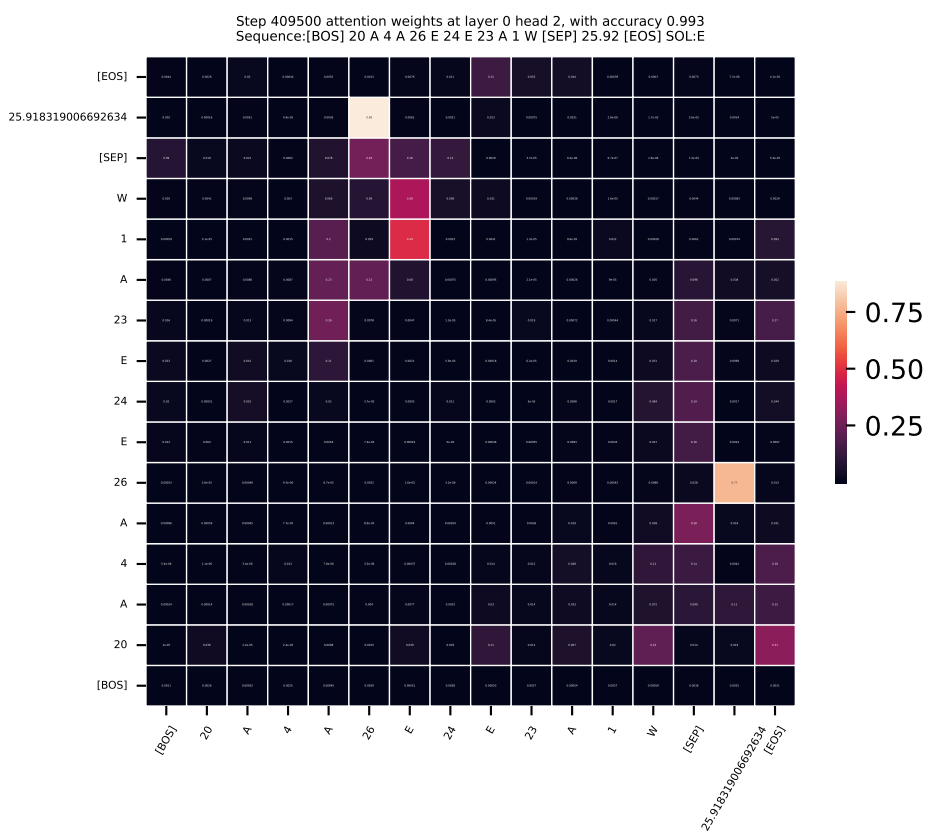

Step 400000 attention weights at layer 1 head 1, with accuracy 0.991
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

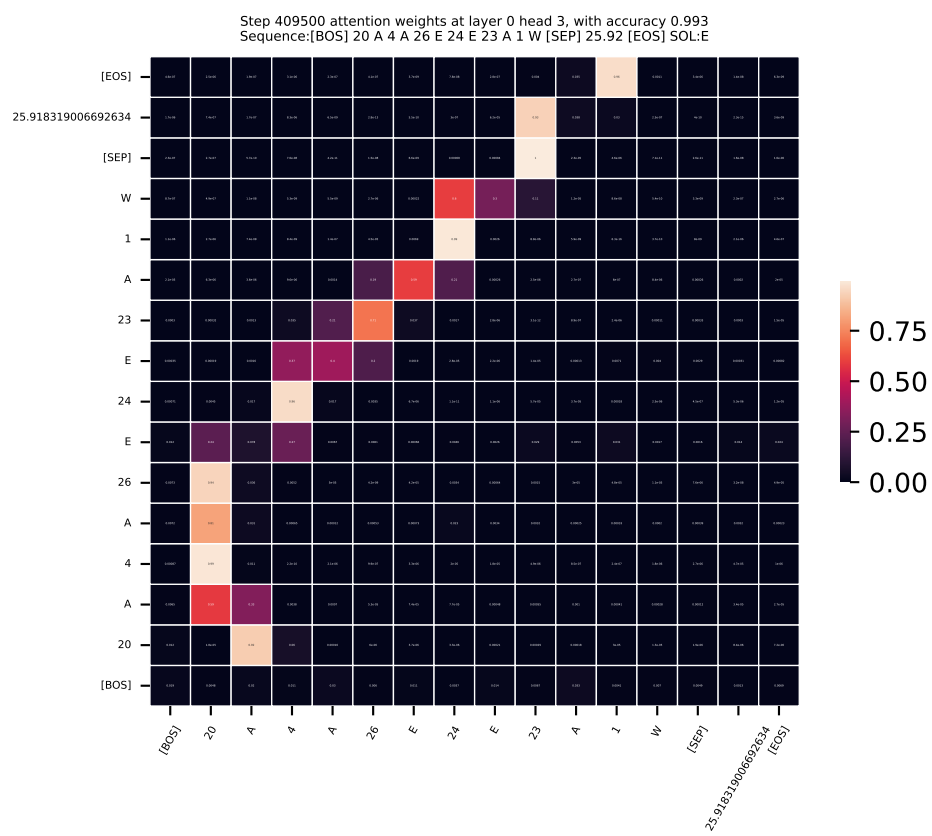

Step 400000 attention weights at layer 1 head 2, with accuracy 0.991
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

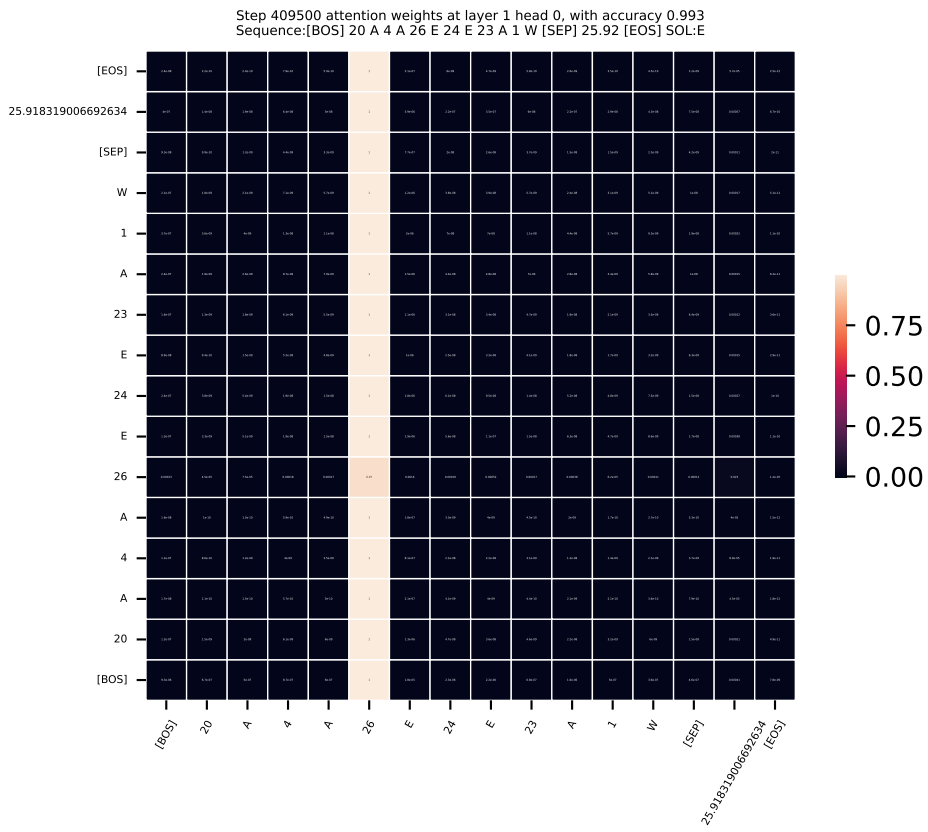

Step 400000 attention weights at layer 1 head 3, with accuracy 0.991
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

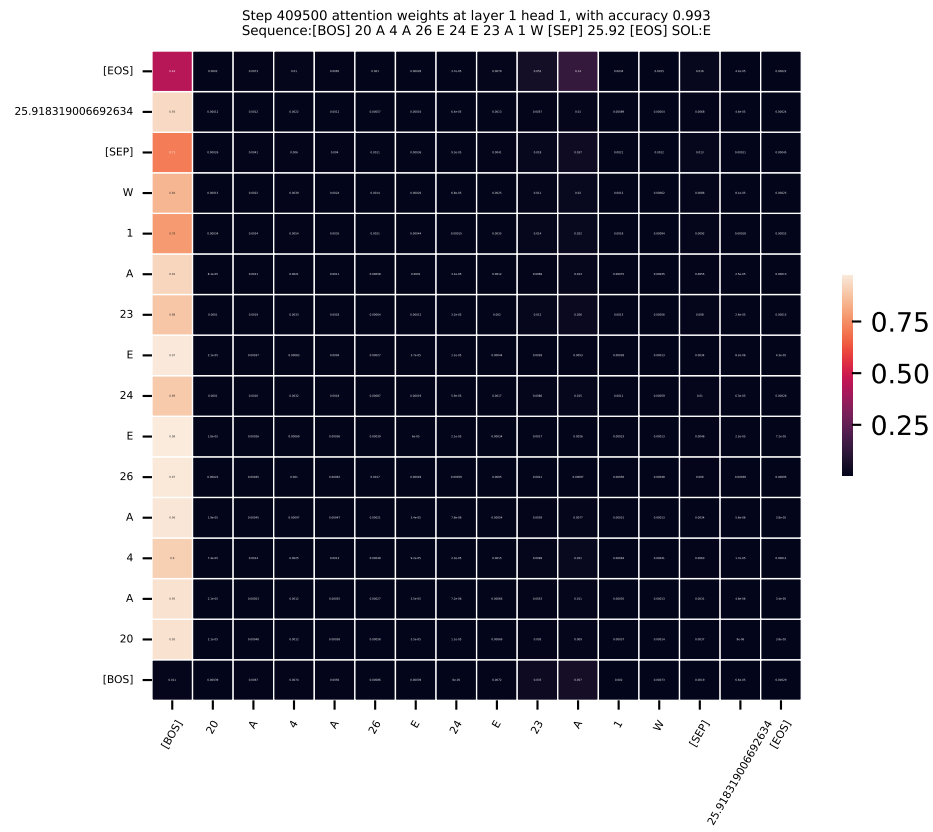

Step 409500 attention weights at layer 0 head 0, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

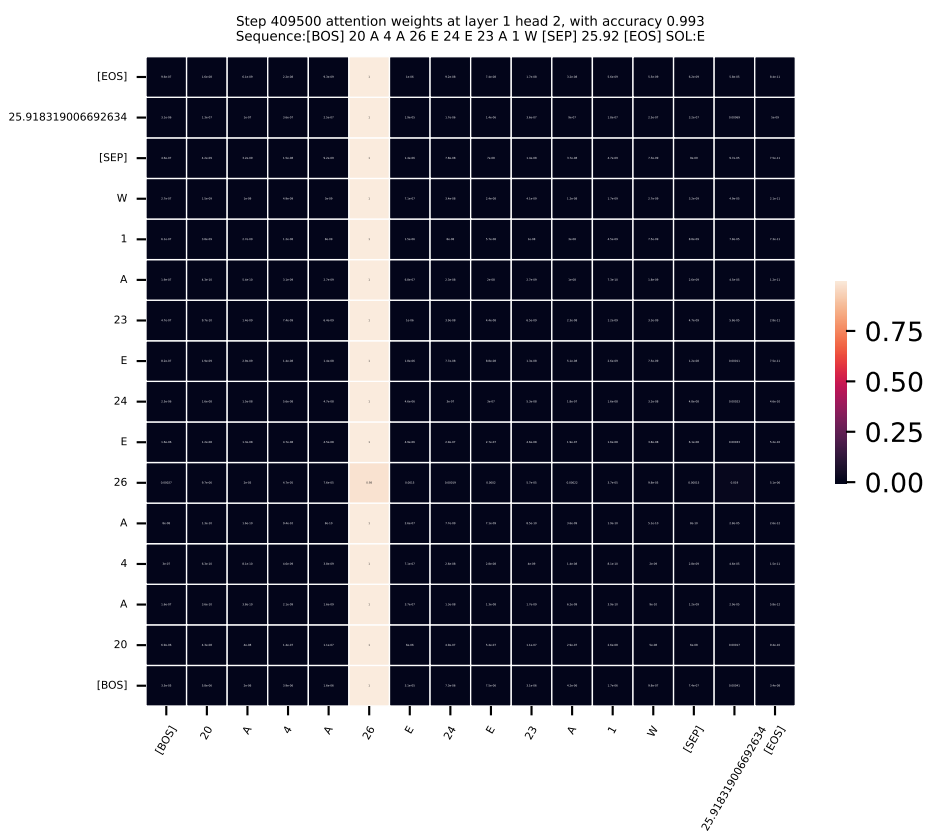

Step 409500 attention weights at layer 0 head 1, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

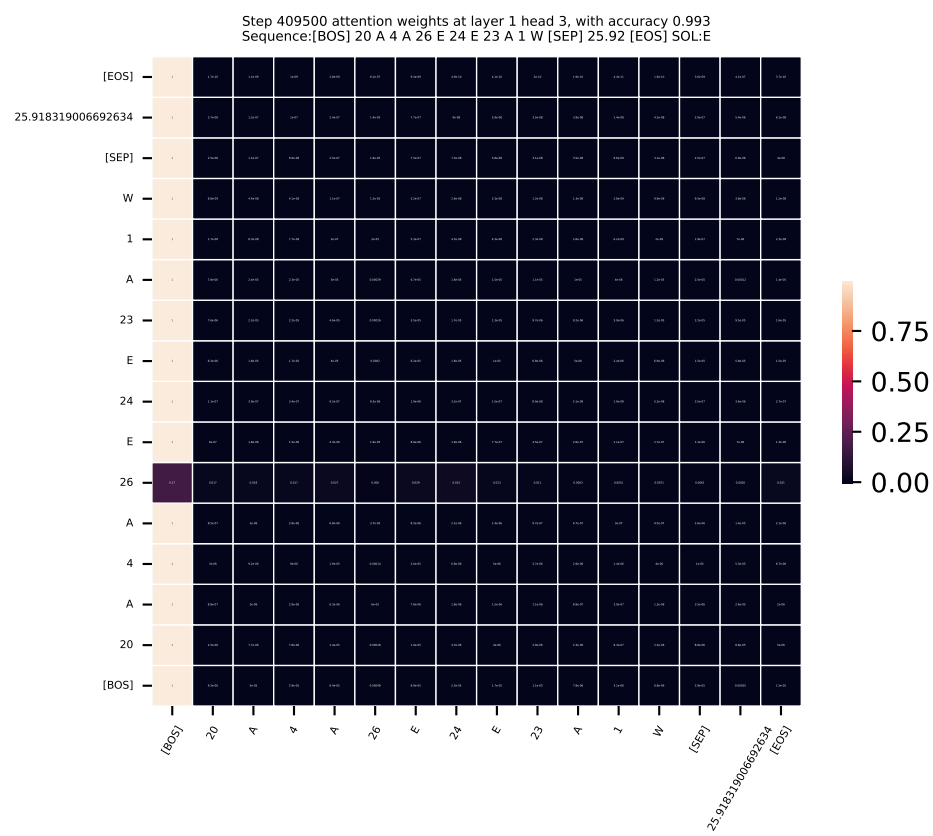

Step 409500 attention weights at layer 0 head 2, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

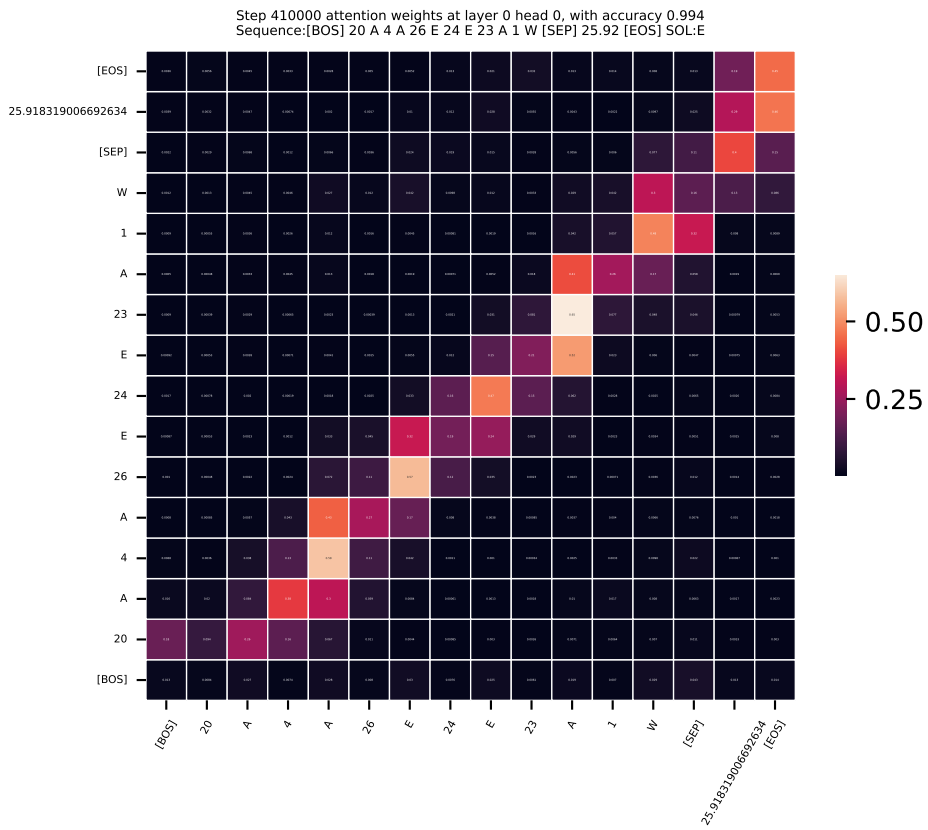

Step 409500 attention weights at layer 0 head 3, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

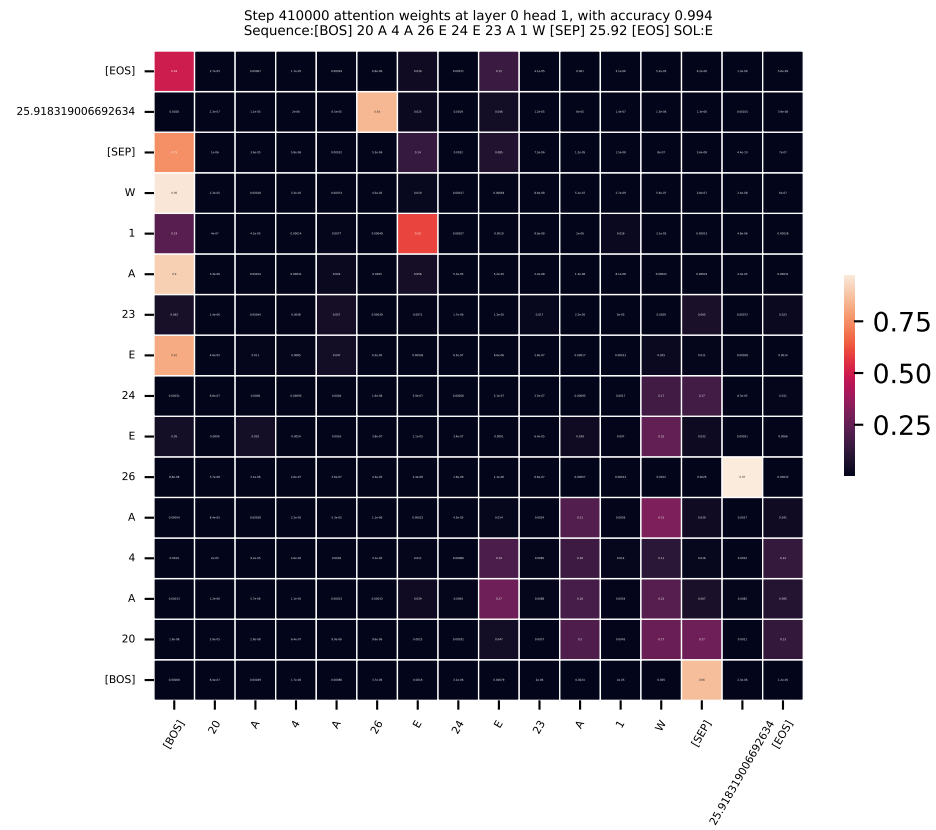

Step 409500 attention weights at layer 1 head 0, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

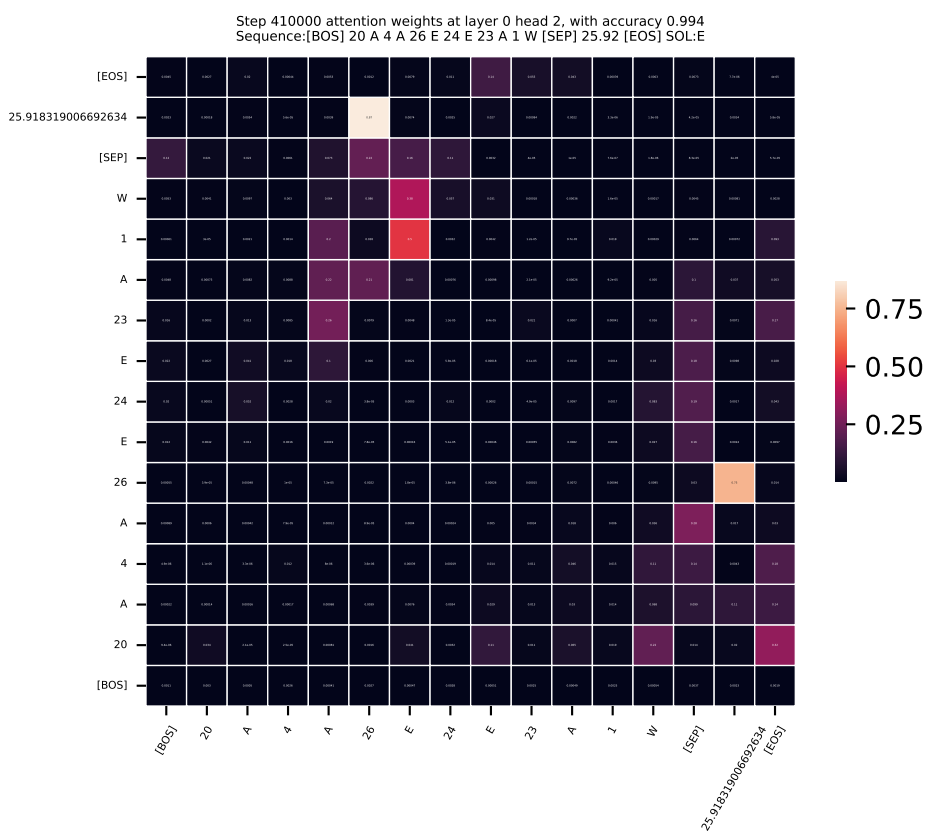

Step 409500 attention weights at layer 1 head 1, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

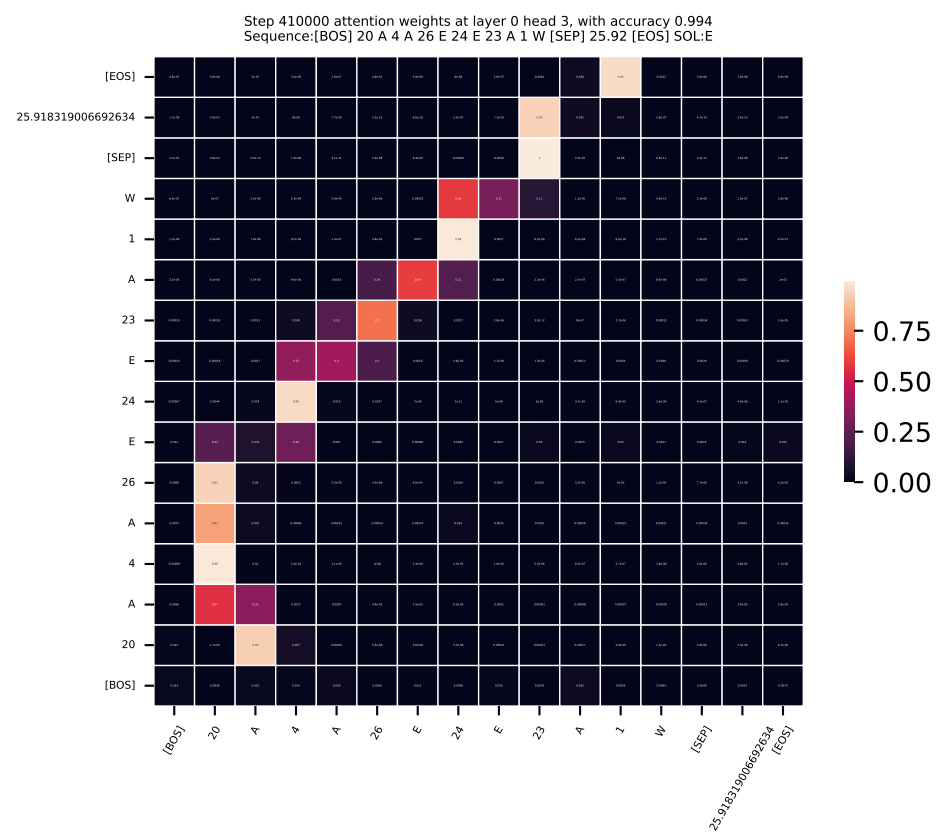

Step 409500 attention weights at layer 1 head 2, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

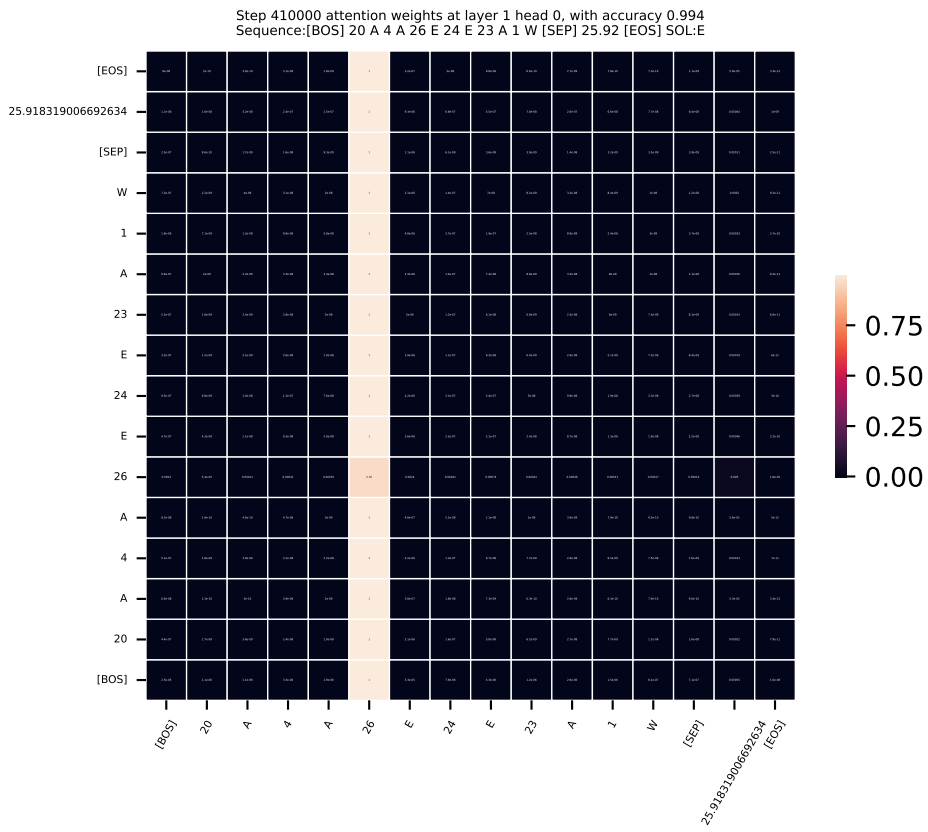

Step 409500 attention weights at layer 1 head 3, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

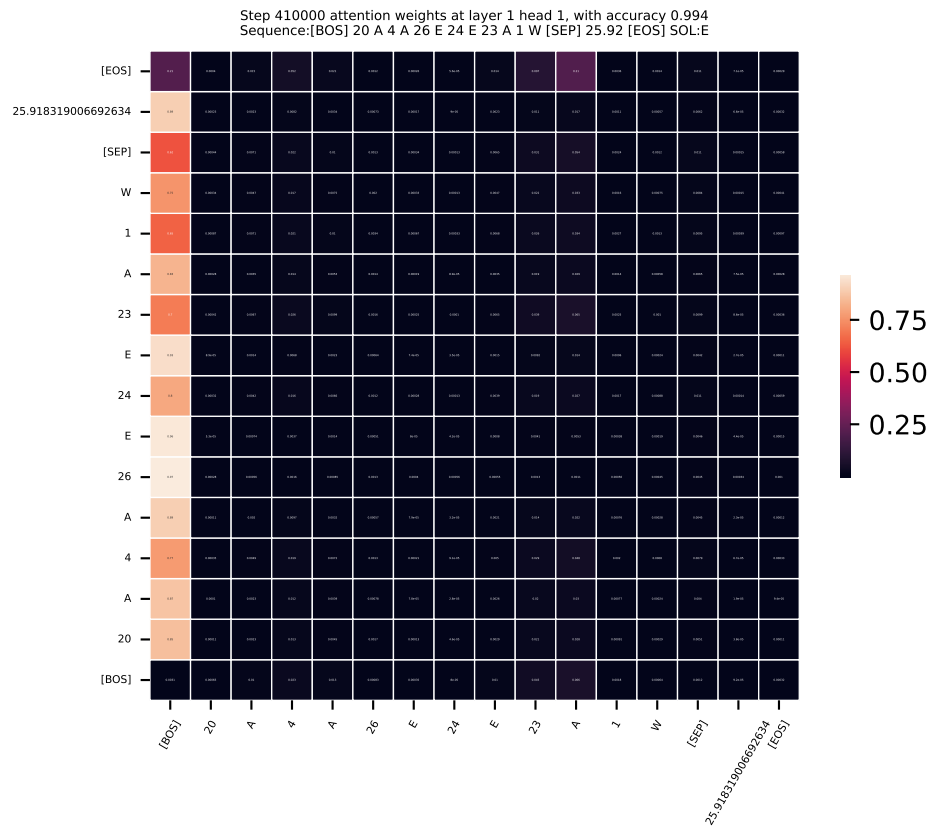

Step 410000 attention weights at layer 0 head 0, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

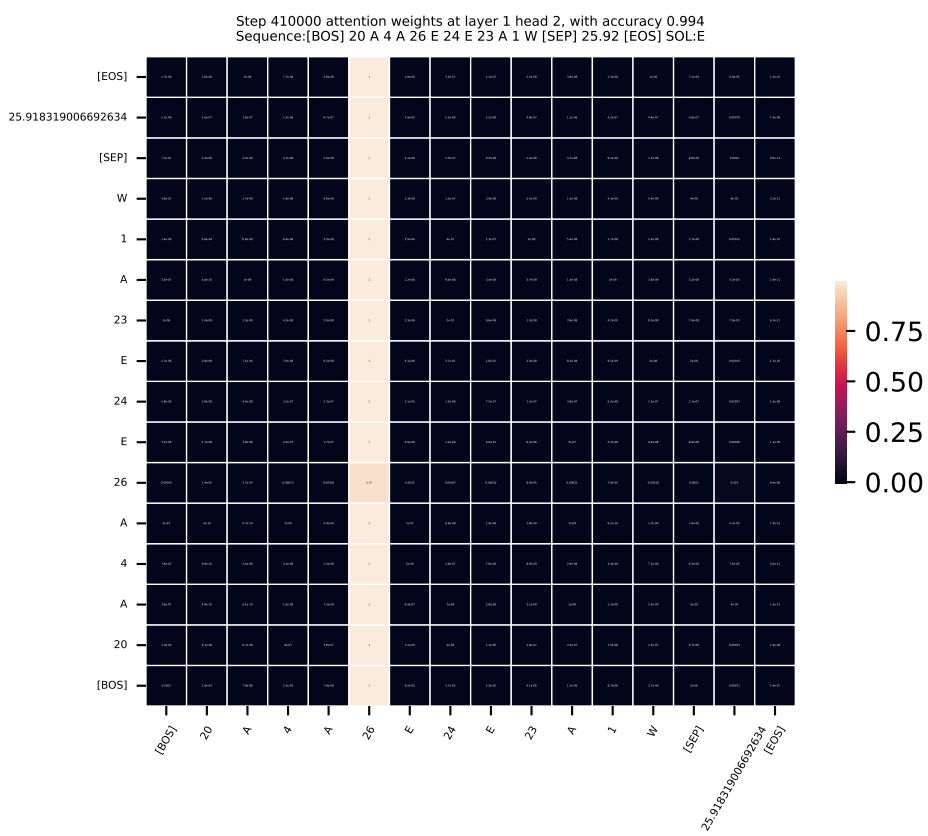

Step 410000 attention weights at layer 0 head 1, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

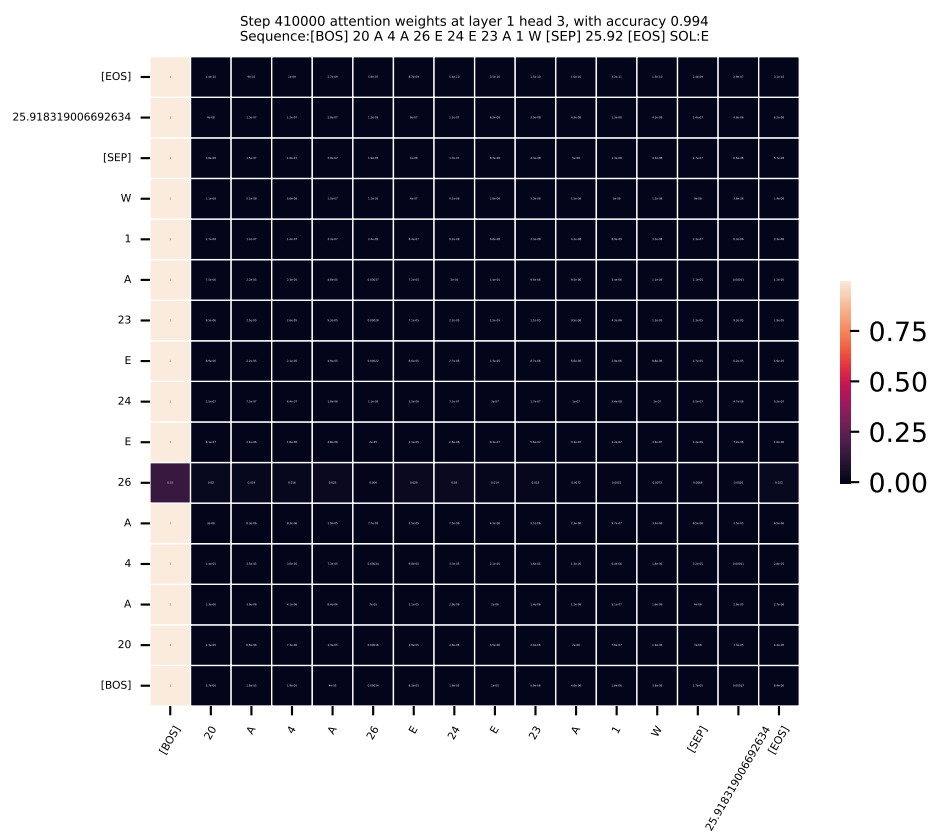

Step 410000 attention weights at layer 0 head 2, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

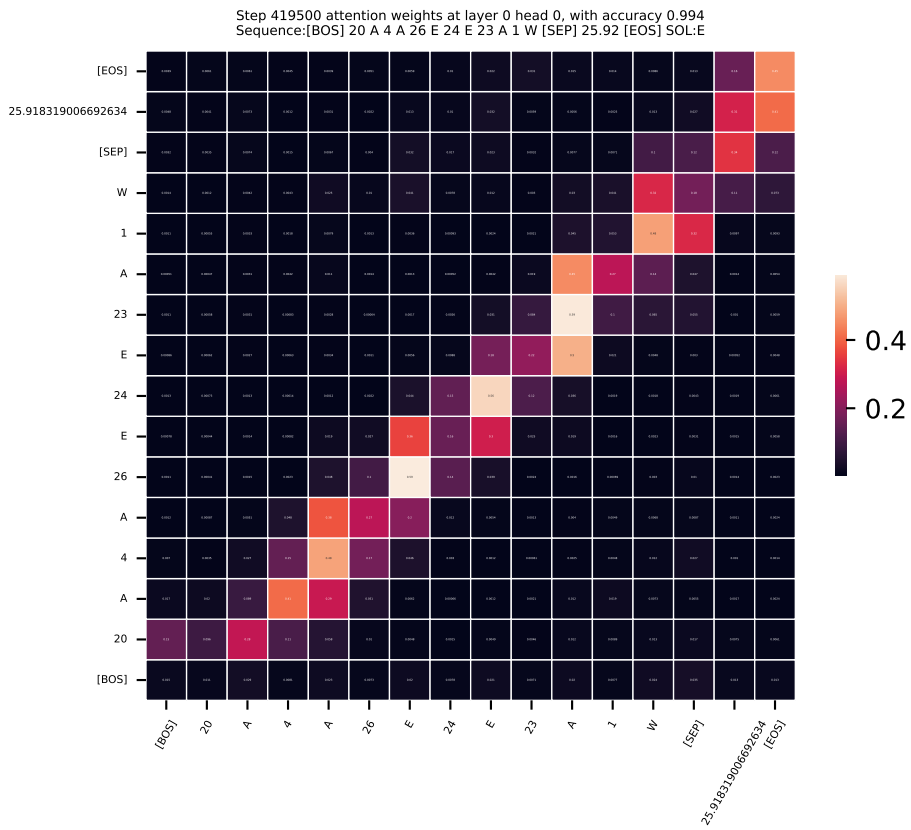

Step 410000 attention weights at layer 0 head 3, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

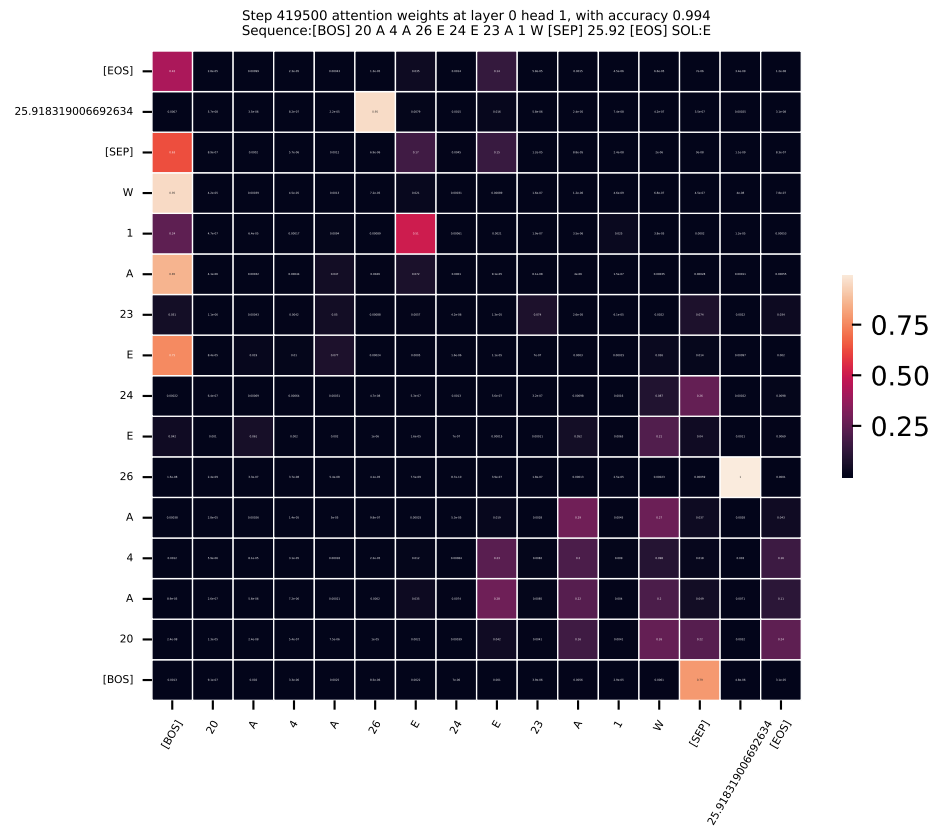

Step 410000 attention weights at layer 1 head 0, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

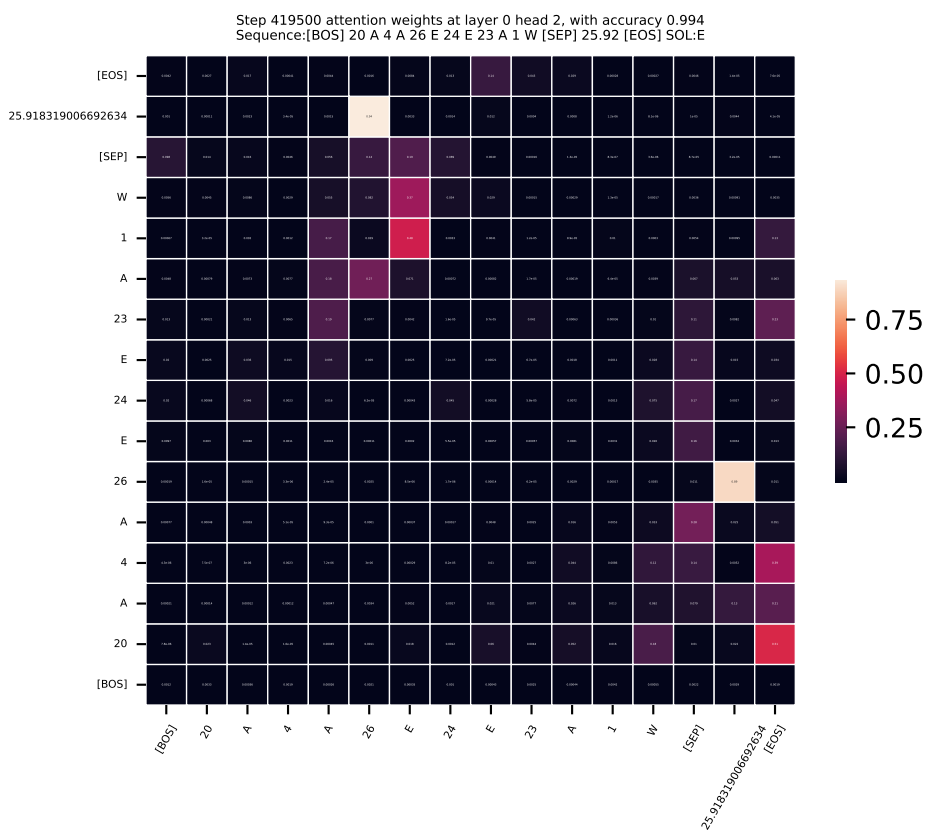

Step 410000 attention weights at layer 1 head 1, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

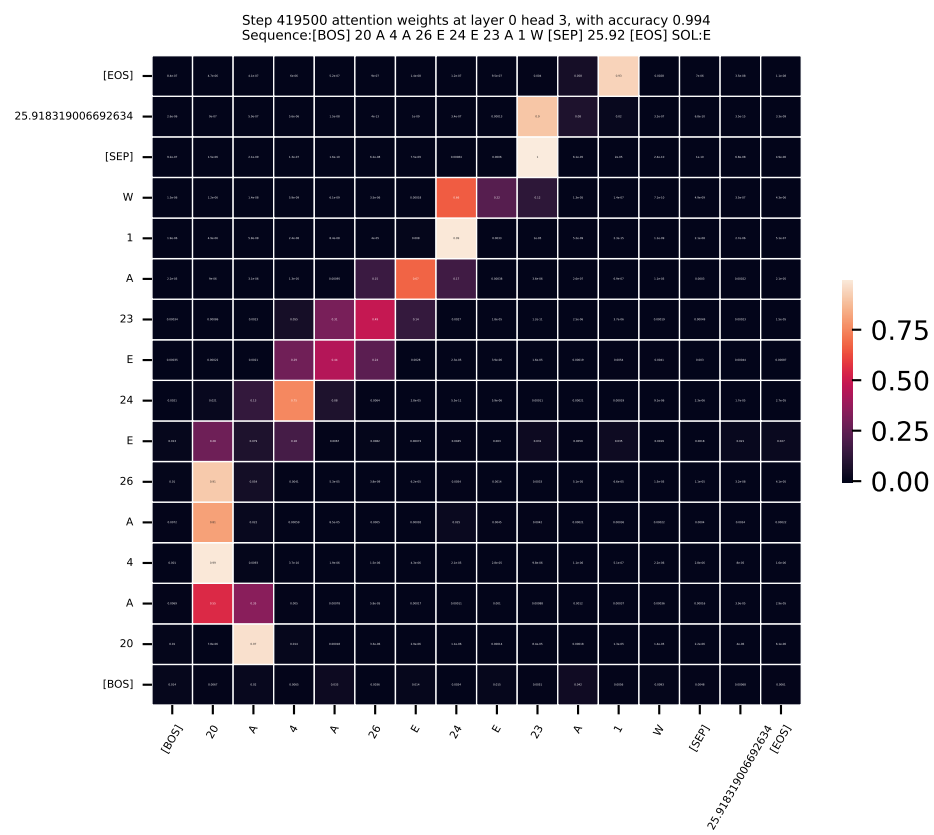

Step 410000 attention weights at layer 1 head 2, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

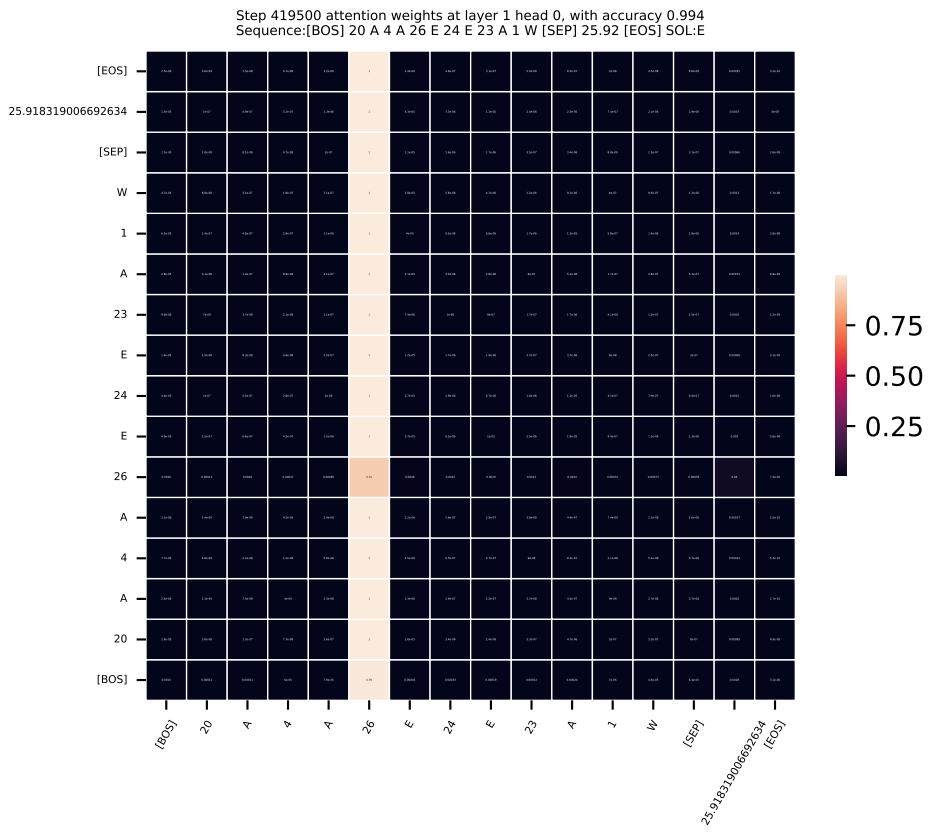

Step 410000 attention weights at layer 1 head 3, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

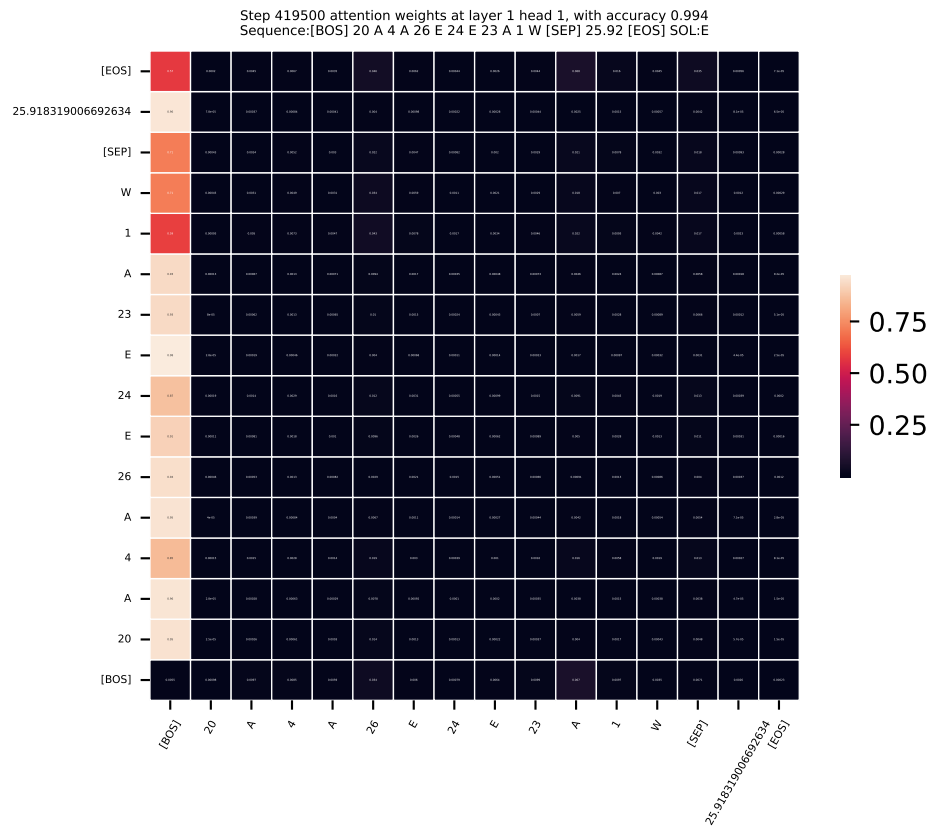

Step 419500 attention weights at layer 0 head 0, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

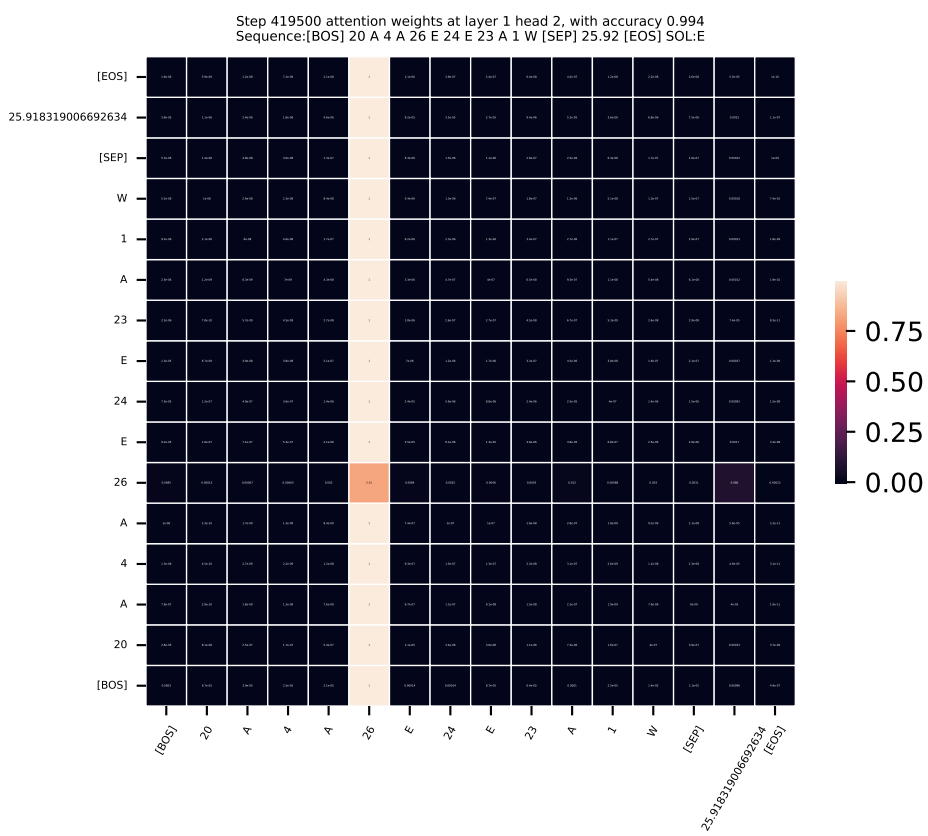

Step 419500 attention weights at layer 0 head 1, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

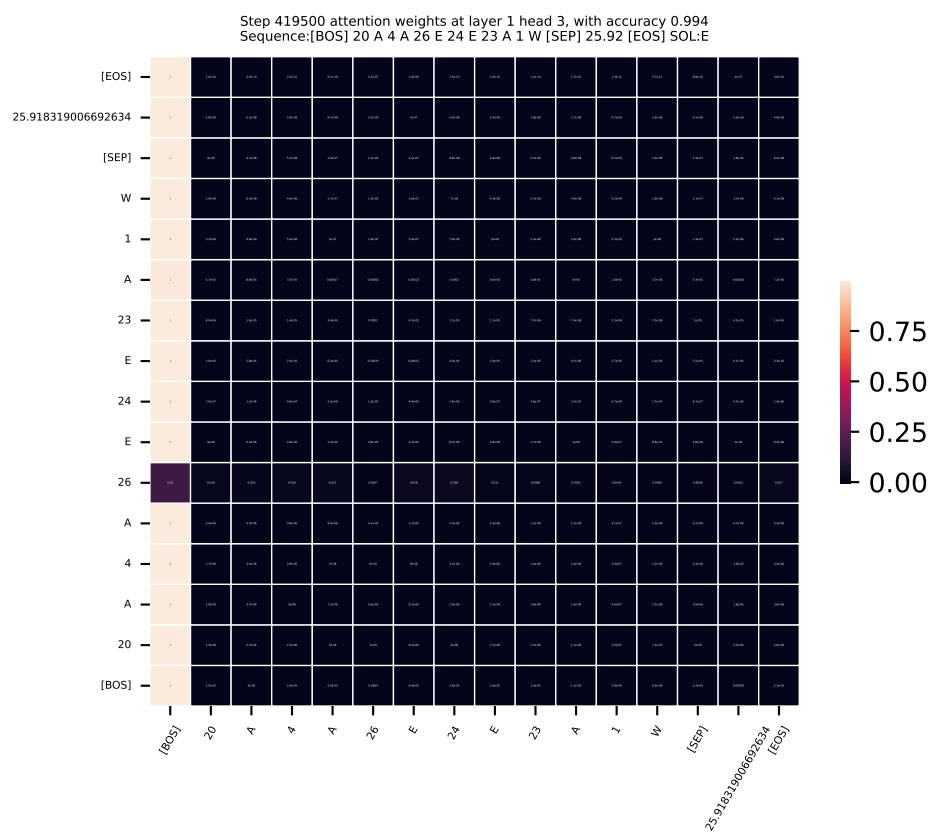

Step 419500 attention weights at layer 0 head 2, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

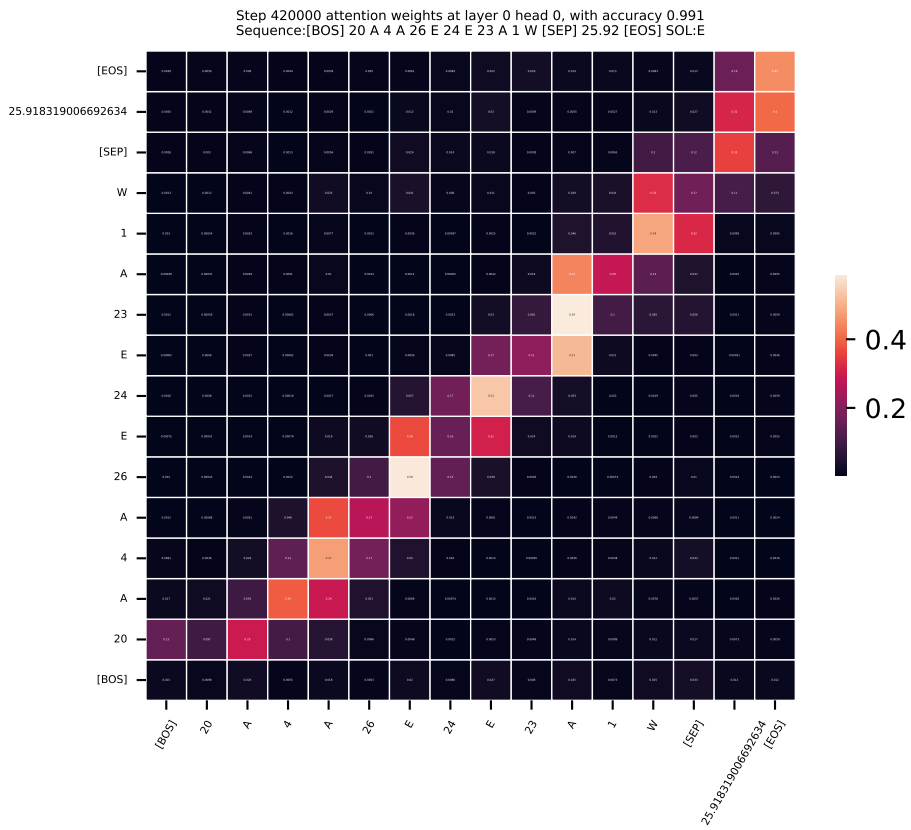

Step 419500 attention weights at layer 0 head 3, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

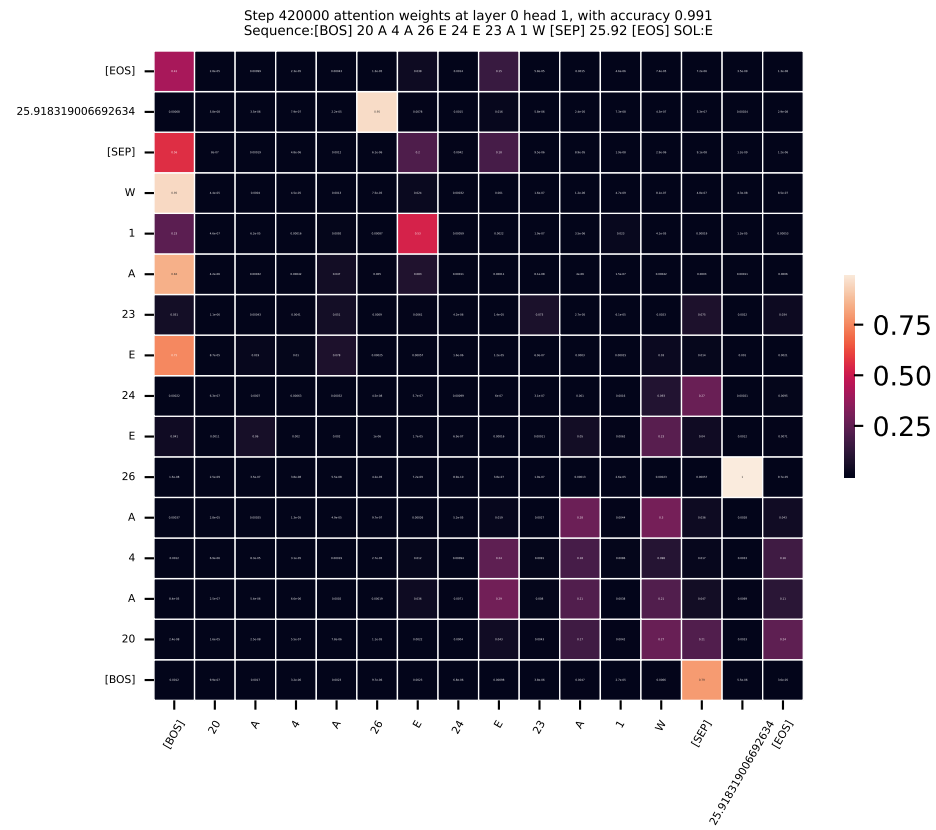

Step 419500 attention weights at layer 1 head 0, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

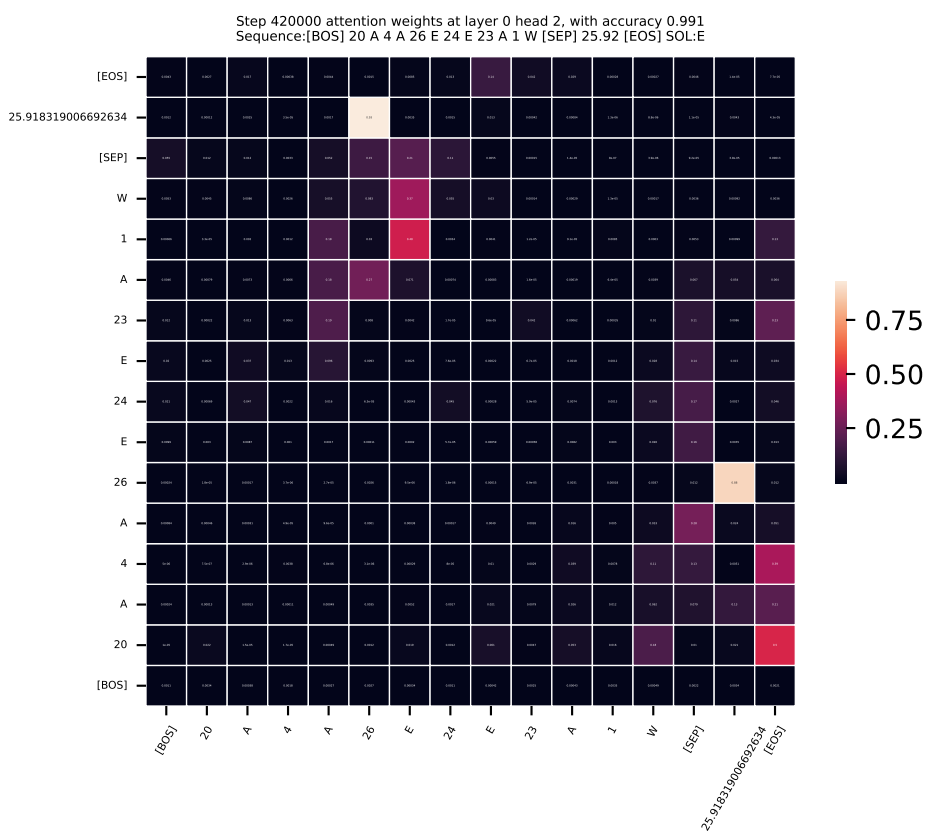

Step 419500 attention weights at layer 1 head 1, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

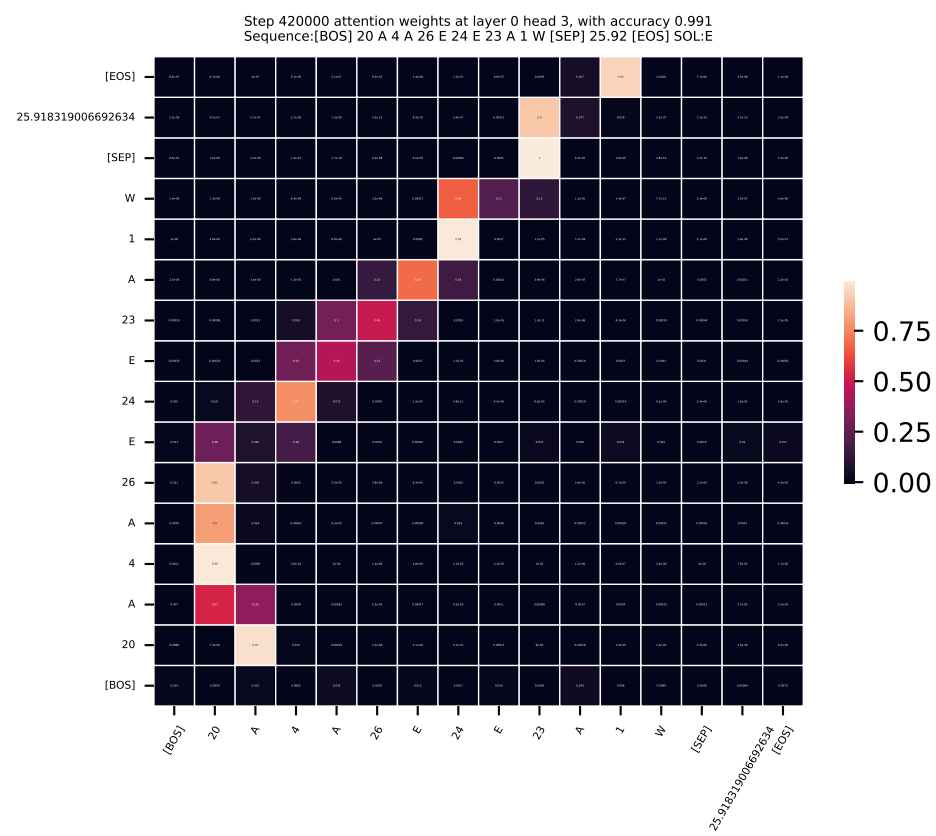

Step 419500 attention weights at layer 1 head 2, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

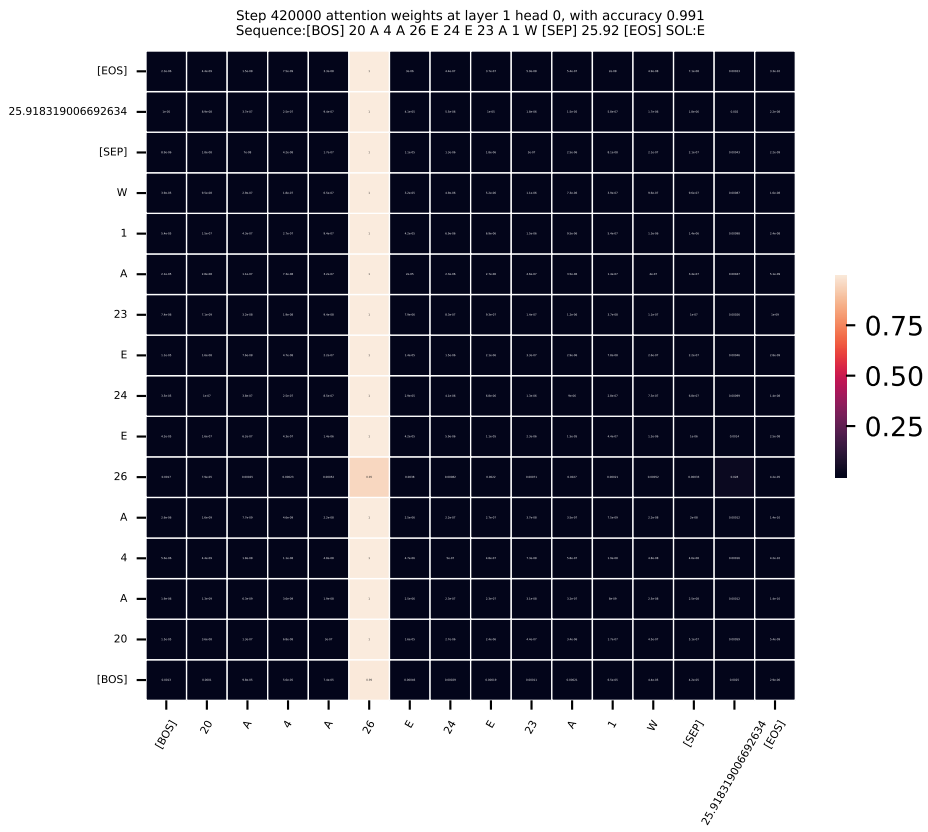

Step 419500 attention weights at layer 1 head 3, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

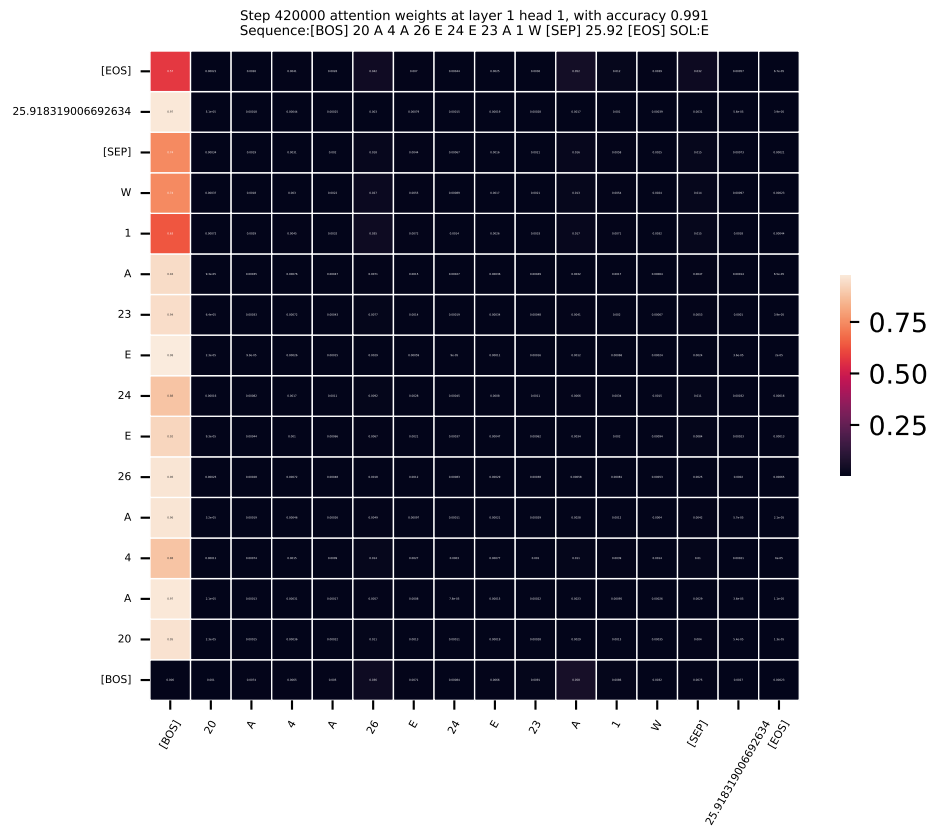

Step 420000 attention weights at layer 0 head 0, with accuracy 0.991
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.4
0.2

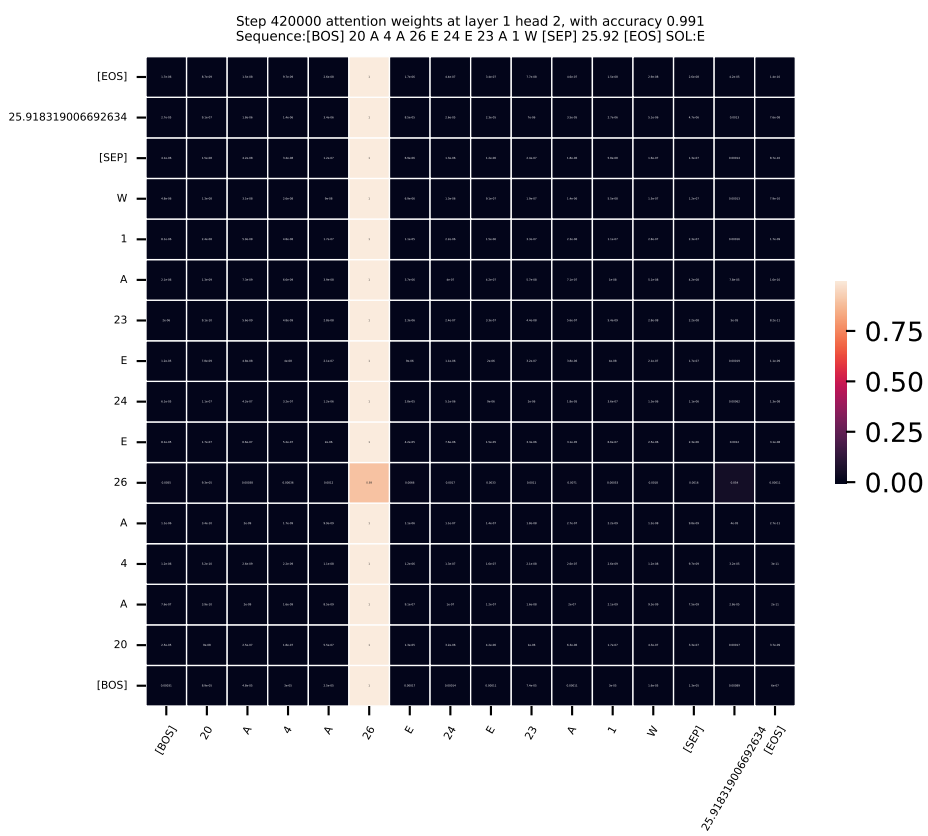

Step 420000 attention weights at layer 0 head 1, with accuracy 0.991
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

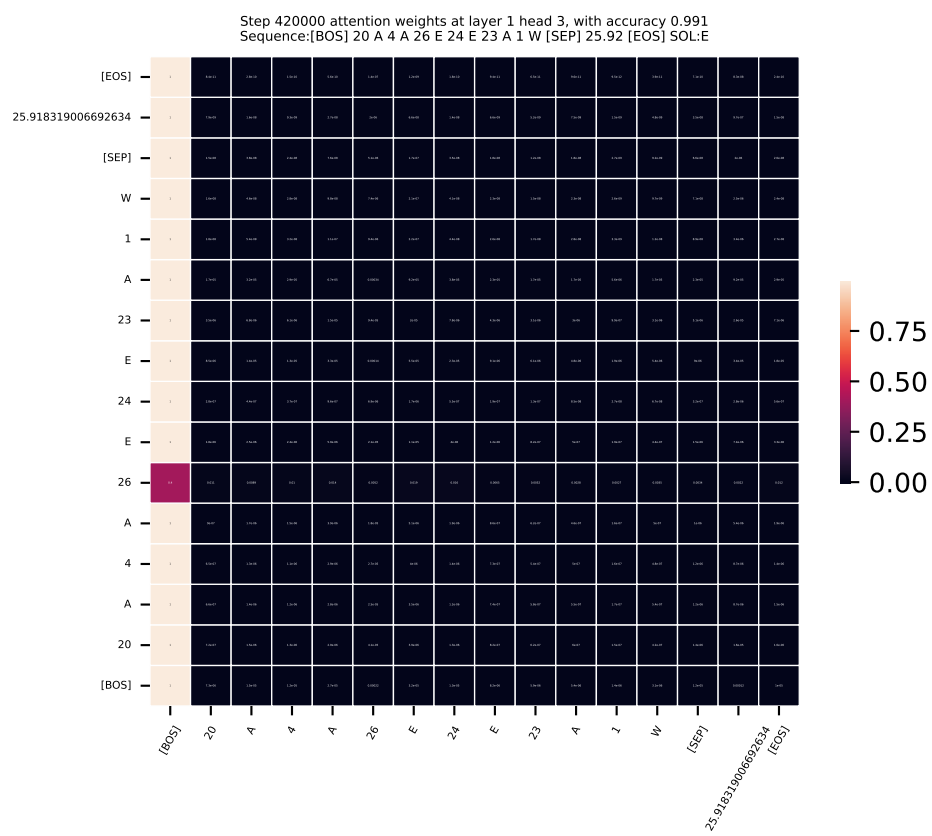

Step 420000 attention weights at layer 0 head 2, with accuracy 0.991
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

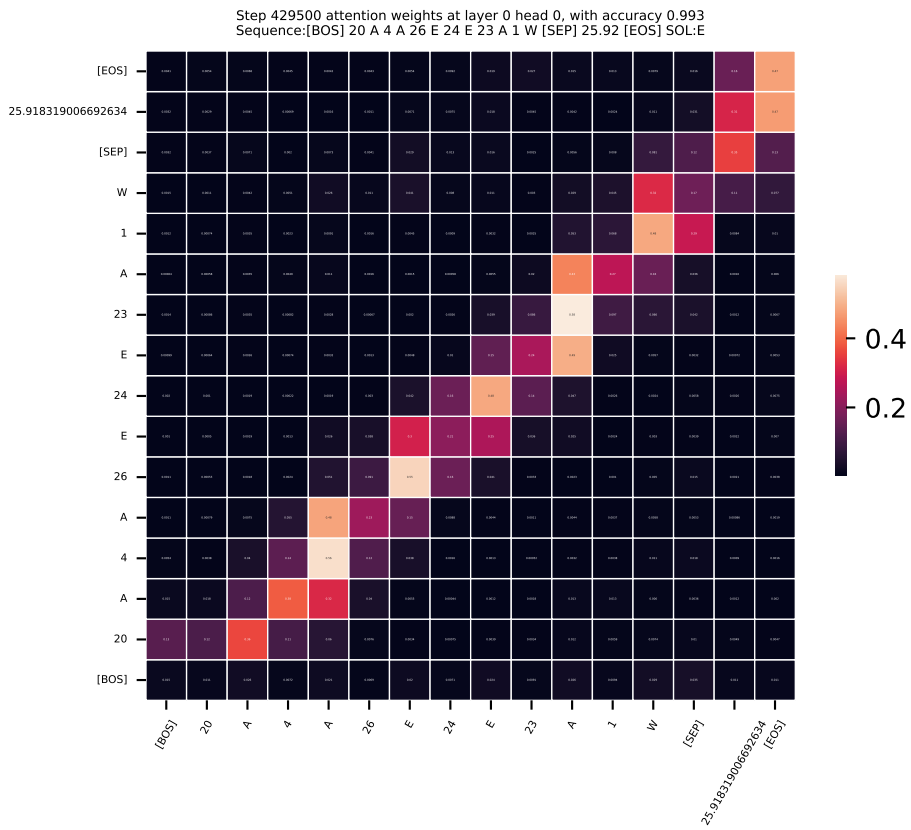

Step 420000 attention weights at layer 0 head 3, with accuracy 0.991
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25
0.00

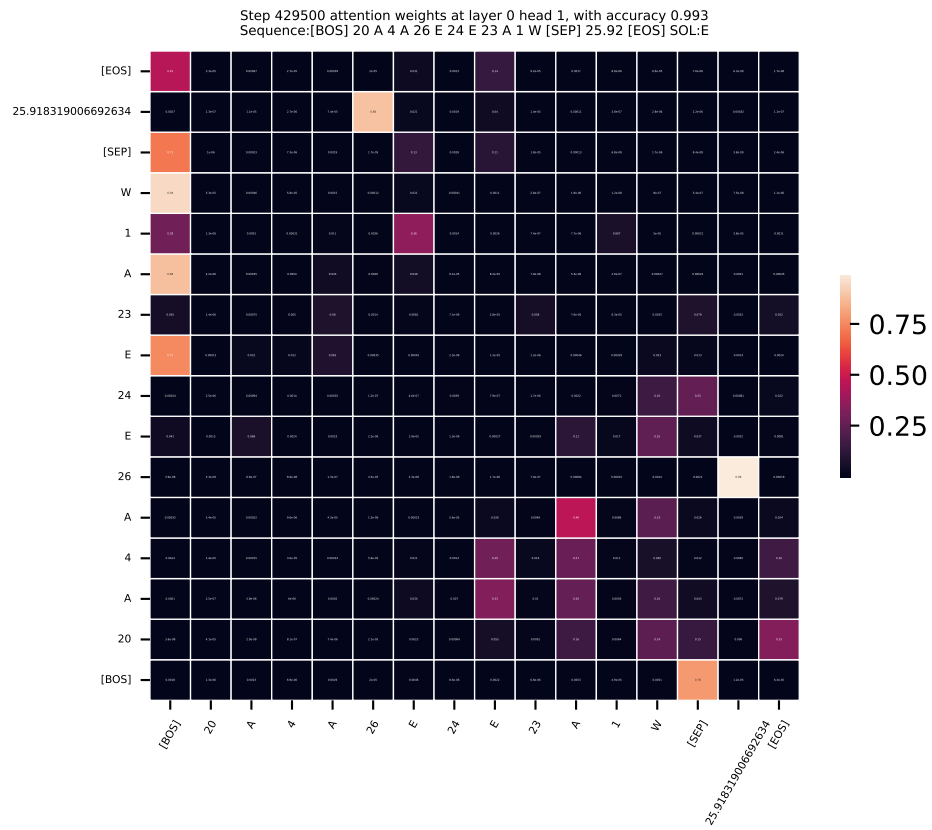

Step 420000 attention weights at layer 1 head 0, with accuracy 0.991
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

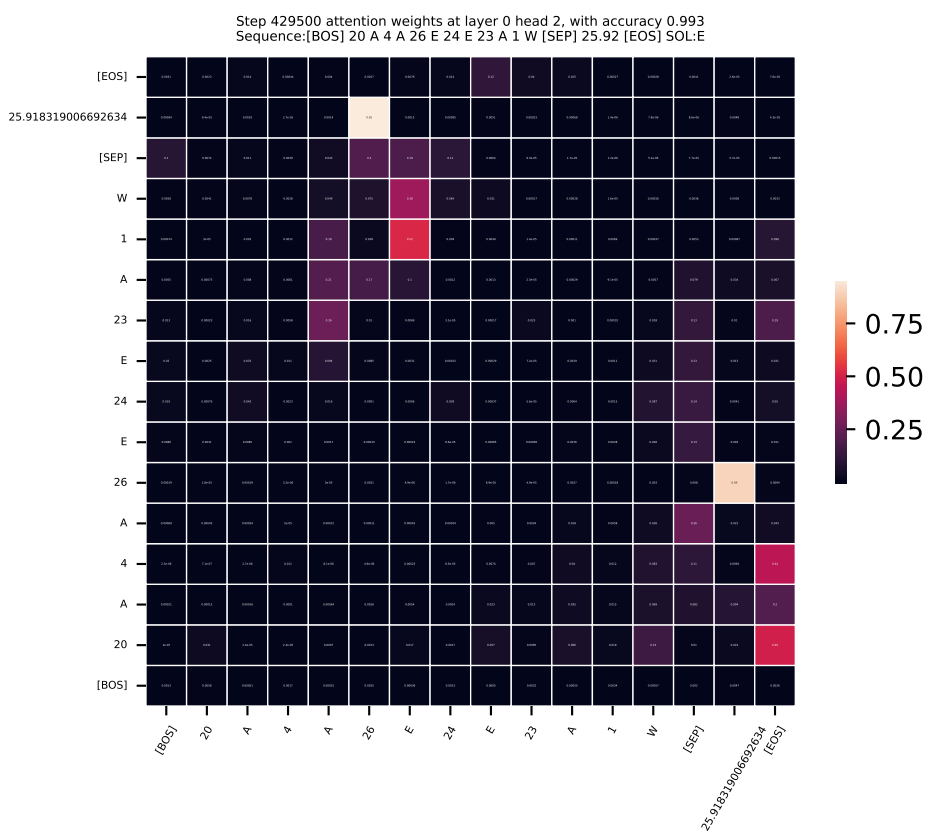

Step 420000 attention weights at layer 1 head 1, with accuracy 0.991
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

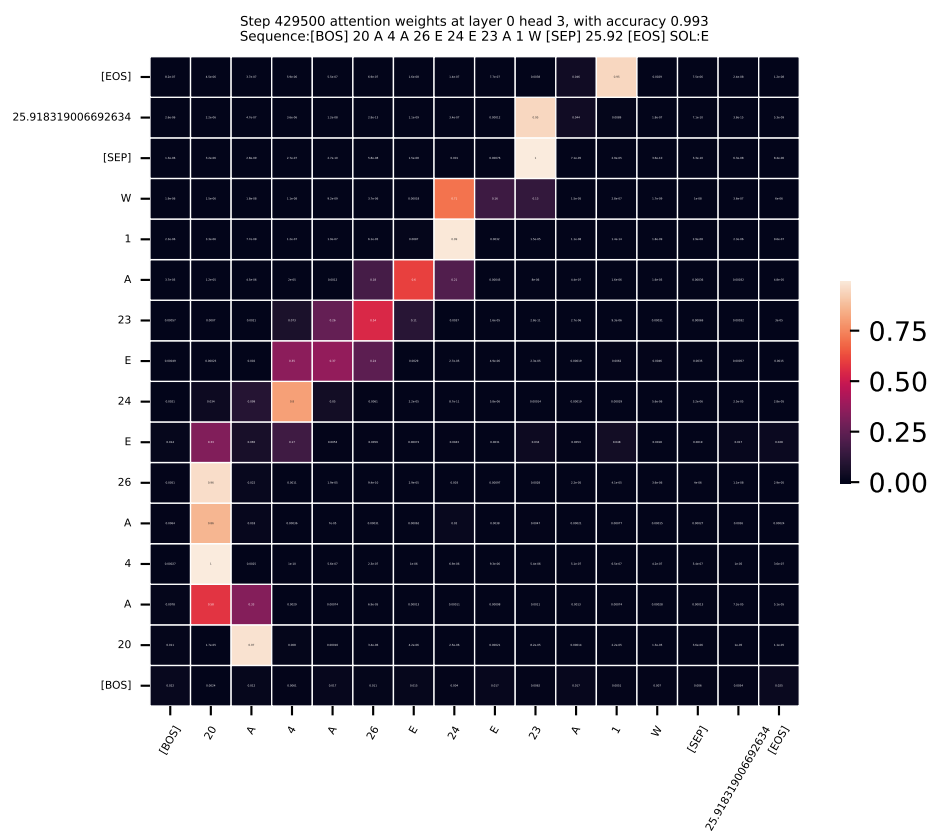

Step 420000 attention weights at layer 1 head 2, with accuracy 0.991
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

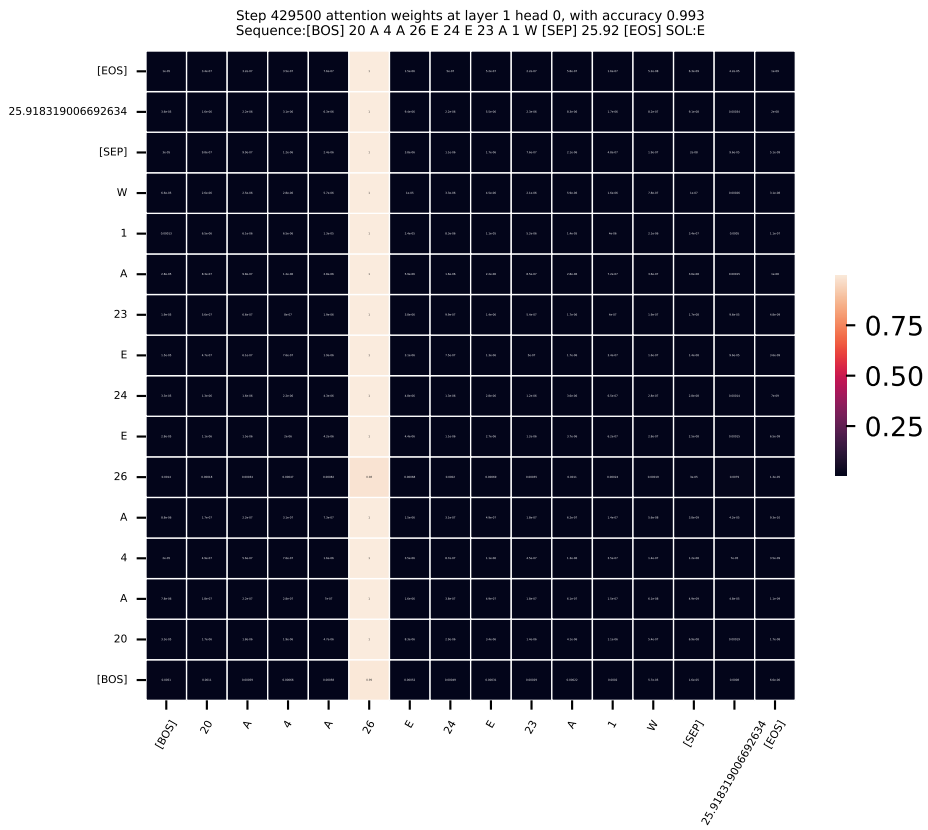

Step 420000 attention weights at layer 1 head 3, with accuracy 0.991
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

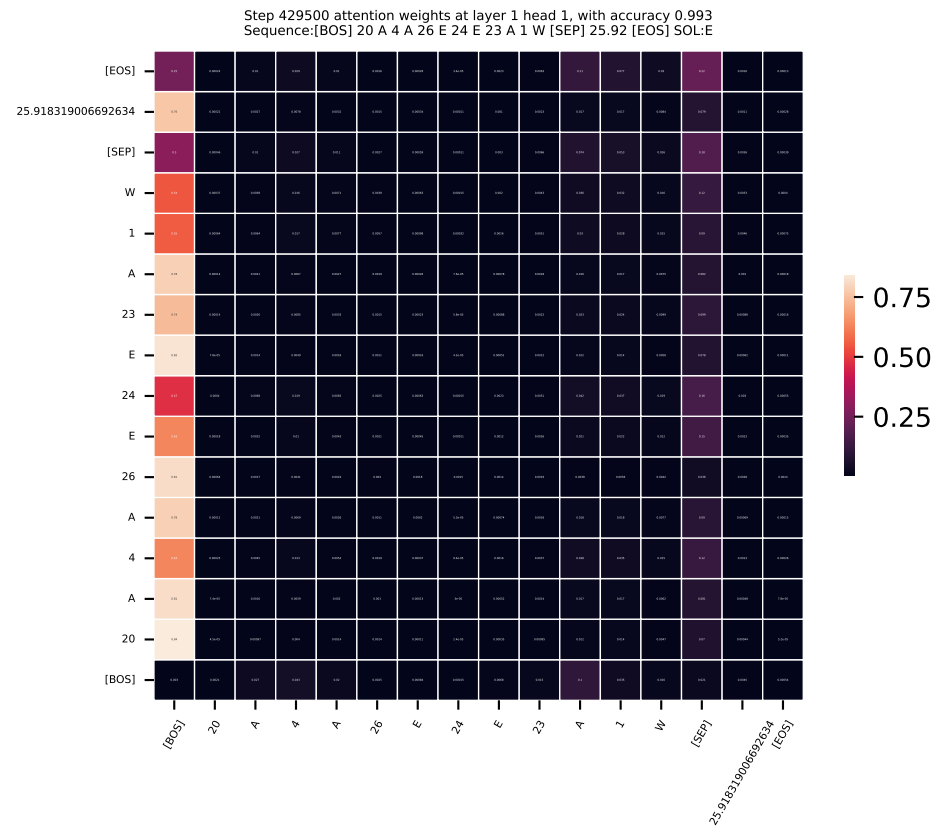

Step 429500 attention weights at layer 0 head 0, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.4
0.2

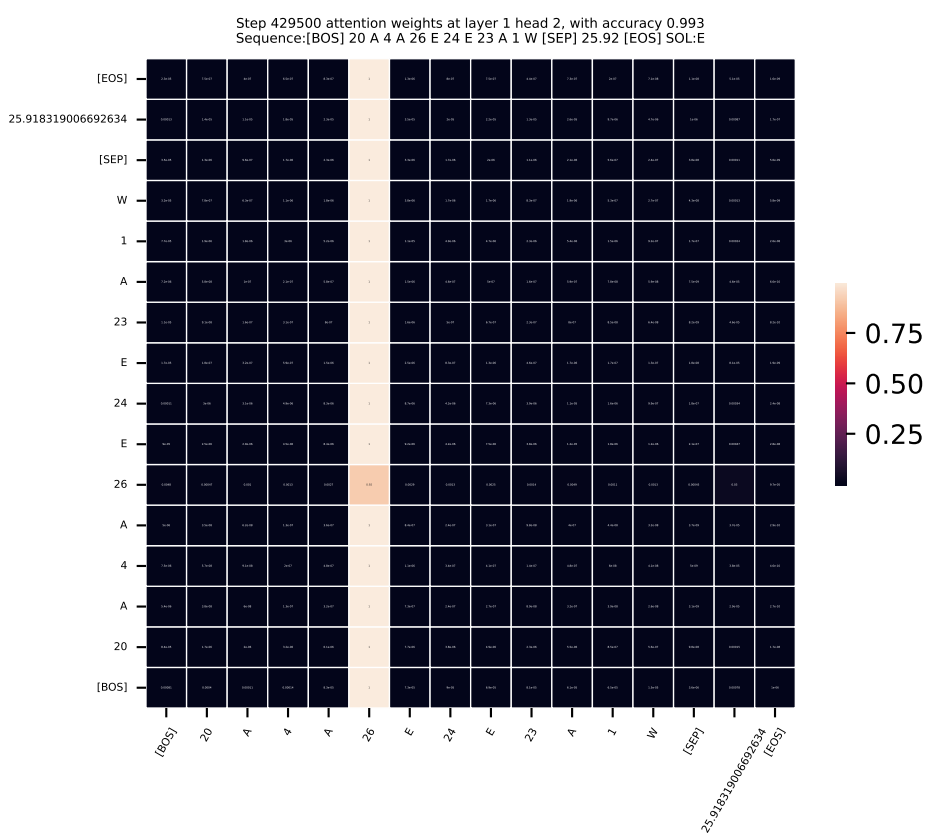

Step 429500 attention weights at layer 0 head 1, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

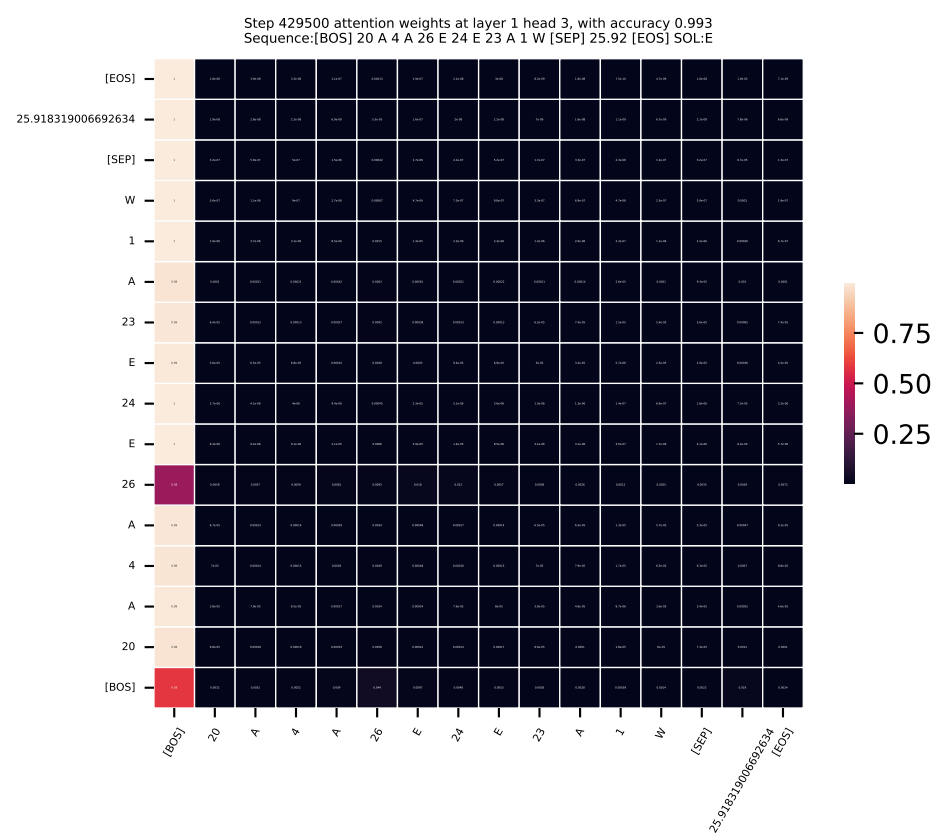

Step 429500 attention weights at layer 0 head 2, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

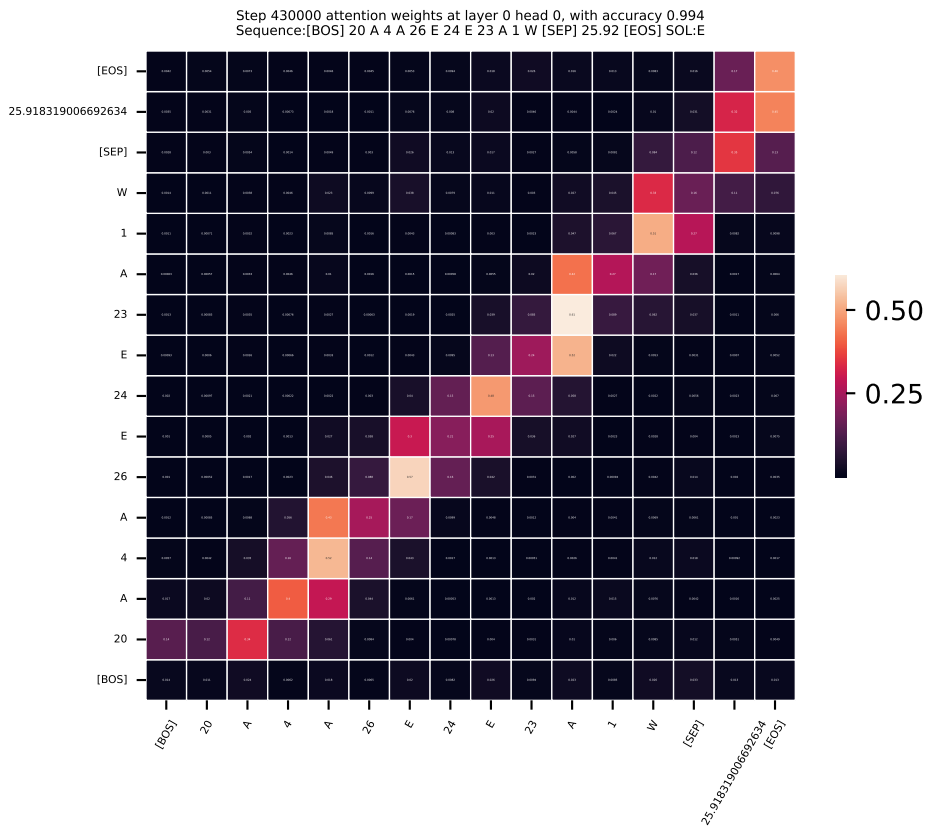

Step 429500 attention weights at layer 0 head 3, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

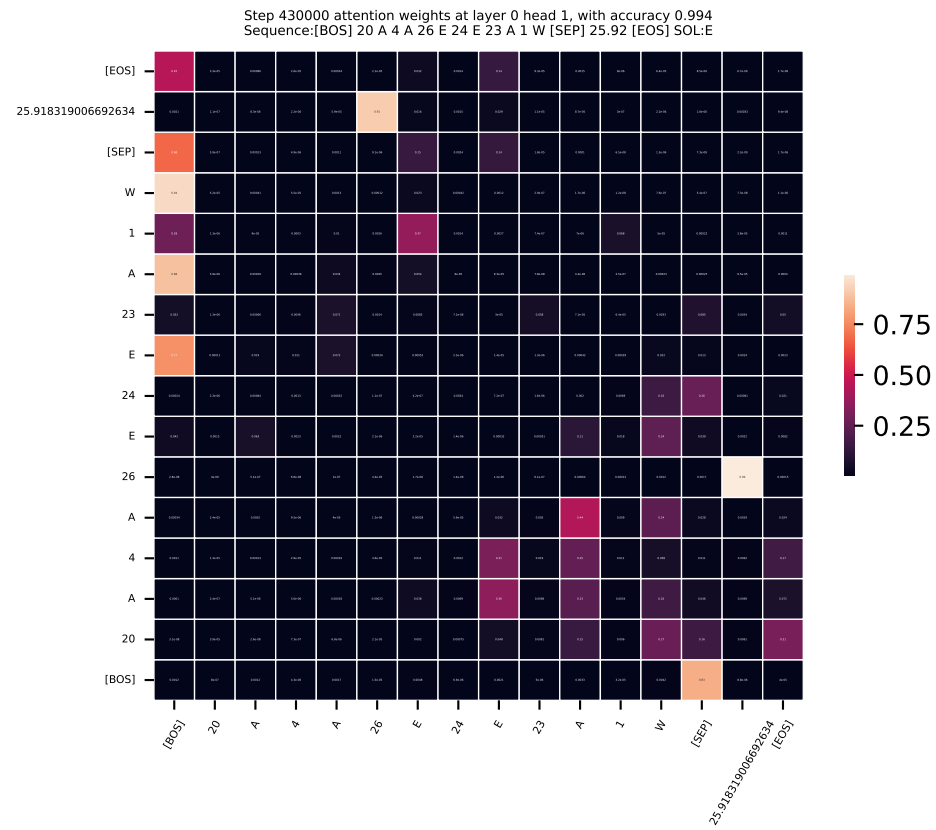

Step 429500 attention weights at layer 1 head 0, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

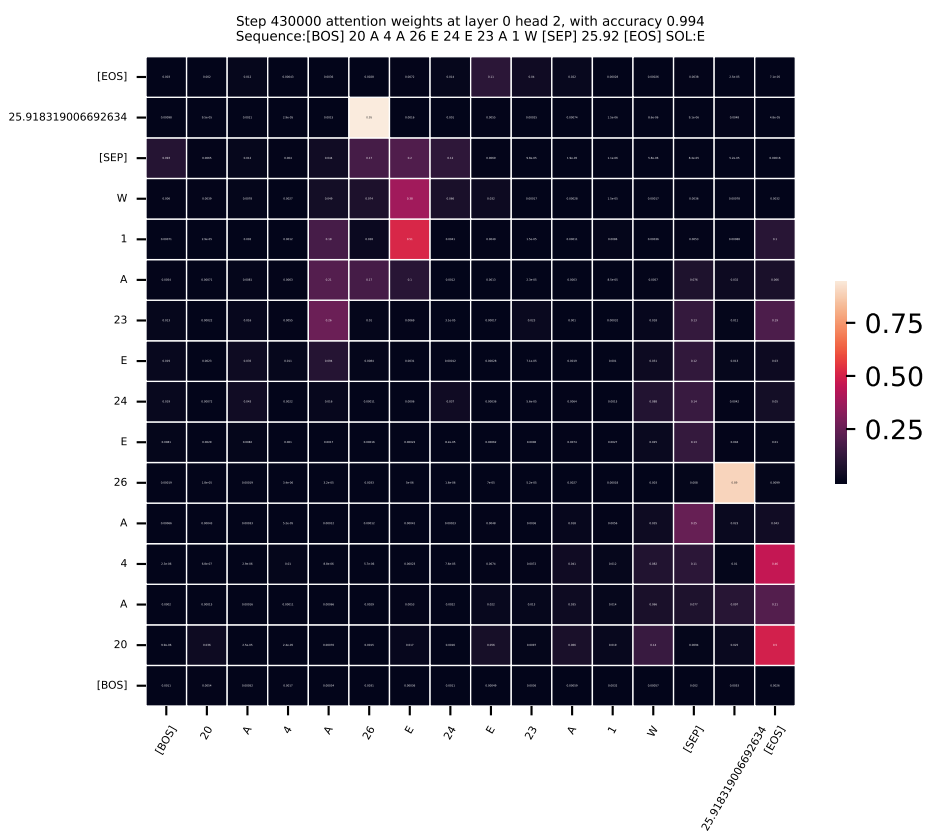

Step 429500 attention weights at layer 1 head 1, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

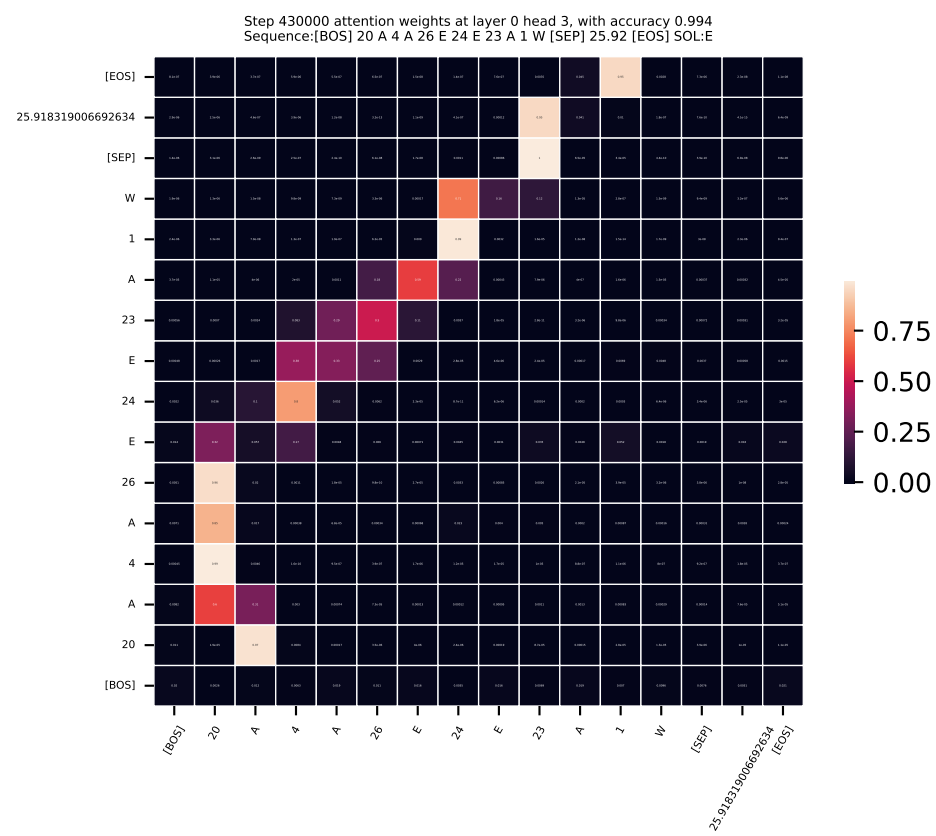

Step 429500 attention weights at layer 1 head 2, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

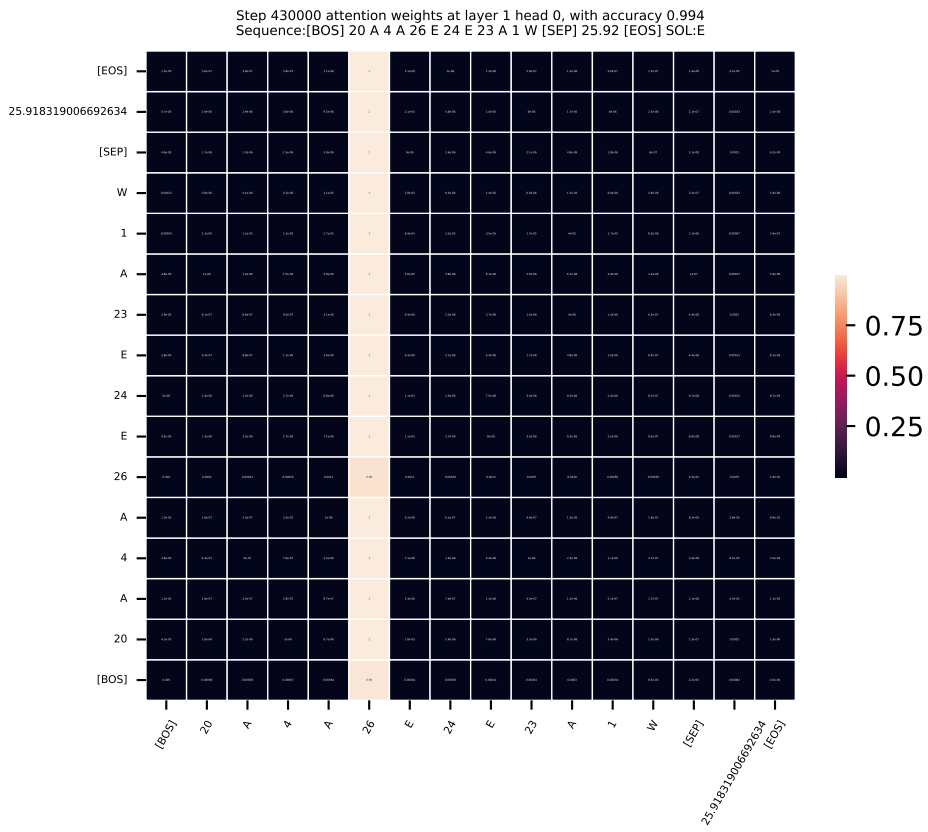

Step 429500 attention weights at layer 1 head 3, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

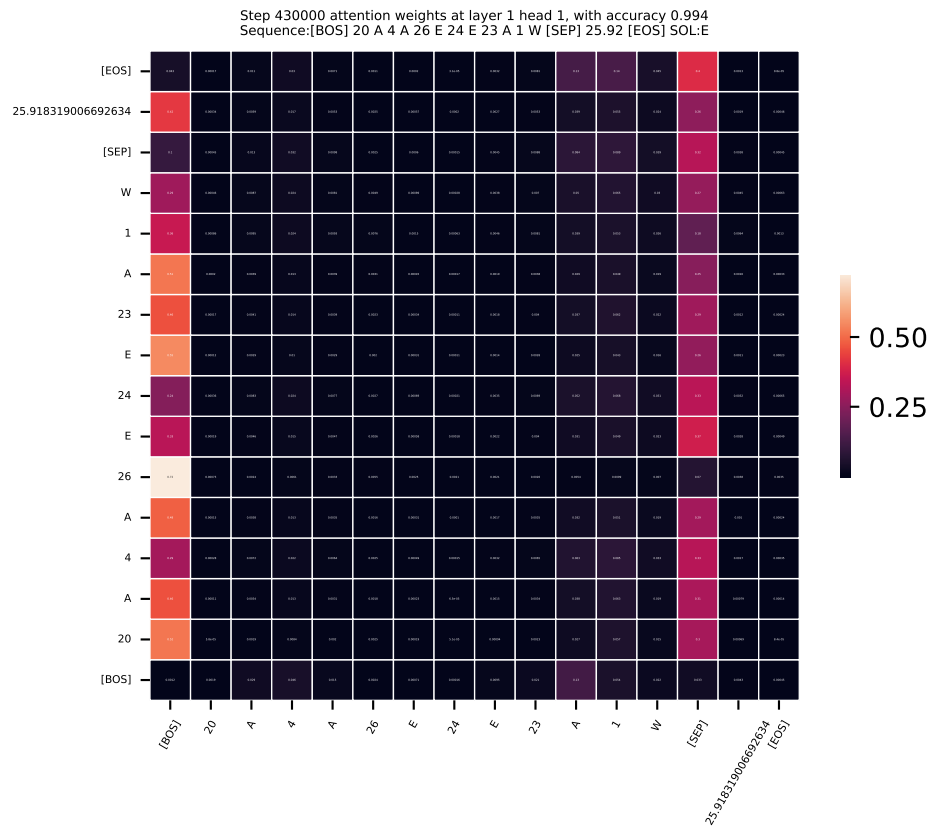

Step 430000 attention weights at layer 0 head 0, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

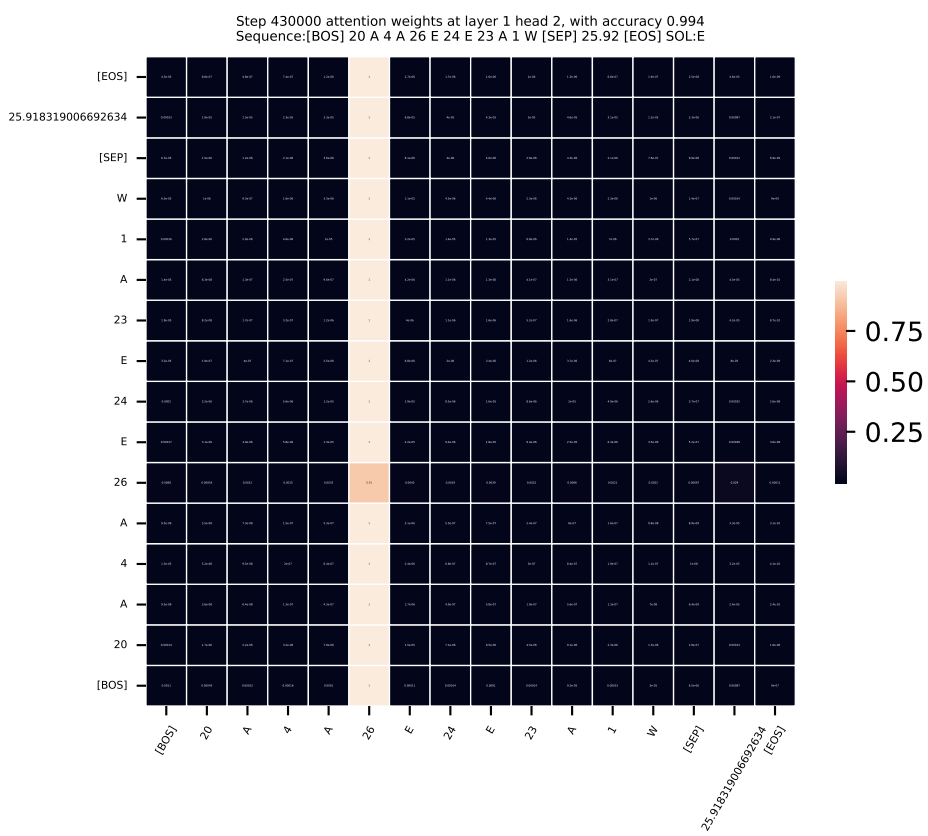

Step 430000 attention weights at layer 0 head 1, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

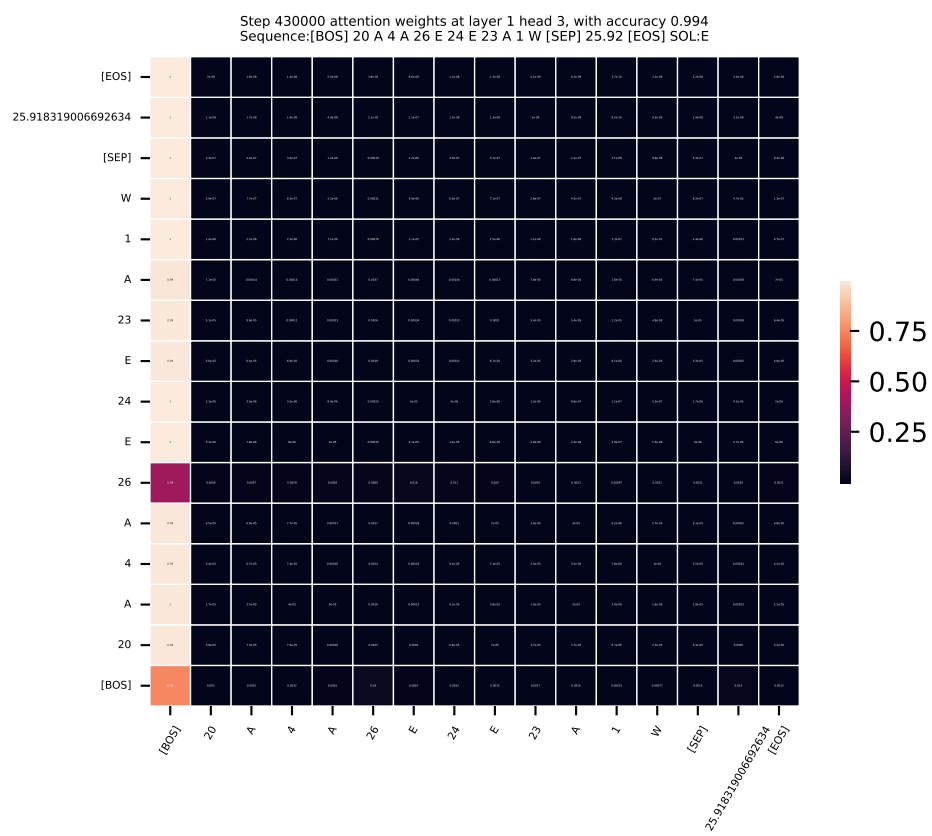

Step 430000 attention weights at layer 0 head 2, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

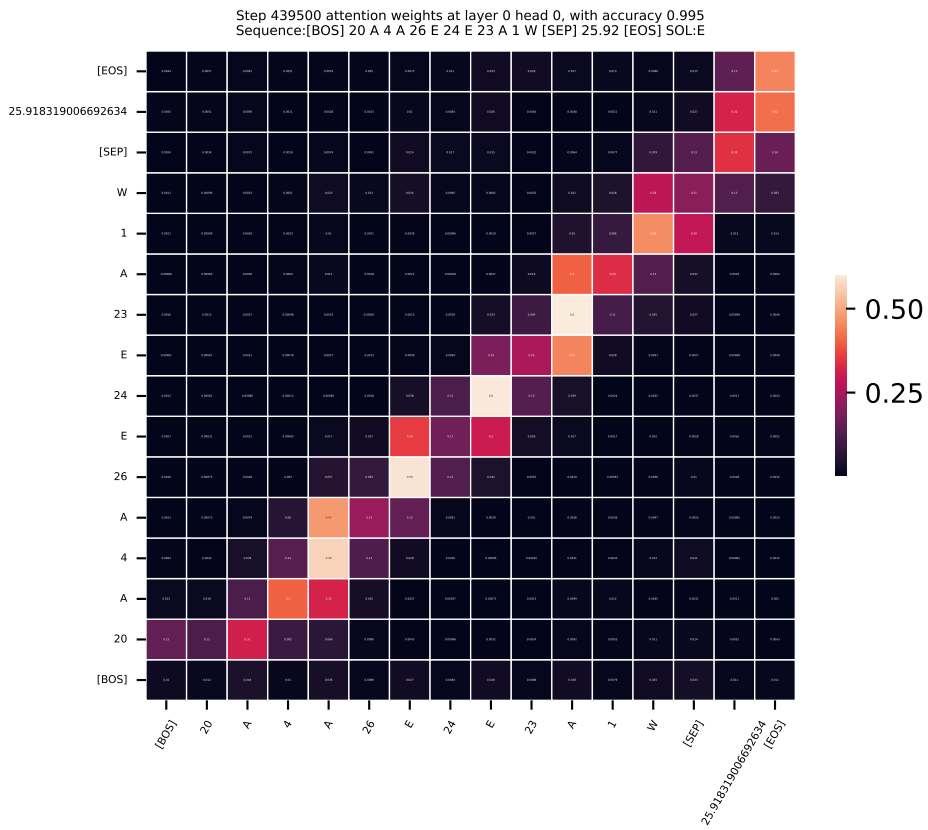

Step 430000 attention weights at layer 0 head 3, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

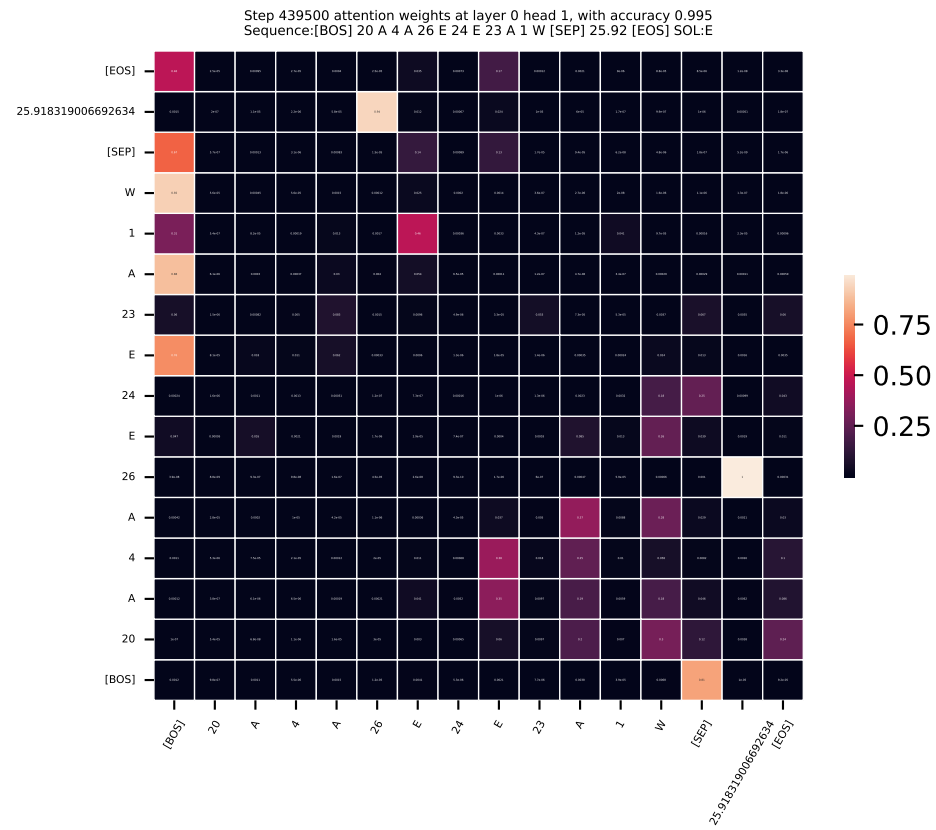

Step 430000 attention weights at layer 1 head 0, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

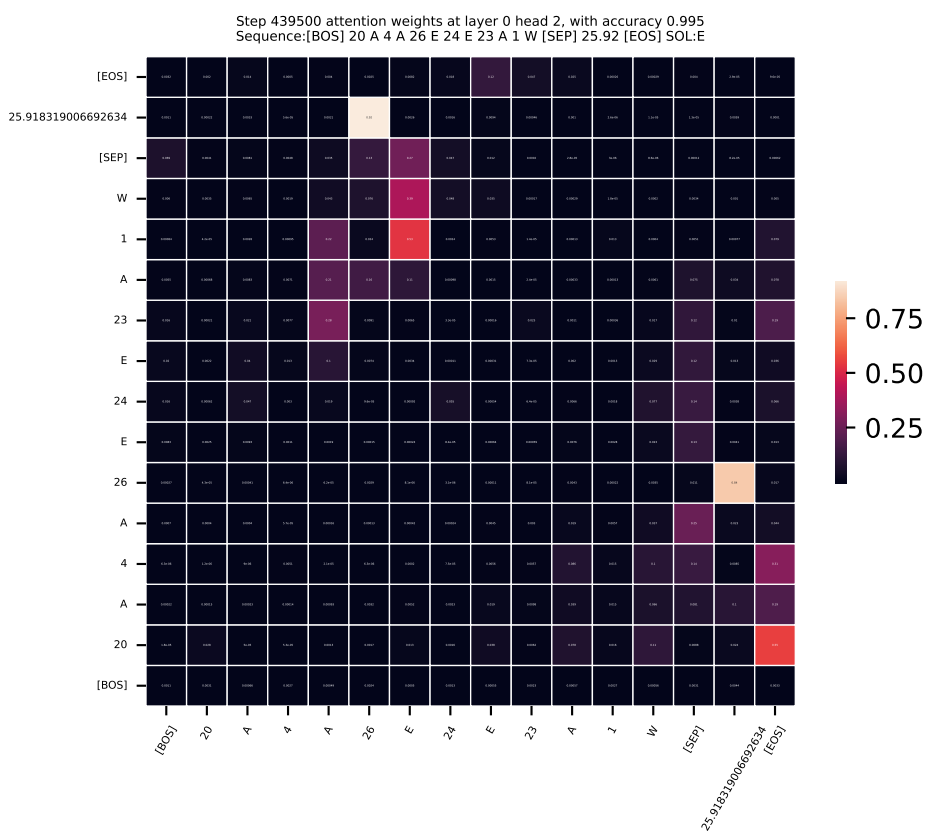

Step 430000 attention weights at layer 1 head 1, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

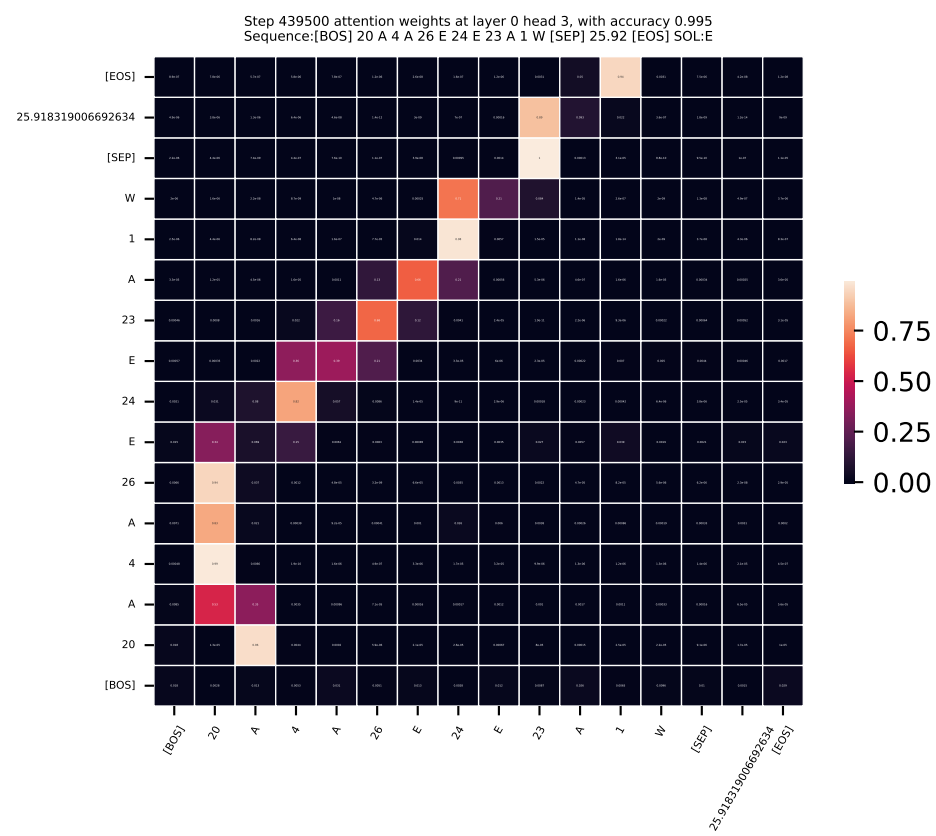

Step 430000 attention weights at layer 1 head 2, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

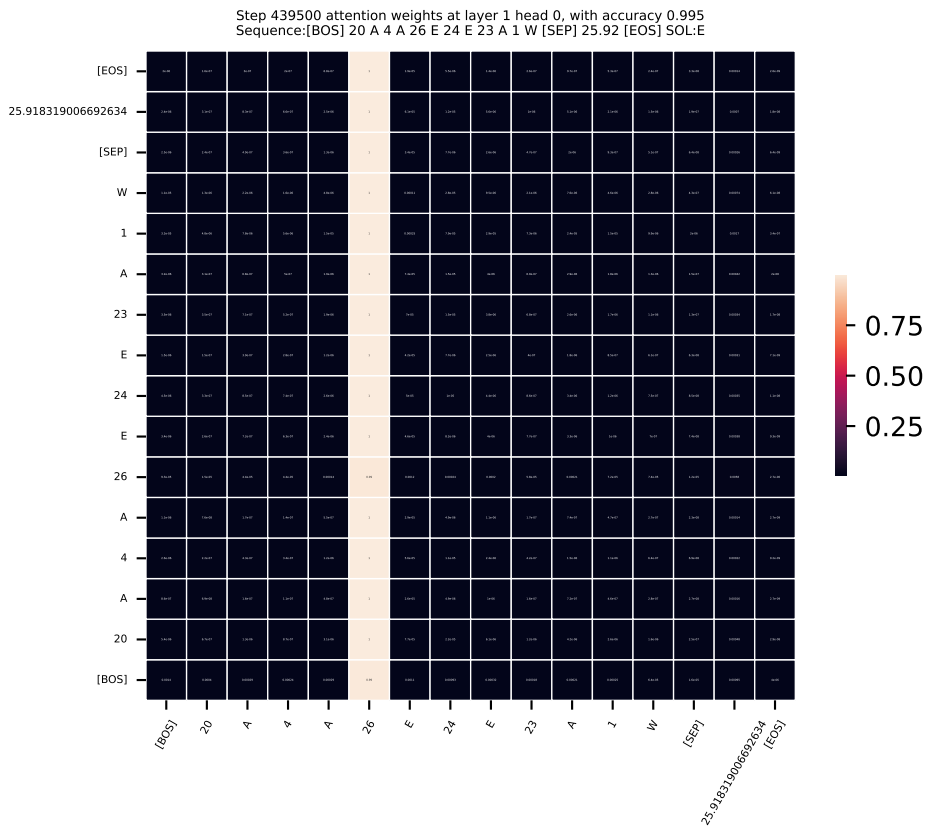

Step 430000 attention weights at layer 1 head 3, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

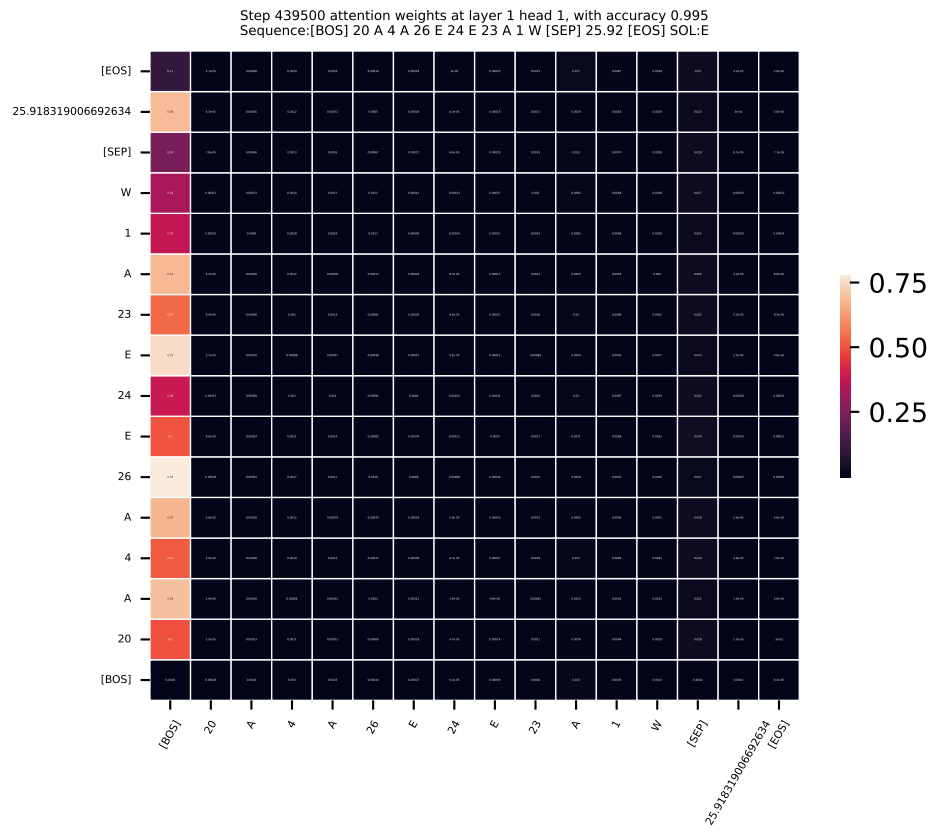

Step 439500 attention weights at layer 0 head 0, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

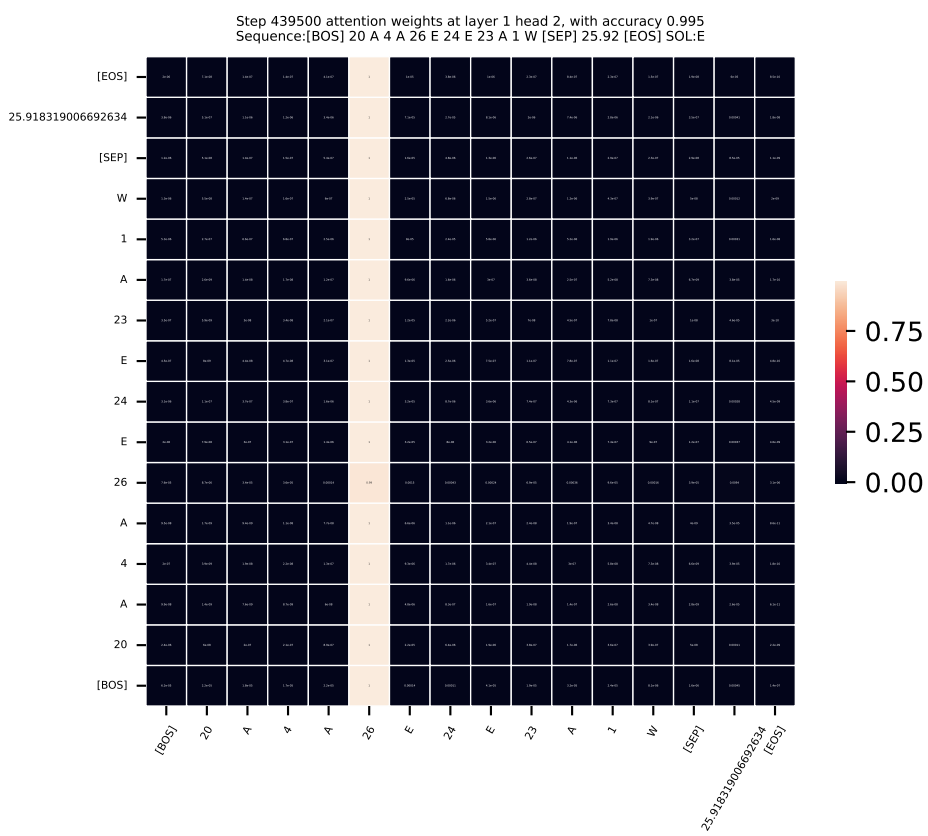

Step 439500 attention weights at layer 0 head 1, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

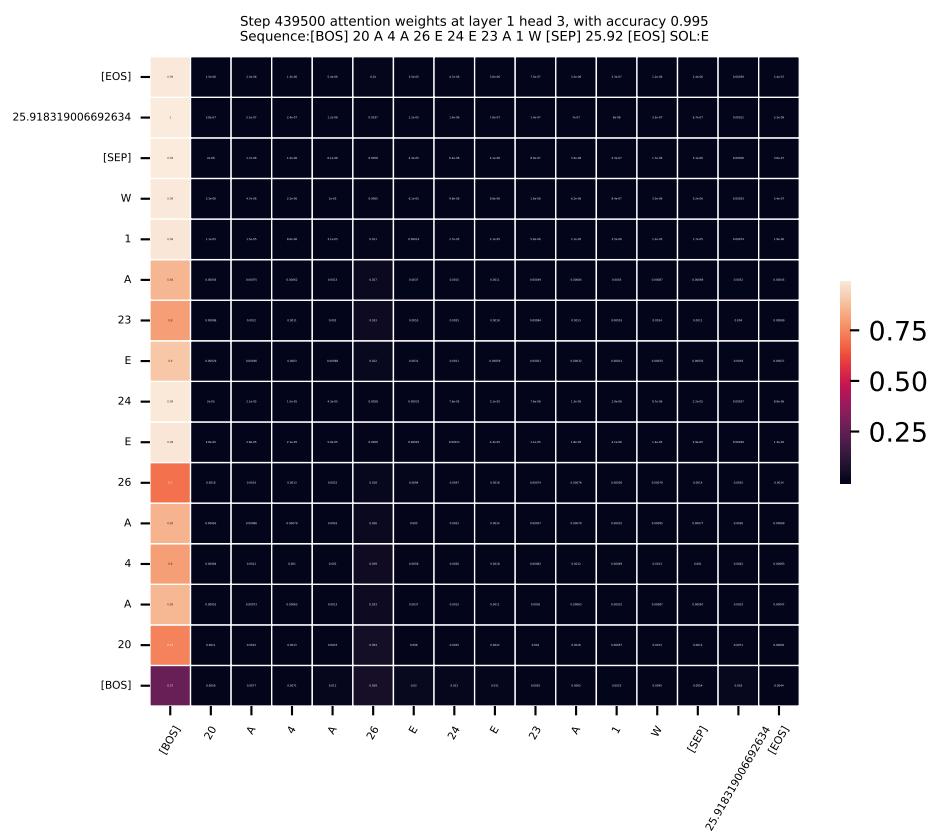

Step 439500 attention weights at layer 0 head 2, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

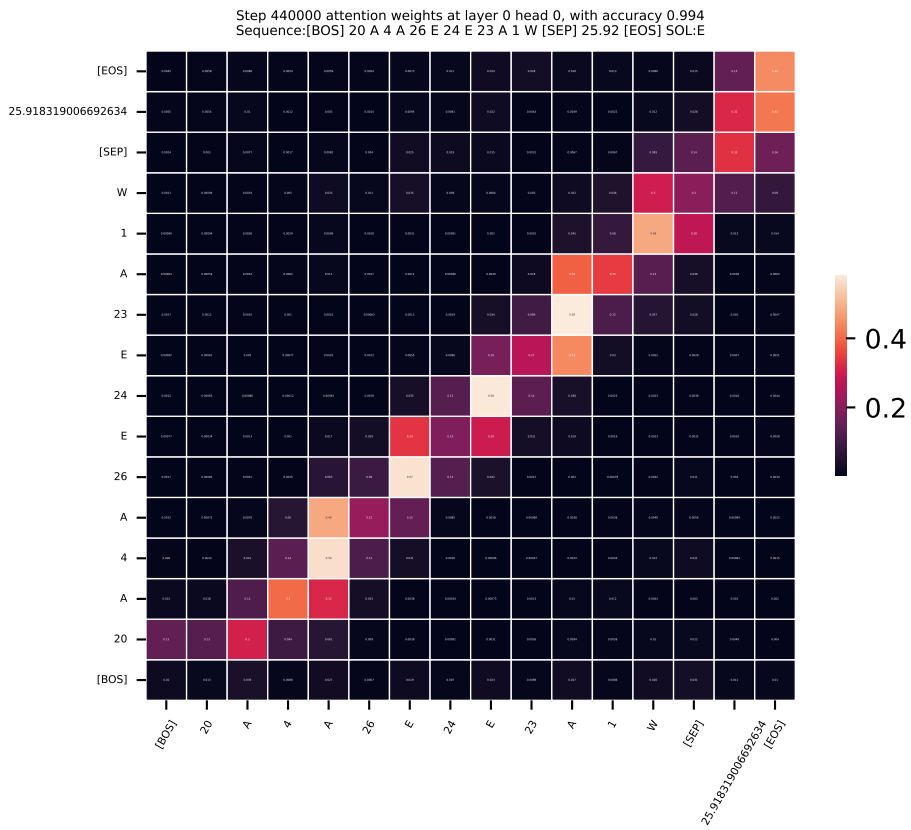

Step 439500 attention weights at layer 0 head 3, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

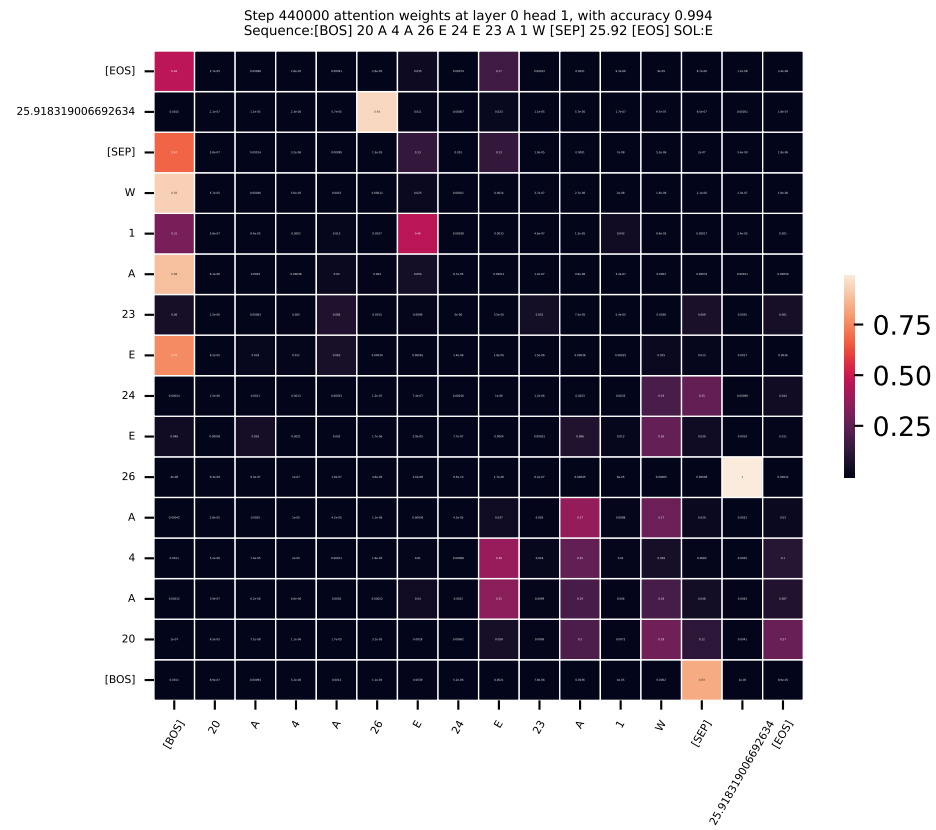

Step 439500 attention weights at layer 1 head 0, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

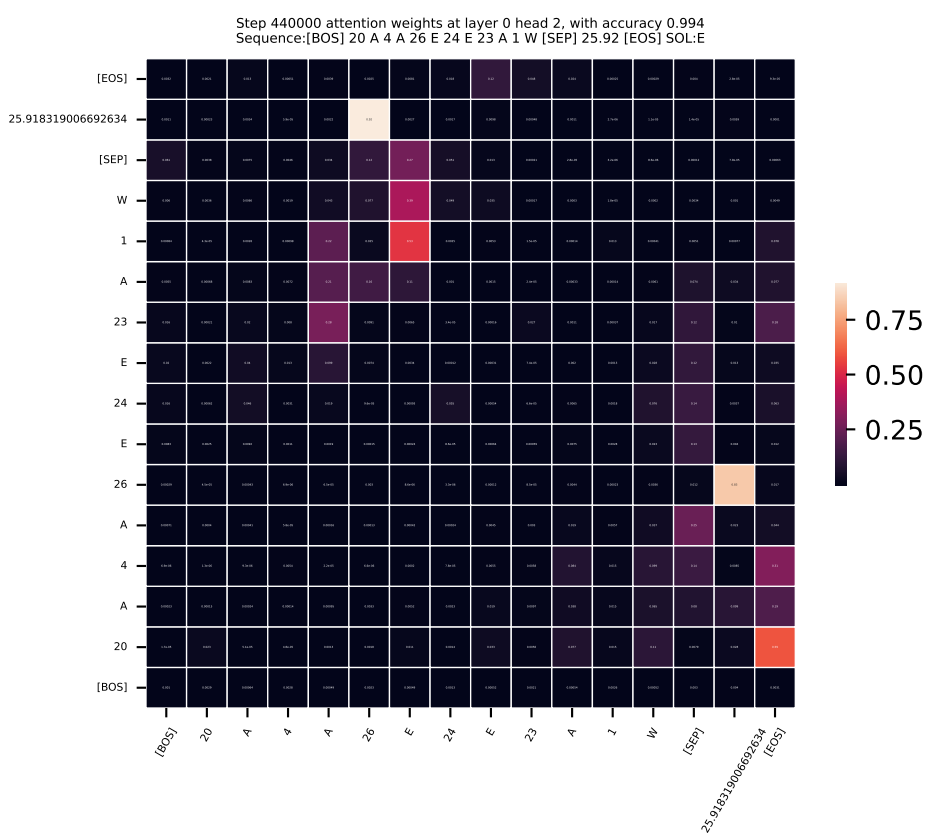

Step 439500 attention weights at layer 1 head 1, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

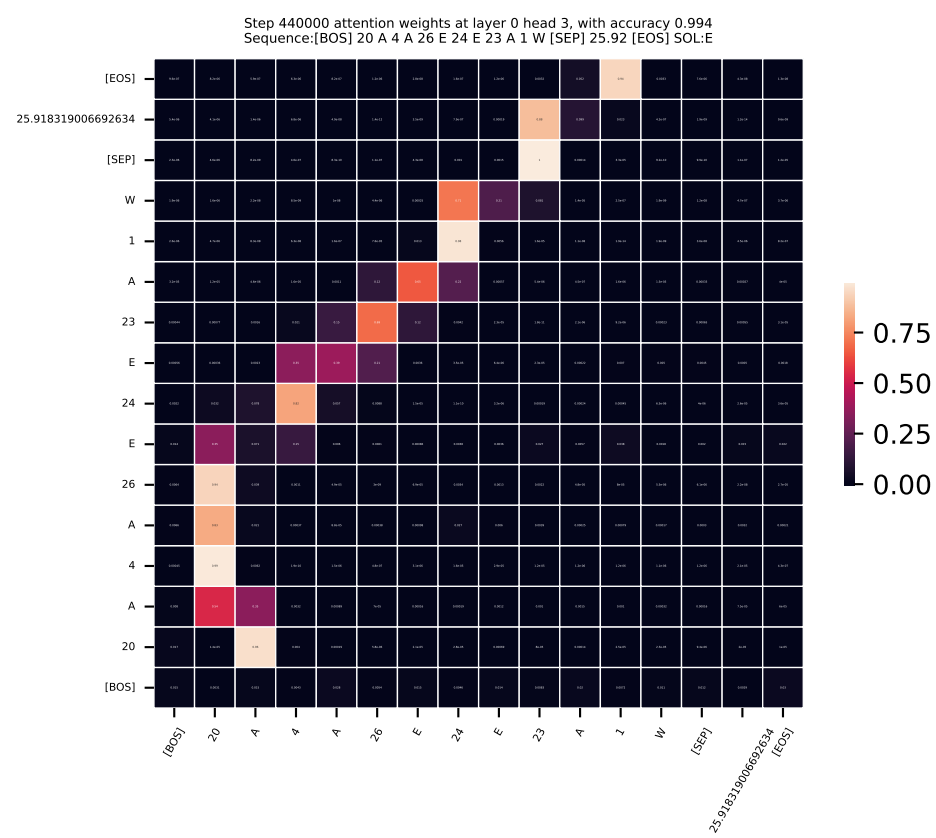

Step 439500 attention weights at layer 1 head 2, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

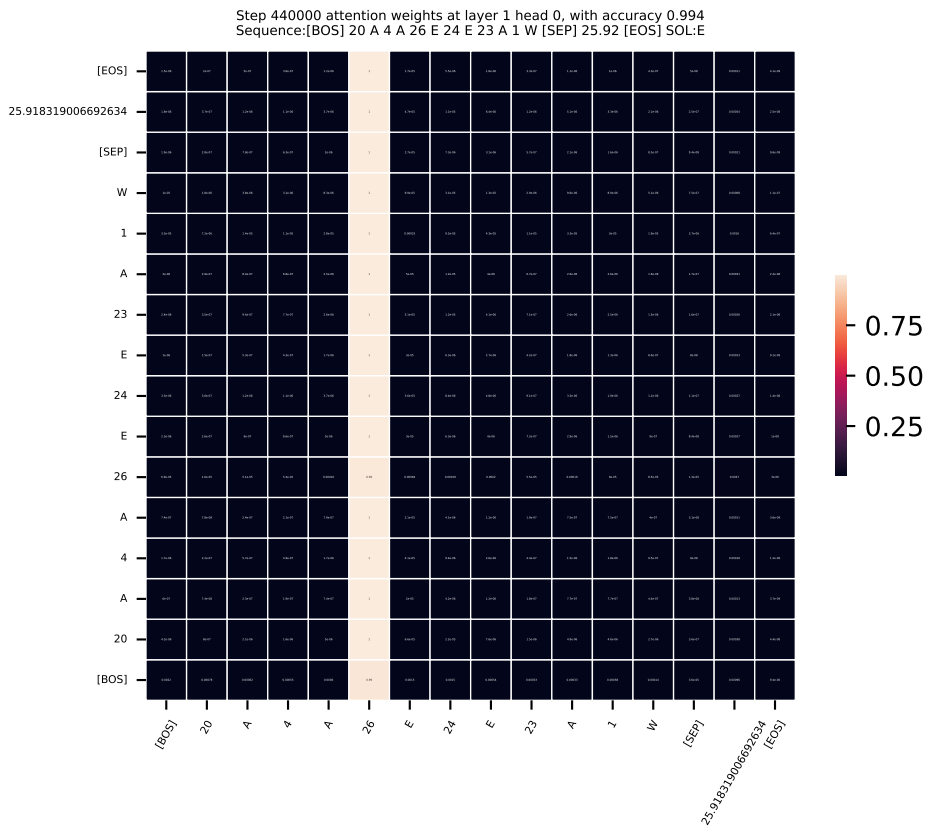

Step 439500 attention weights at layer 1 head 3, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

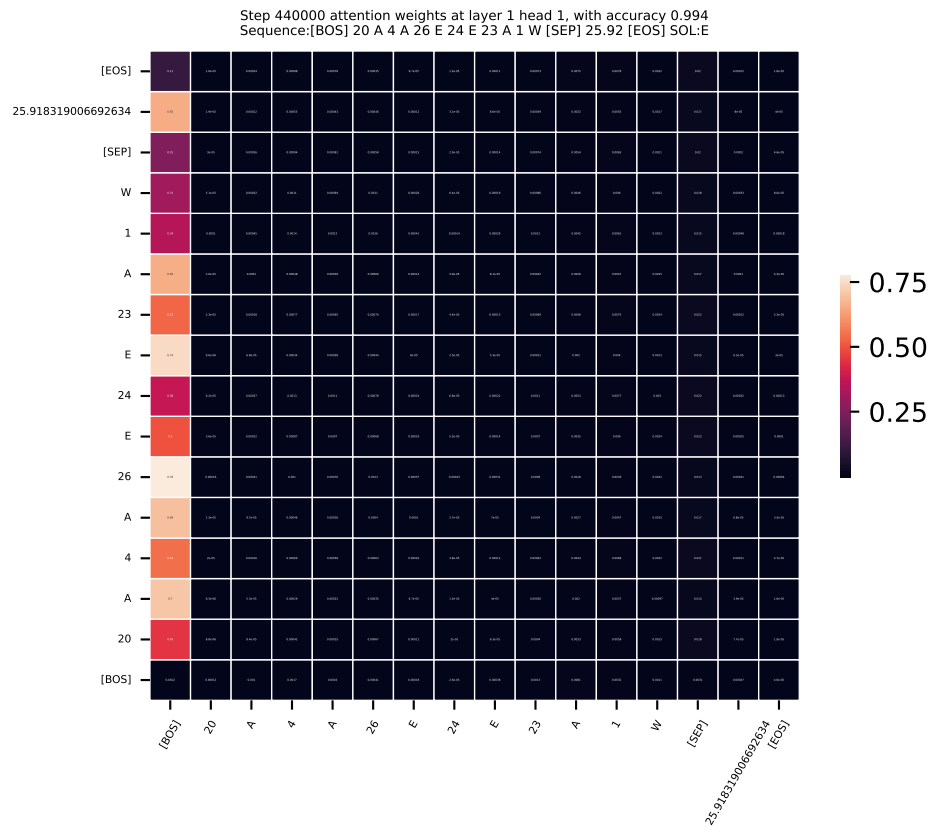

Step 440000 attention weights at layer 0 head 0, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

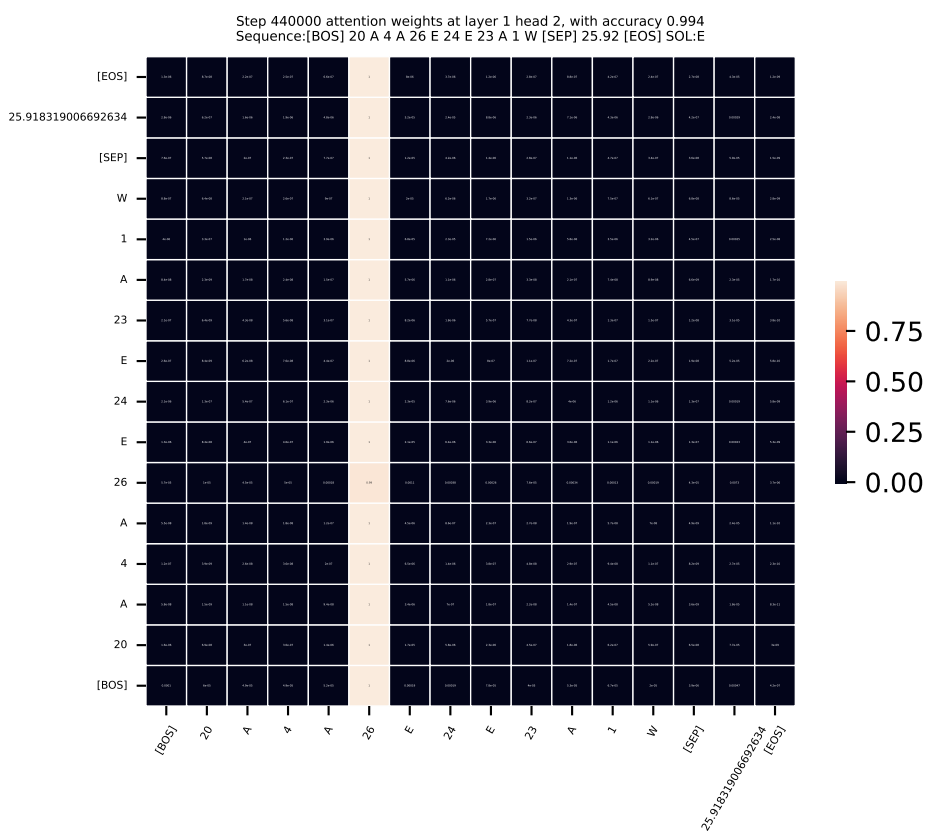

Step 440000 attention weights at layer 0 head 1, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

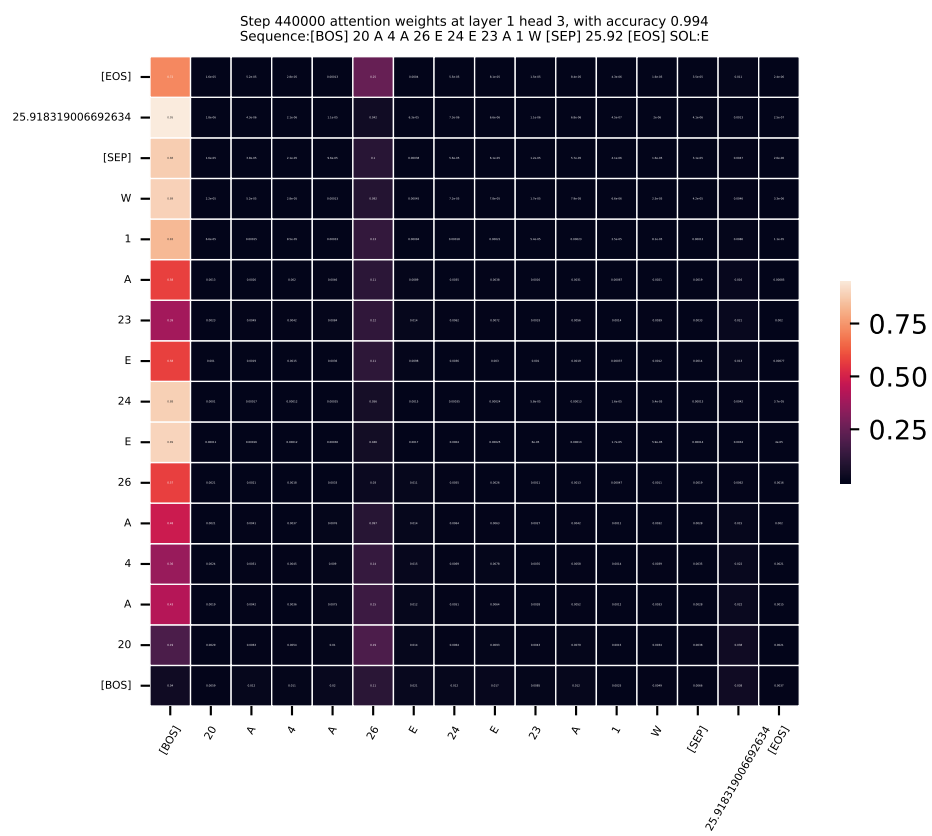

Step 440000 attention weights at layer 0 head 2, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

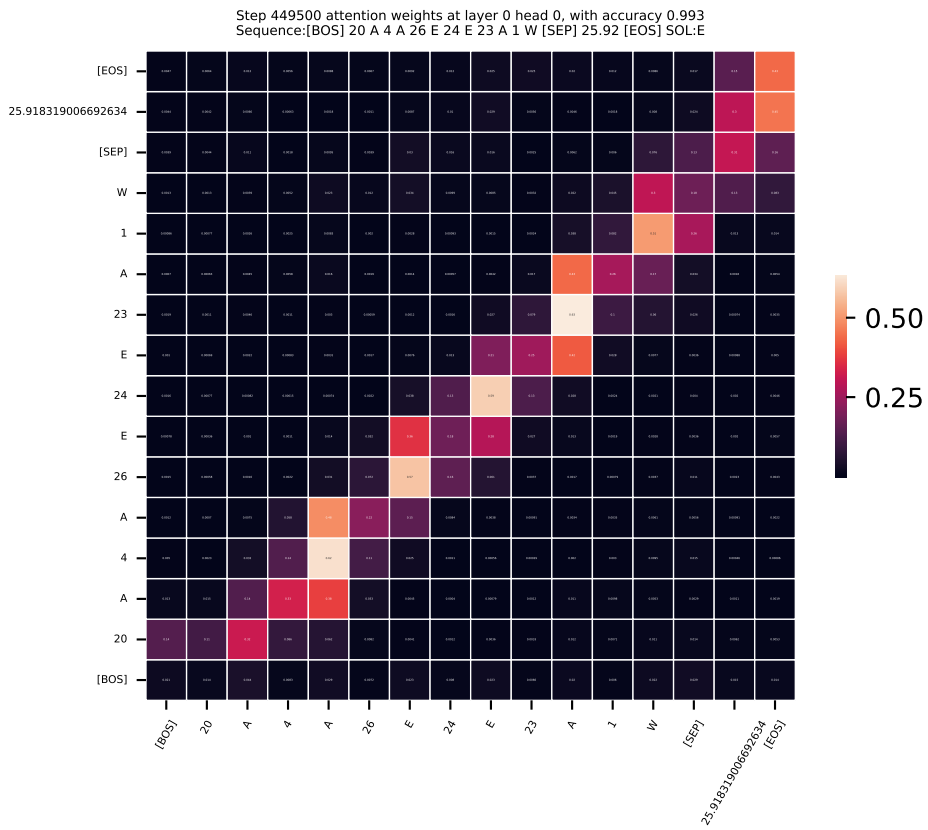

Step 440000 attention weights at layer 0 head 3, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

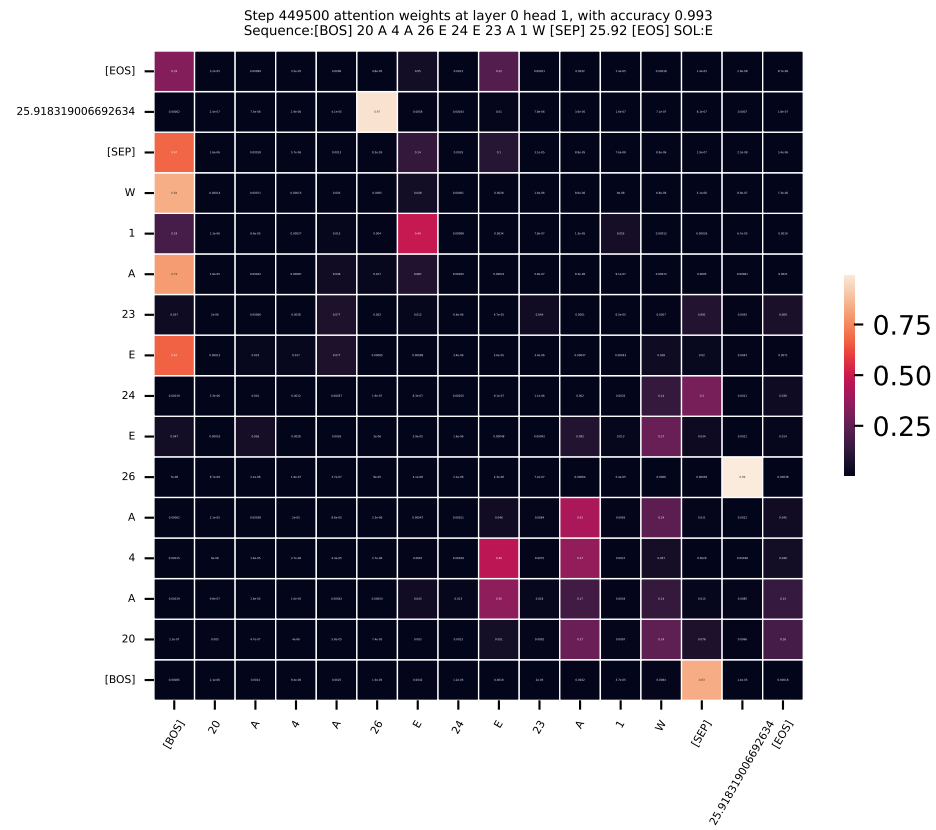

Step 440000 attention weights at layer 1 head 0, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

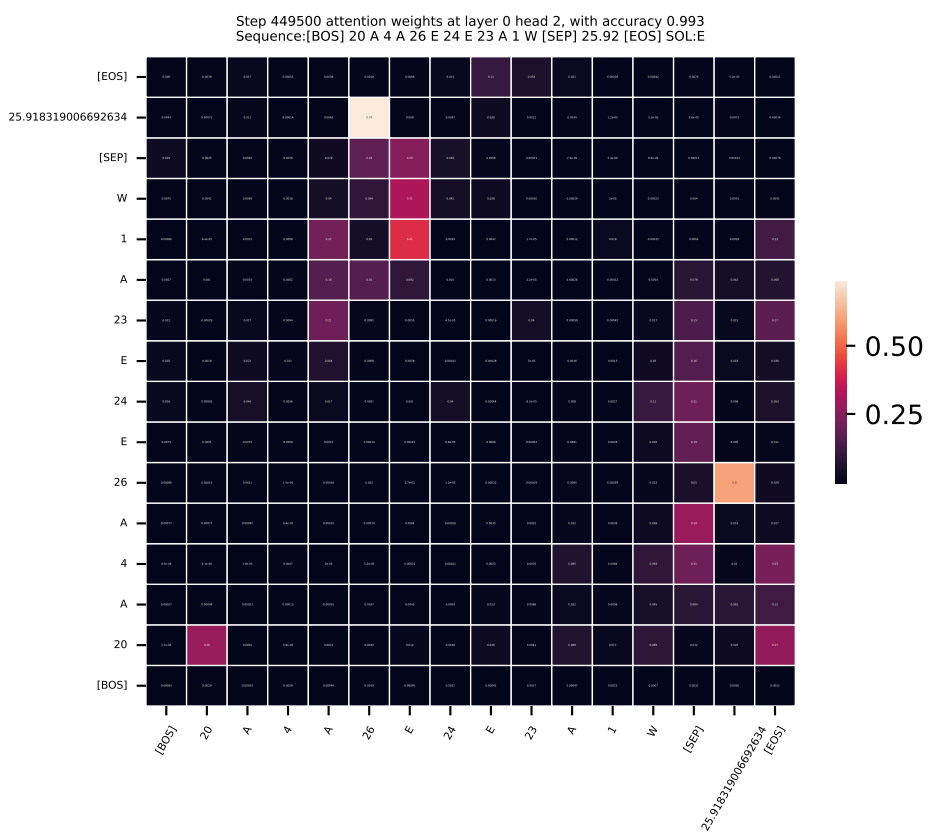

Step 440000 attention weights at layer 1 head 1, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

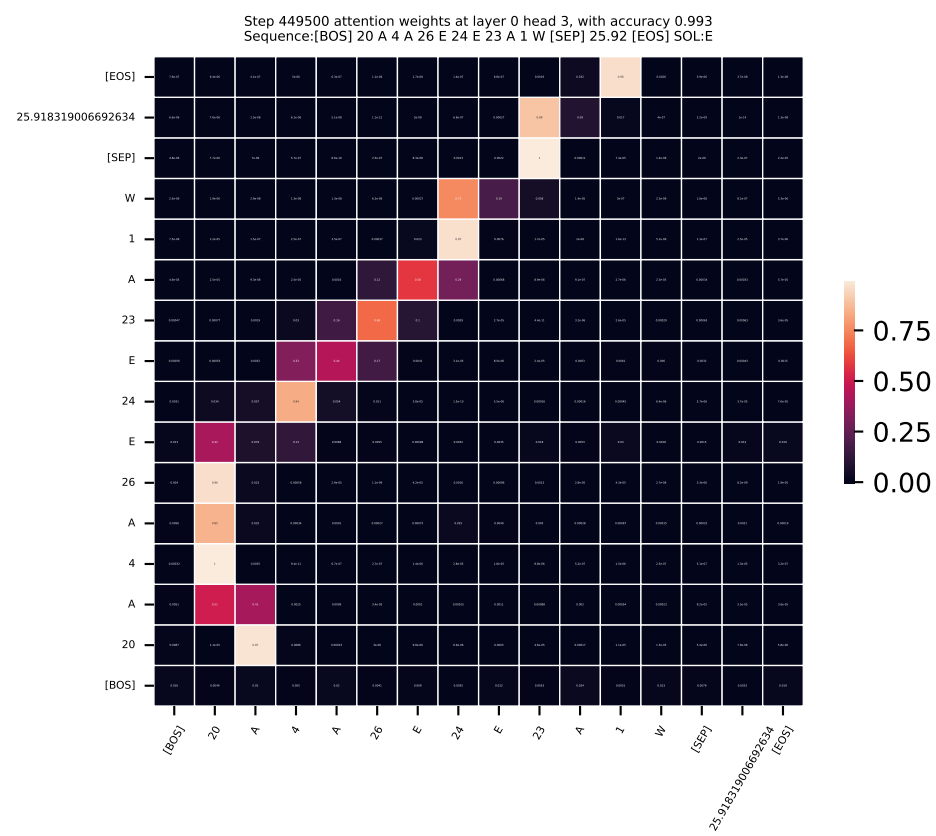

Step 440000 attention weights at layer 1 head 2, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

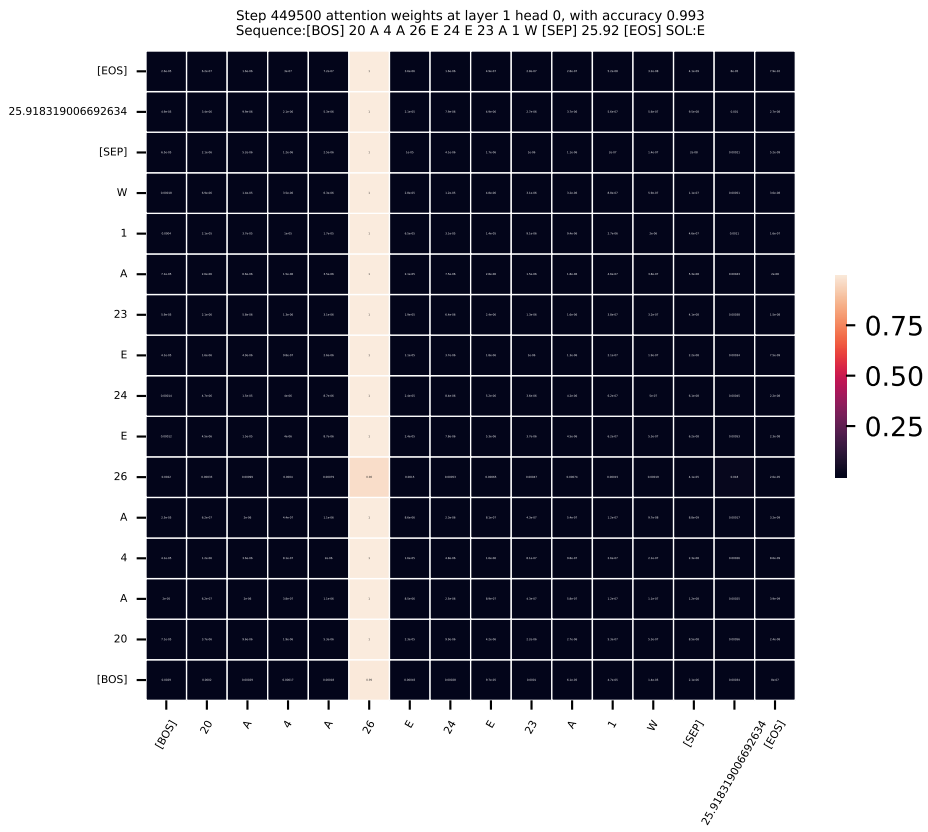

Step 440000 attention weights at layer 1 head 3, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

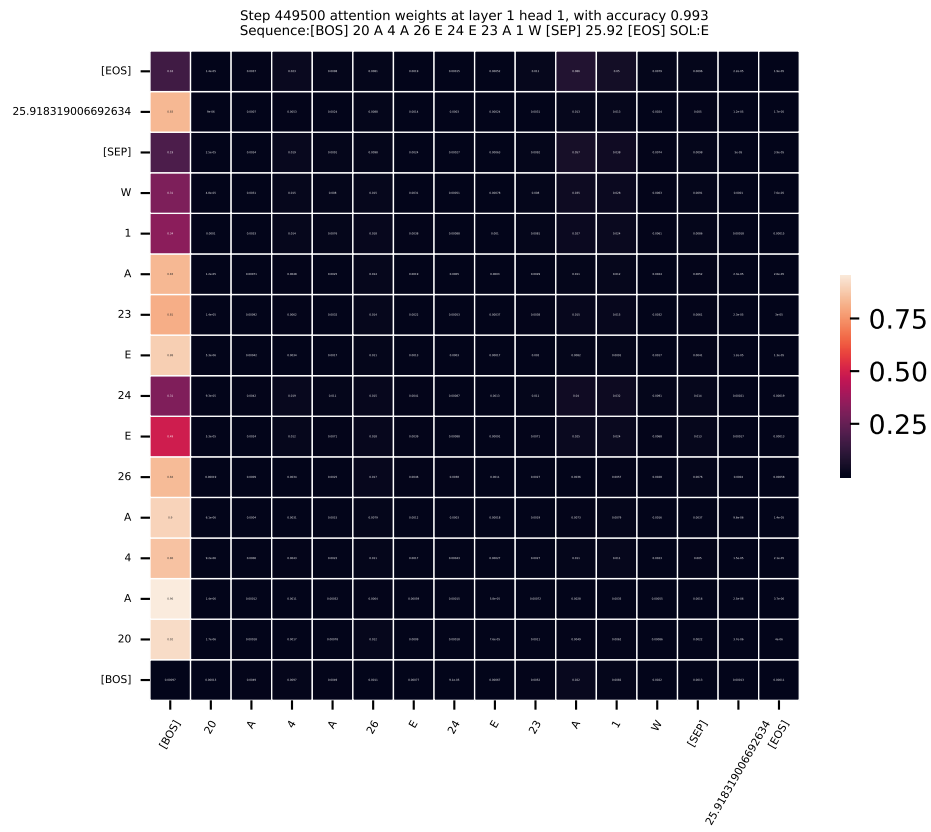

Step 449500 attention weights at layer 0 head 0, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

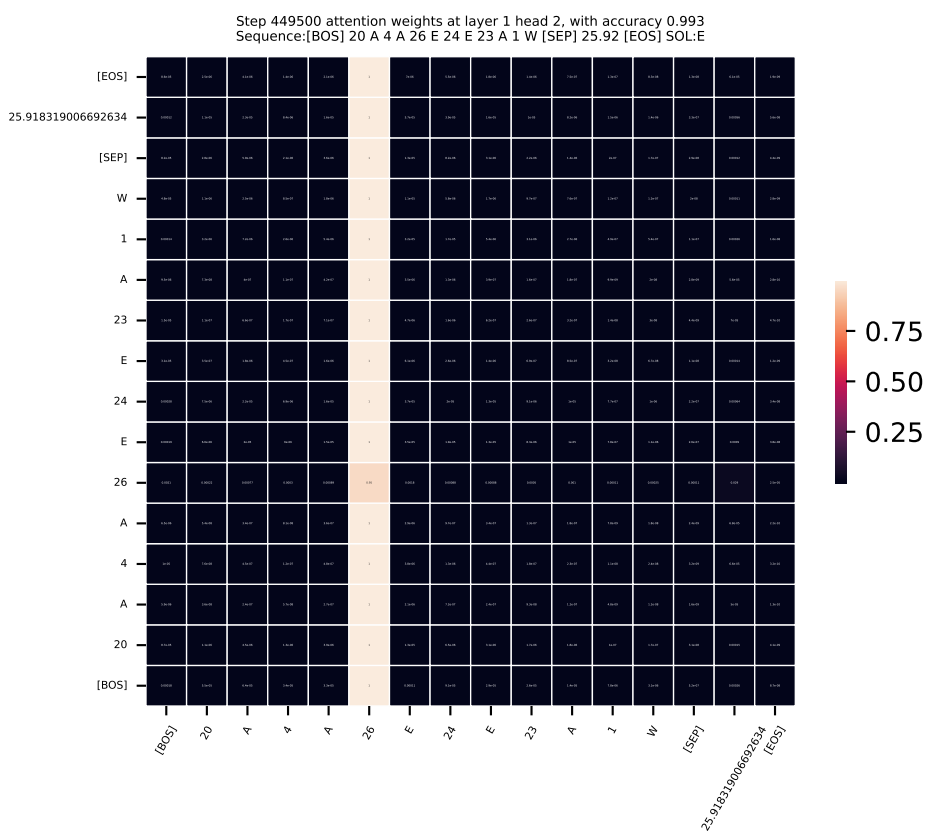

Step 449500 attention weights at layer 0 head 1, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

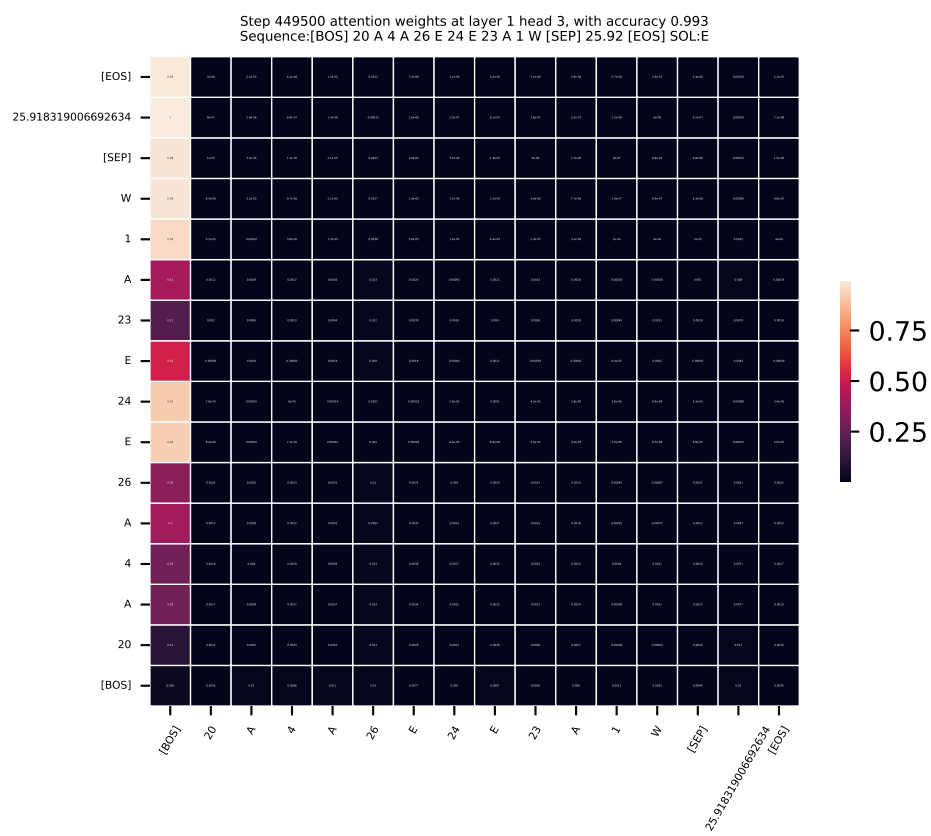

Step 449500 attention weights at layer 0 head 2, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

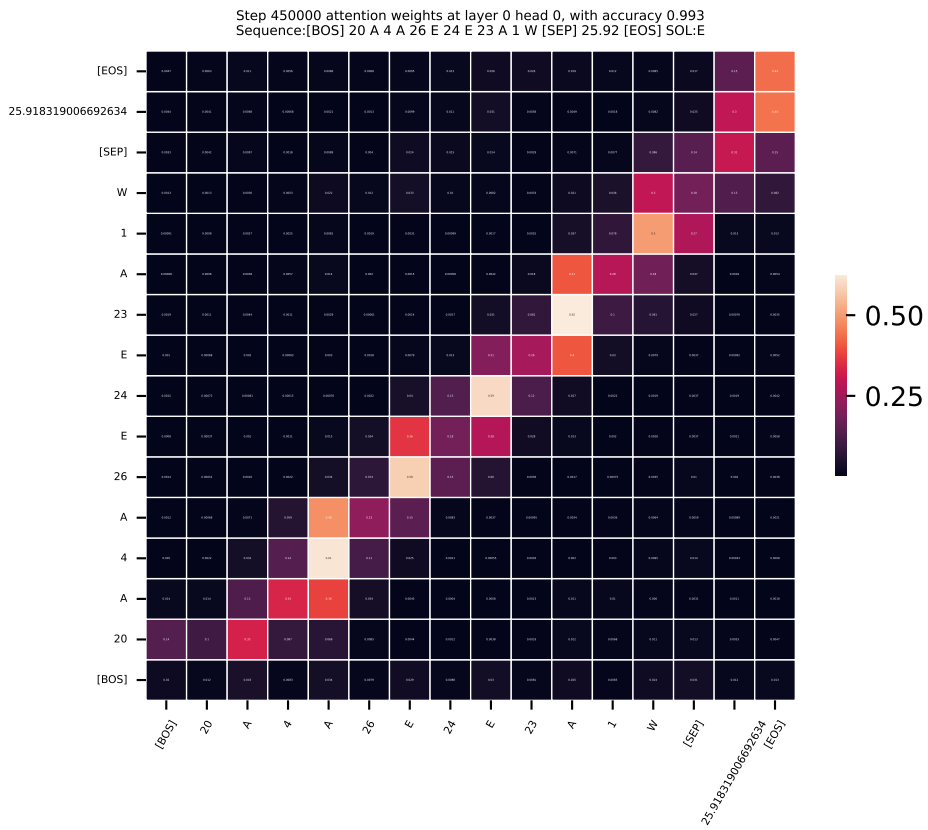

Step 449500 attention weights at layer 0 head 3, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

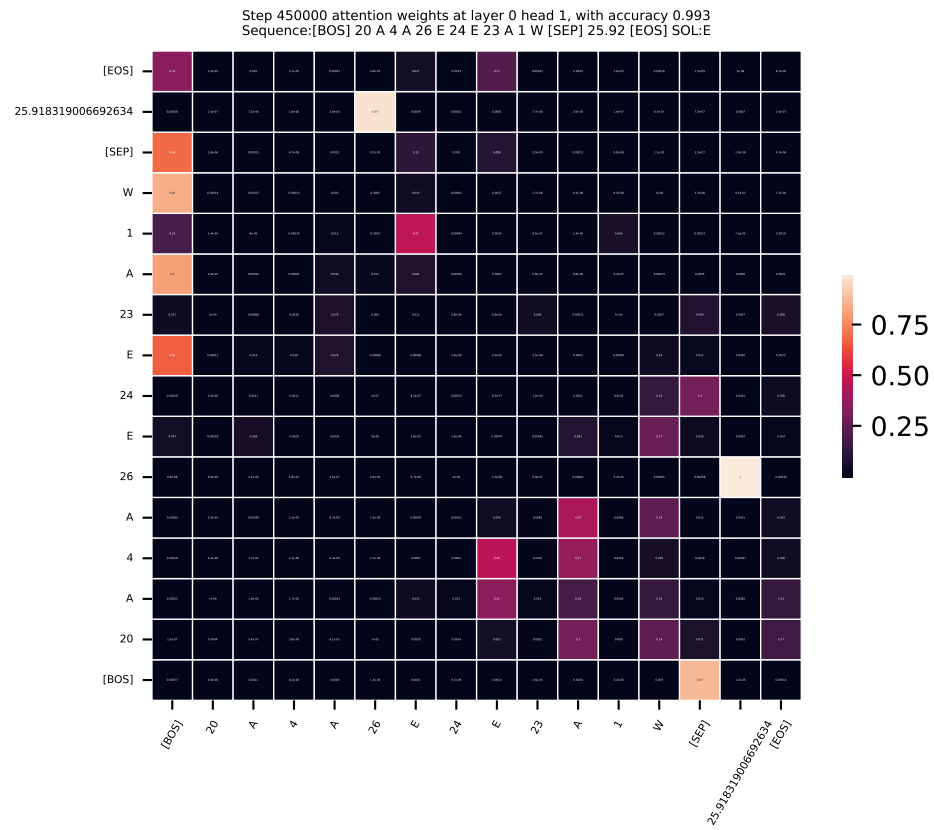

Step 449500 attention weights at layer 1 head 0, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

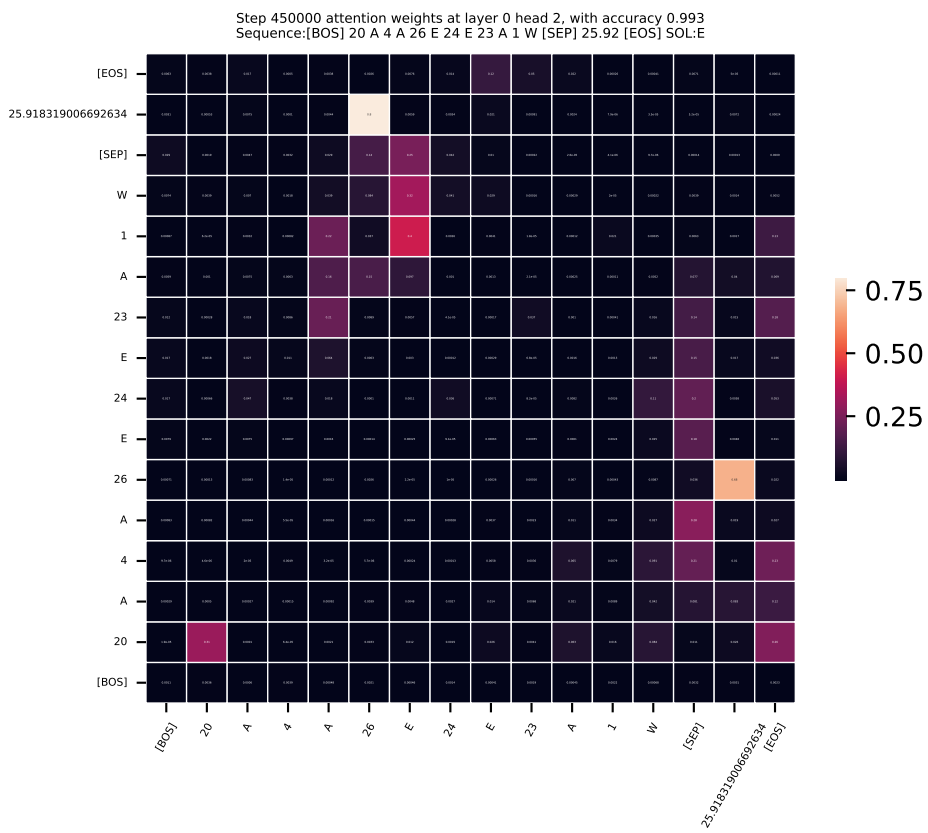

Step 449500 attention weights at layer 1 head 1, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

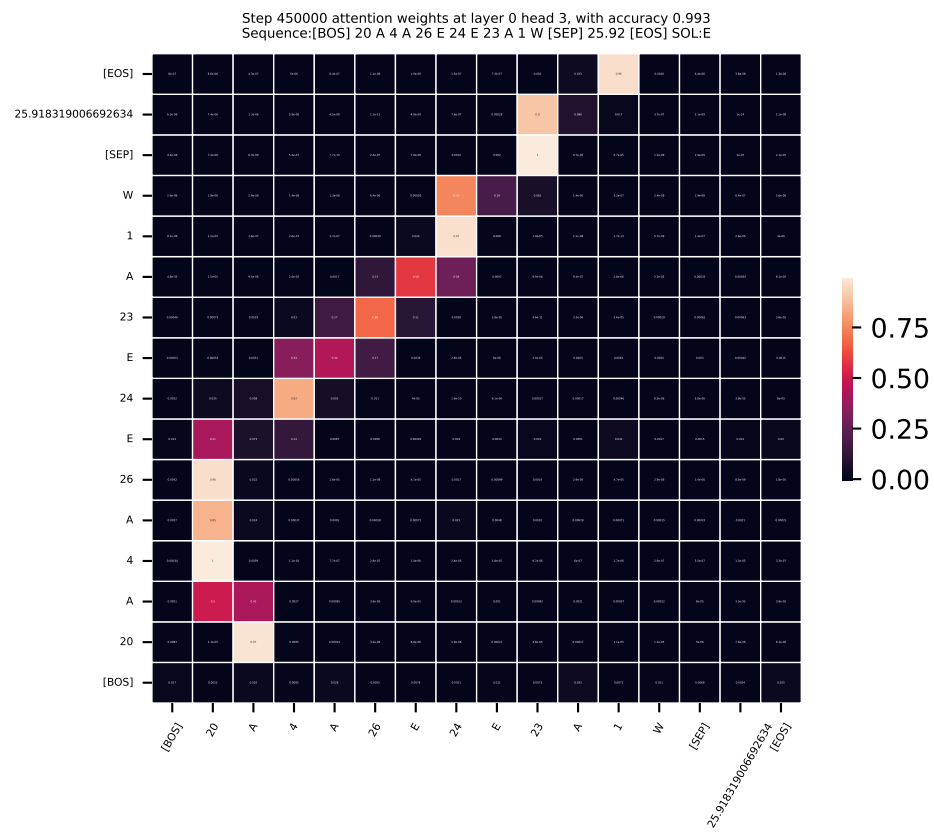

Step 449500 attention weights at layer 1 head 2, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

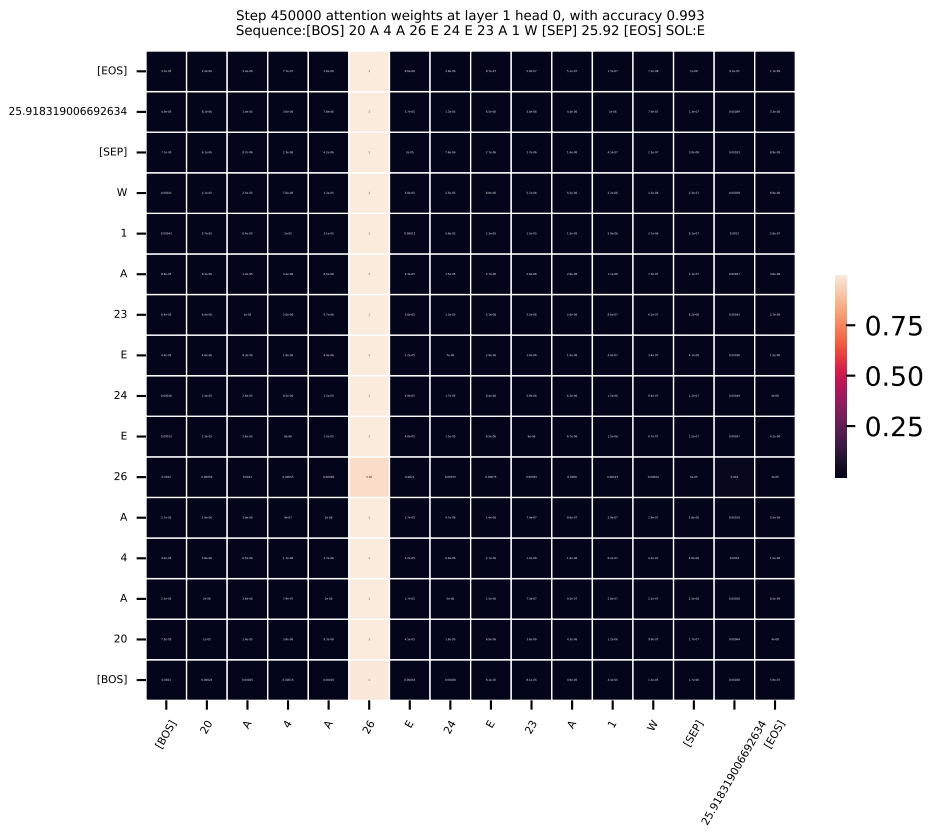

Step 449500 attention weights at layer 1 head 3, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

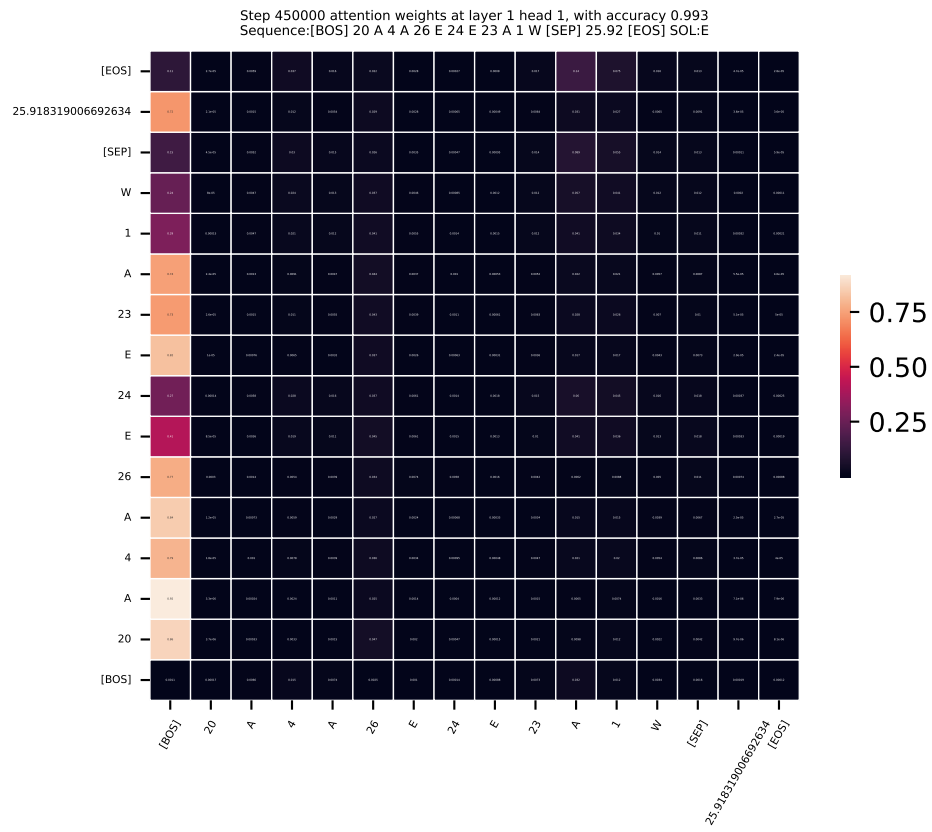

Step 450000 attention weights at layer 0 head 0, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.50
0.25

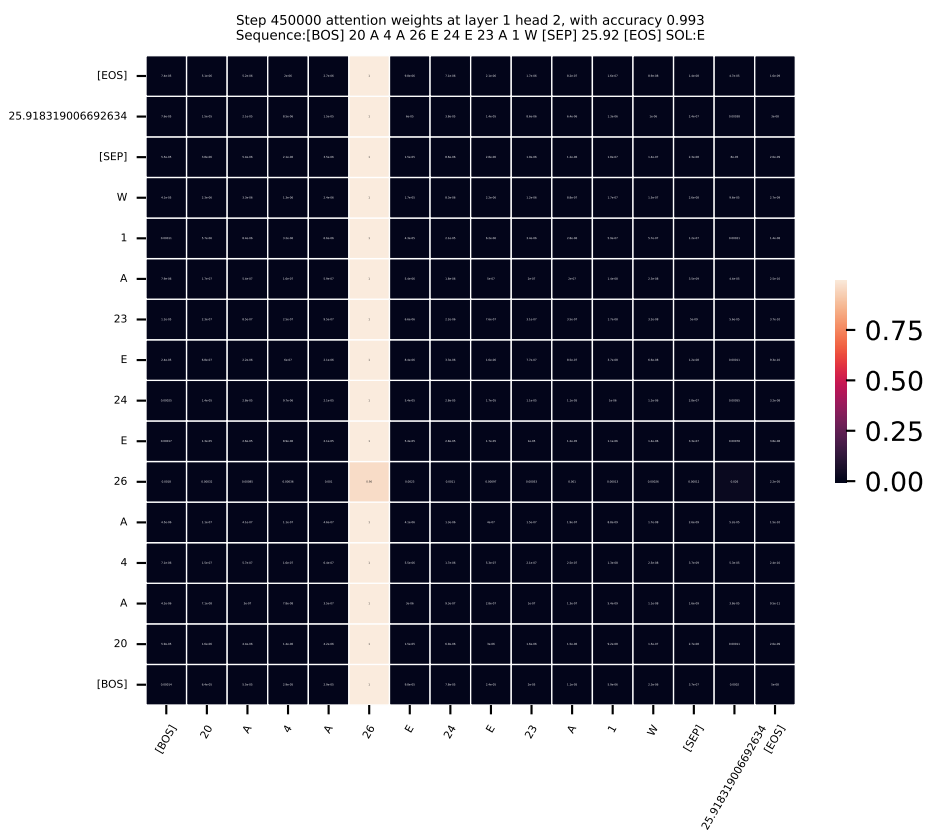

Step 450000 attention weights at layer 0 head 1, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

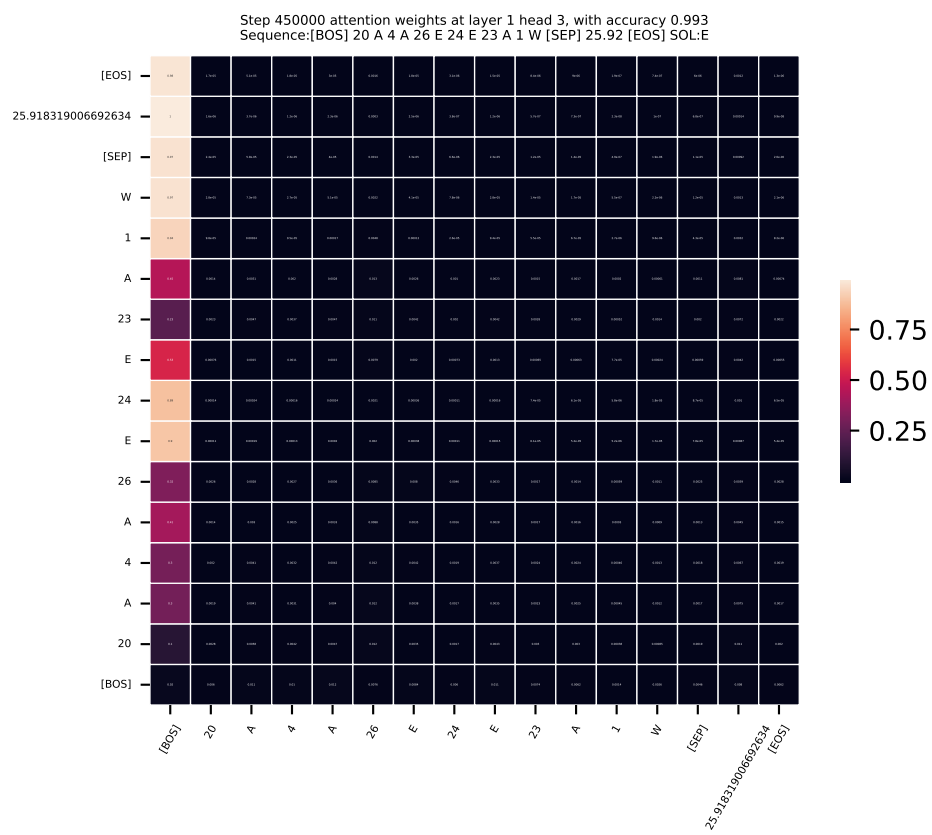

Step 450000 attention weights at layer 0 head 2, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

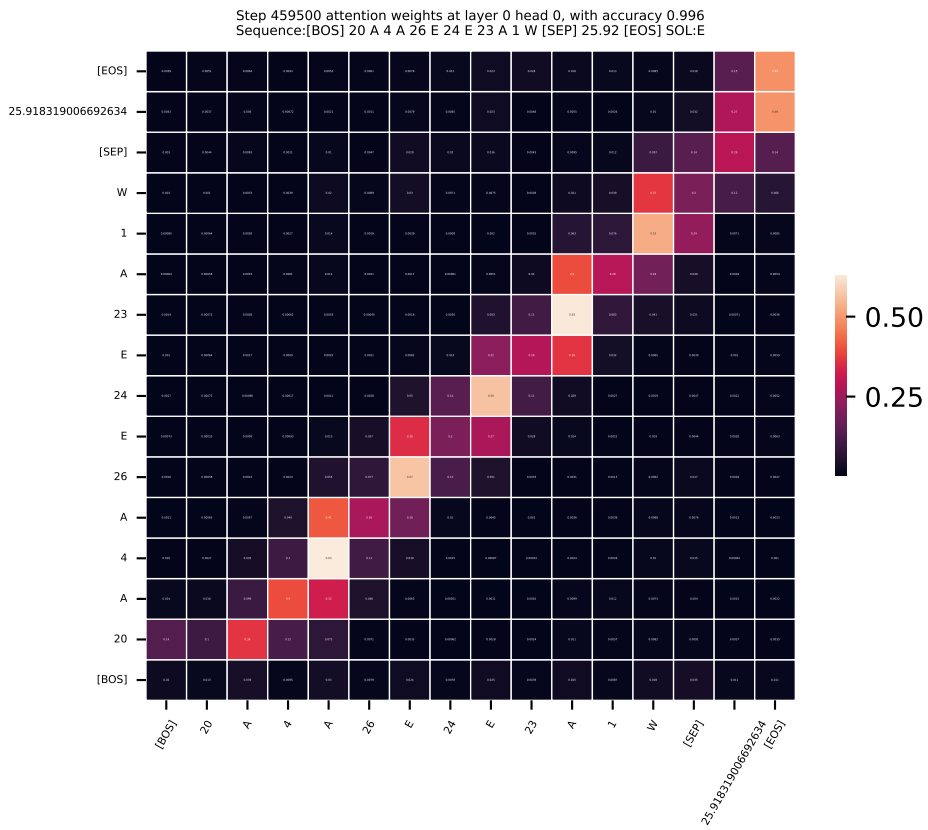

Step 450000 attention weights at layer 0 head 3, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25
0.00

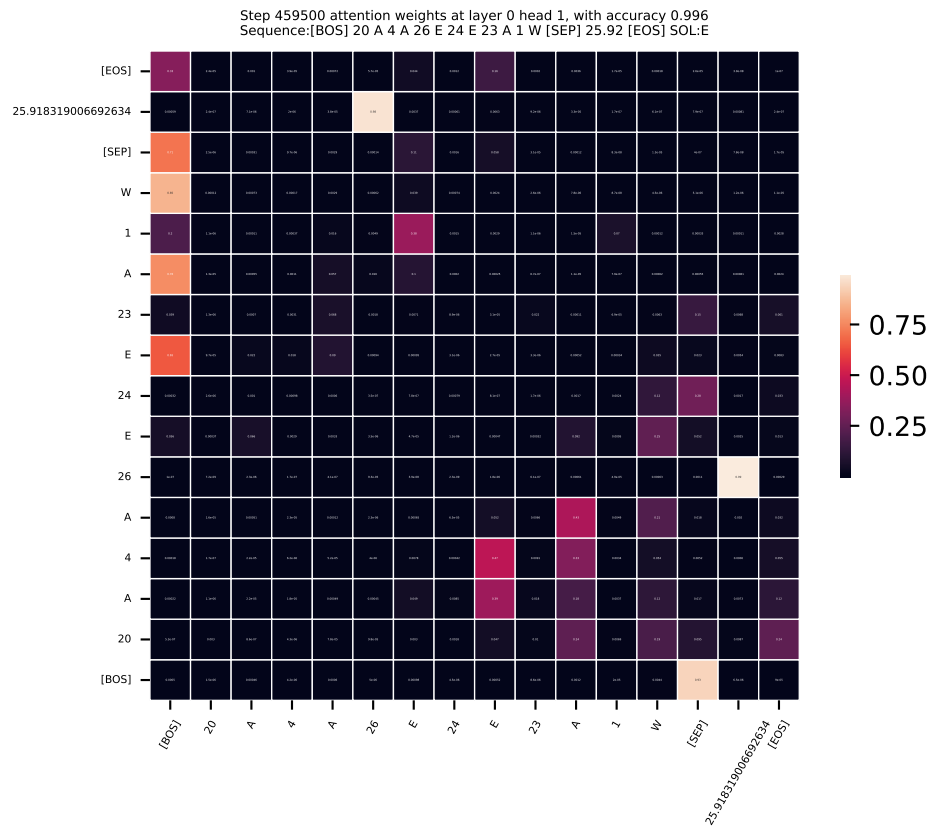

Step 450000 attention weights at layer 1 head 0, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

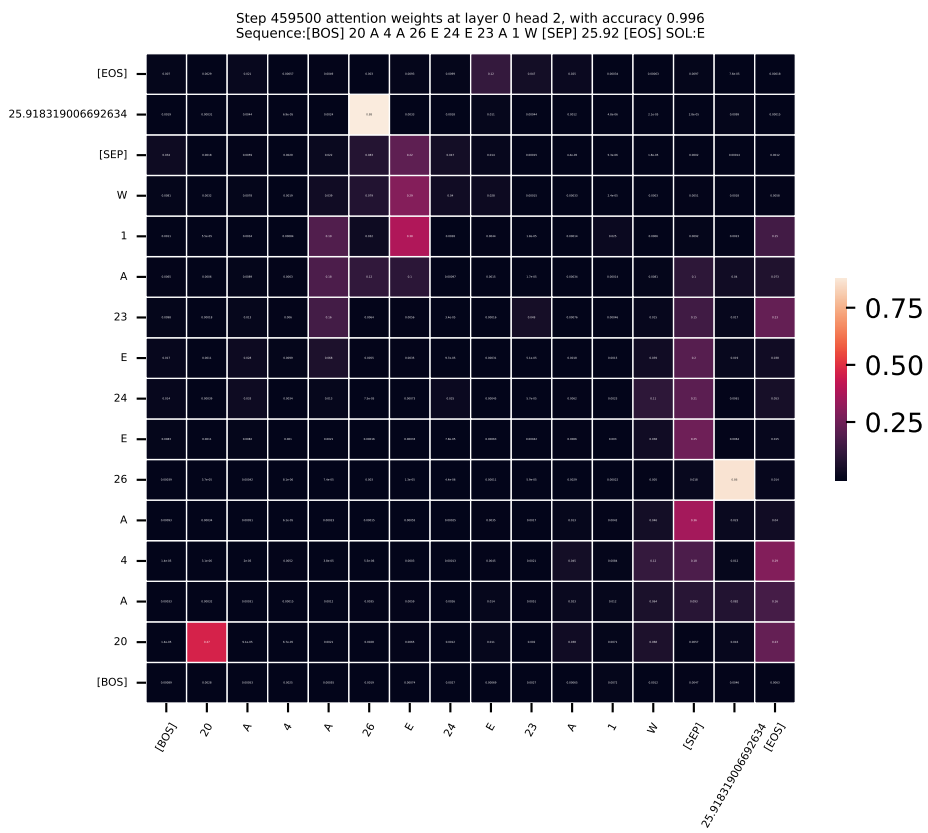

Step 450000 attention weights at layer 1 head 1, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

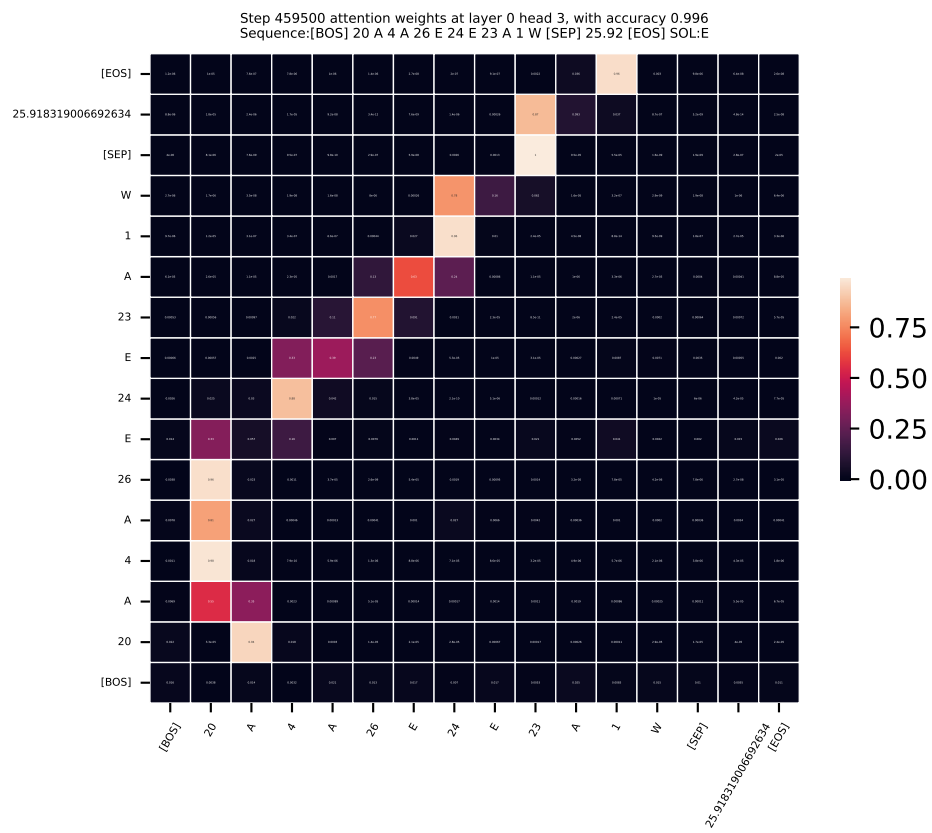

Step 450000 attention weights at layer 1 head 2, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

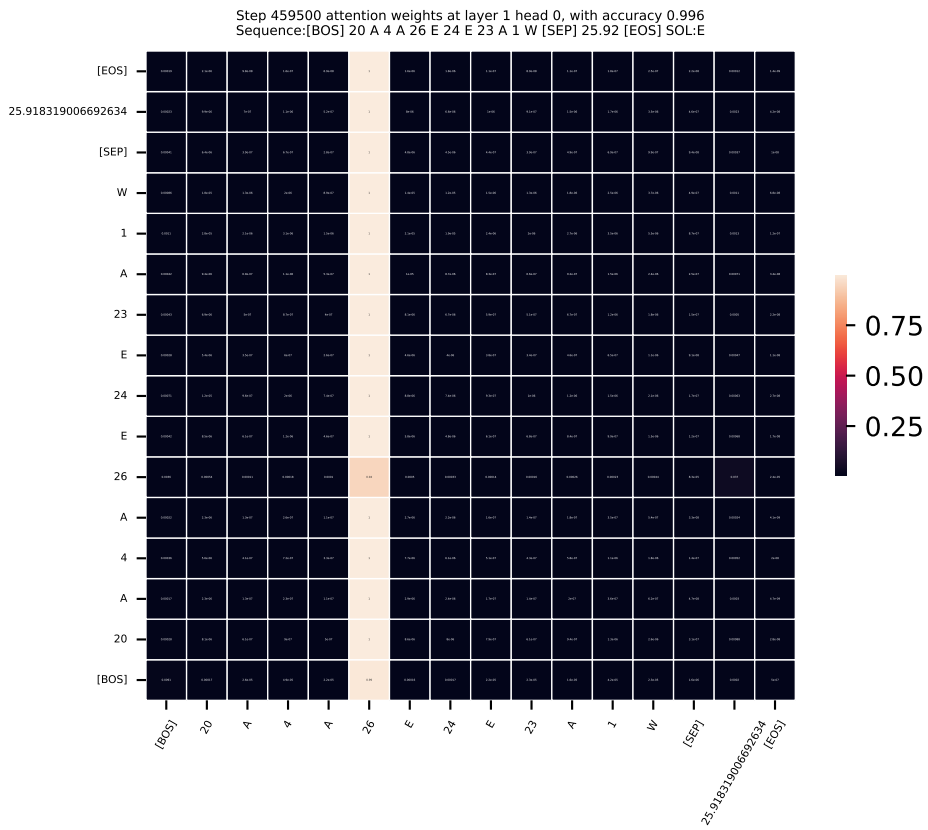

Step 450000 attention weights at layer 1 head 3, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

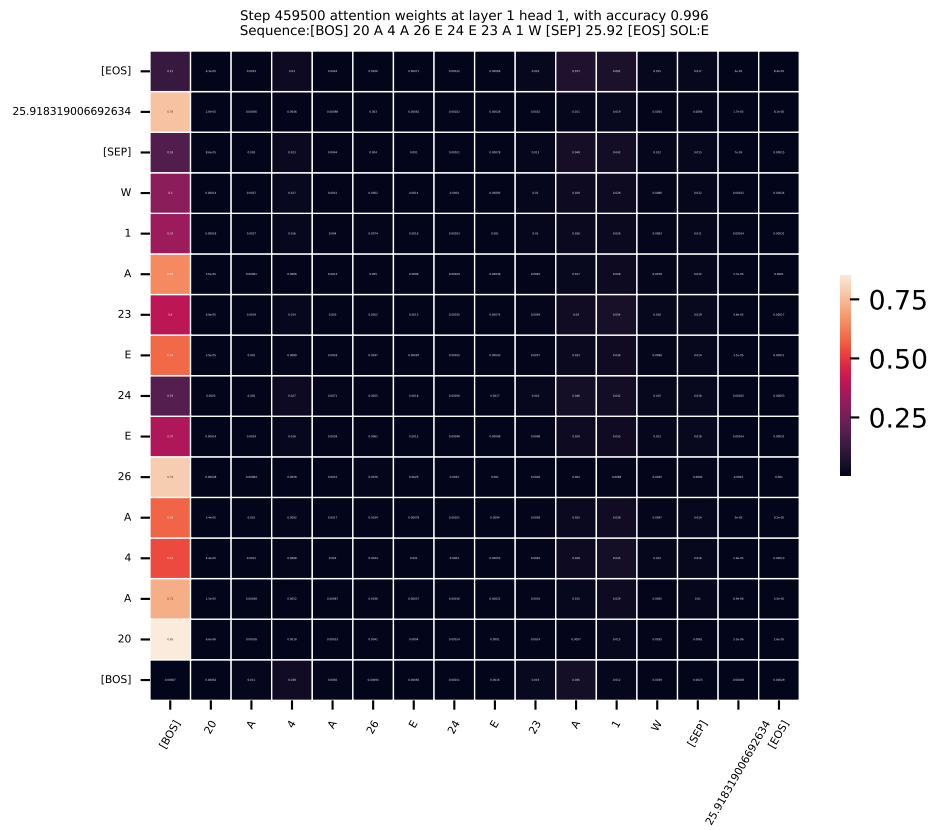

Step 459500 attention weights at layer 0 head 0, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

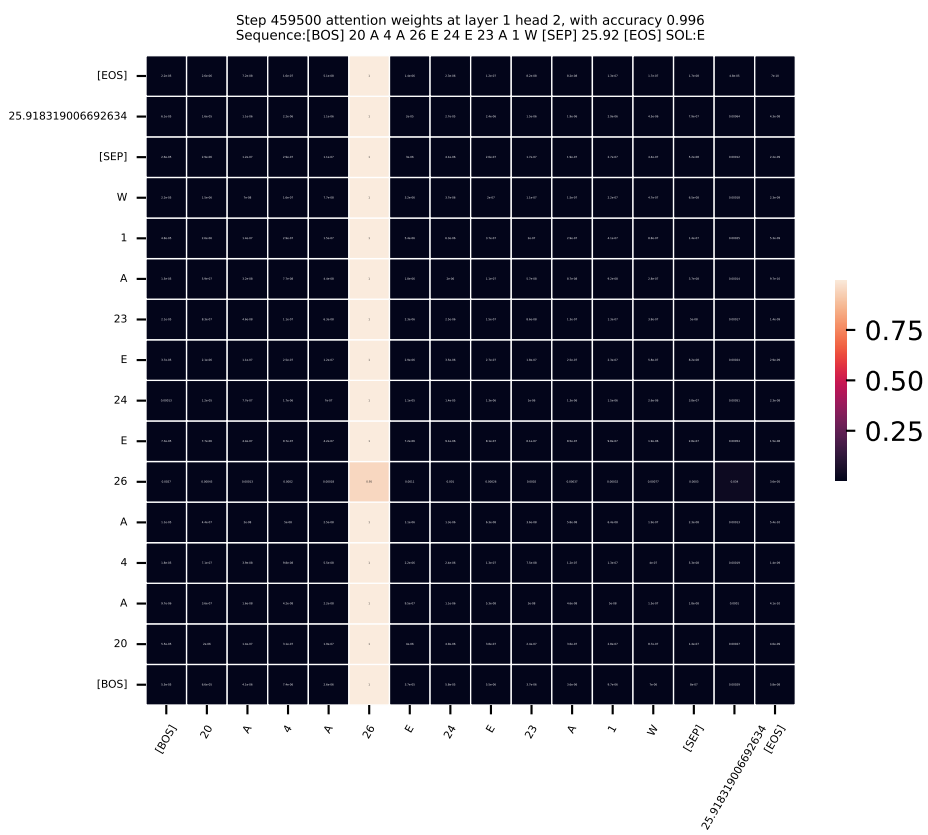

Step 459500 attention weights at layer 0 head 1, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

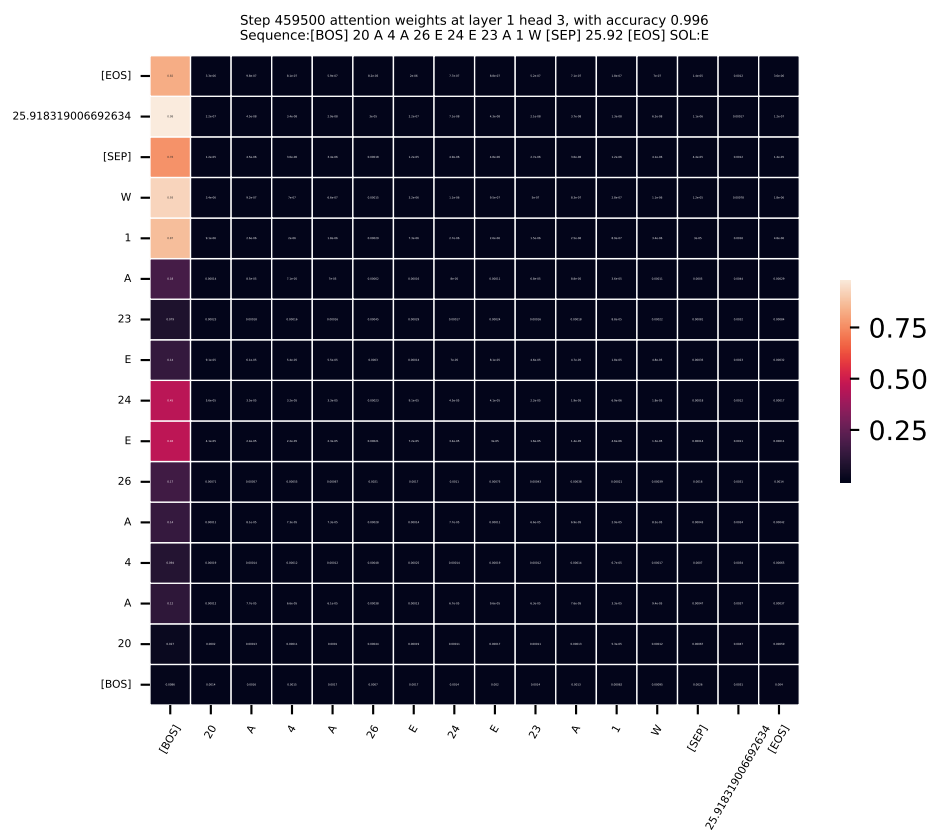

Step 459500 attention weights at layer 0 head 2, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

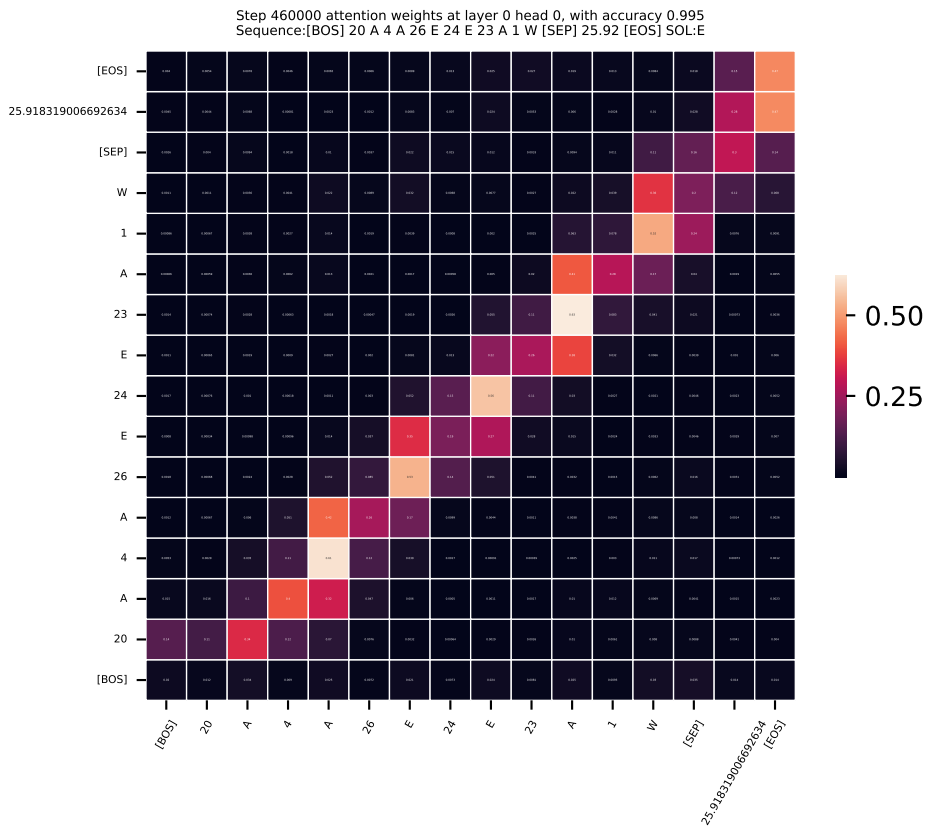

Step 459500 attention weights at layer 0 head 3, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

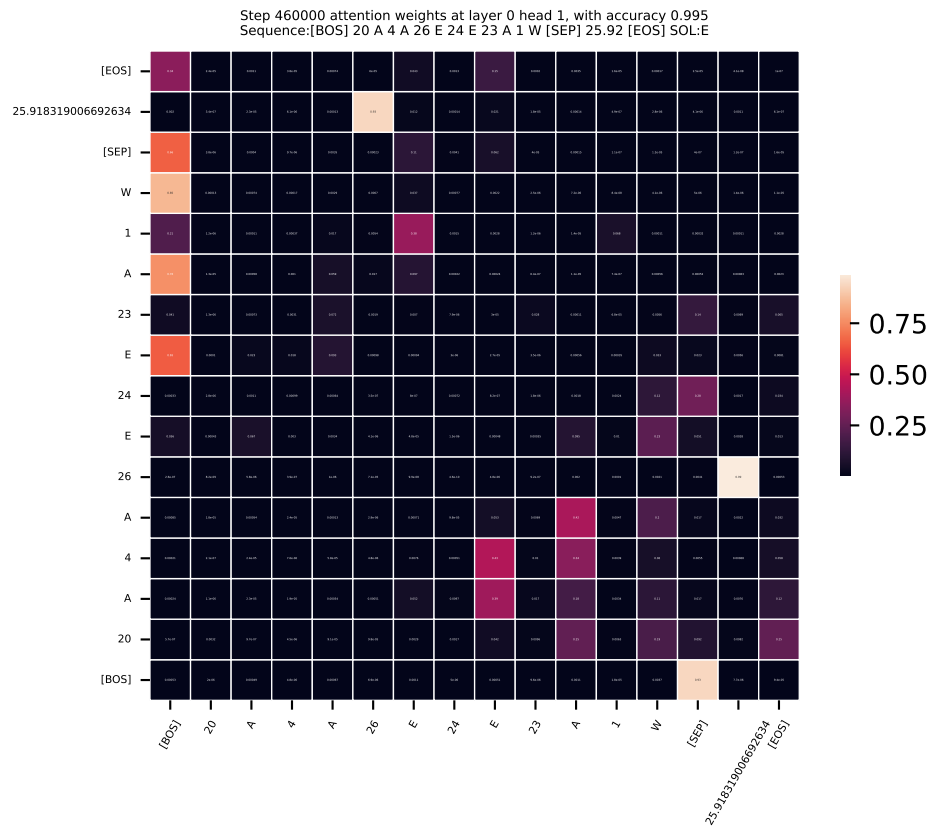

Step 459500 attention weights at layer 1 head 0, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

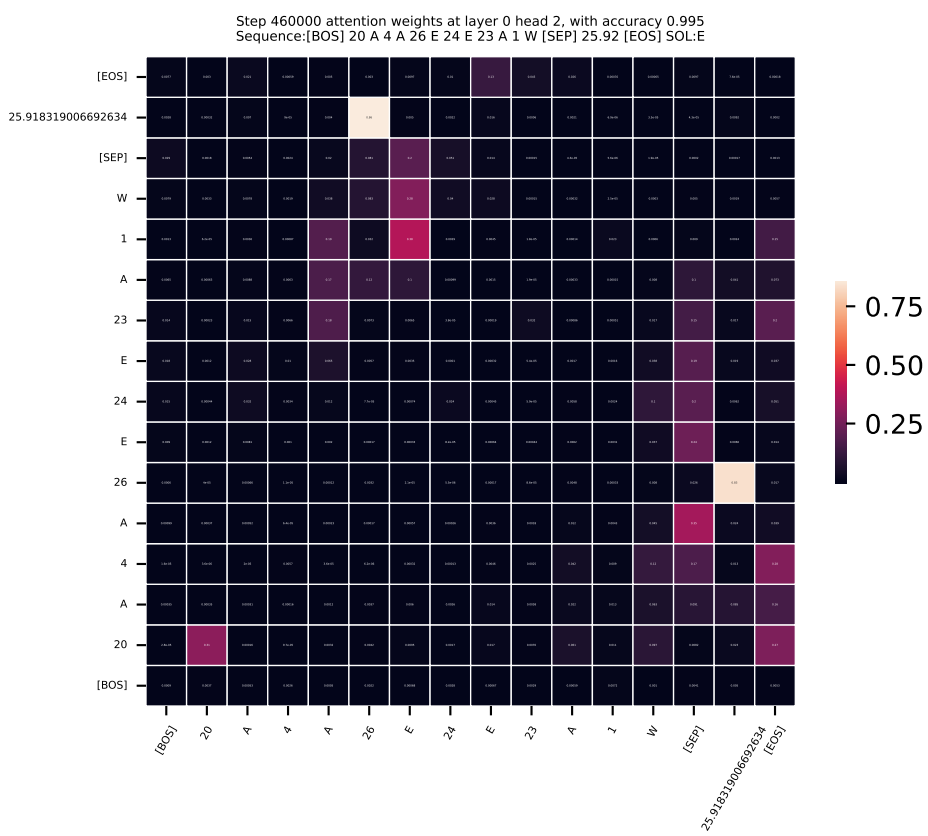

Step 459500 attention weights at layer 1 head 1, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

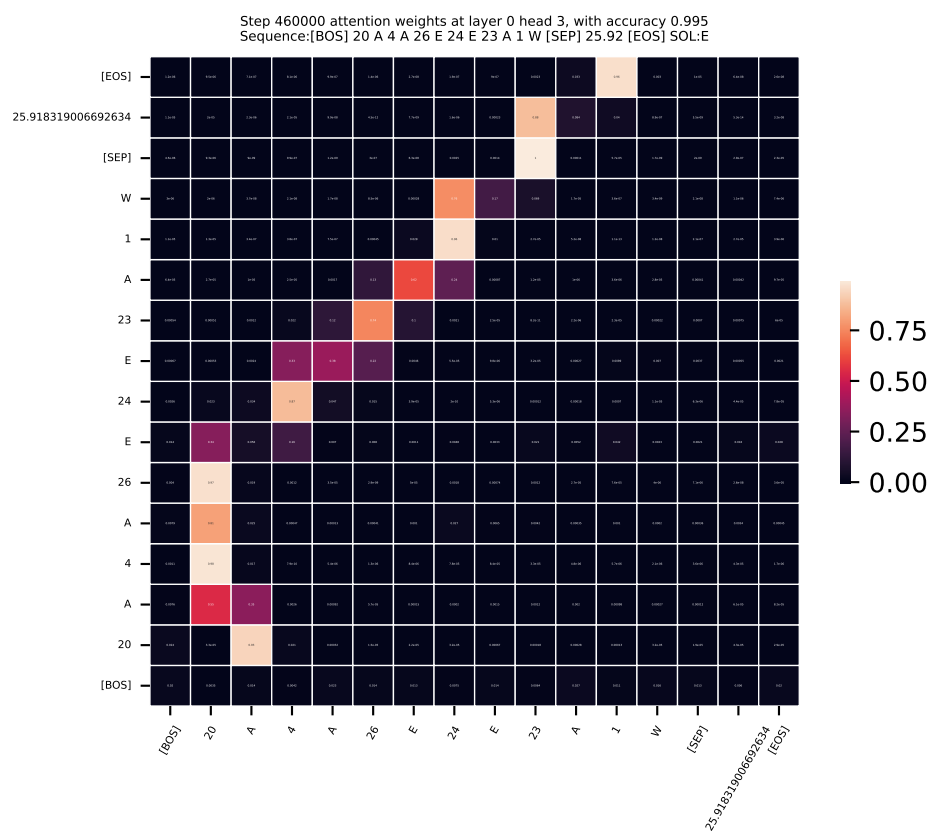

Step 459500 attention weights at layer 1 head 2, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

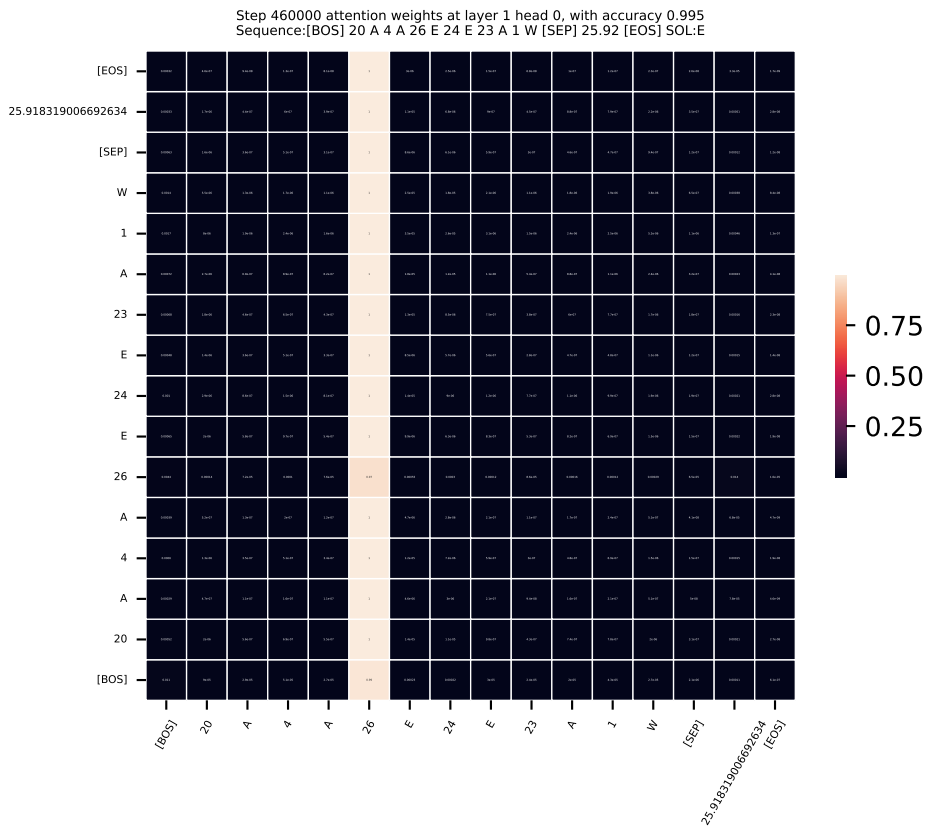

Step 459500 attention weights at layer 1 head 3, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

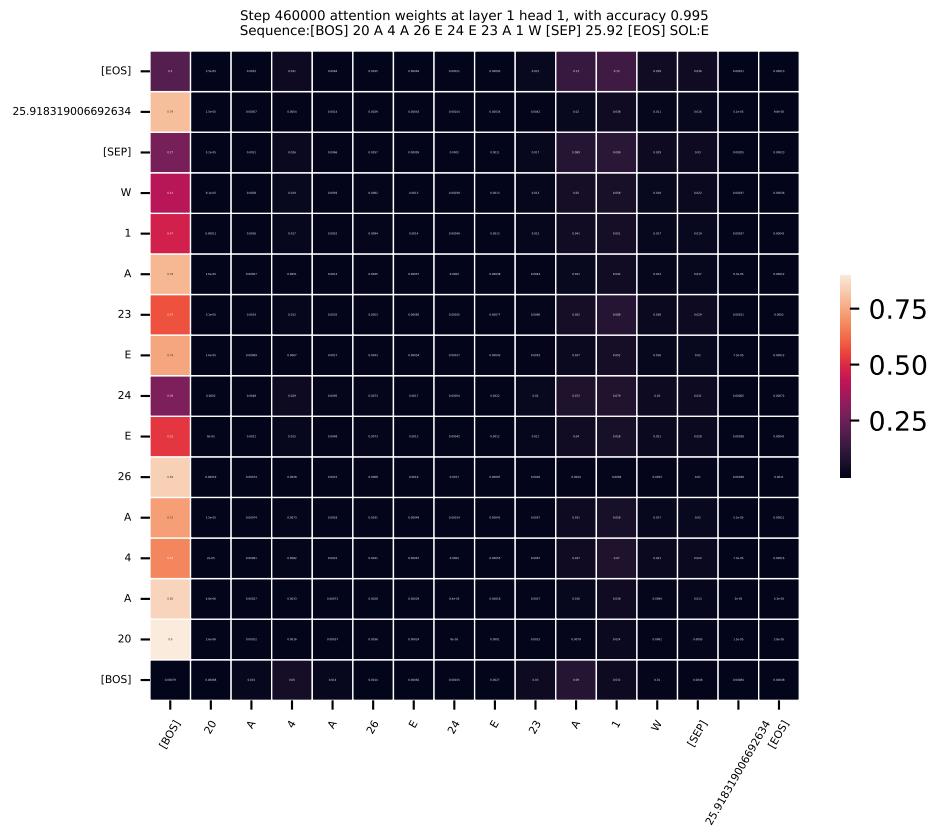

Step 460000 attention weights at layer 0 head 0, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.50
0.25

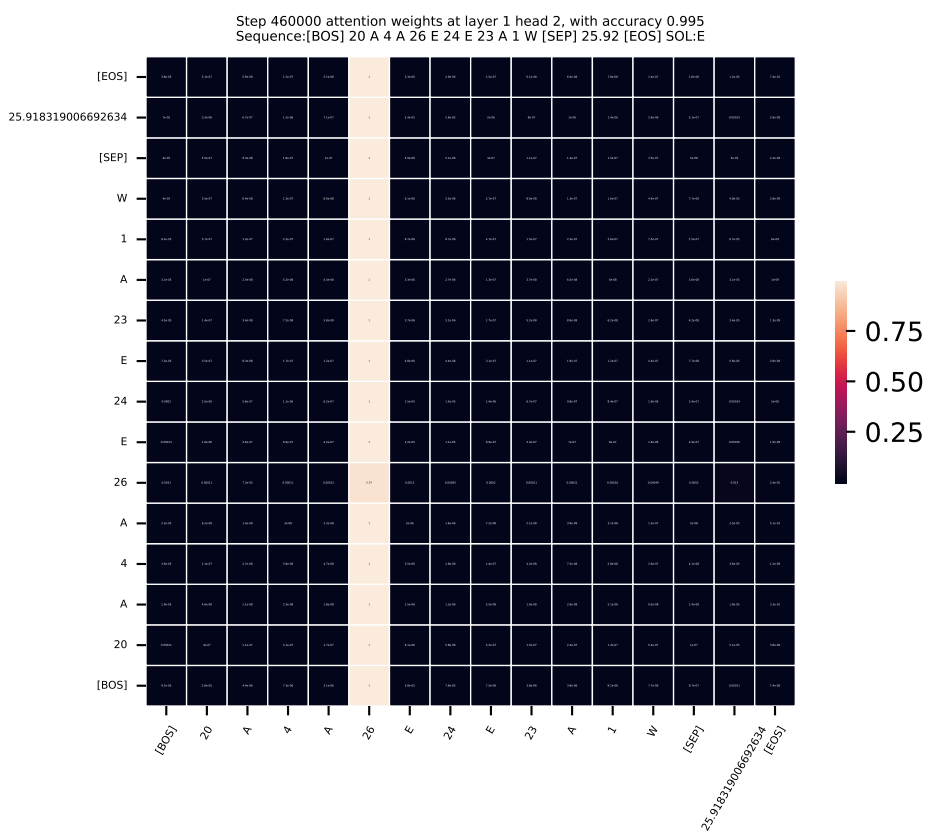

Step 460000 attention weights at layer 0 head 1, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

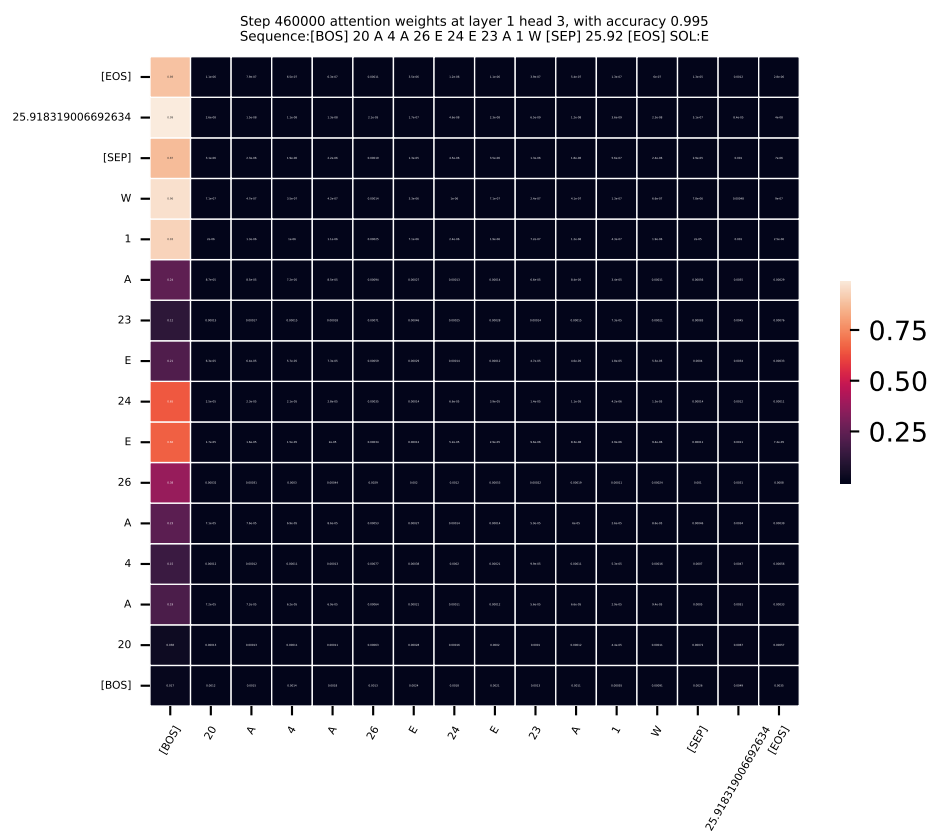

Step 460000 attention weights at layer 0 head 2, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

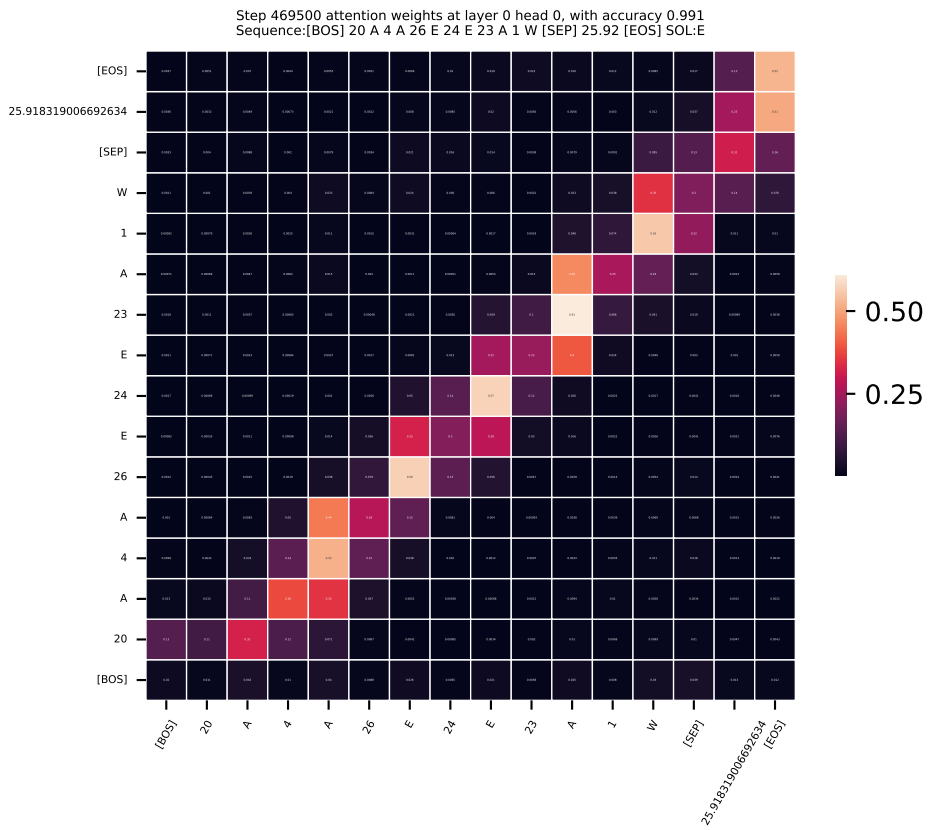

Step 460000 attention weights at layer 0 head 3, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

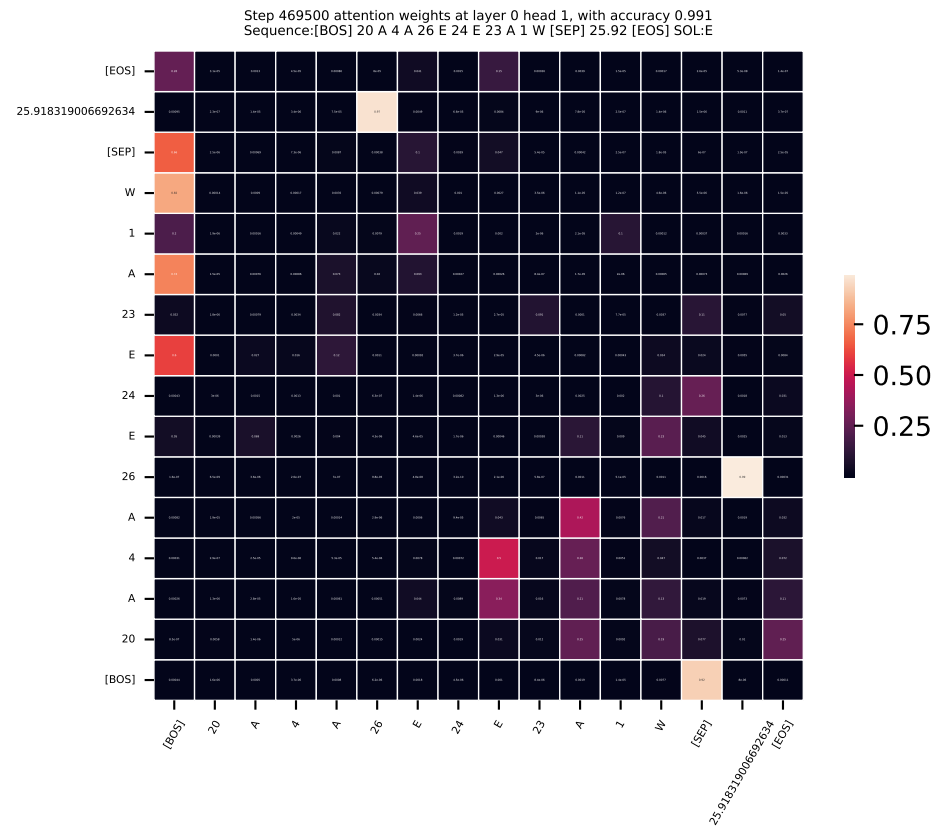

Step 460000 attention weights at layer 1 head 0, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

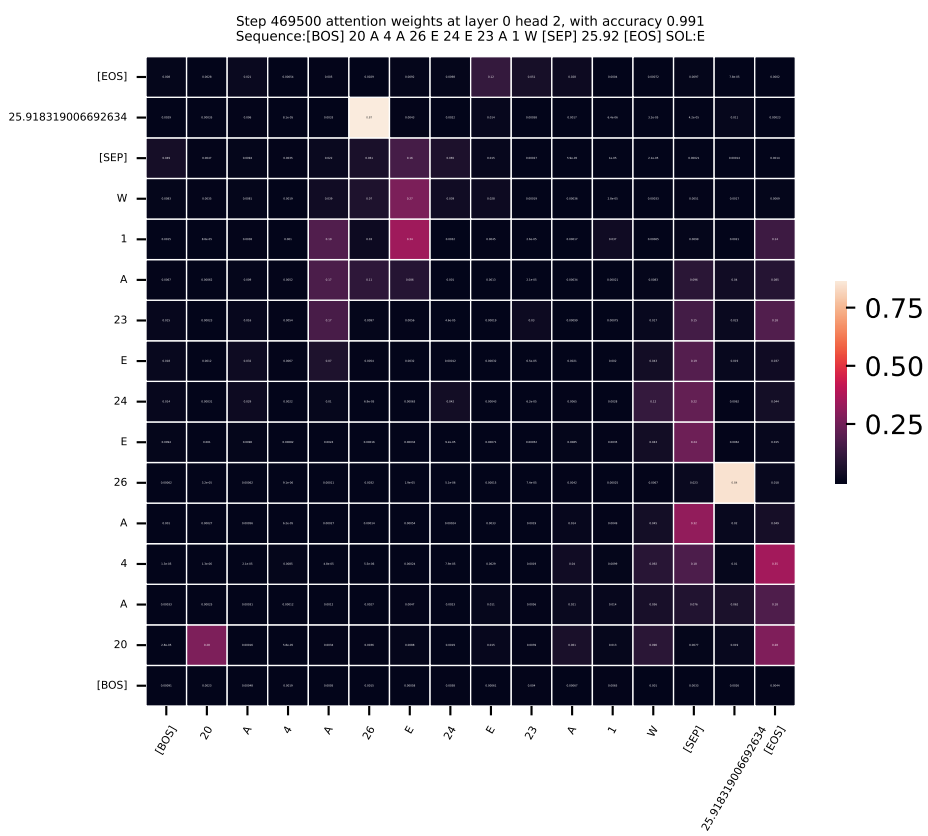

Step 460000 attention weights at layer 1 head 1, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

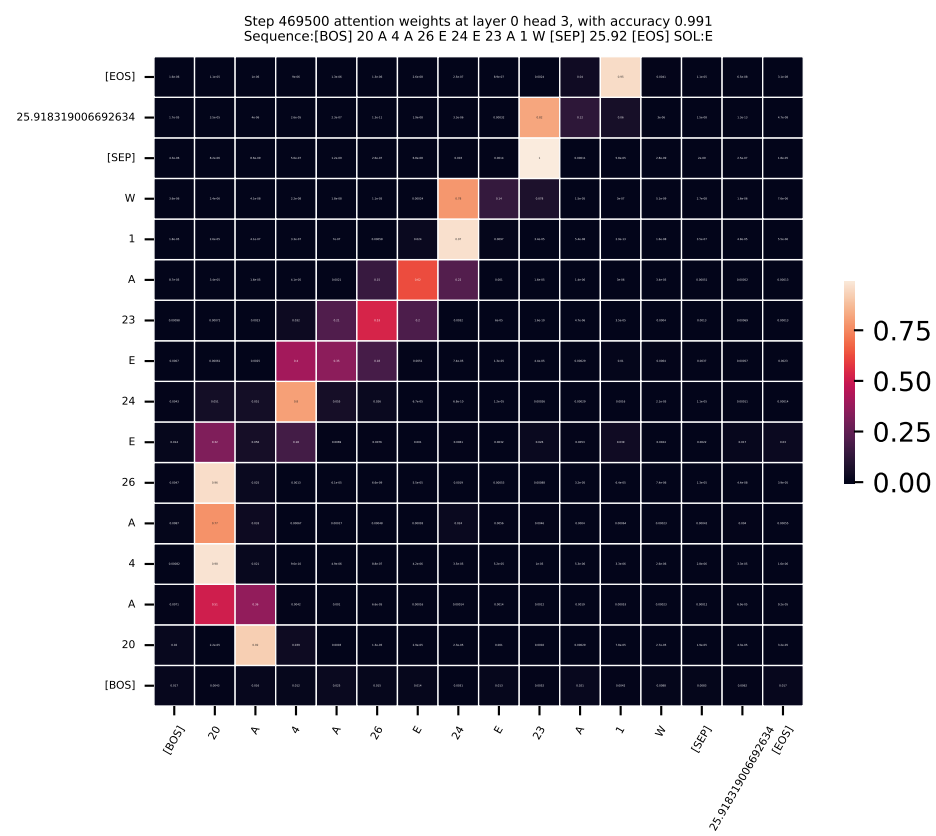

Step 460000 attention weights at layer 1 head 2, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

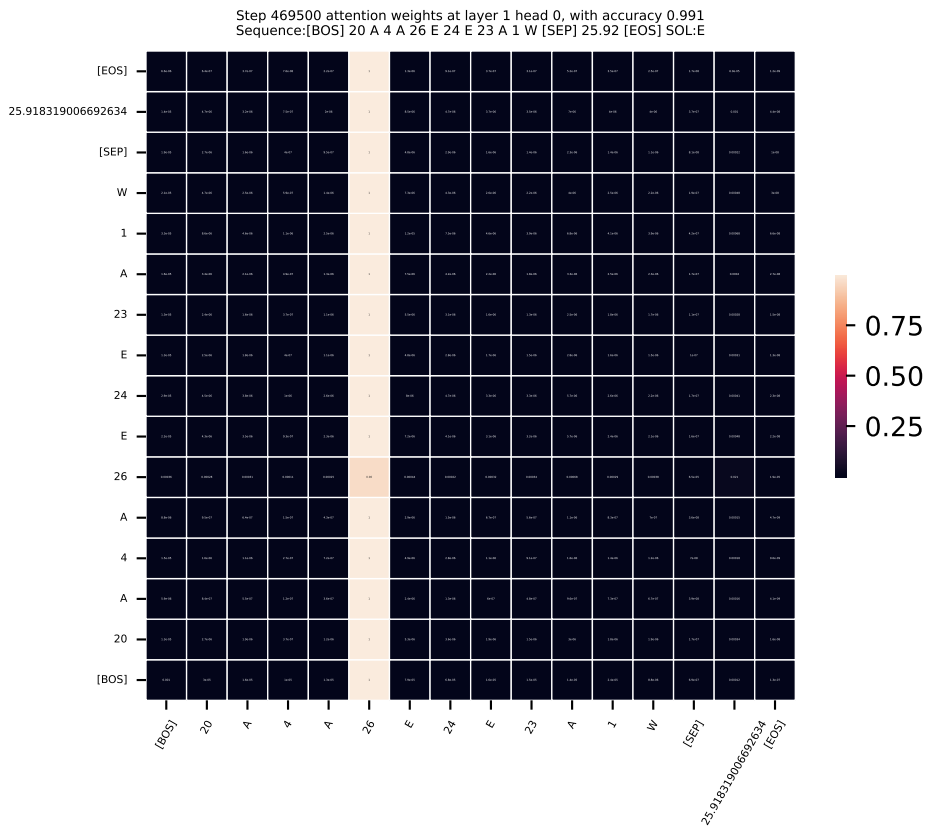

Step 460000 attention weights at layer 1 head 3, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

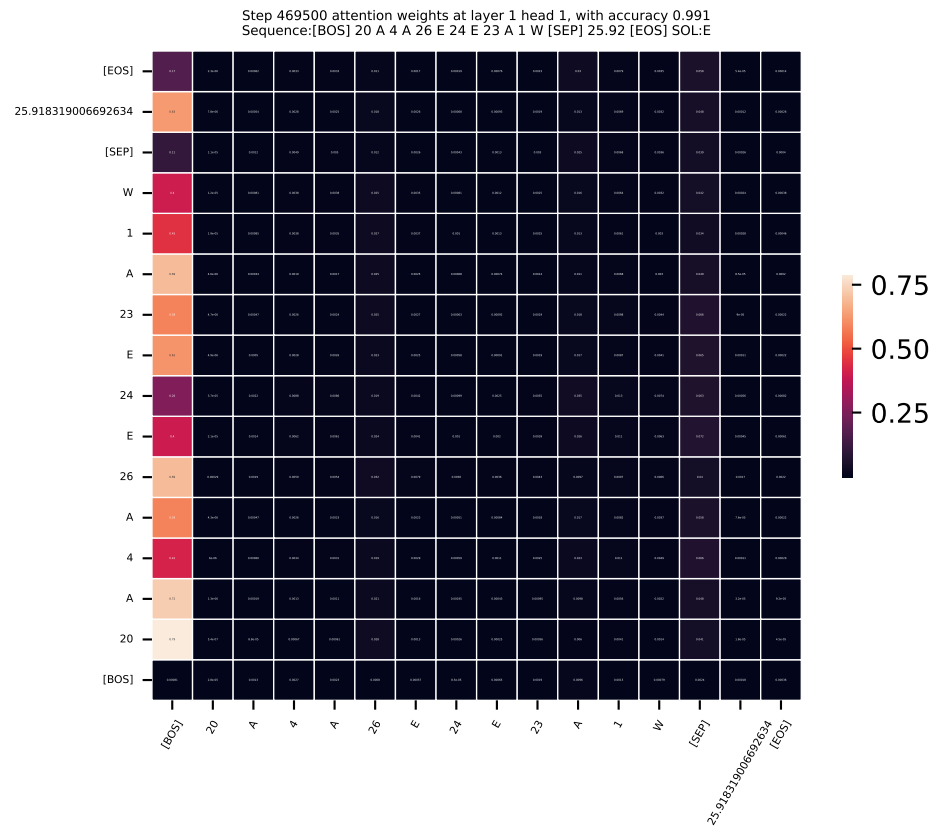

Step 469500 attention weights at layer 0 head 0, with accuracy 0.991
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.50
0.25

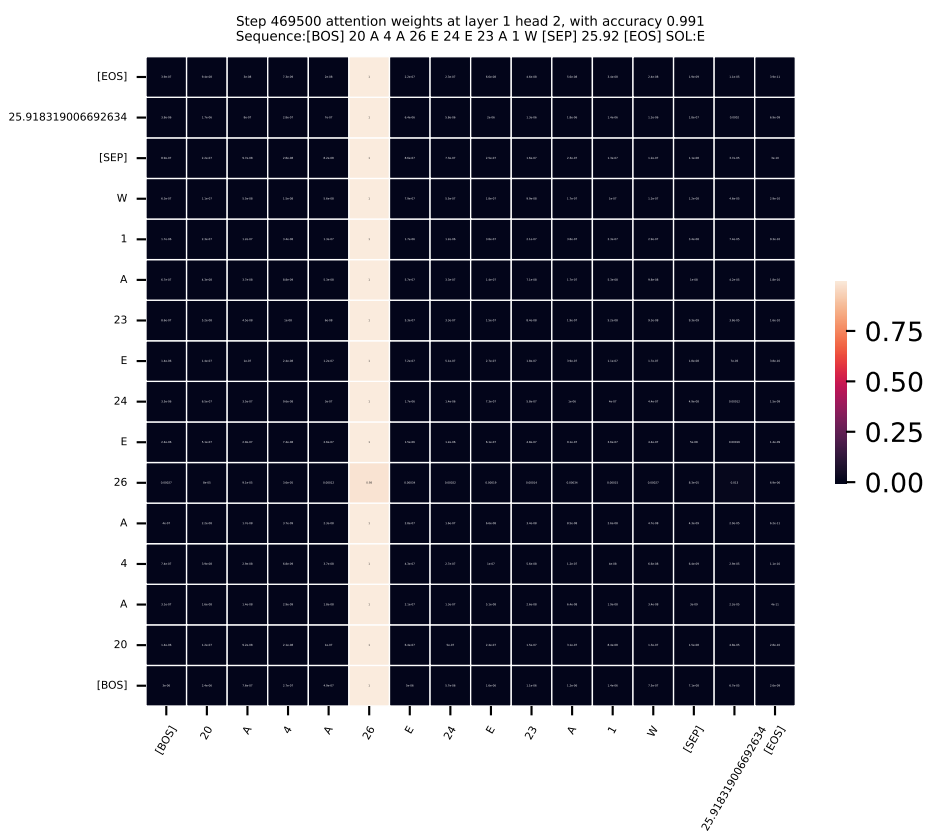

Step 469500 attention weights at layer 0 head 1, with accuracy 0.991
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

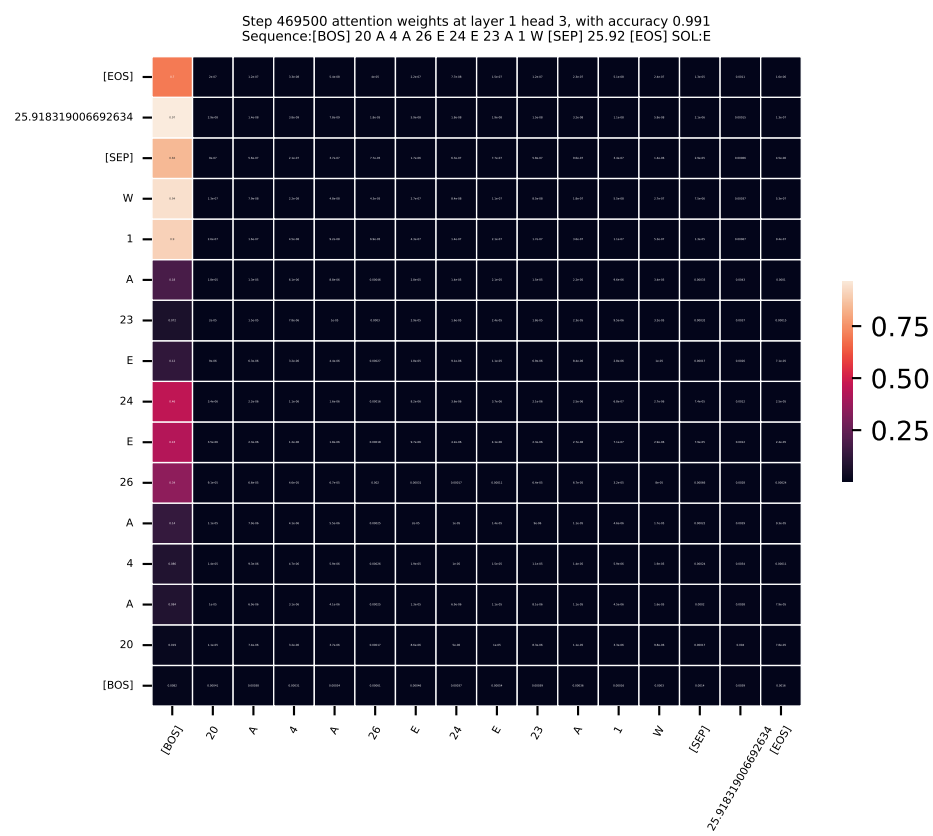

Step 469500 attention weights at layer 0 head 2, with accuracy 0.991
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

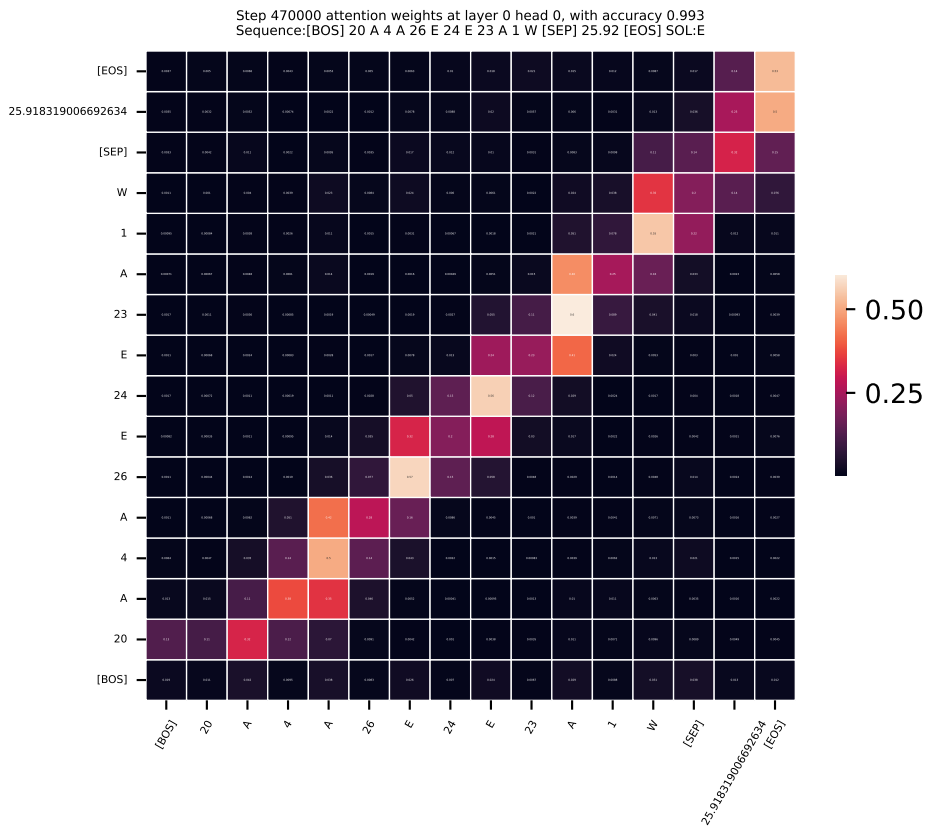

Step 469500 attention weights at layer 0 head 3, with accuracy 0.991
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

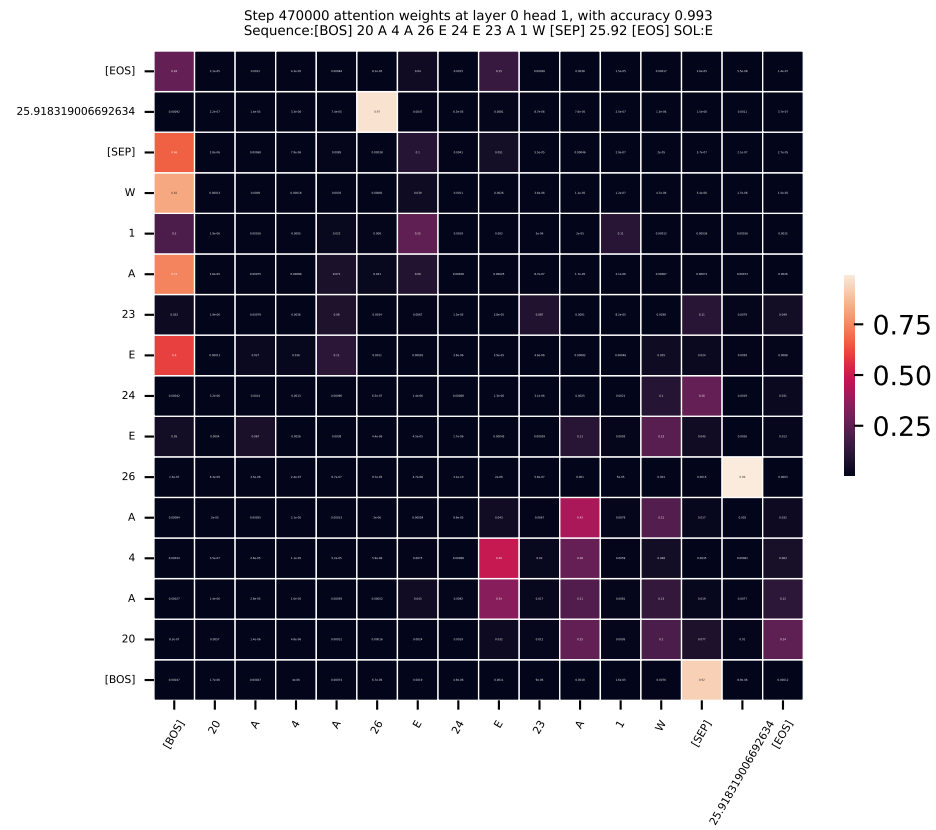

Step 469500 attention weights at layer 1 head 0, with accuracy 0.991
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

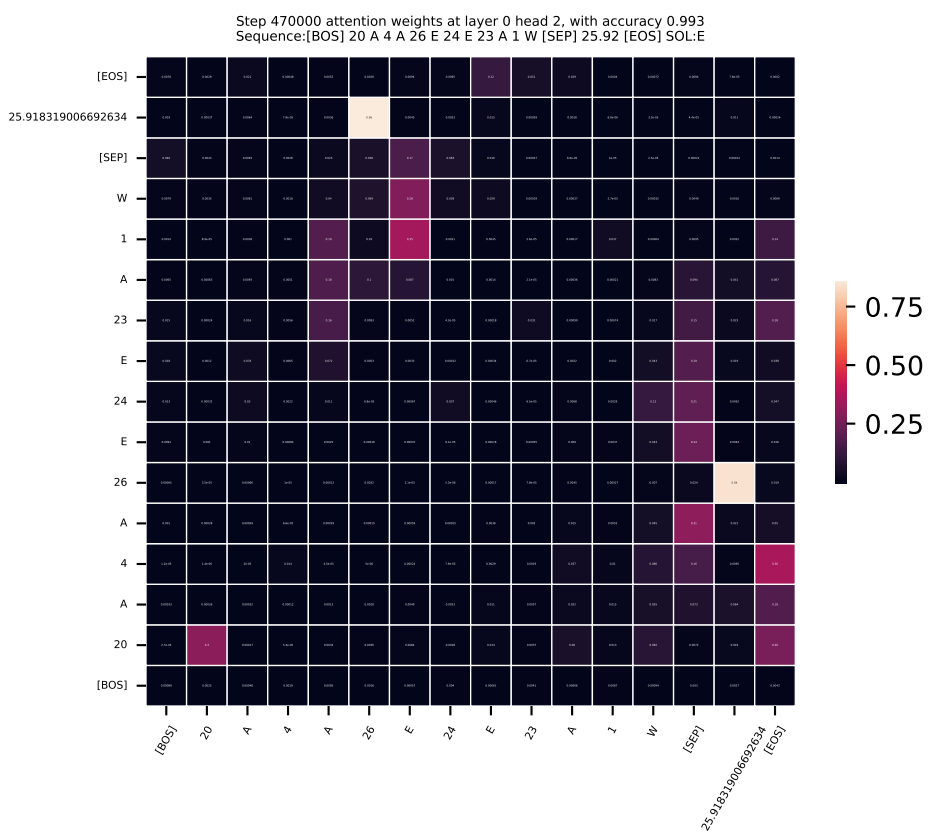

Step 469500 attention weights at layer 1 head 1, with accuracy 0.991
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

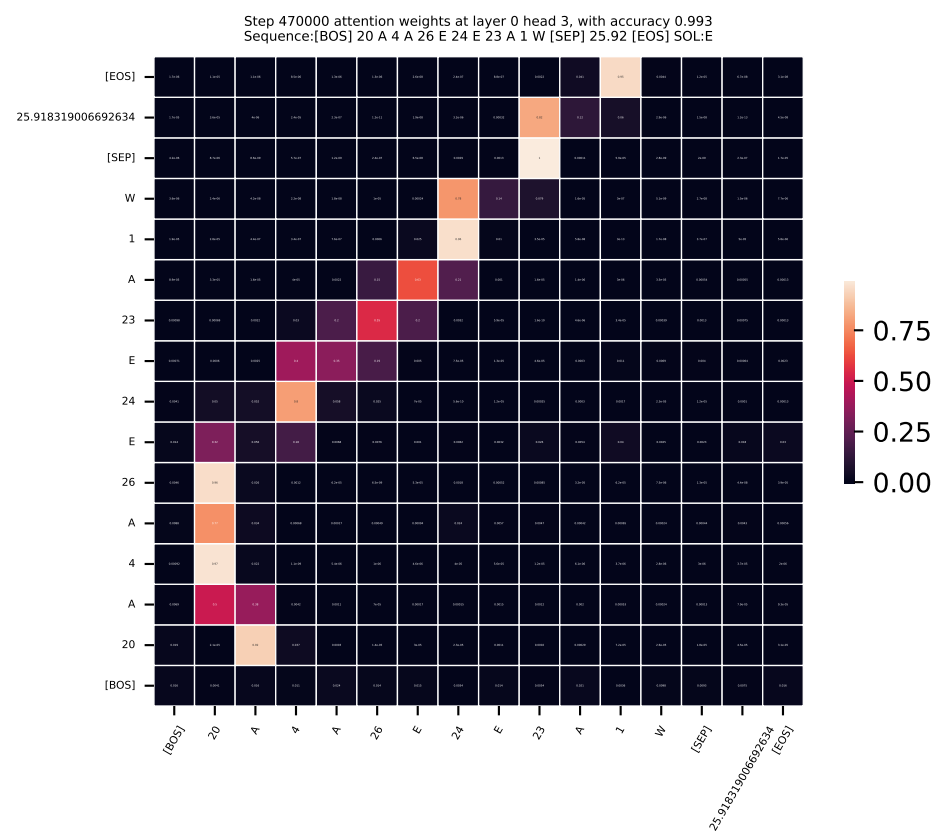

Step 469500 attention weights at layer 1 head 2, with accuracy 0.991
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

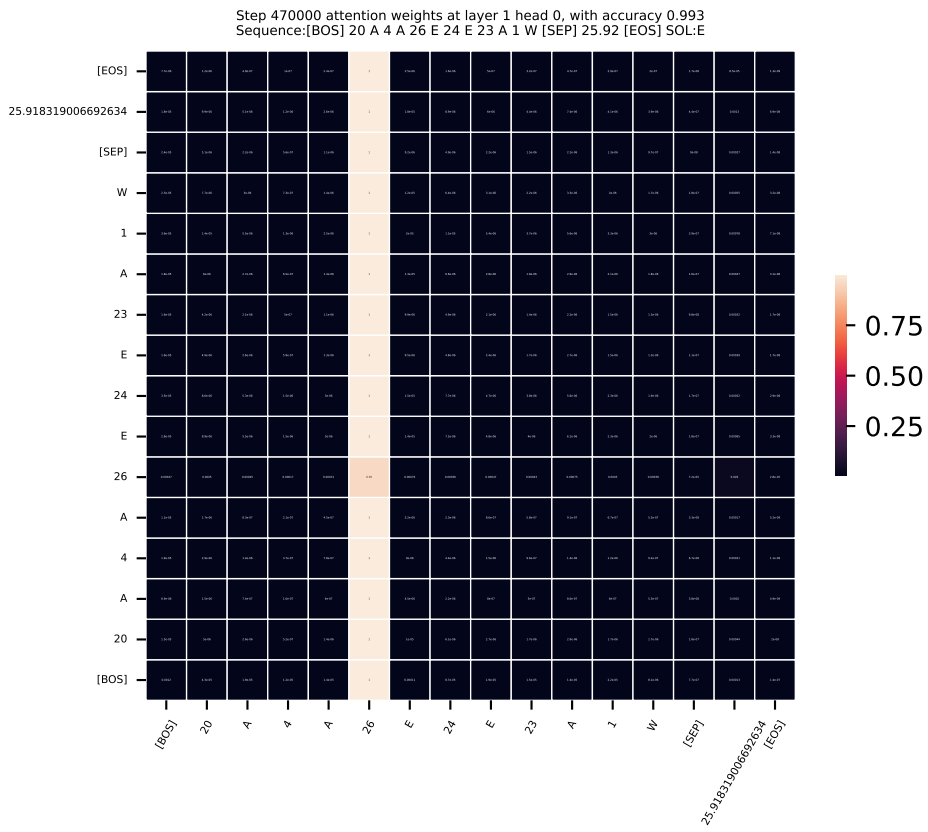

Step 469500 attention weights at layer 1 head 3, with accuracy 0.991
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

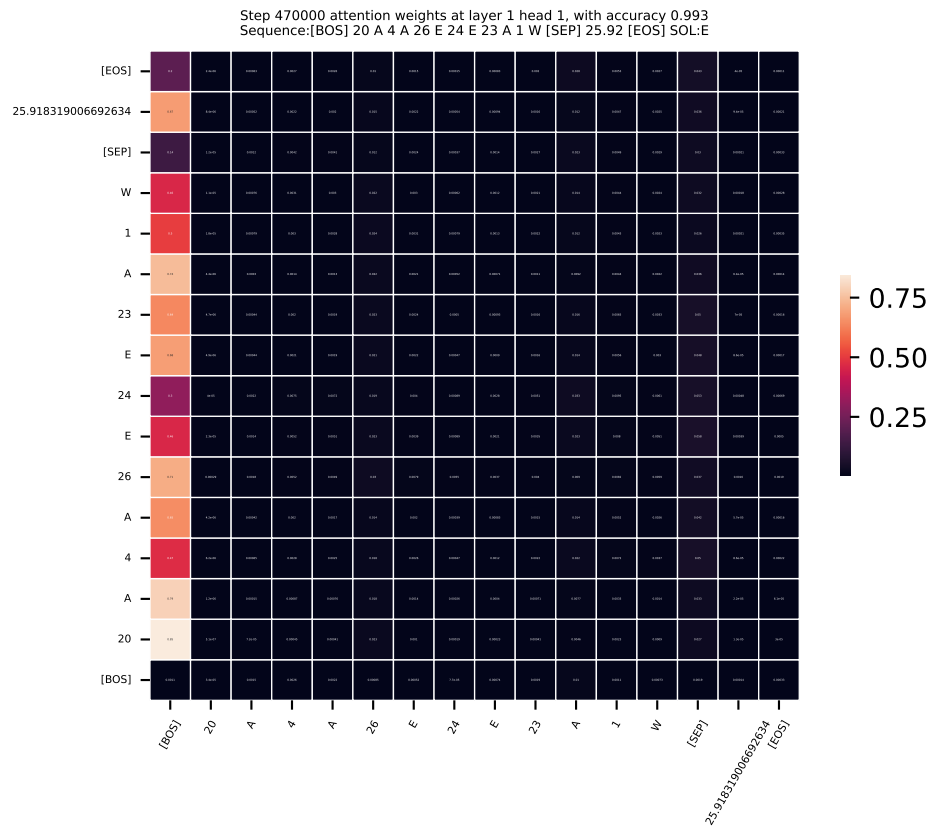

Step 470000 attention weights at layer 0 head 0, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

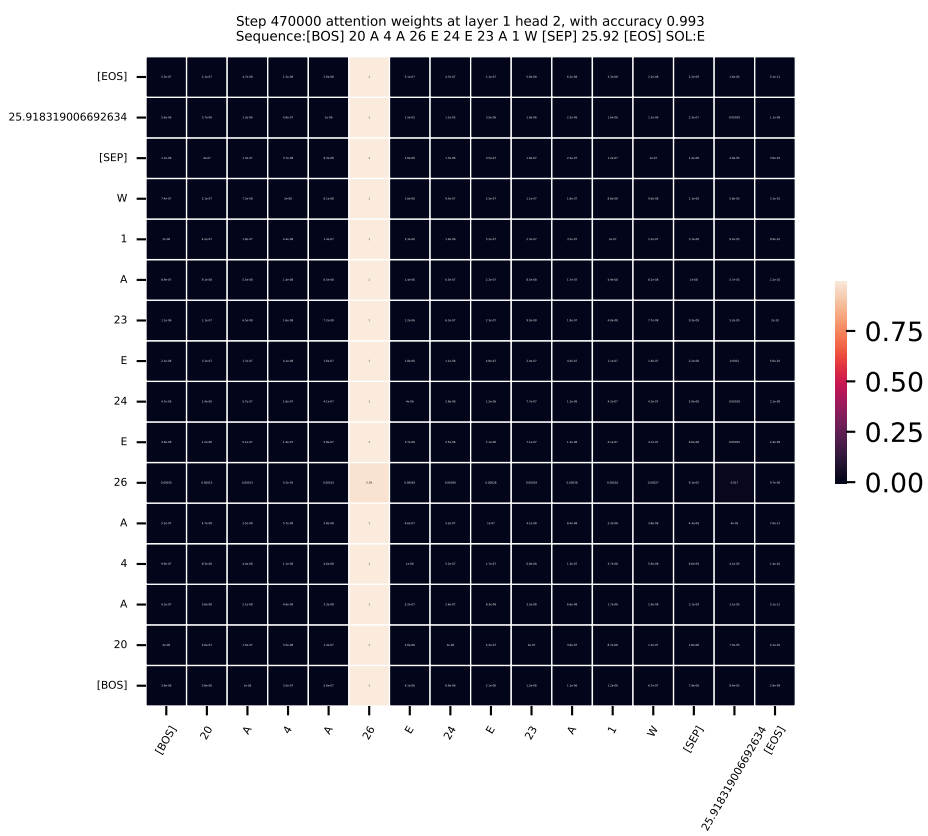

Step 470000 attention weights at layer 0 head 1, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

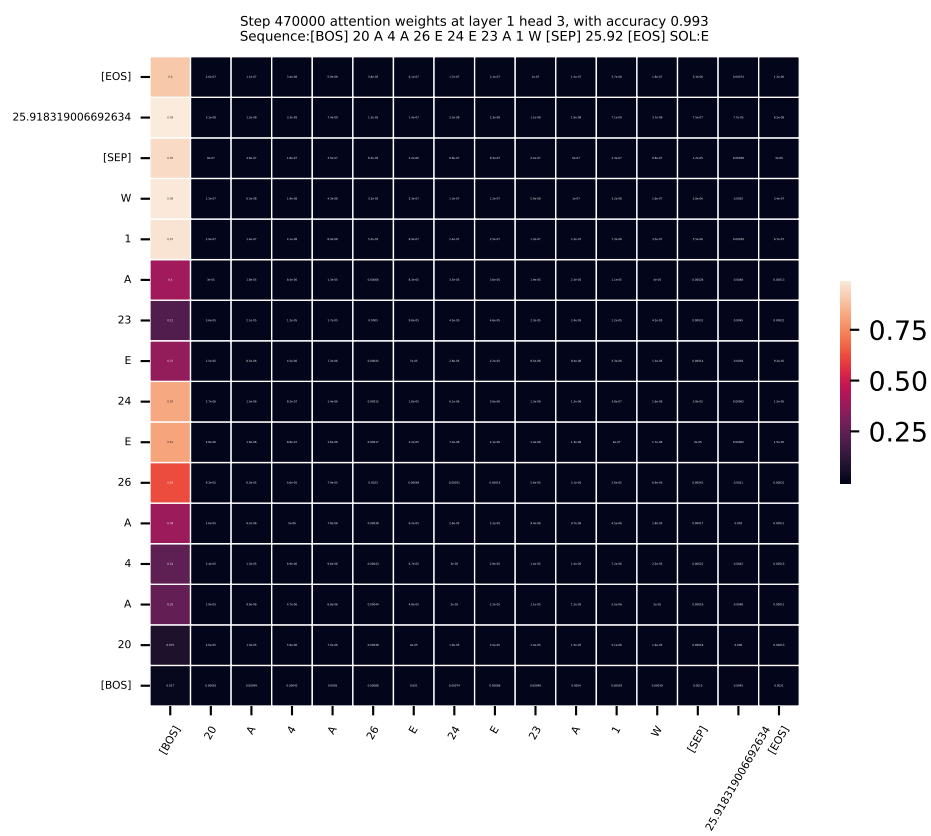

Step 470000 attention weights at layer 0 head 2, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

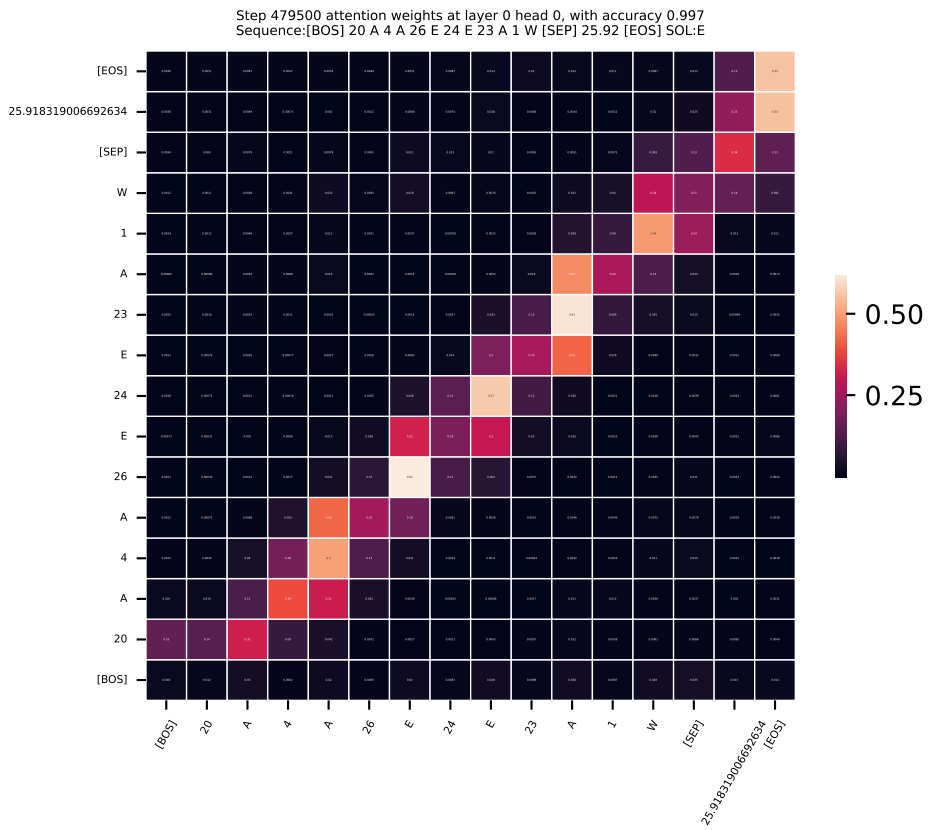

Step 470000 attention weights at layer 0 head 3, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

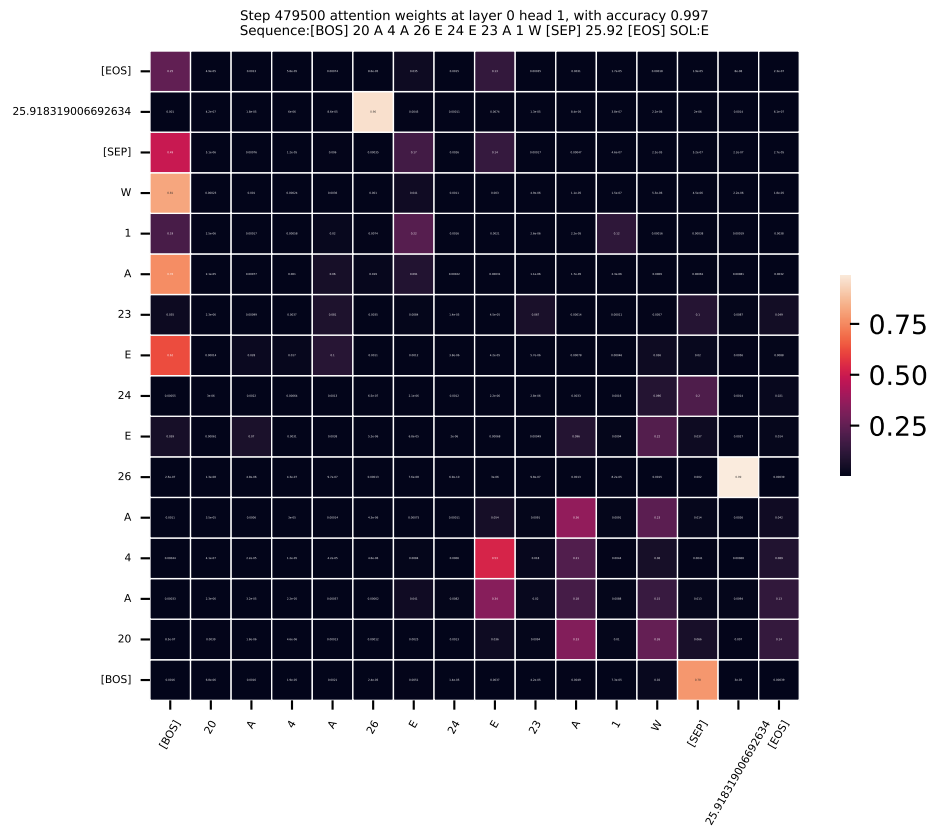

Step 470000 attention weights at layer 1 head 0, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

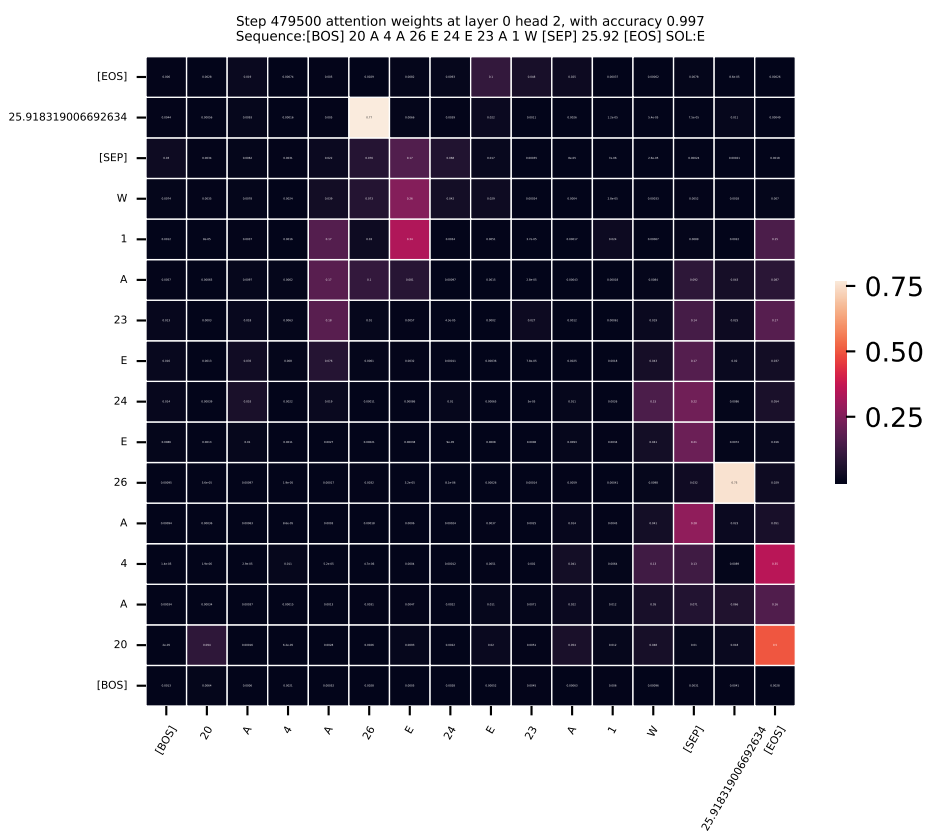

Step 470000 attention weights at layer 1 head 1, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

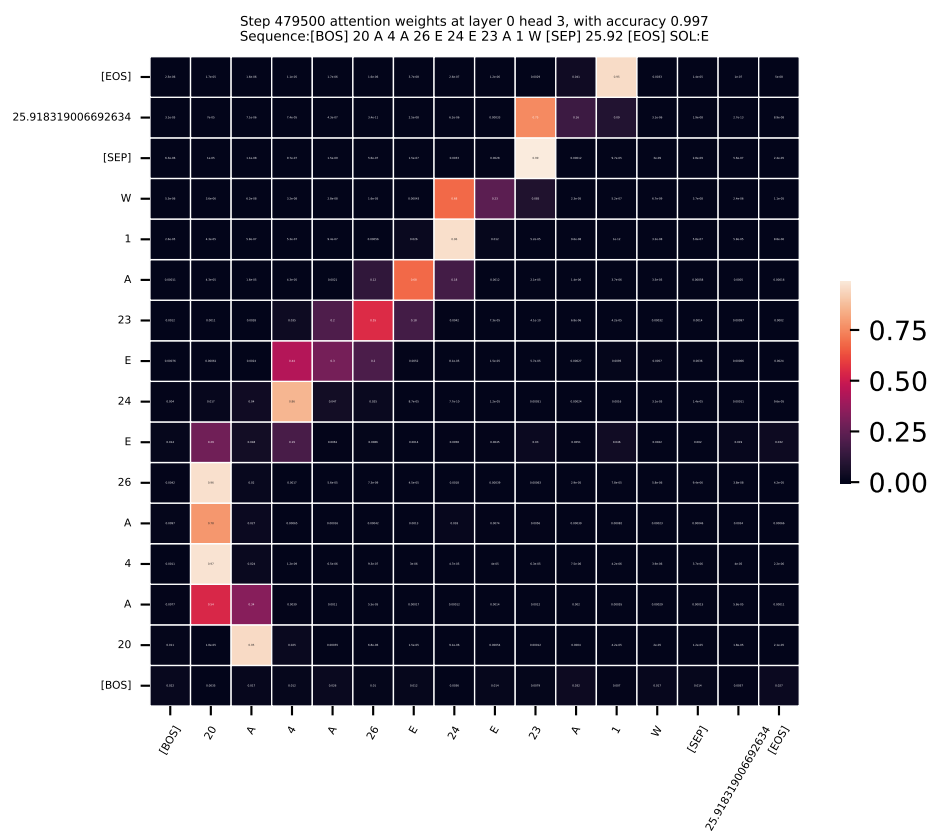

Step 470000 attention weights at layer 1 head 2, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

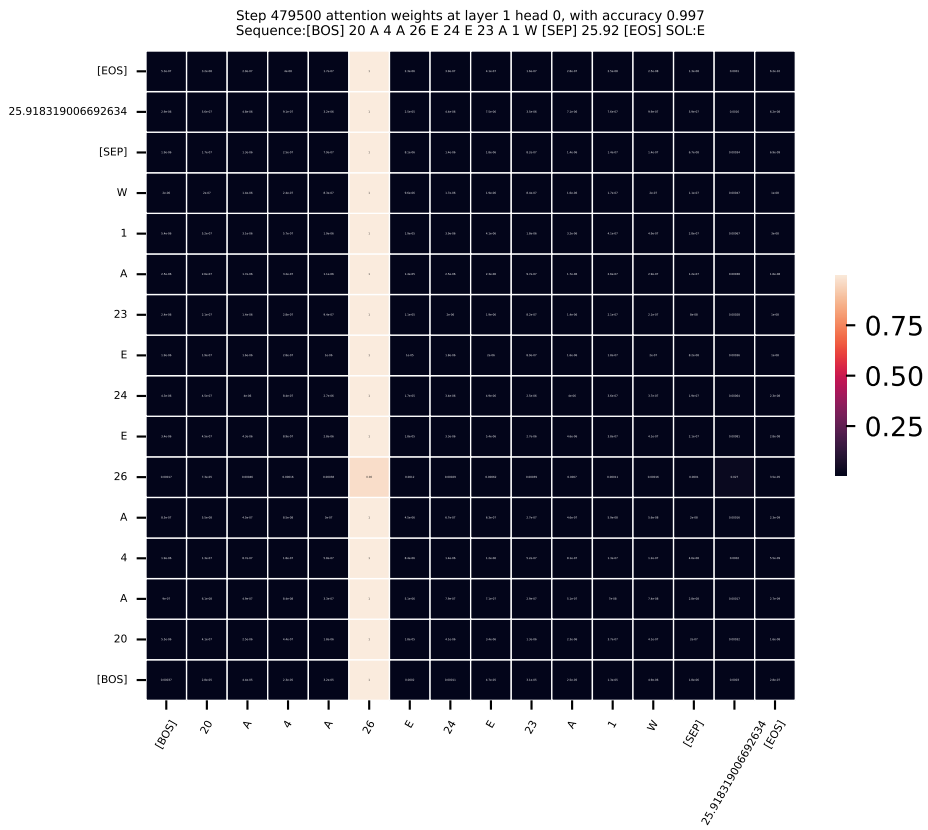

Step 470000 attention weights at layer 1 head 3, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

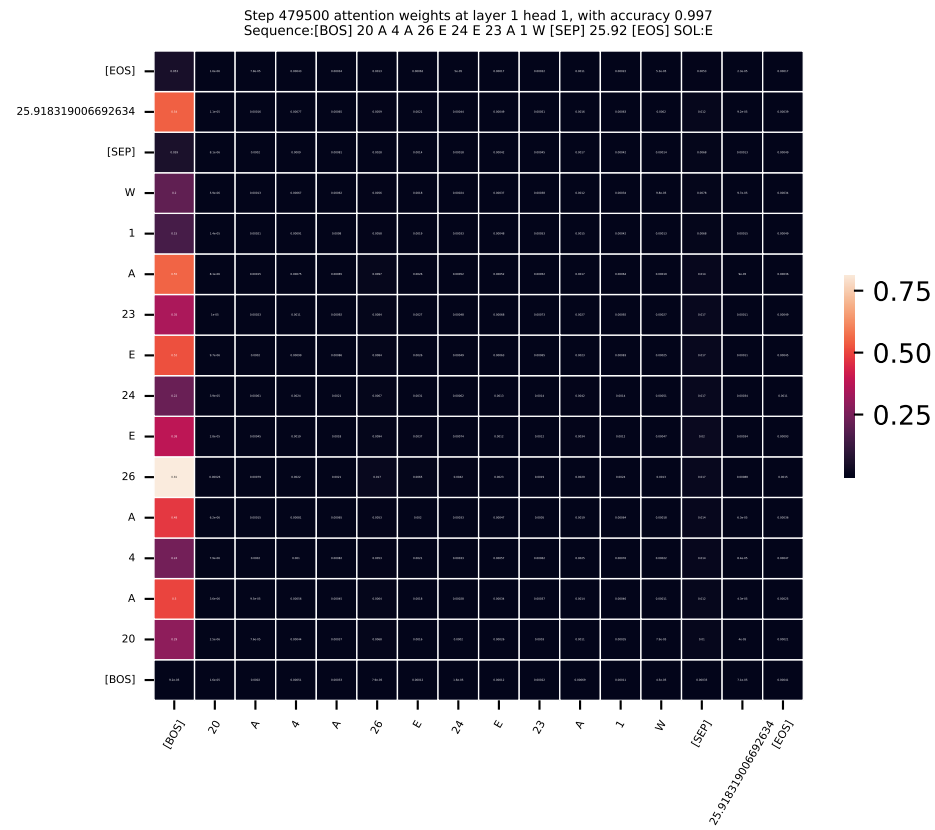

Step 479500 attention weights at layer 0 head 0, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

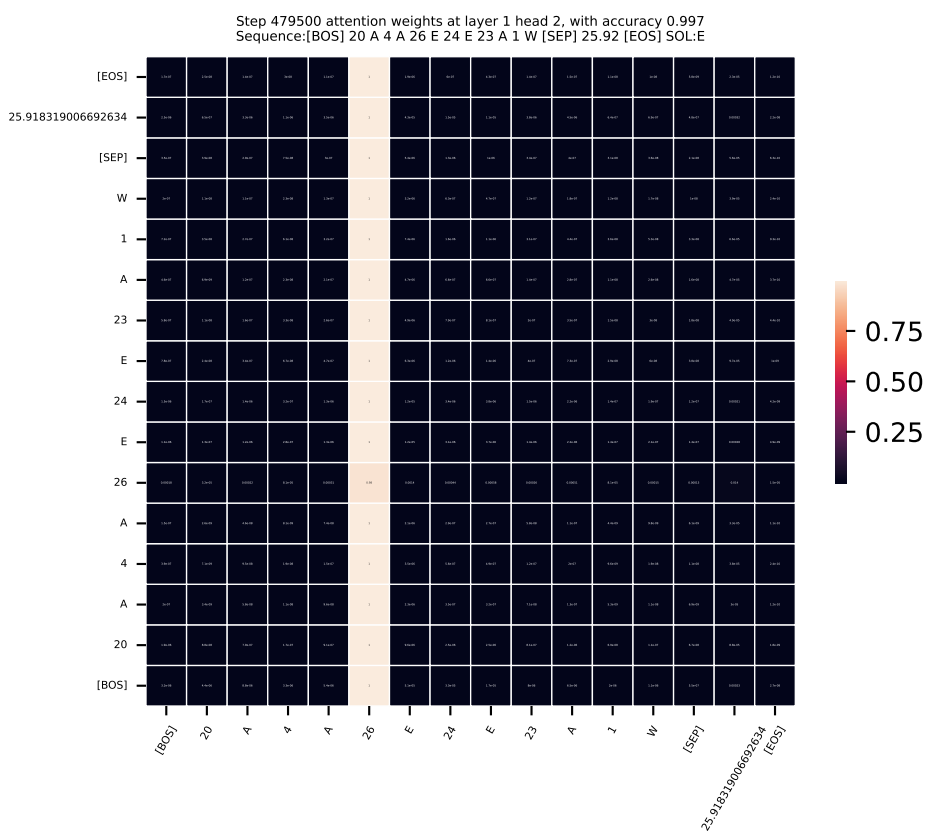

Step 479500 attention weights at layer 0 head 1, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

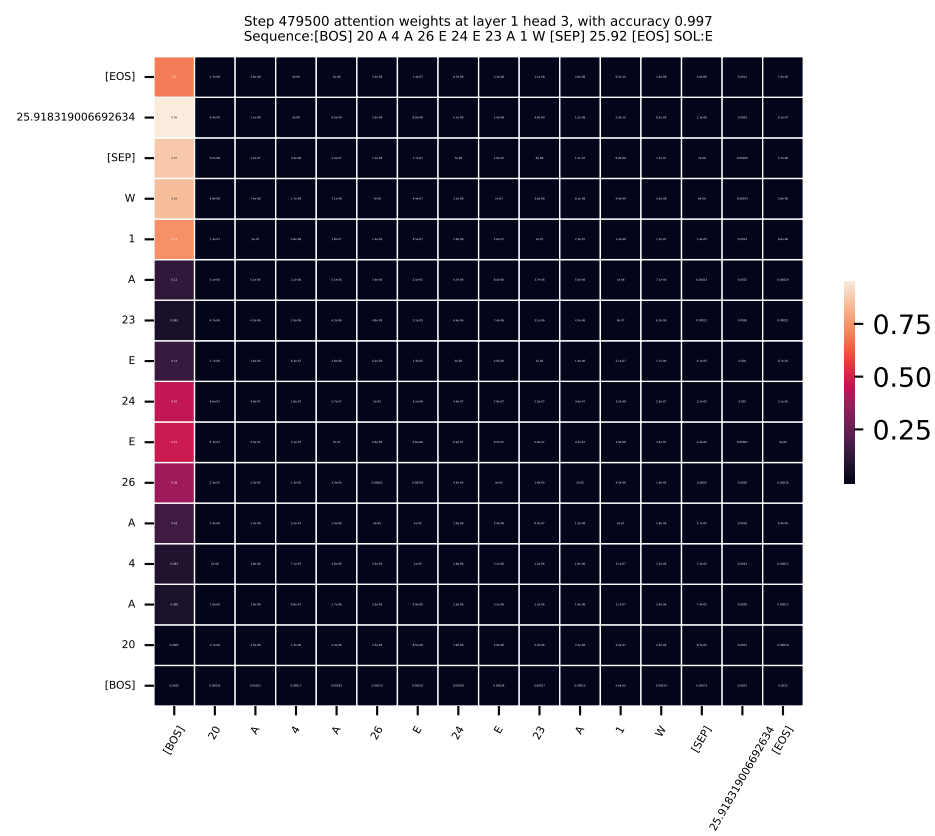

Step 479500 attention weights at layer 0 head 2, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

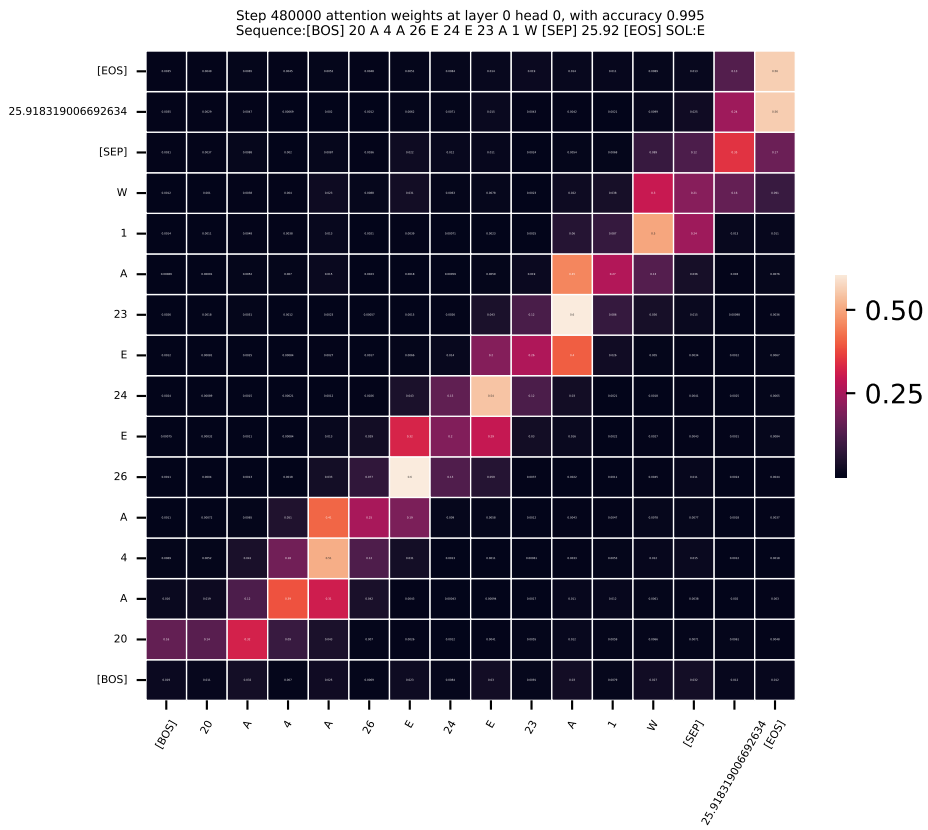

Step 479500 attention weights at layer 0 head 3, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

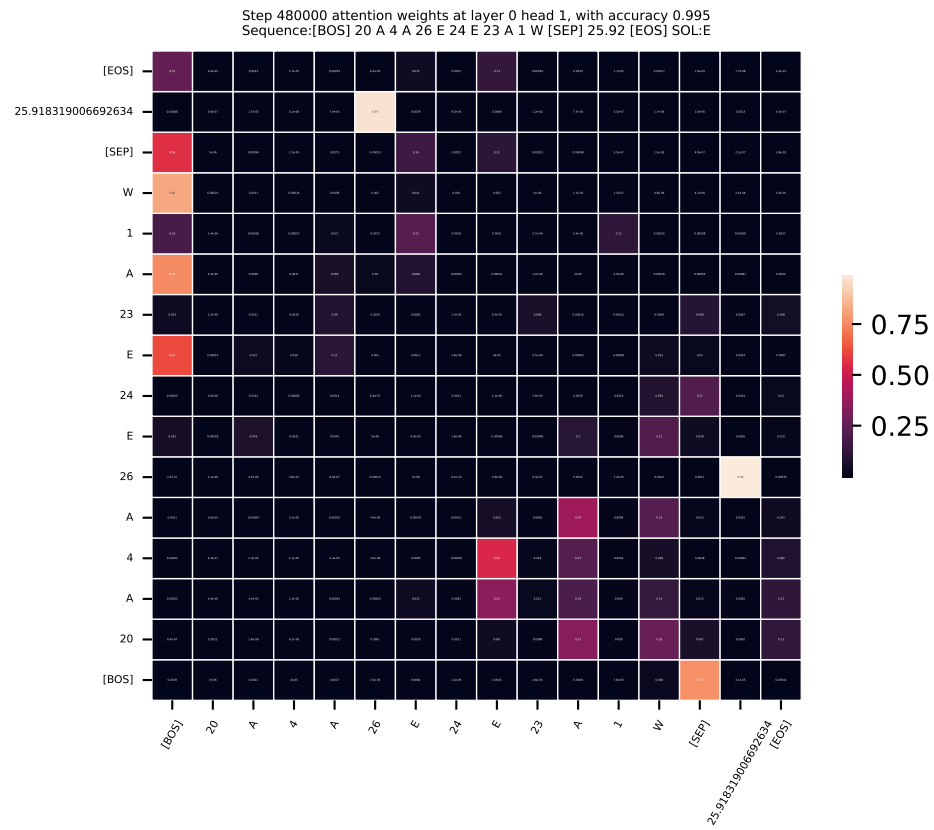

Step 479500 attention weights at layer 1 head 0, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

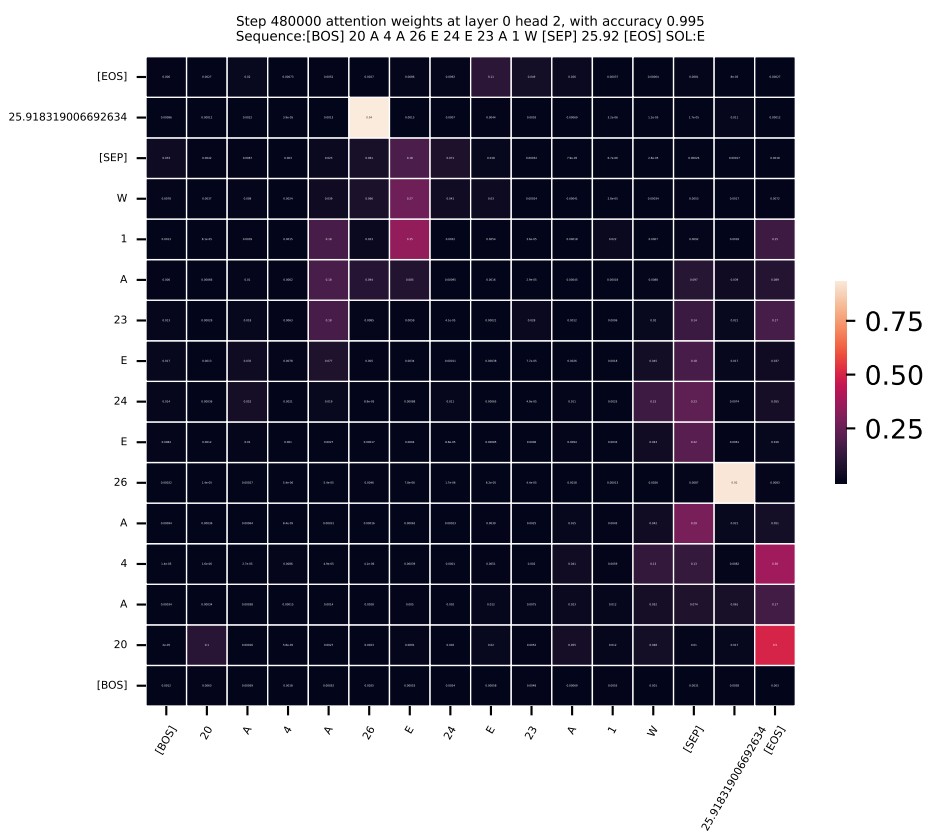

Step 479500 attention weights at layer 1 head 1, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

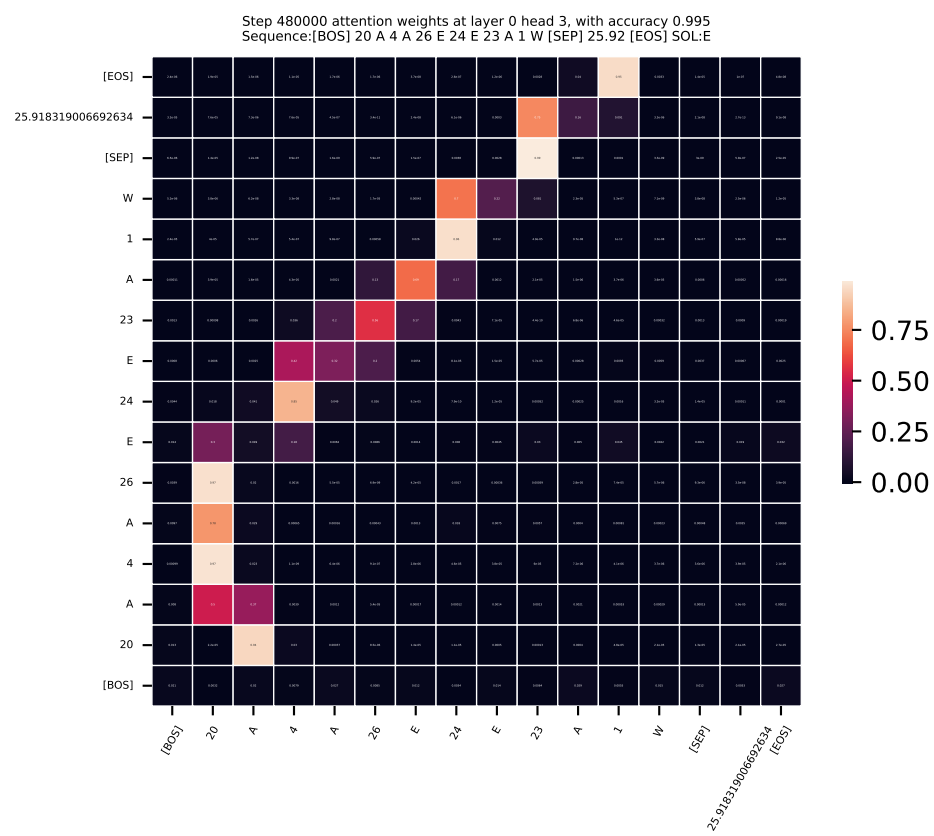

Step 479500 attention weights at layer 1 head 2, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

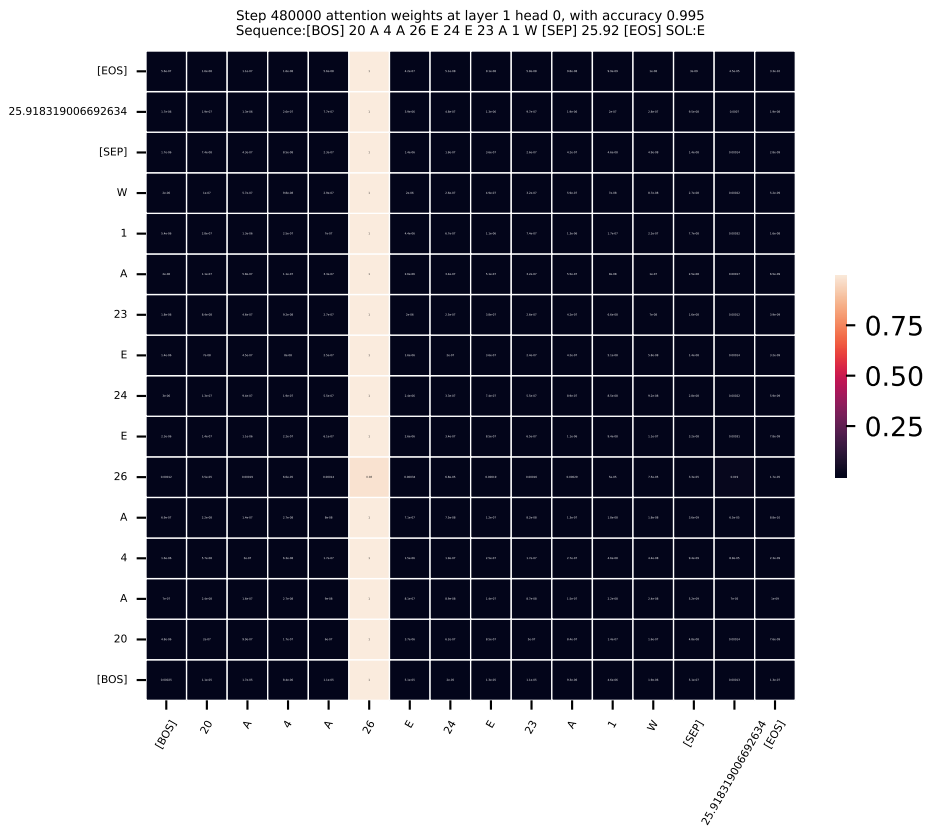

Step 479500 attention weights at layer 1 head 3, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

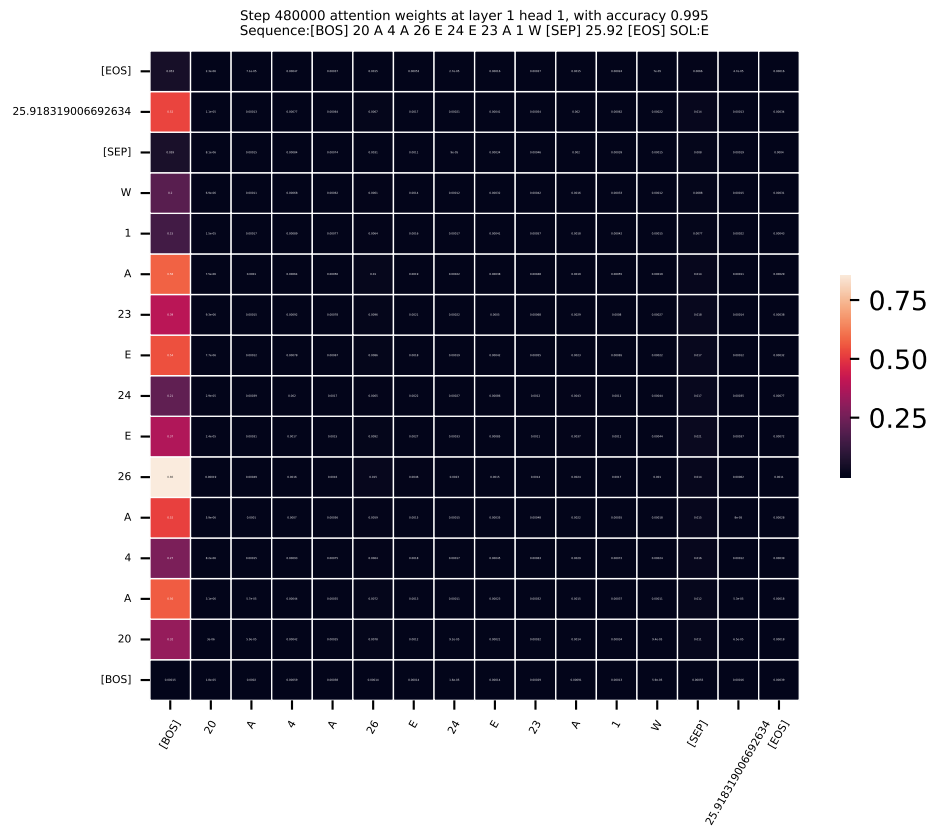

Step 480000 attention weights at layer 0 head 0, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

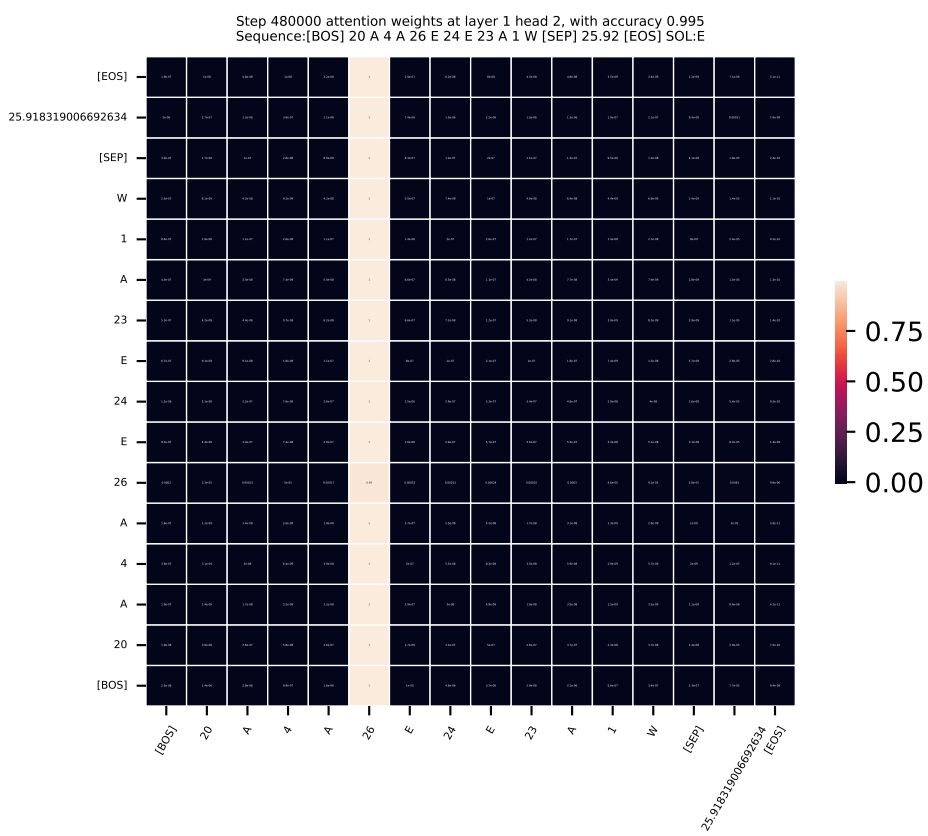

Step 480000 attention weights at layer 0 head 1, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

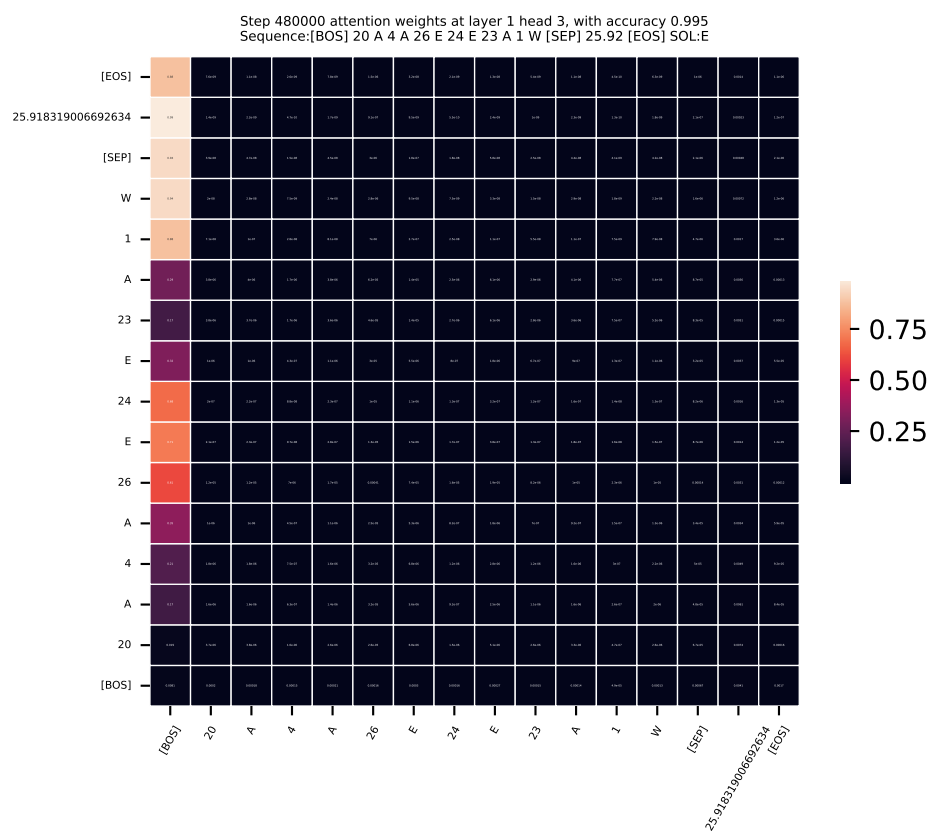

Step 480000 attention weights at layer 0 head 2, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

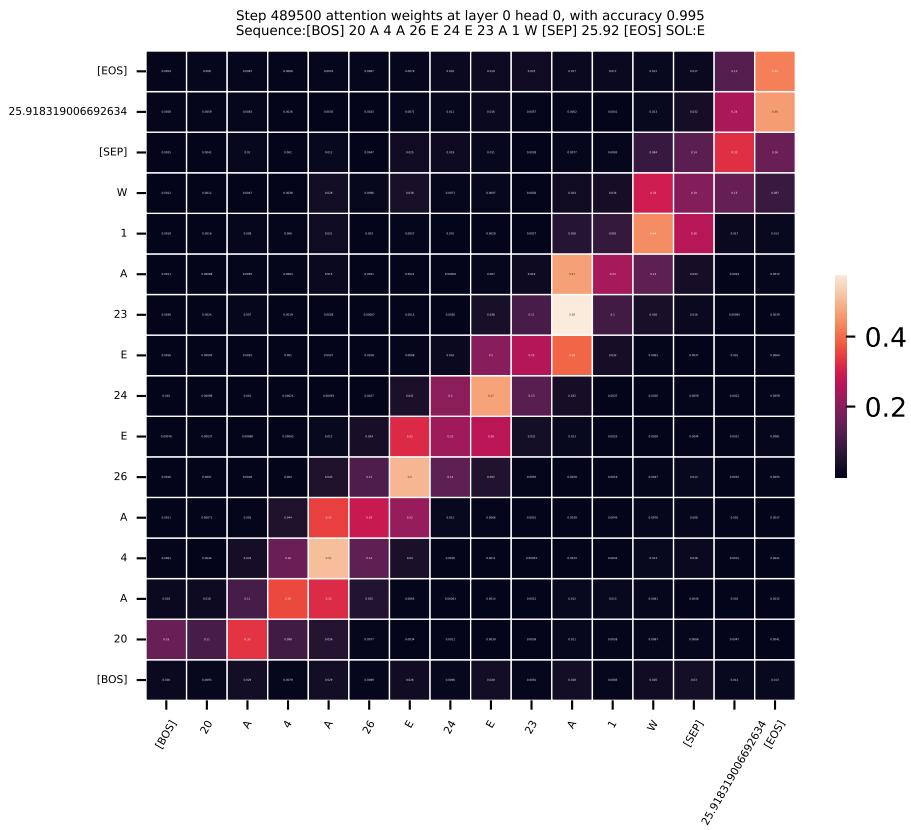

Step 480000 attention weights at layer 0 head 3, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

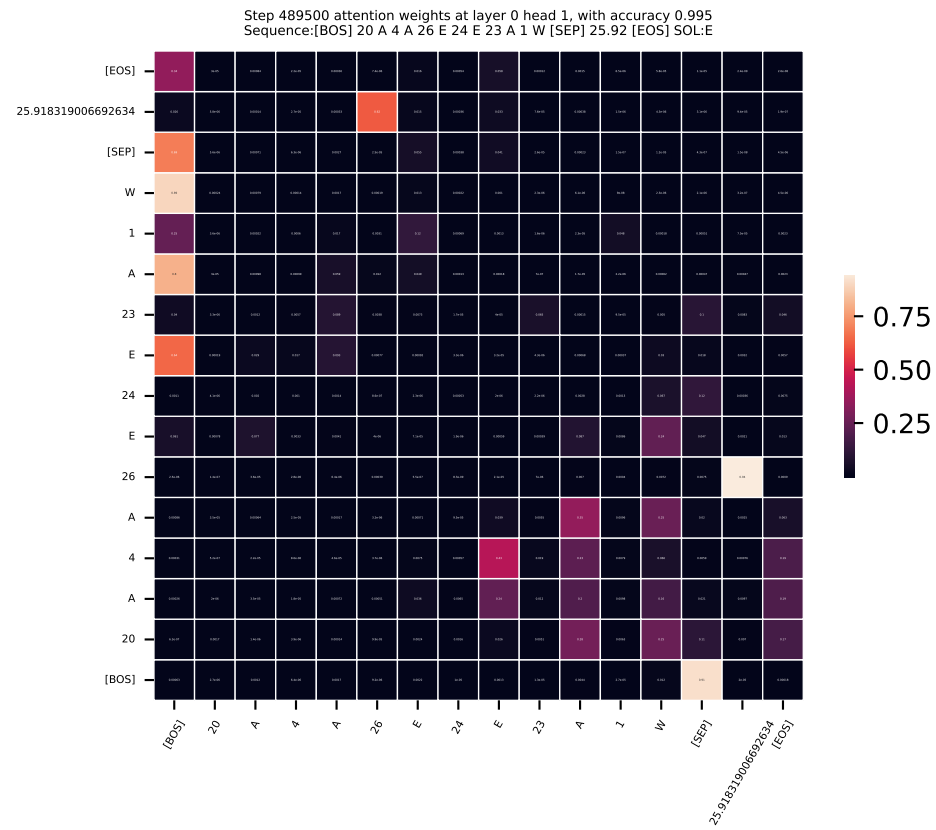

Step 480000 attention weights at layer 1 head 0, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

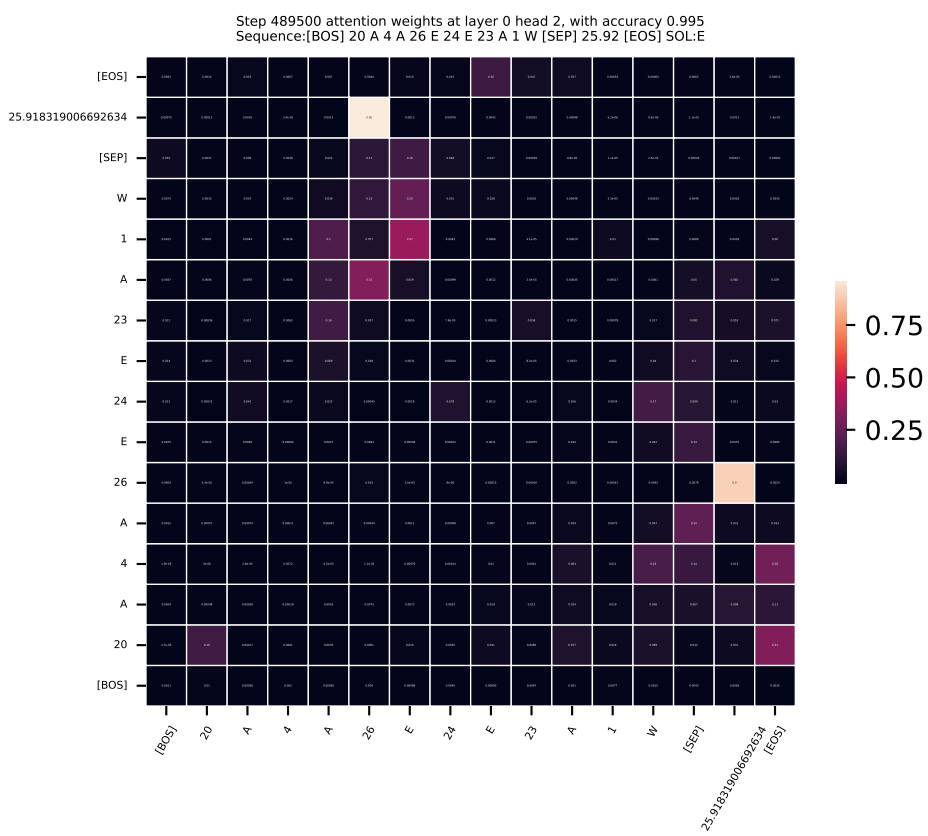

Step 480000 attention weights at layer 1 head 1, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

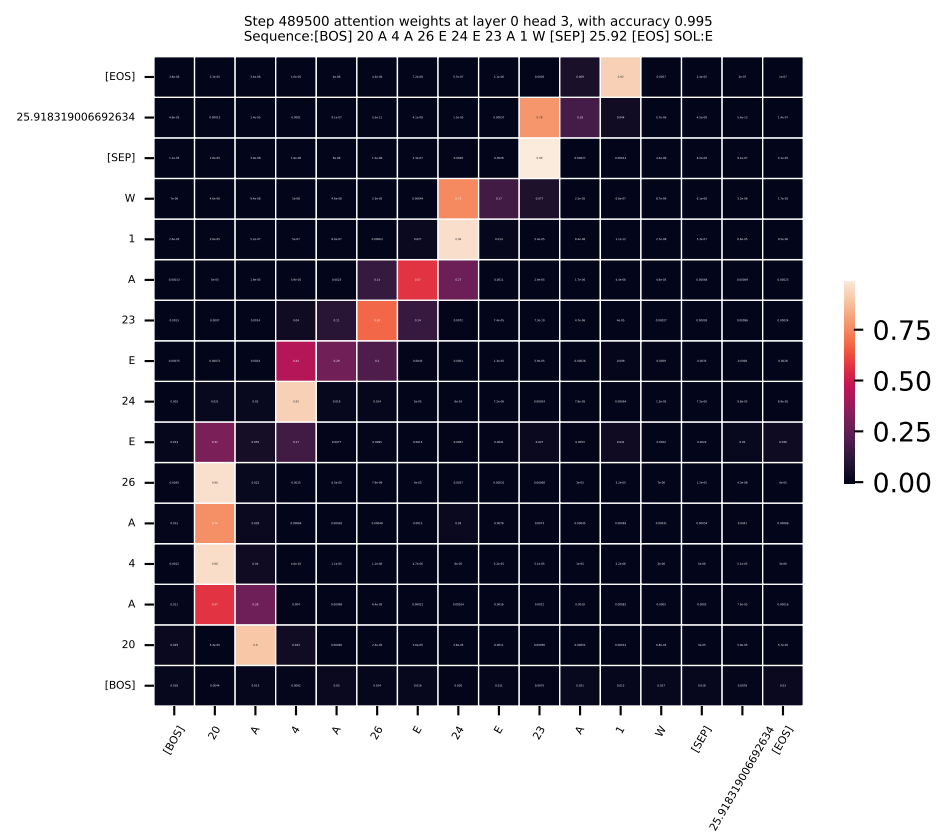

Step 480000 attention weights at layer 1 head 2, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

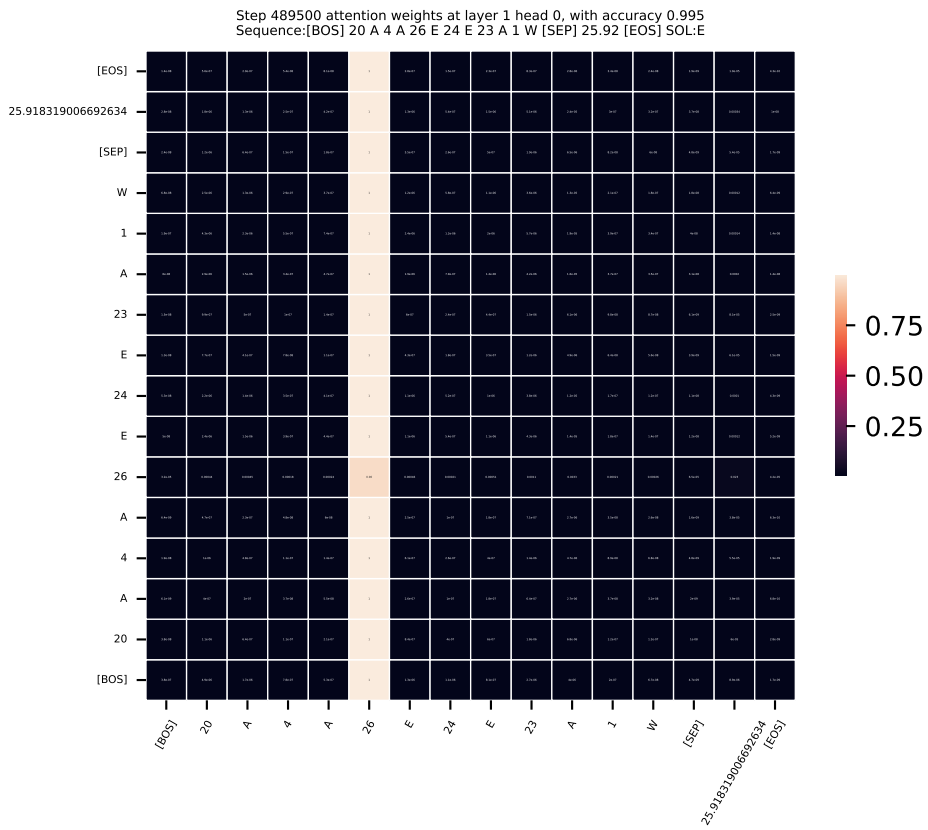

Step 480000 attention weights at layer 1 head 3, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

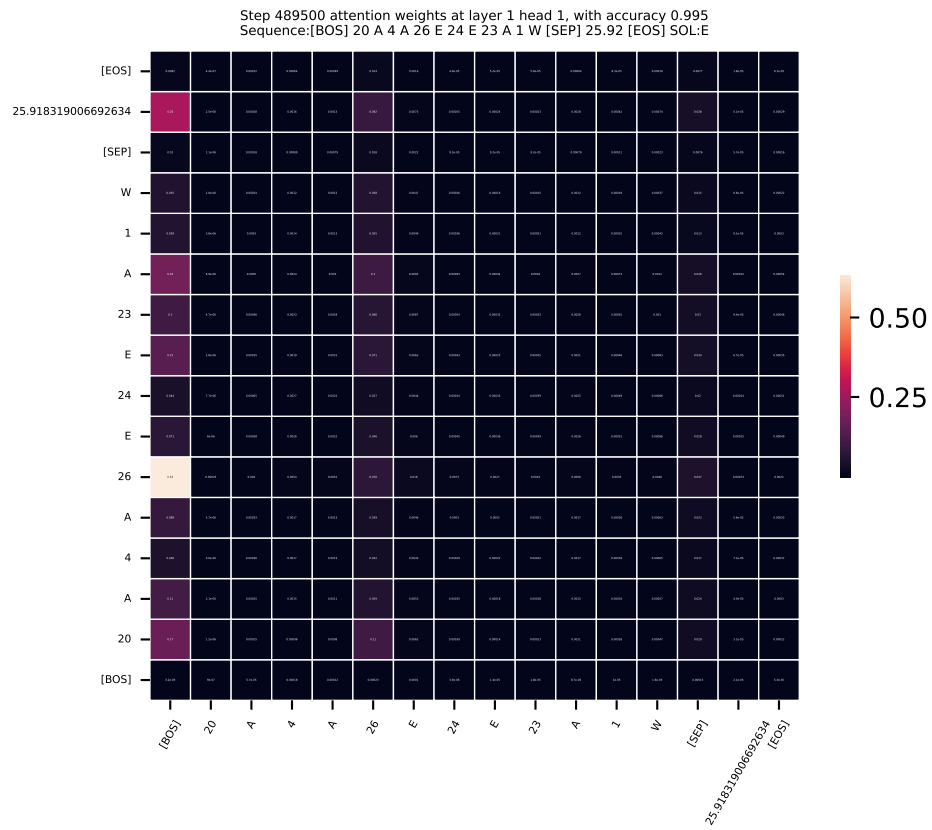

Step 489500 attention weights at layer 0 head 0, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

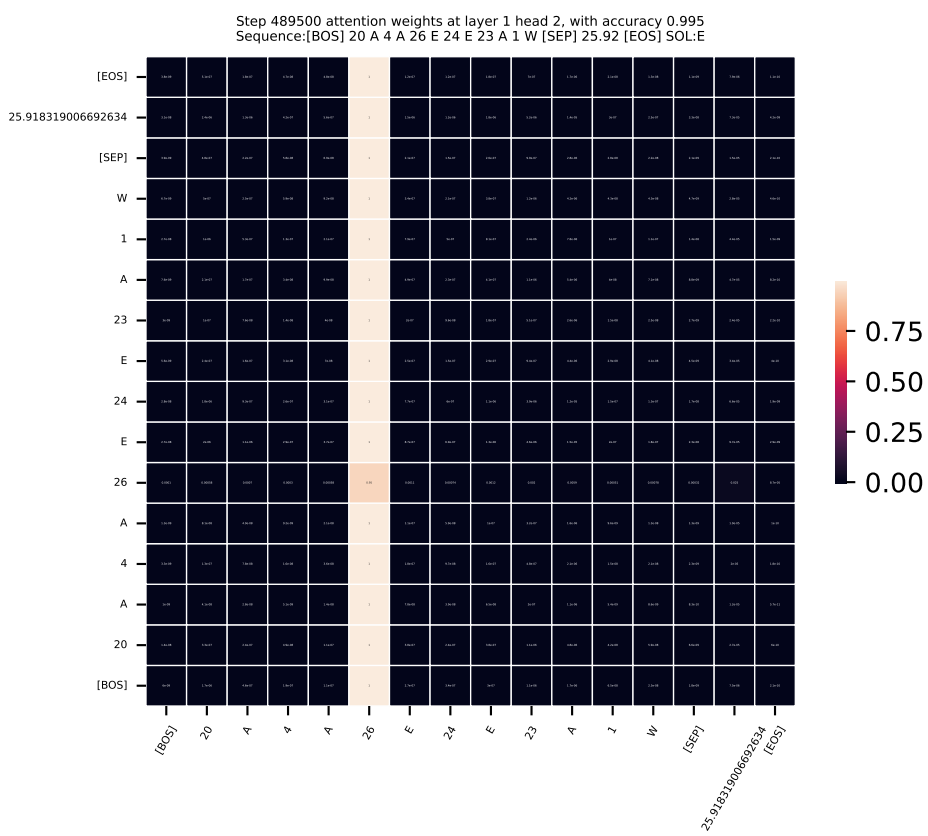

Step 489500 attention weights at layer 0 head 1, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

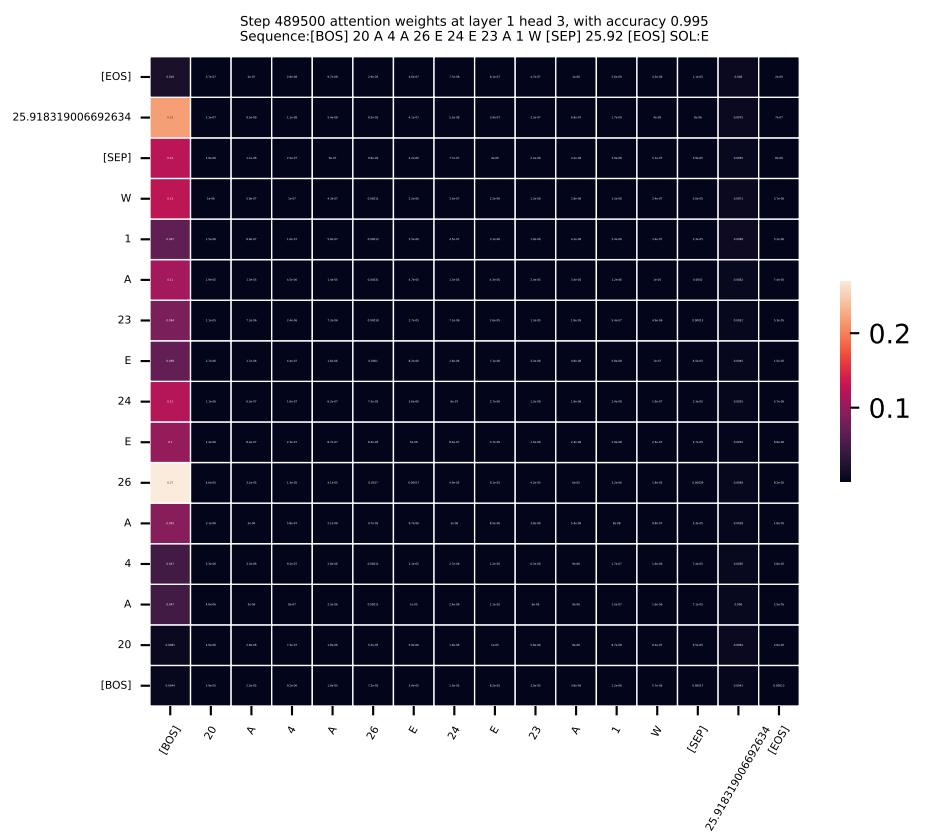

Step 489500 attention weights at layer 0 head 2, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

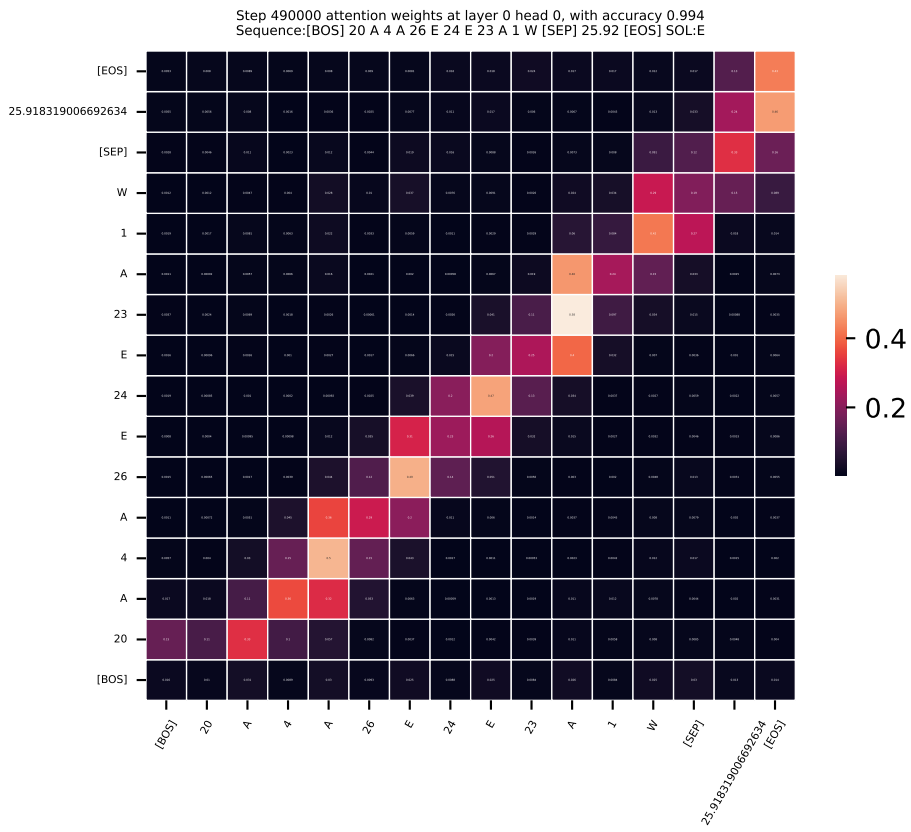

Step 489500 attention weights at layer 0 head 3, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

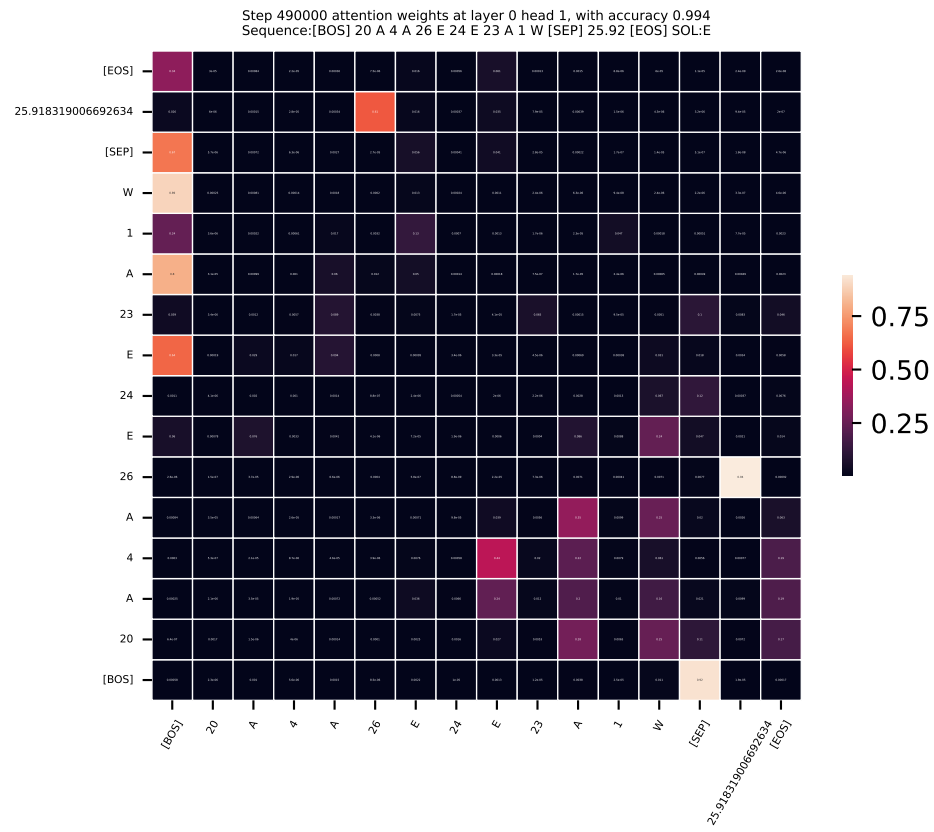

Step 489500 attention weights at layer 1 head 0, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

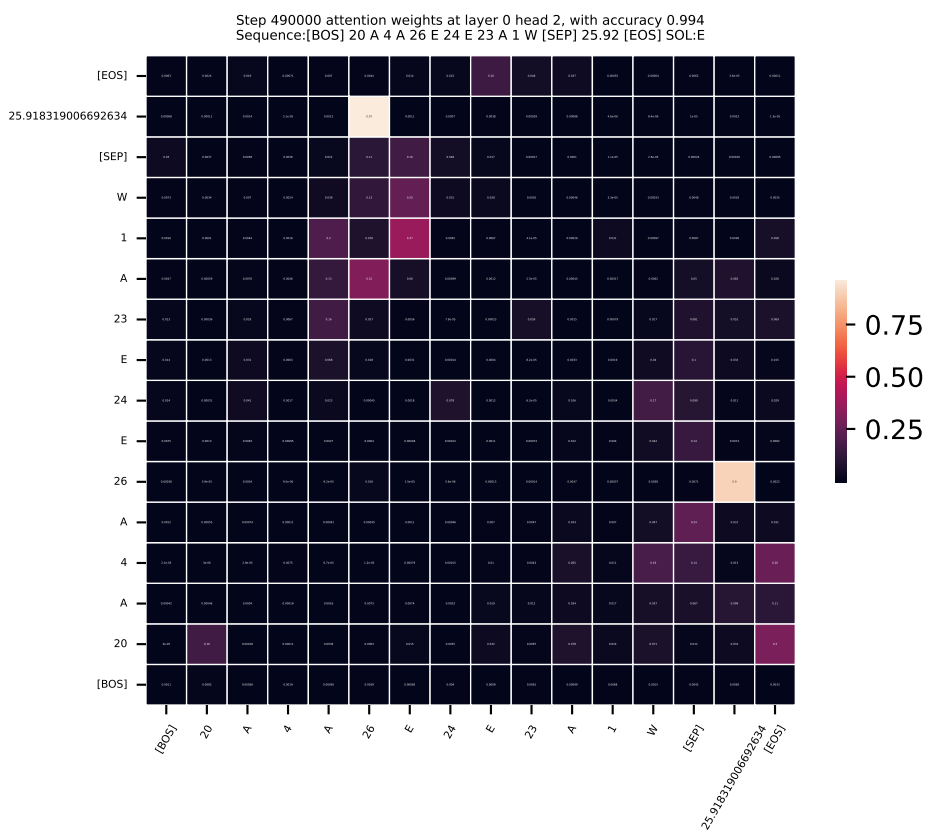

Step 489500 attention weights at layer 1 head 1, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

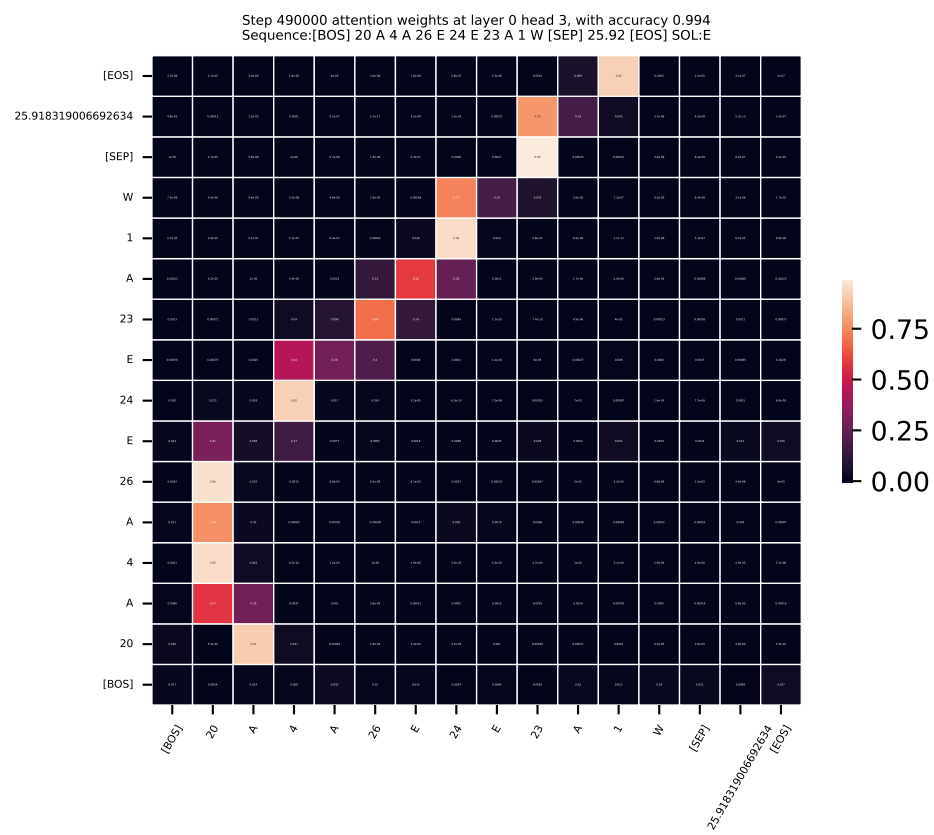

Step 489500 attention weights at layer 1 head 2, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

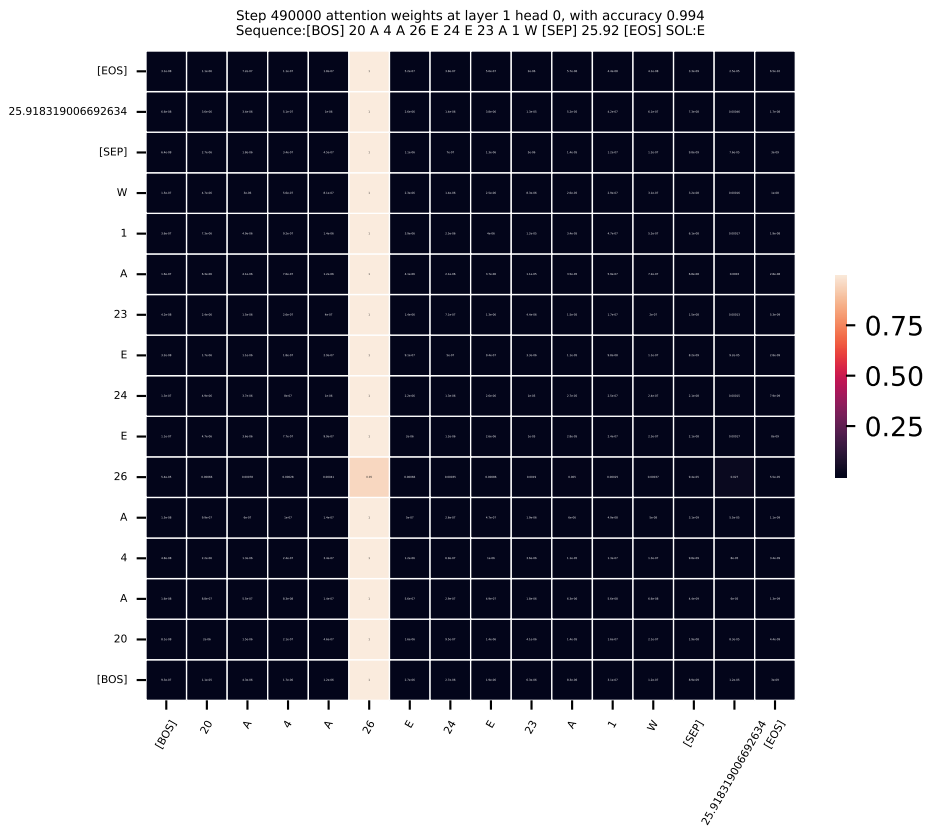

Step 489500 attention weights at layer 1 head 3, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.2
0.1

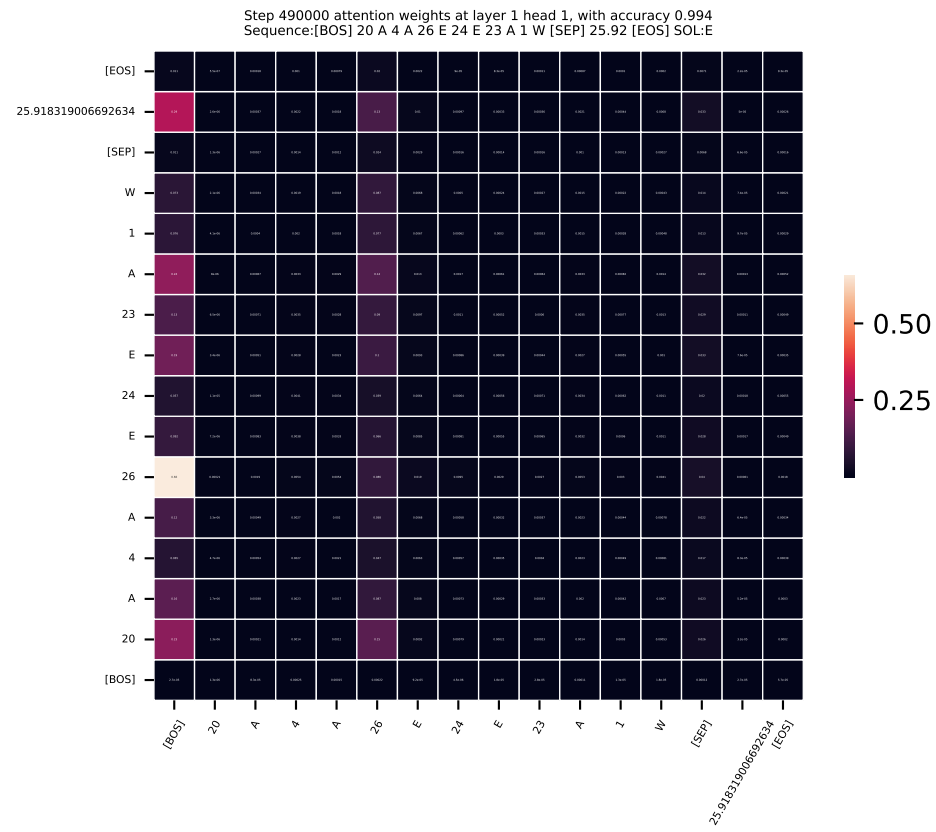

Step 490000 attention weights at layer 0 head 0, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

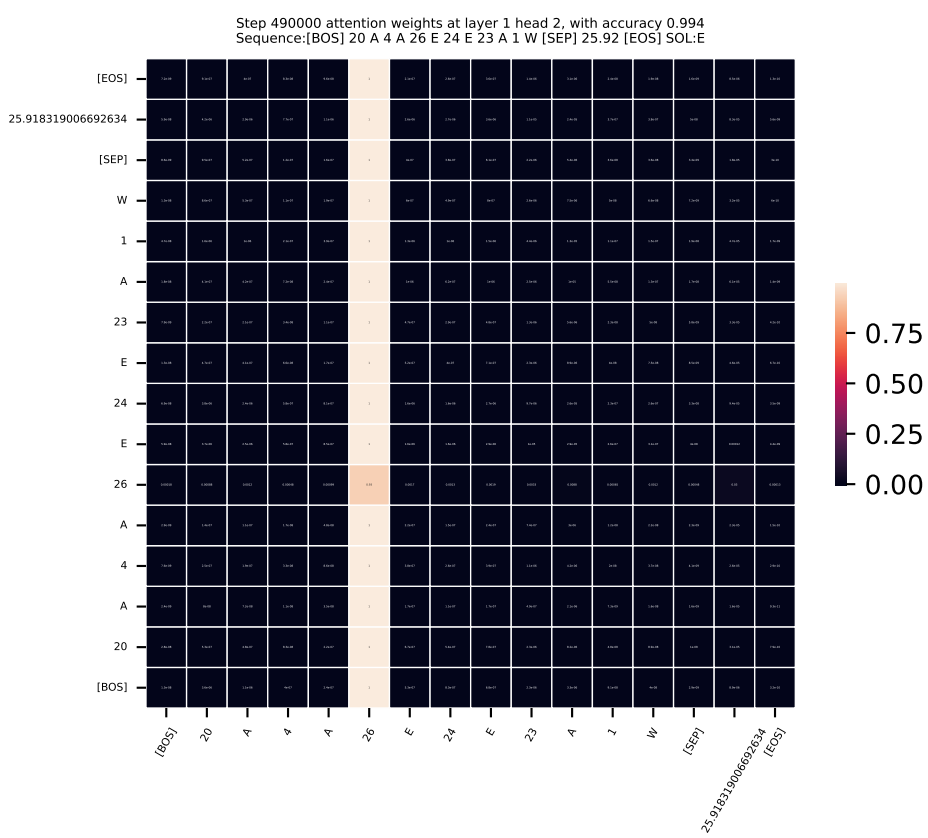

Step 490000 attention weights at layer 0 head 1, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

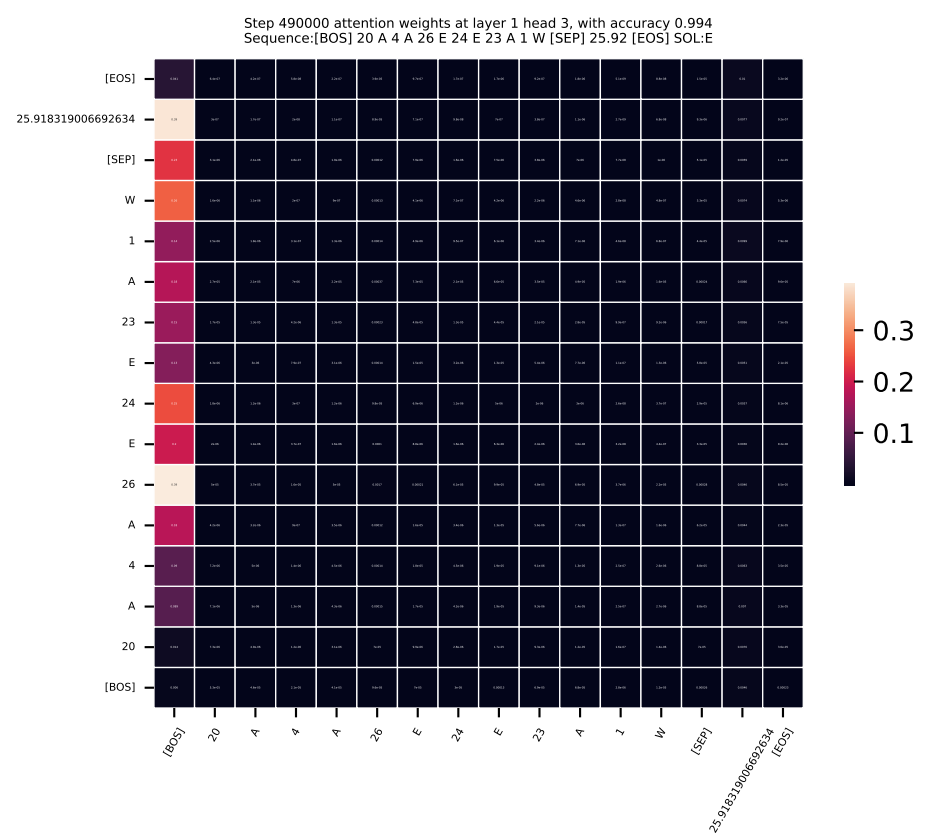

Step 490000 attention weights at layer 0 head 2, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

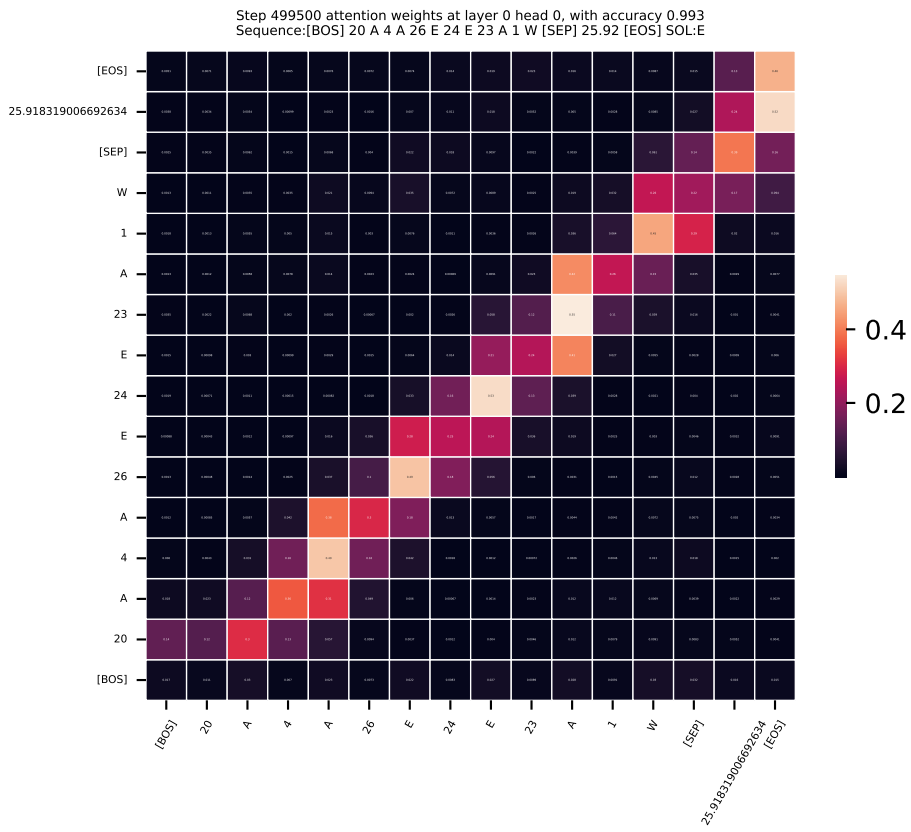

Step 490000 attention weights at layer 0 head 3, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

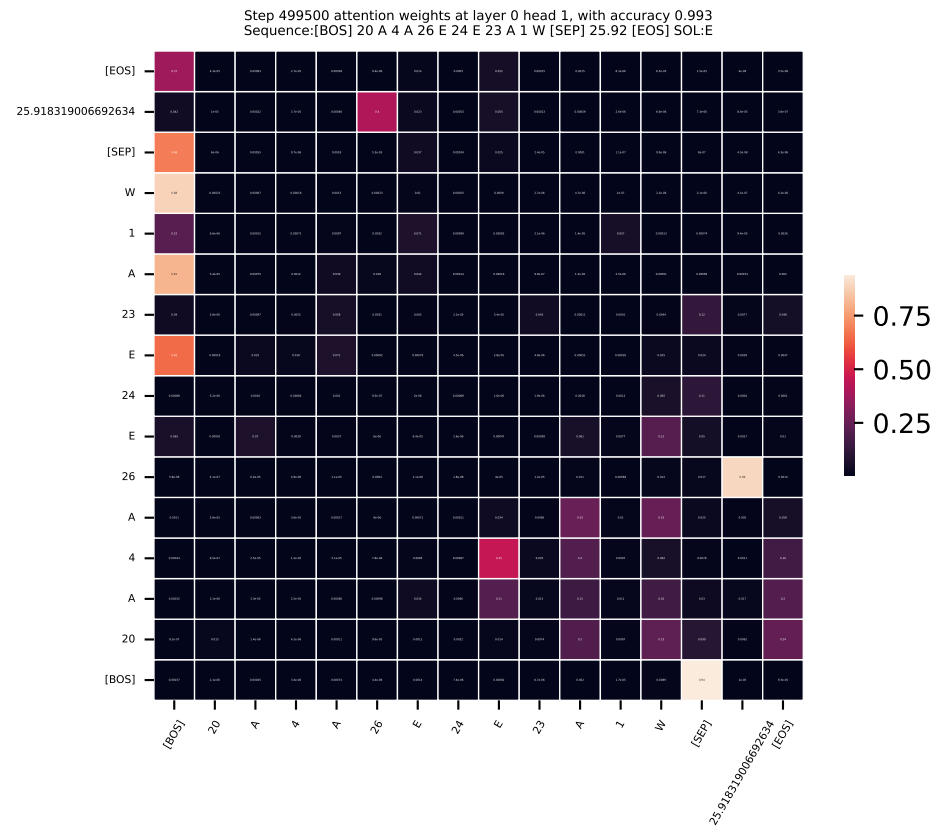

Step 490000 attention weights at layer 1 head 0, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

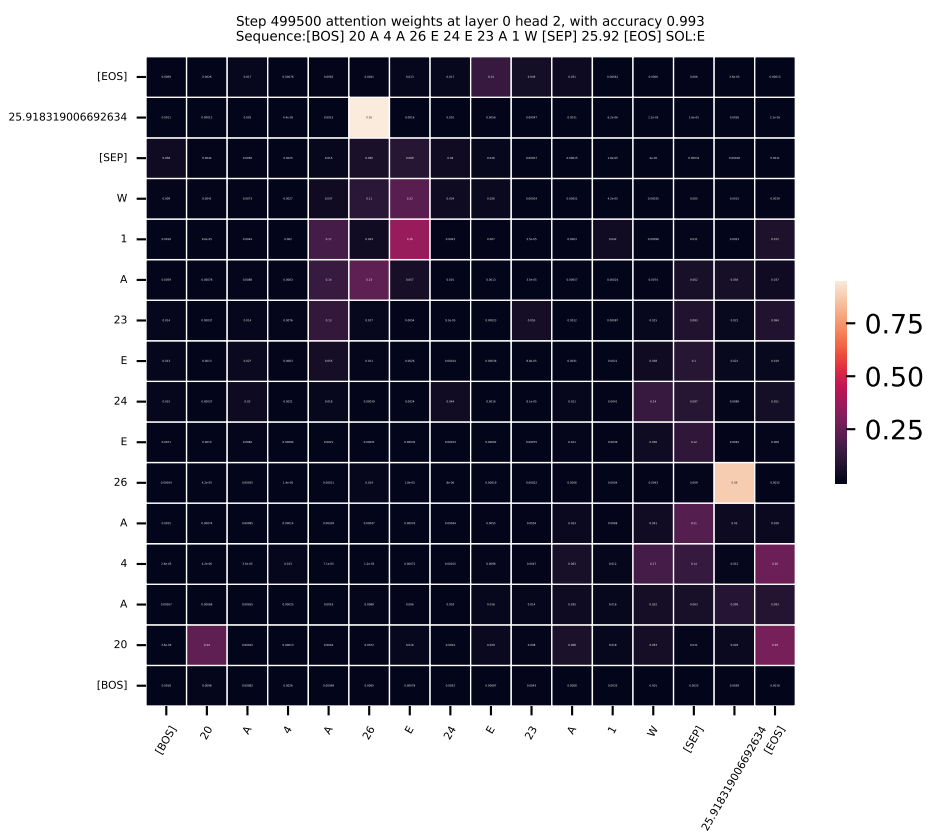

Step 490000 attention weights at layer 1 head 1, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

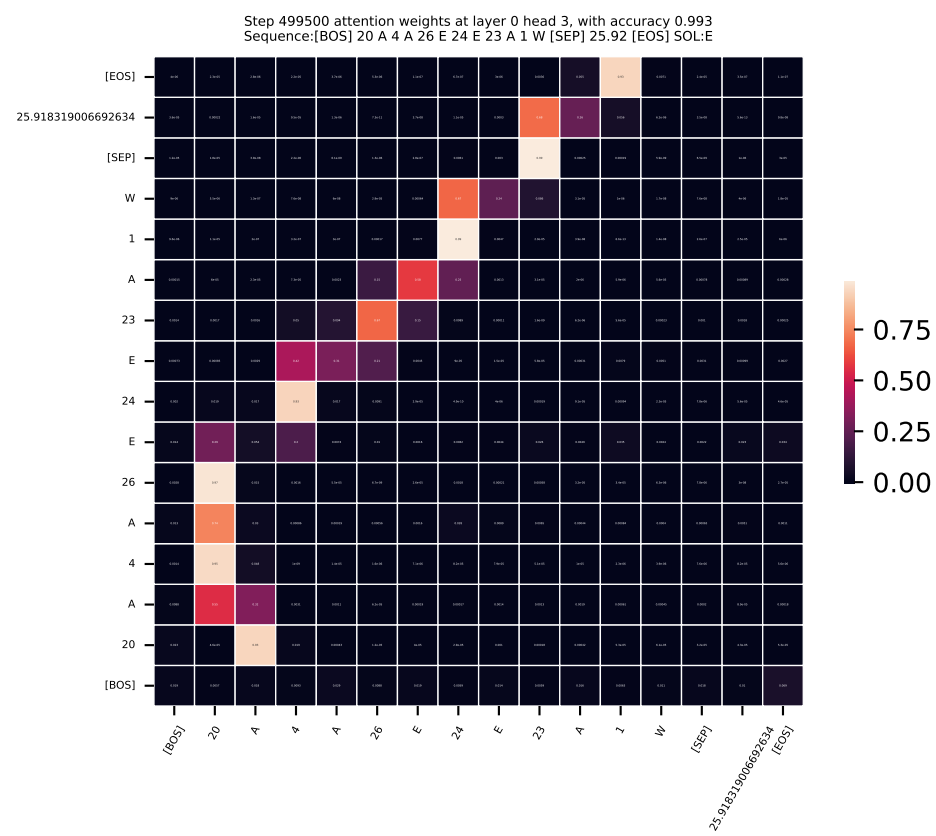

Step 490000 attention weights at layer 1 head 2, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

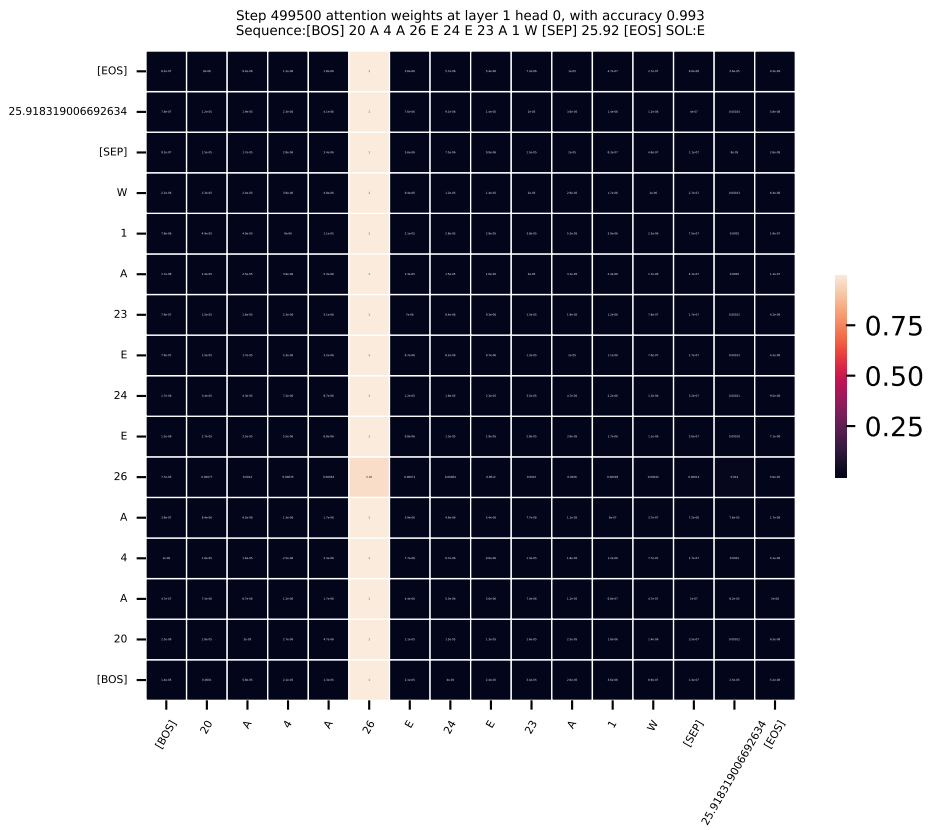

Step 490000 attention weights at layer 1 head 3, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

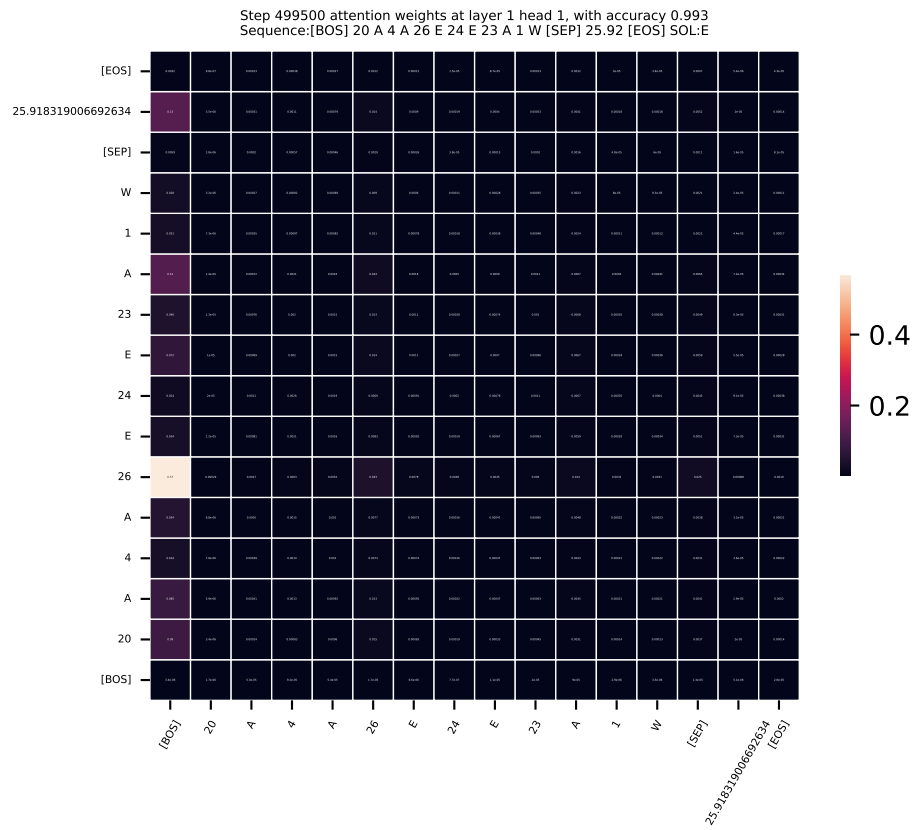

Step 499500 attention weights at layer 0 head 0, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

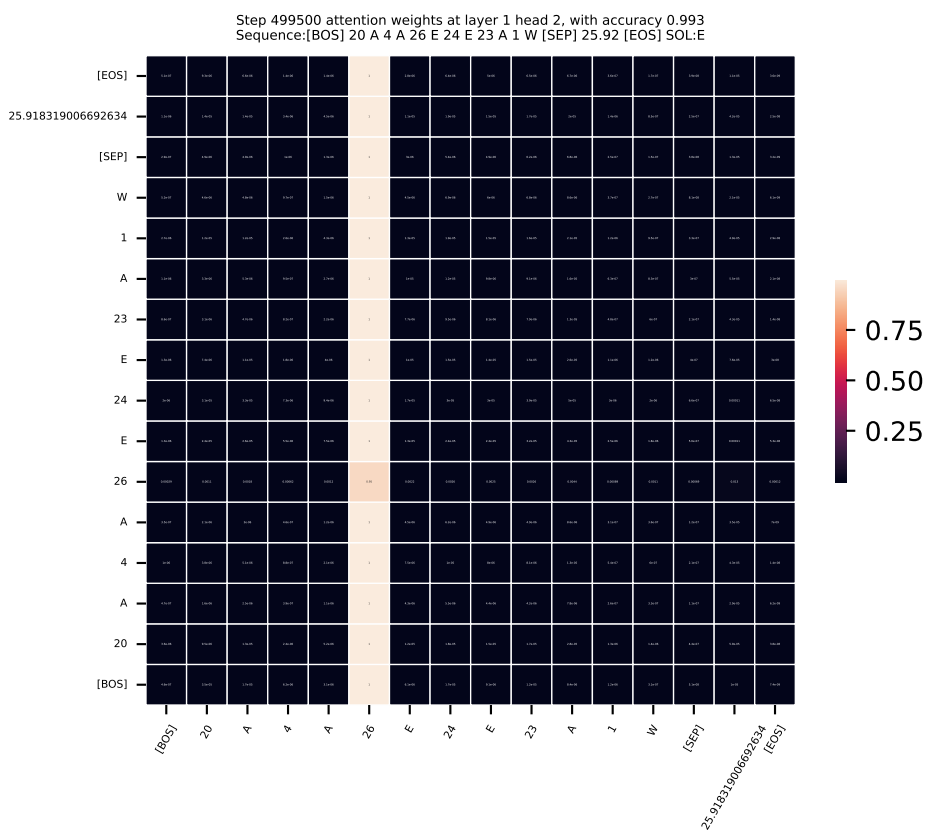

Step 499500 attention weights at layer 0 head 1, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

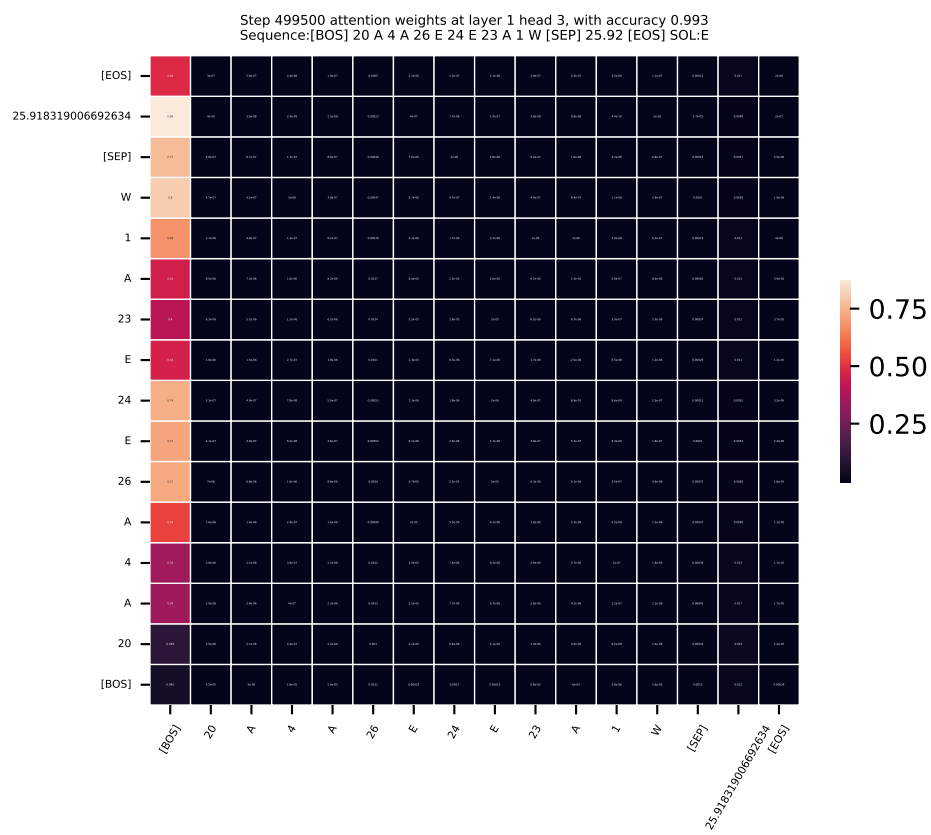

Step 499500 attention weights at layer 0 head 2, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

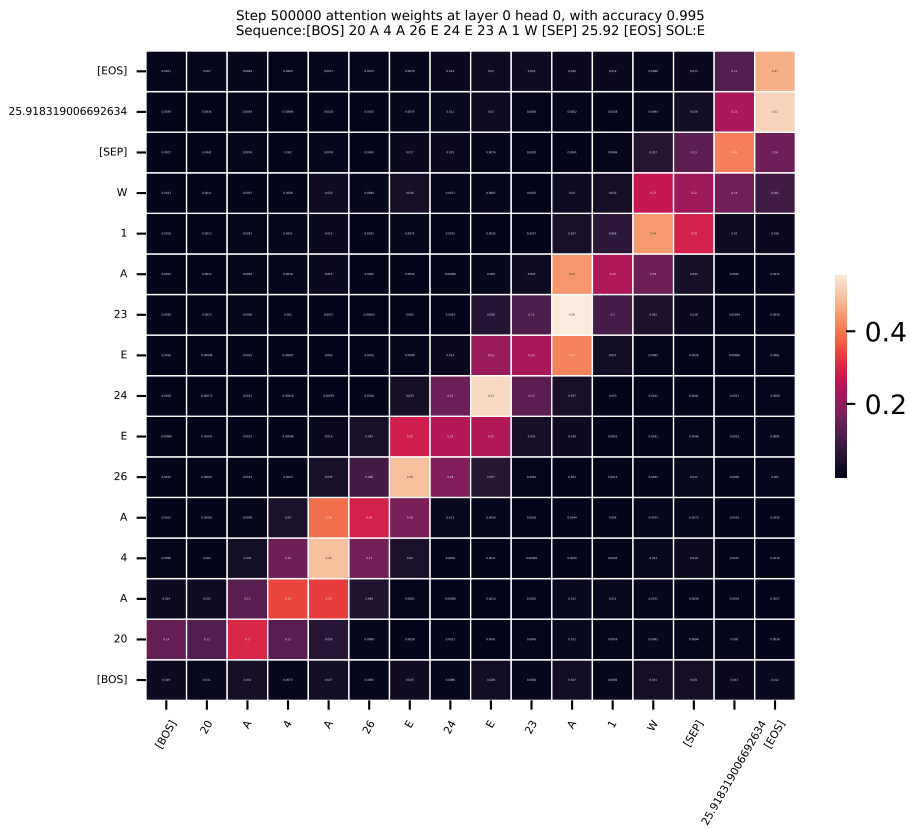

Step 499500 attention weights at layer 0 head 3, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

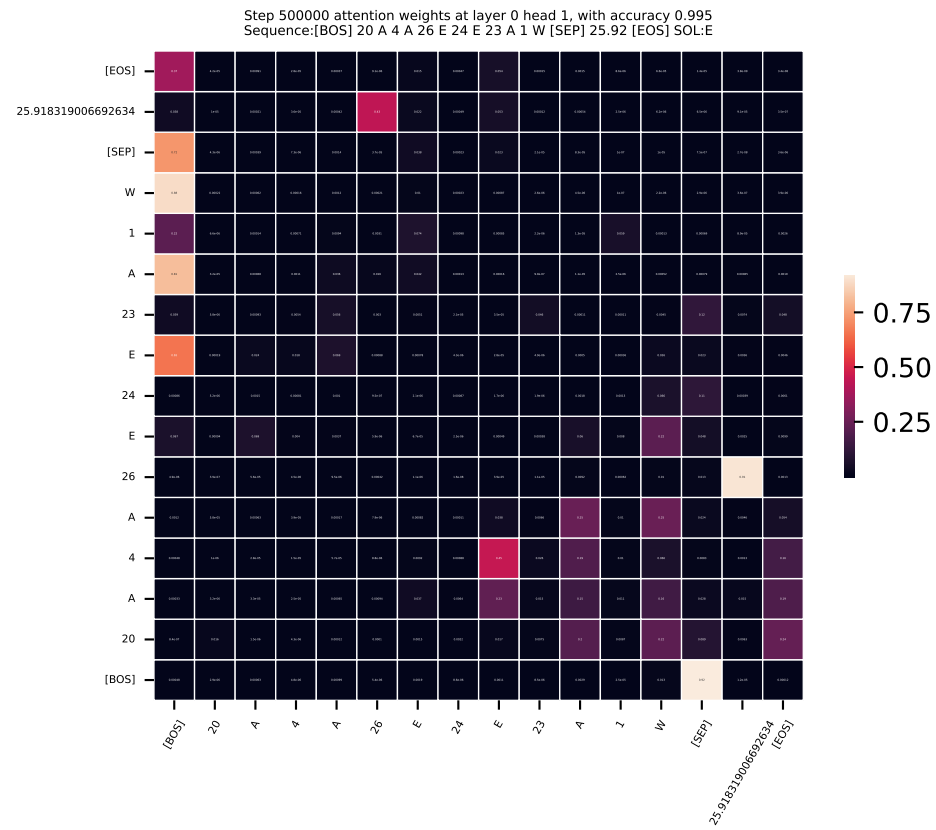

Step 499500 attention weights at layer 1 head 0, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

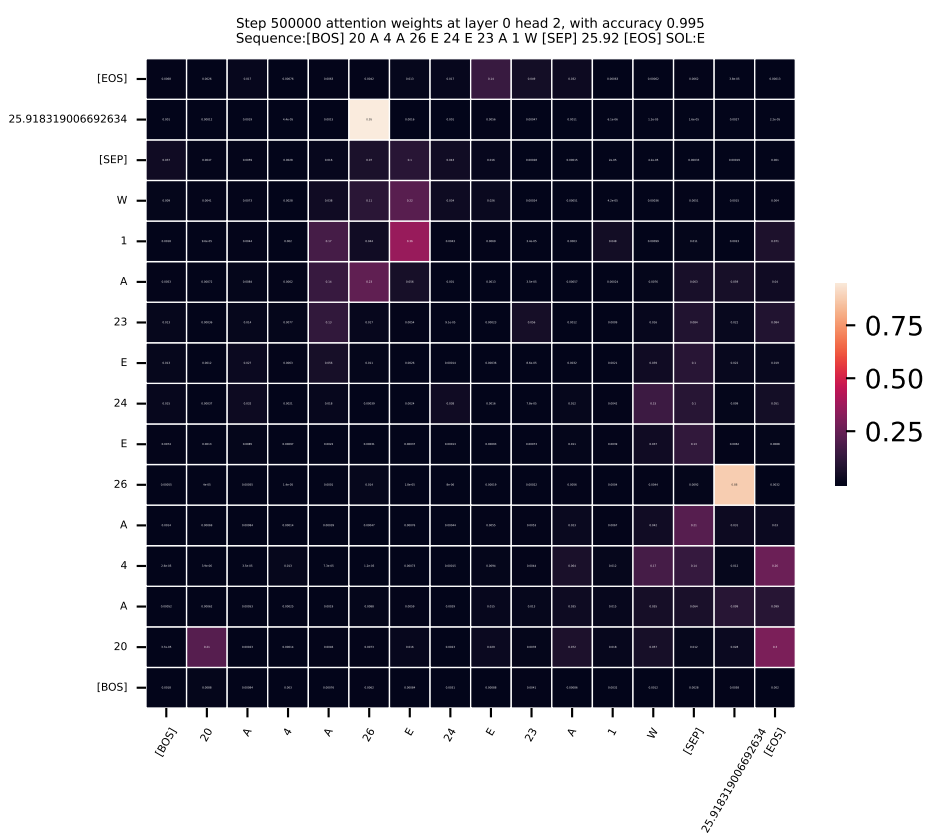

Step 499500 attention weights at layer 1 head 1, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

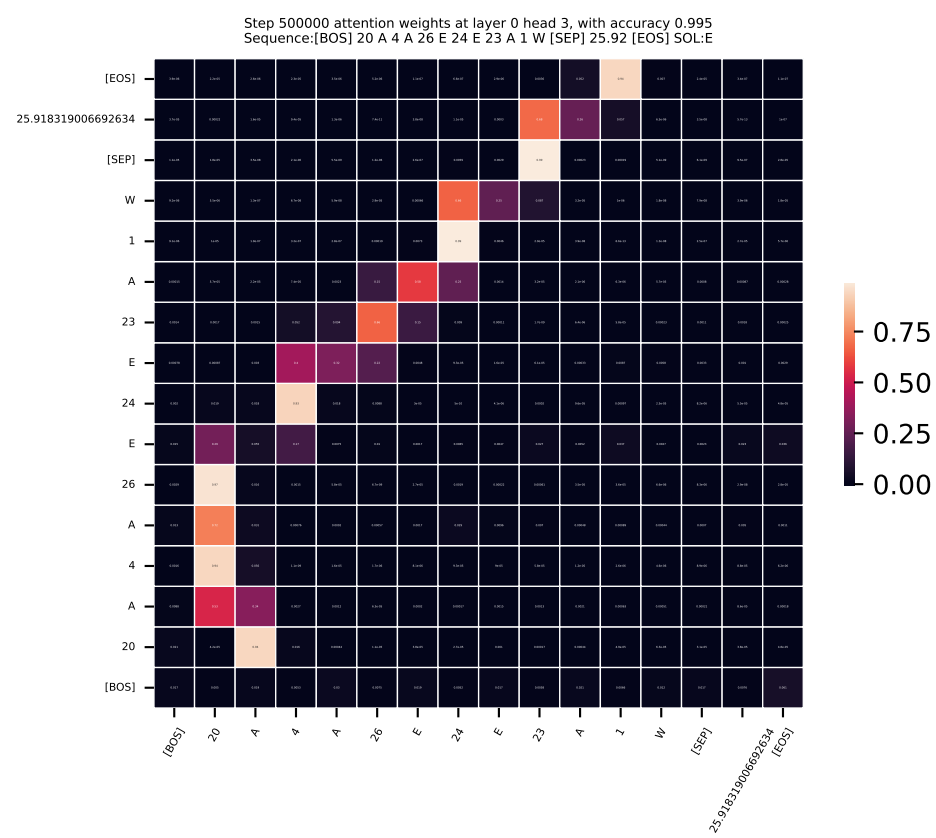

Step 499500 attention weights at layer 1 head 2, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

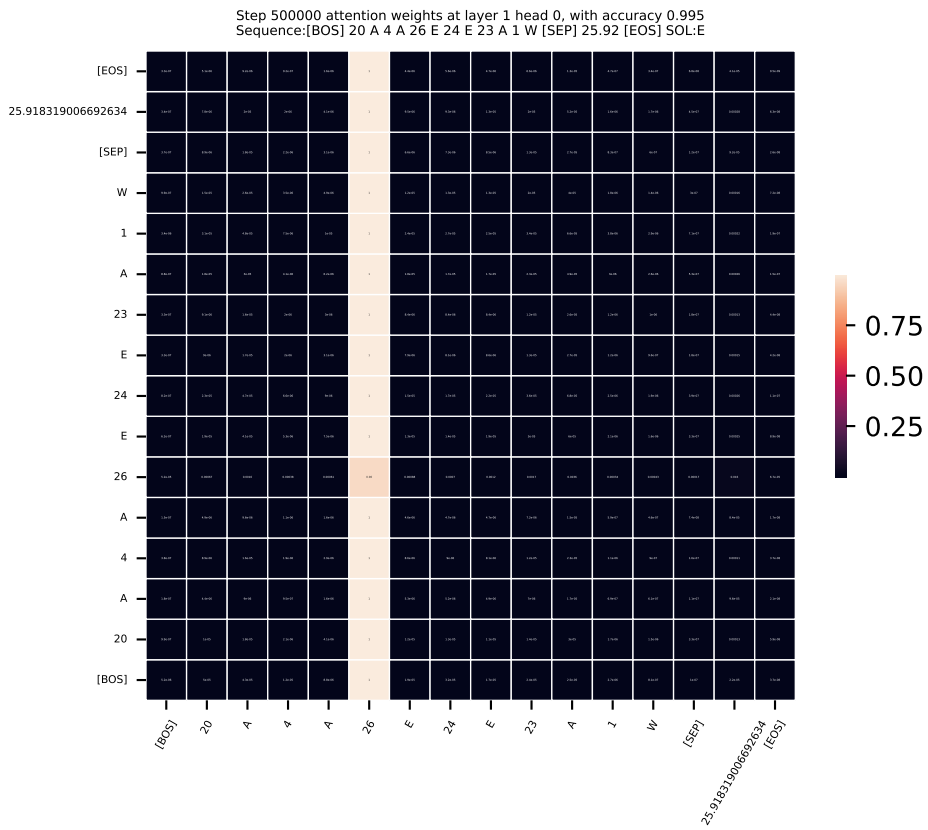

Step 499500 attention weights at layer 1 head 3, with accuracy 0.993
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

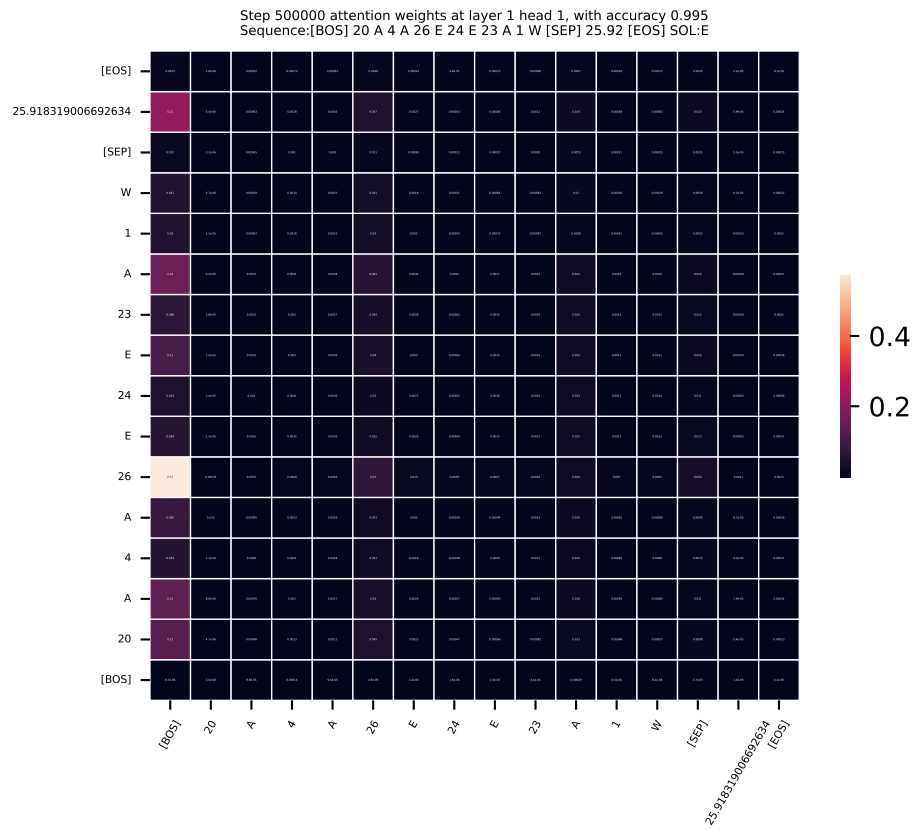

Step 500000 attention weights at layer 0 head 0, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.4
0.2

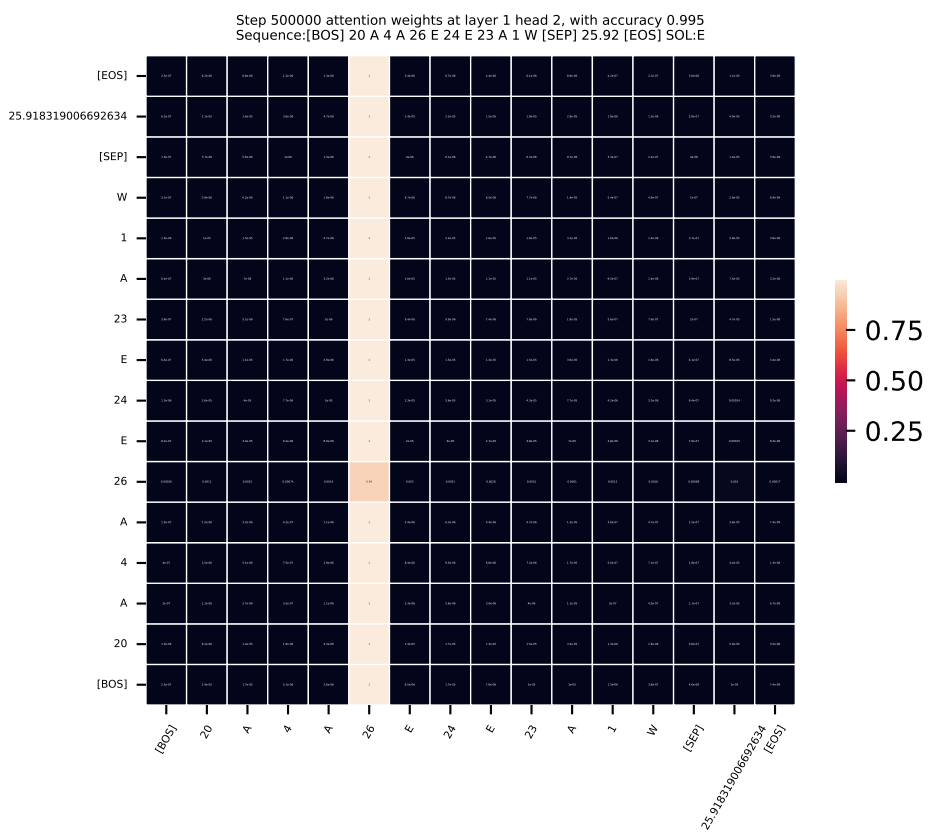

Step 500000 attention weights at layer 0 head 1, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

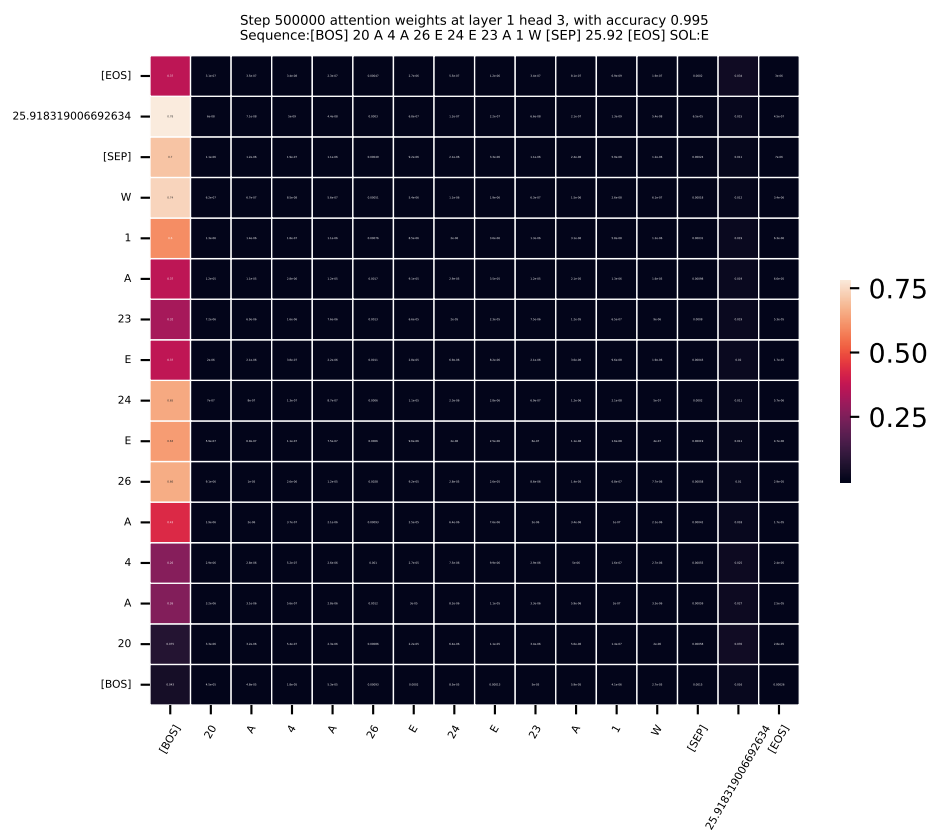

Step 500000 attention weights at layer 0 head 2, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

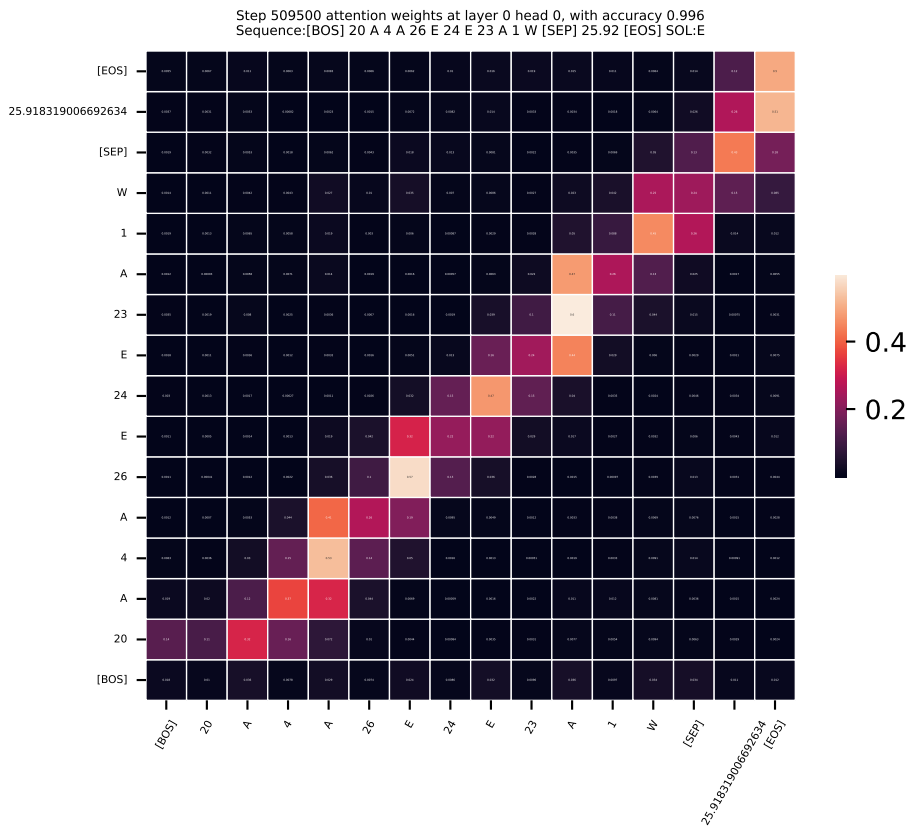

Step 500000 attention weights at layer 0 head 3, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

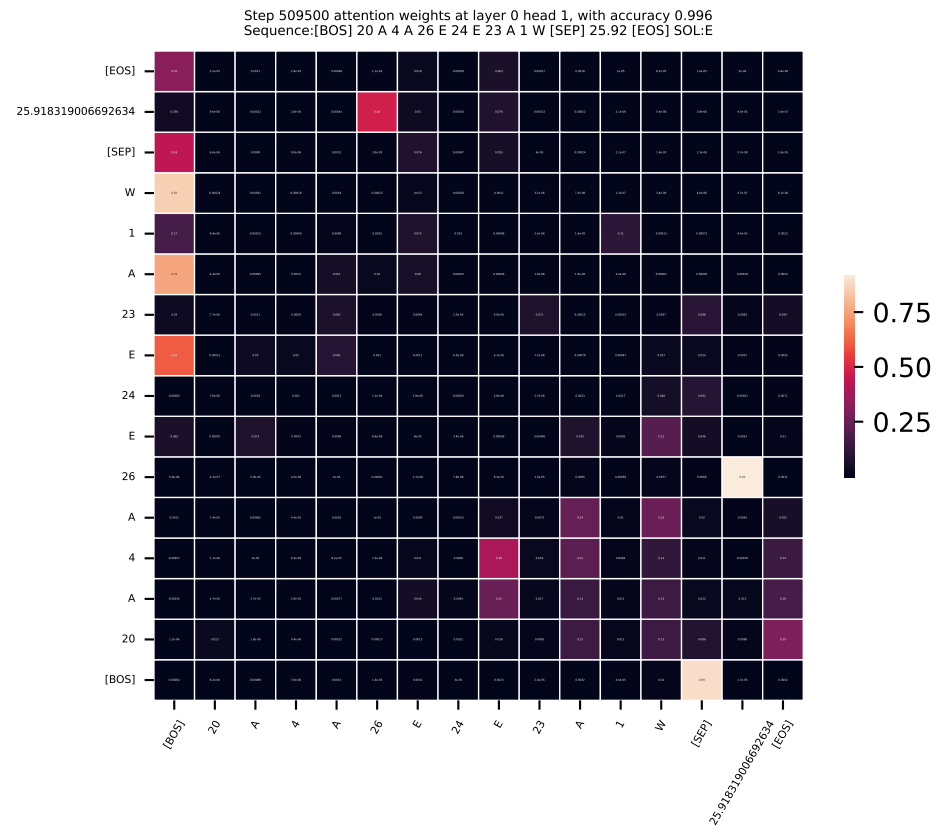

Step 500000 attention weights at layer 1 head 0, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

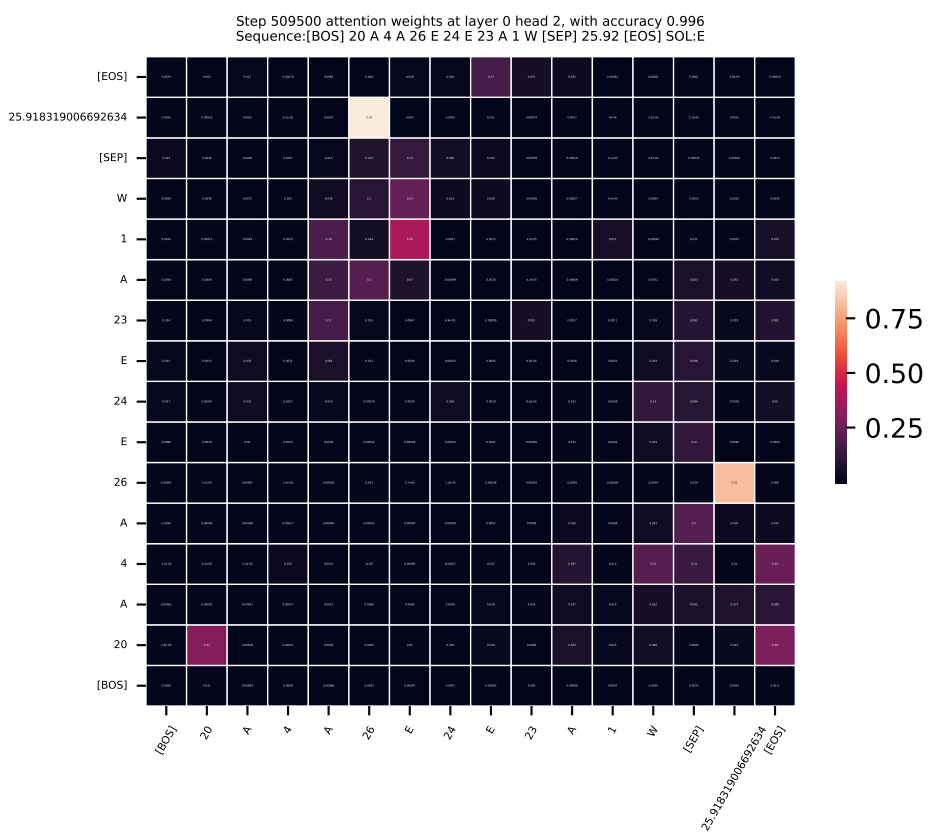

Step 500000 attention weights at layer 1 head 1, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

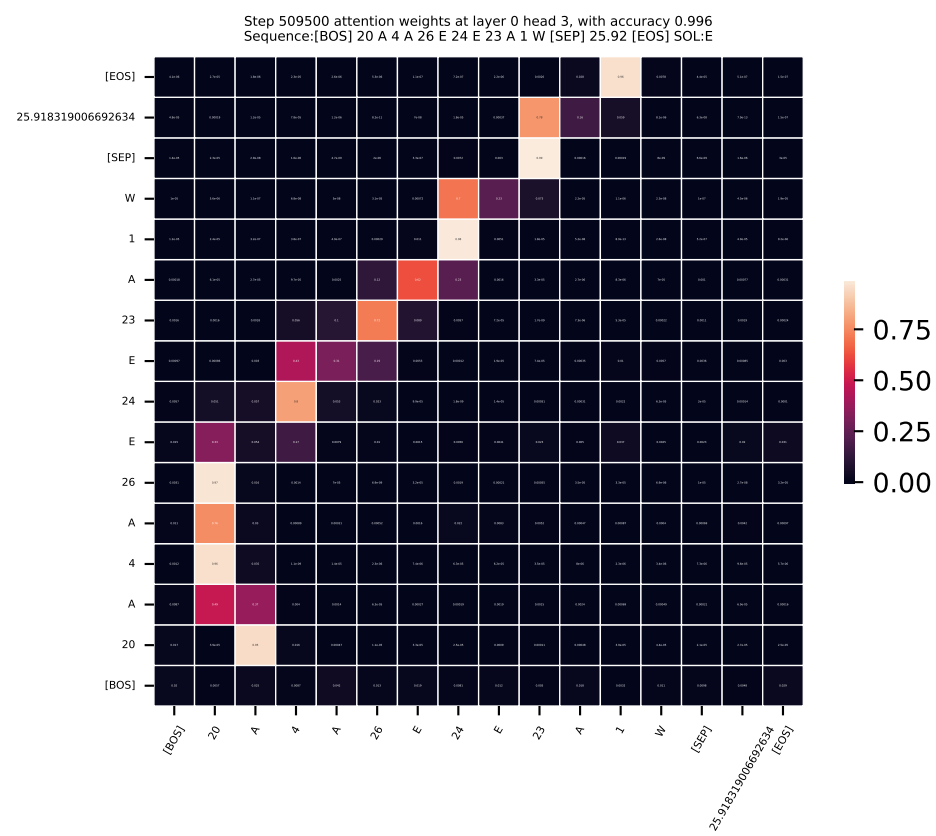

Step 500000 attention weights at layer 1 head 2, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

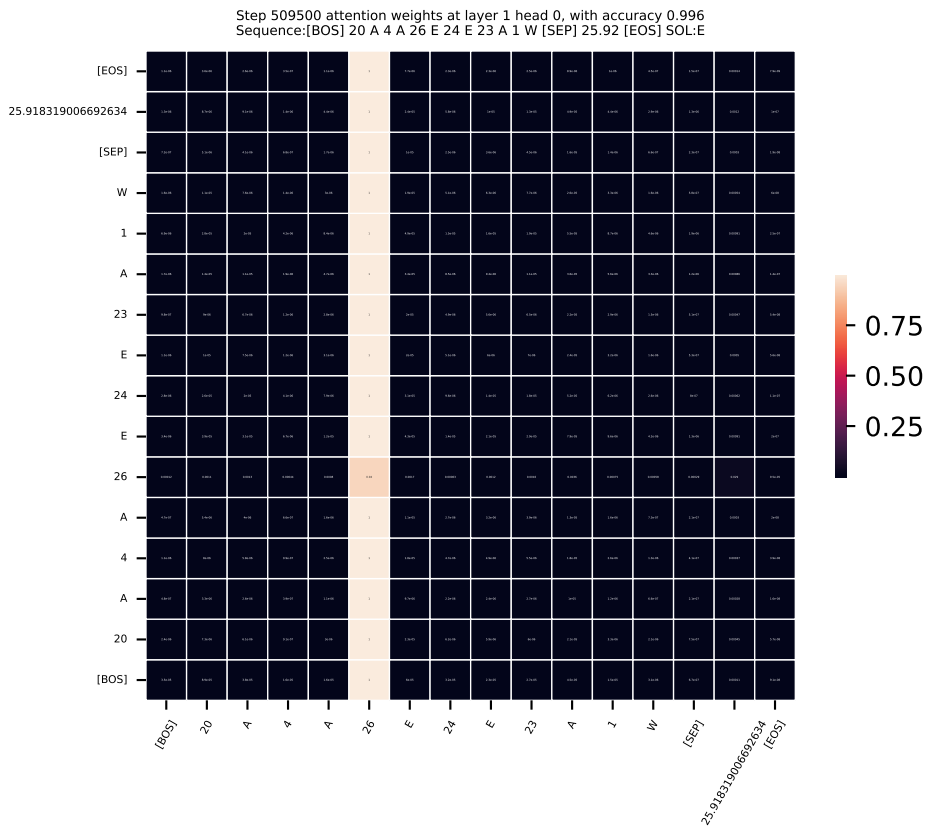

Step 500000 attention weights at layer 1 head 3, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

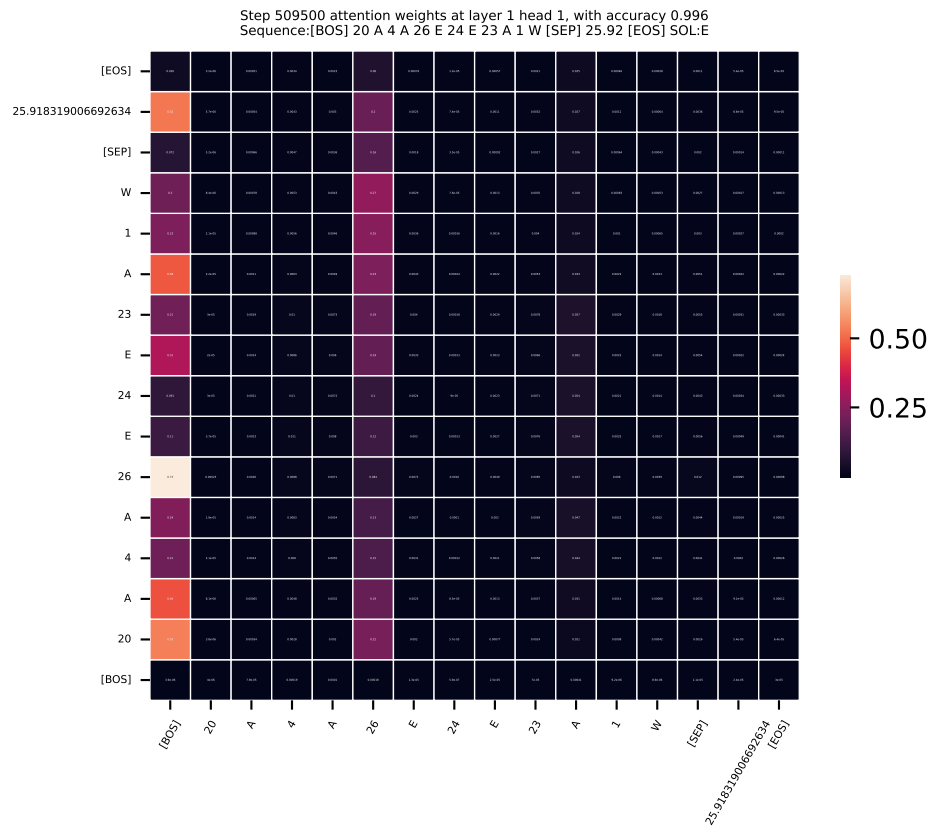

Step 509500 attention weights at layer 0 head 0, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

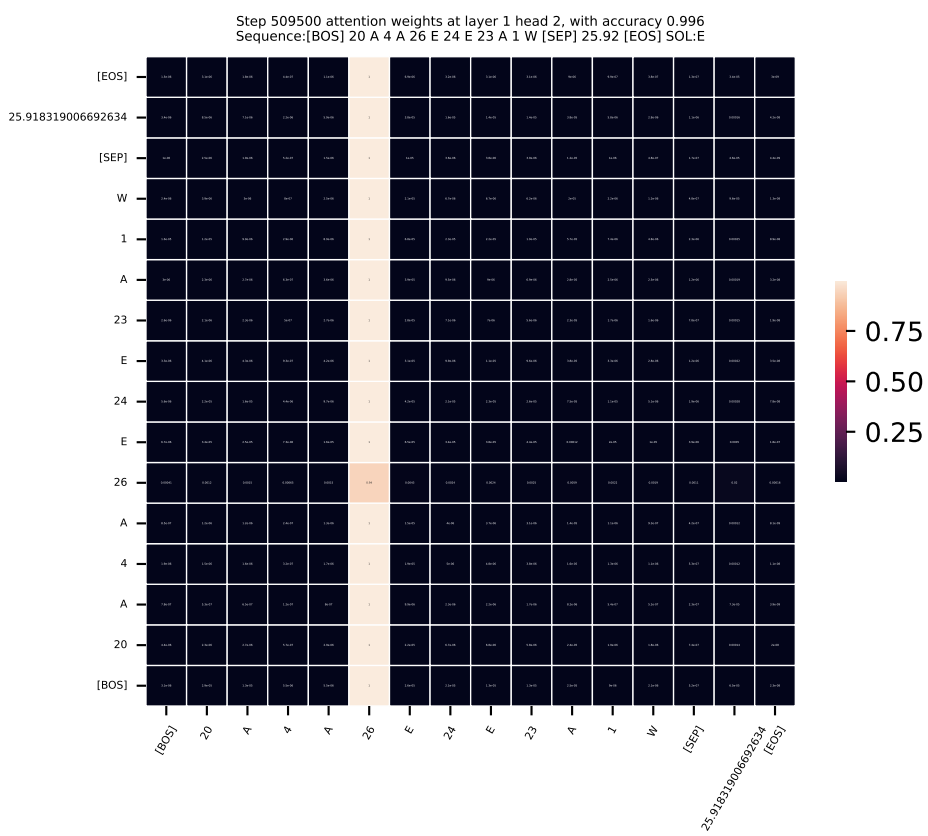

Step 509500 attention weights at layer 0 head 1, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

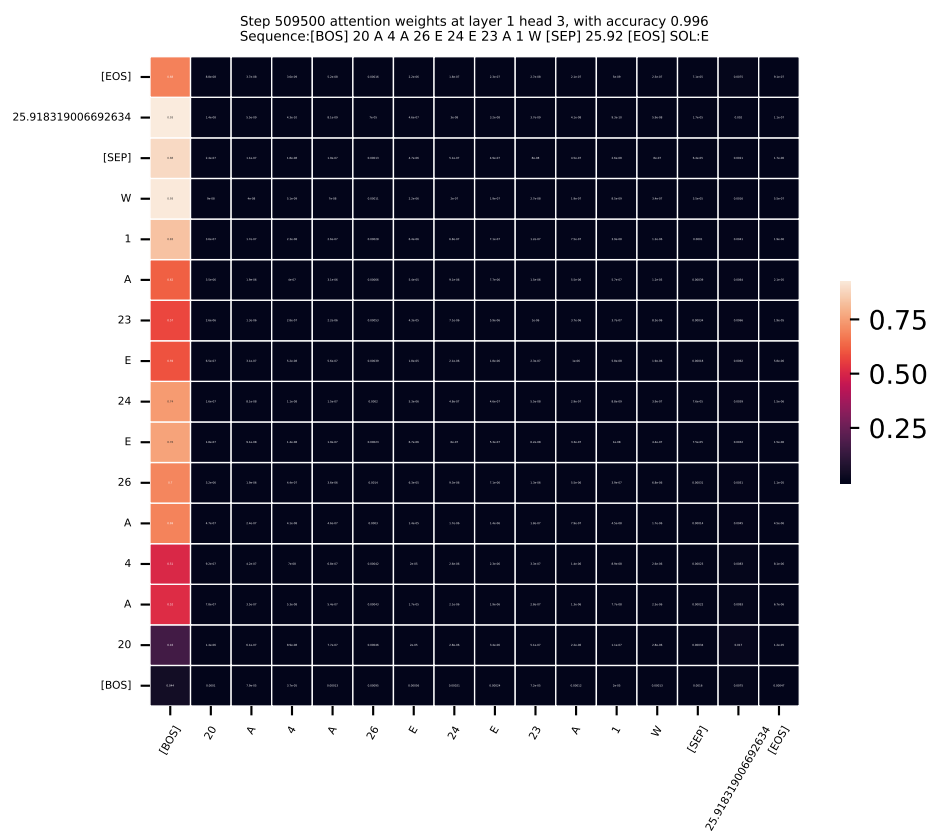

Step 509500 attention weights at layer 0 head 2, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

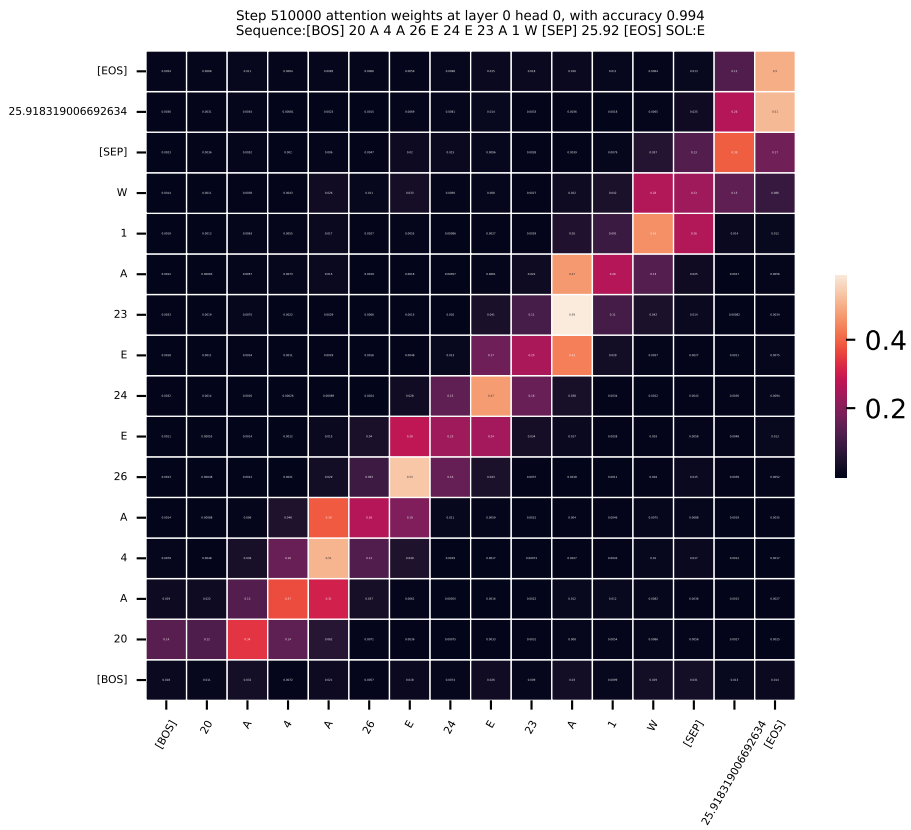

Step 509500 attention weights at layer 0 head 3, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

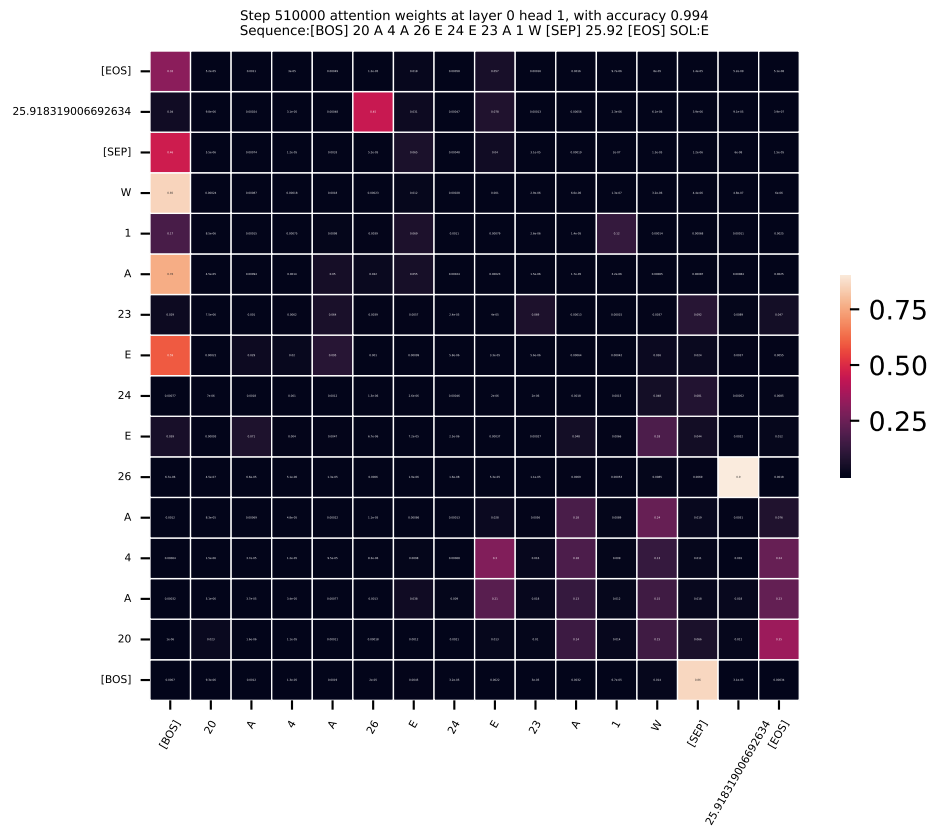

Step 509500 attention weights at layer 1 head 0, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

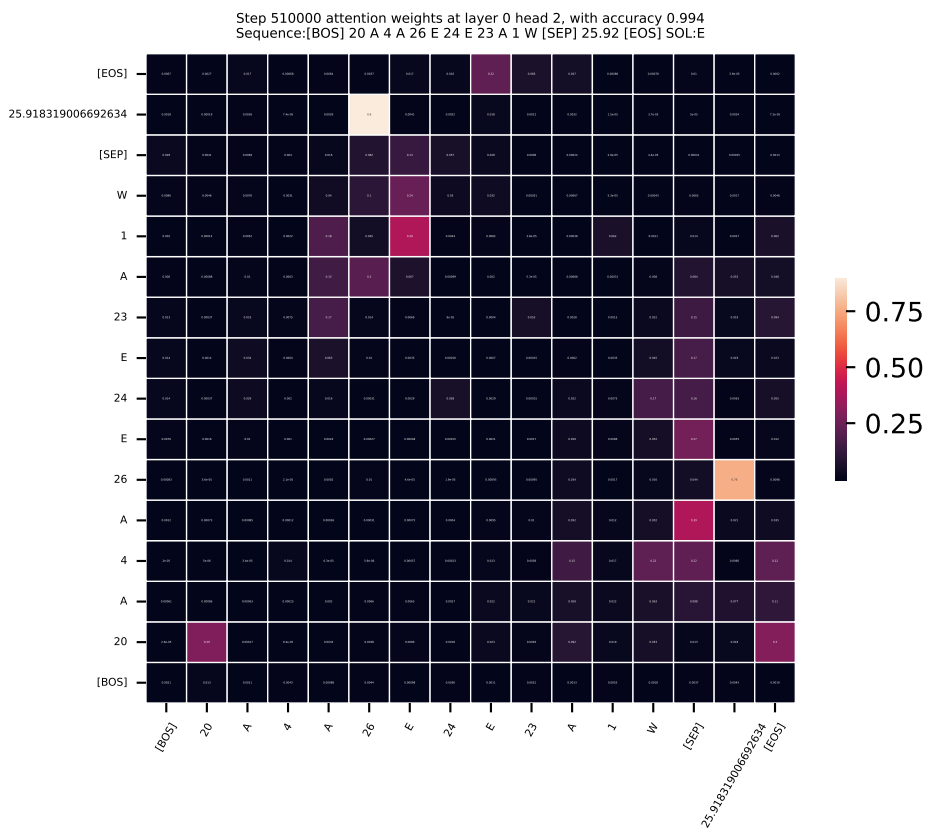

Step 509500 attention weights at layer 1 head 1, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

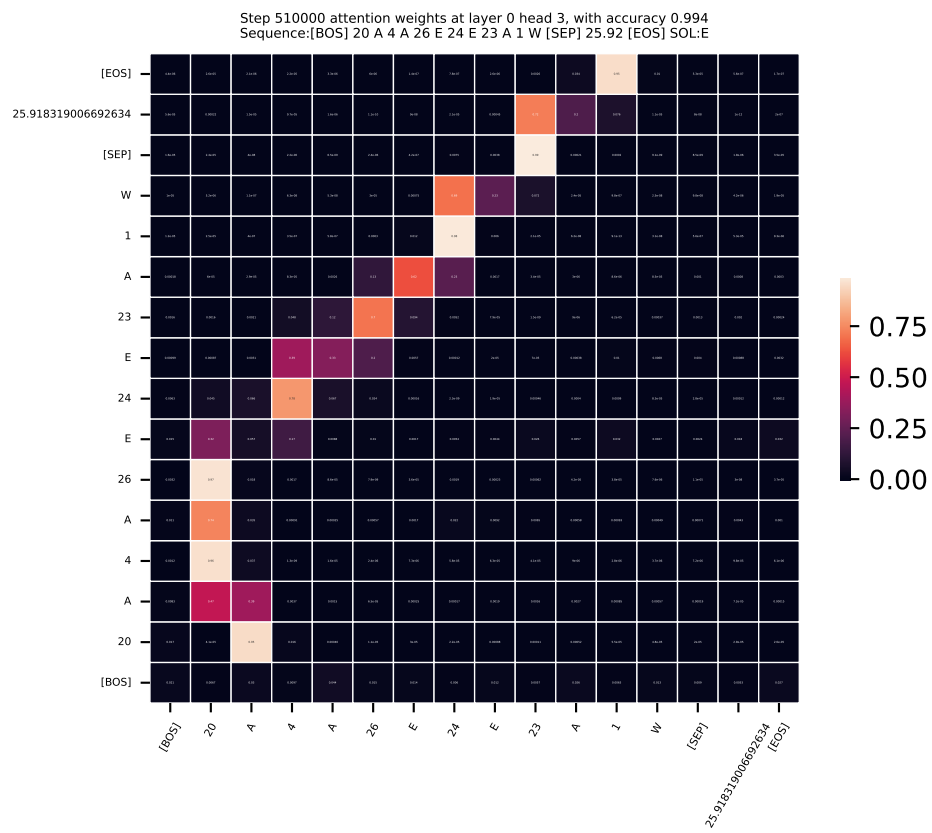

Step 509500 attention weights at layer 1 head 2, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

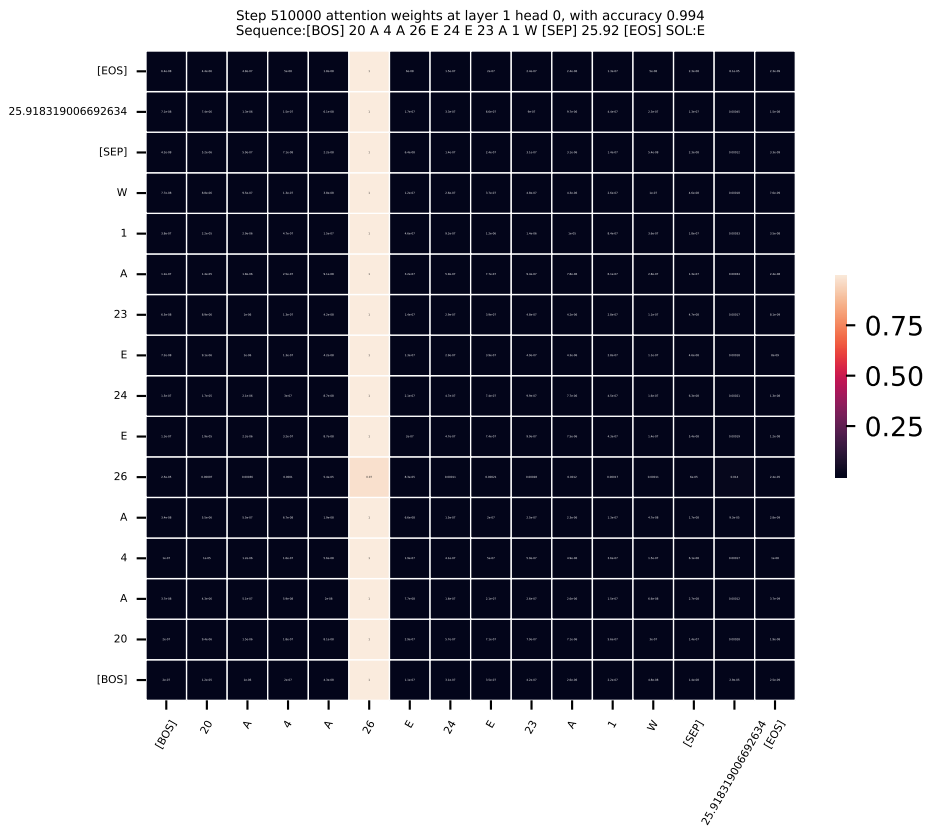

Step 509500 attention weights at layer 1 head 3, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

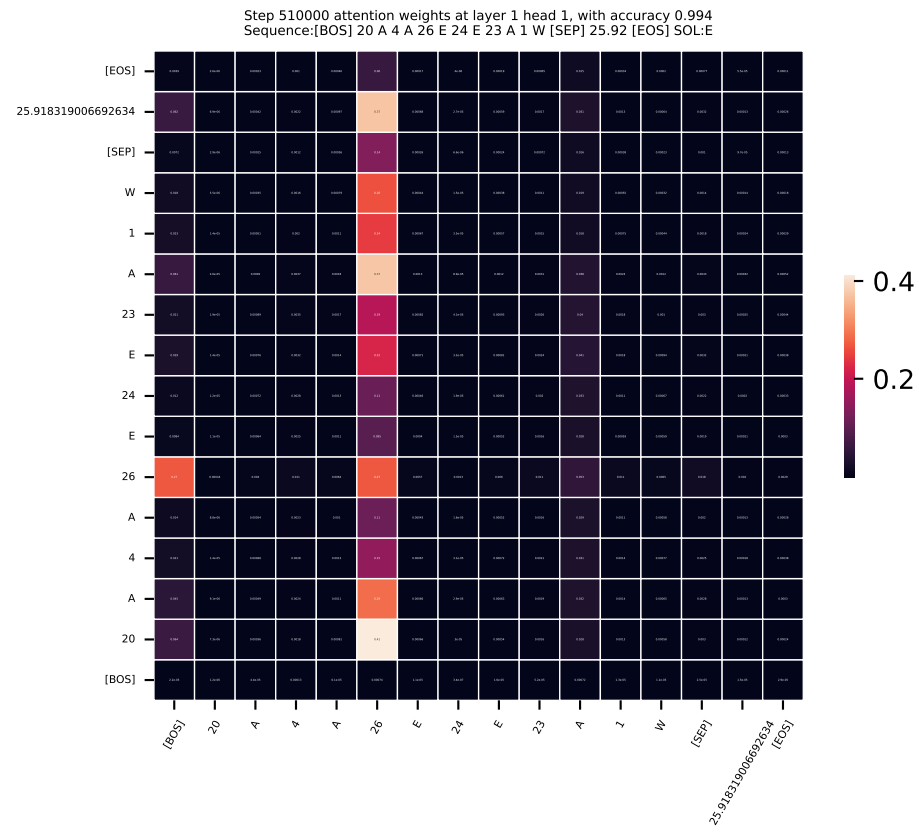

Step 510000 attention weights at layer 0 head 0, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

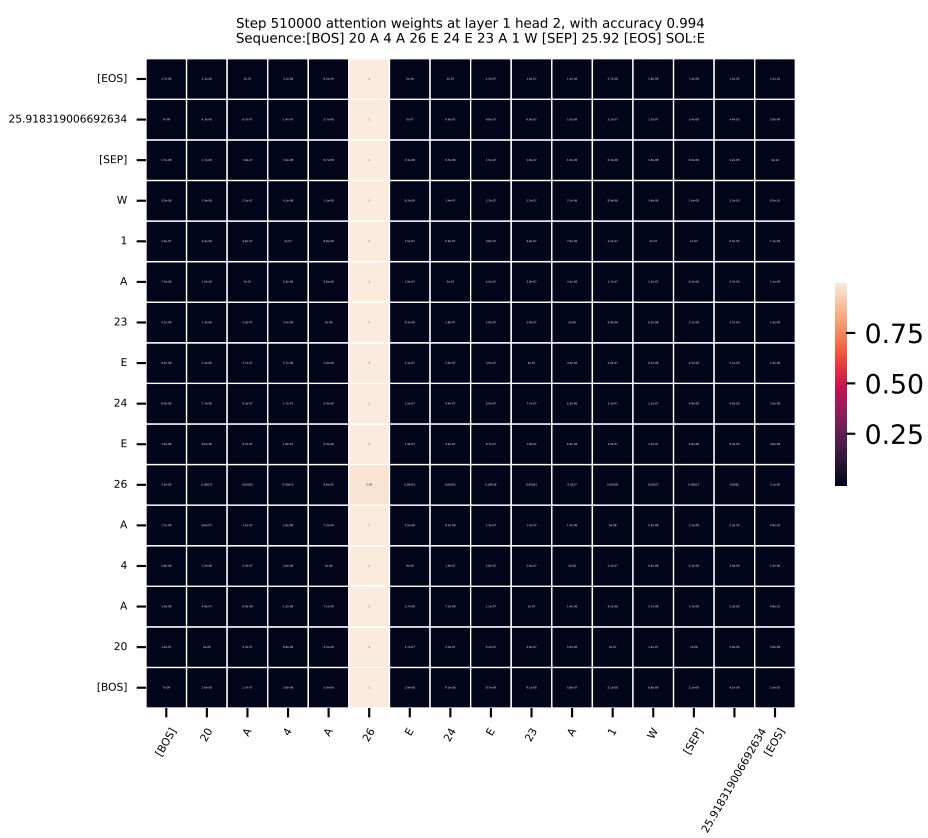

Step 510000 attention weights at layer 0 head 1, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

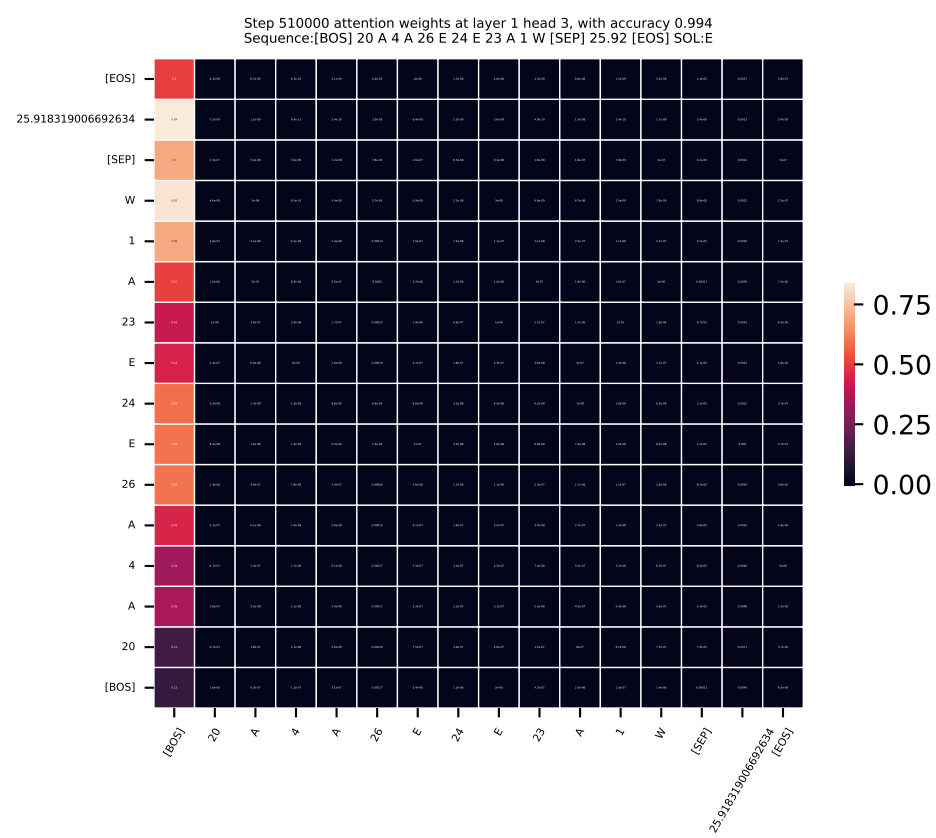

Step 510000 attention weights at layer 0 head 2, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

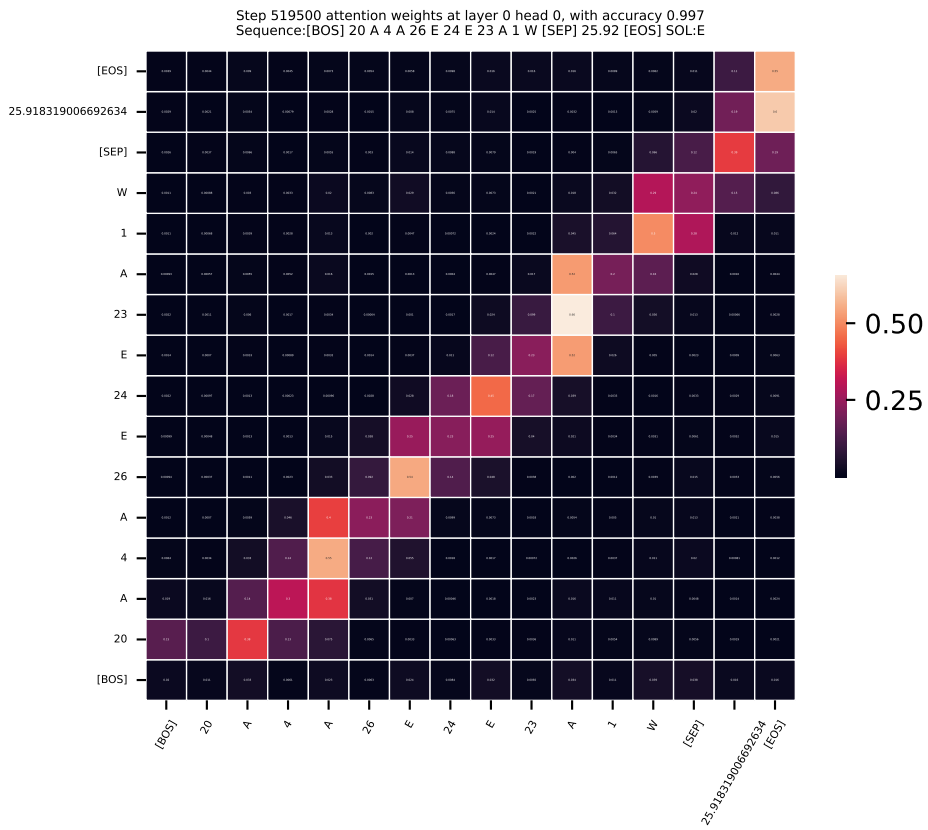

Step 510000 attention weights at layer 0 head 3, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

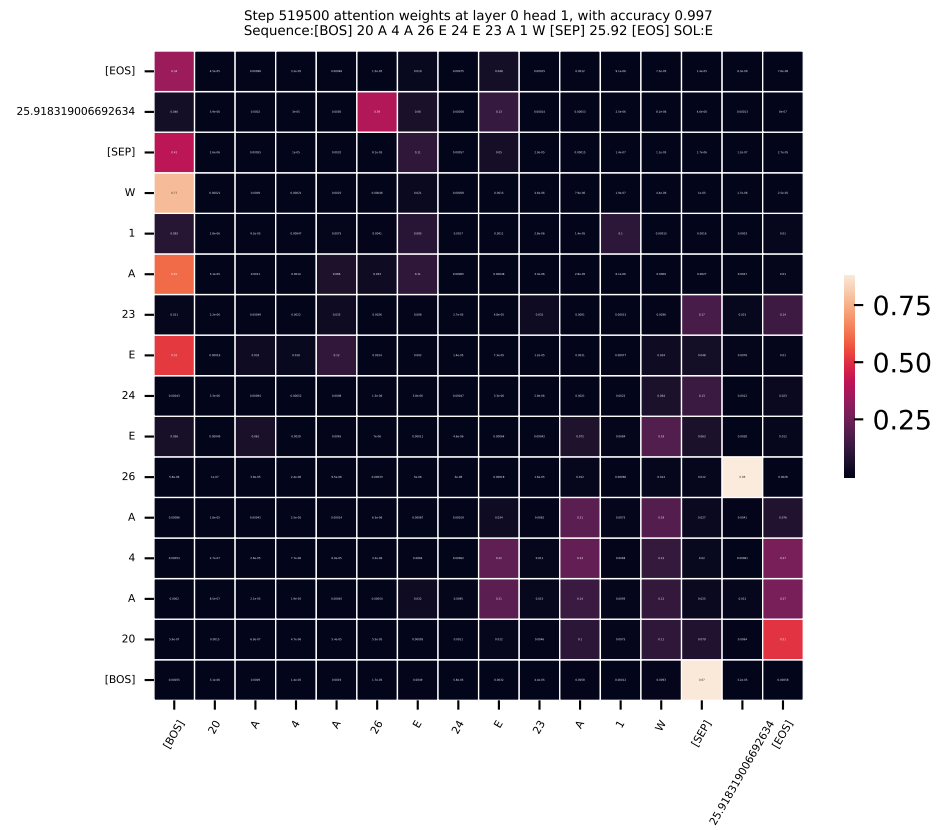

Step 510000 attention weights at layer 1 head 0, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

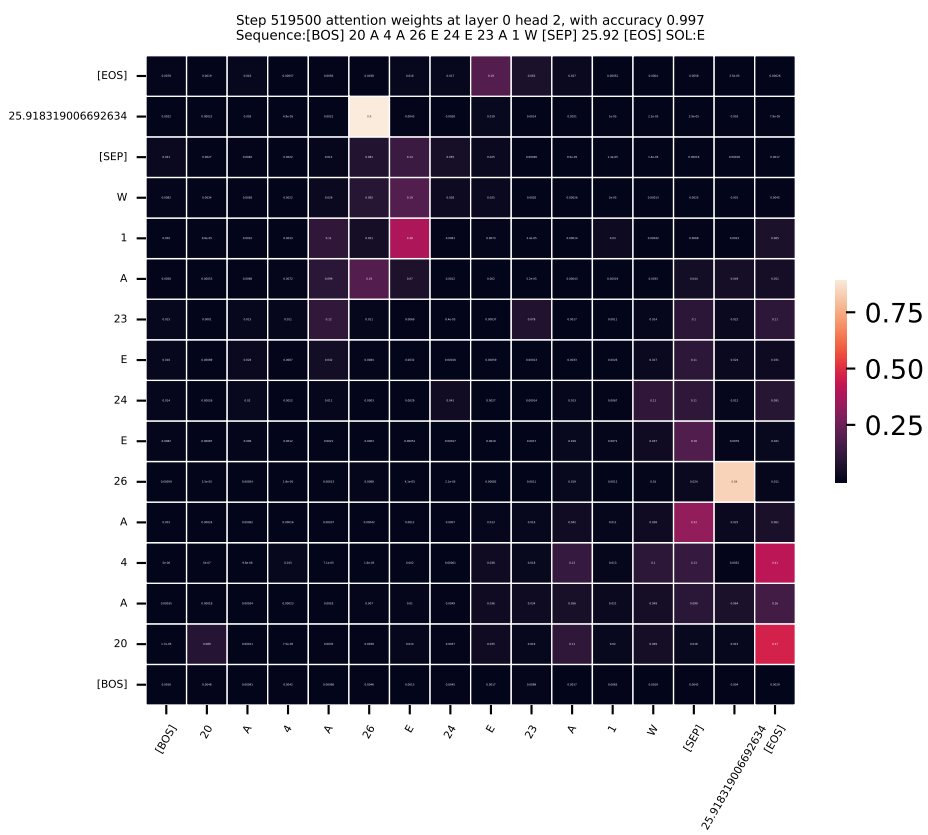

Step 510000 attention weights at layer 1 head 1, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

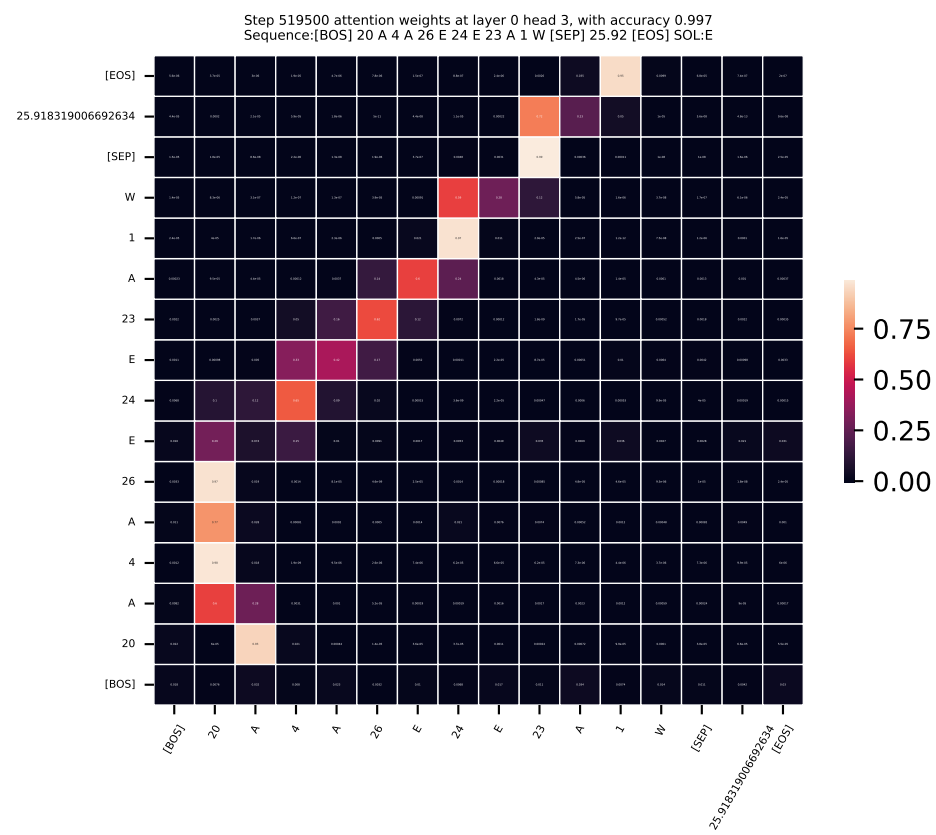

Step 510000 attention weights at layer 1 head 2, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

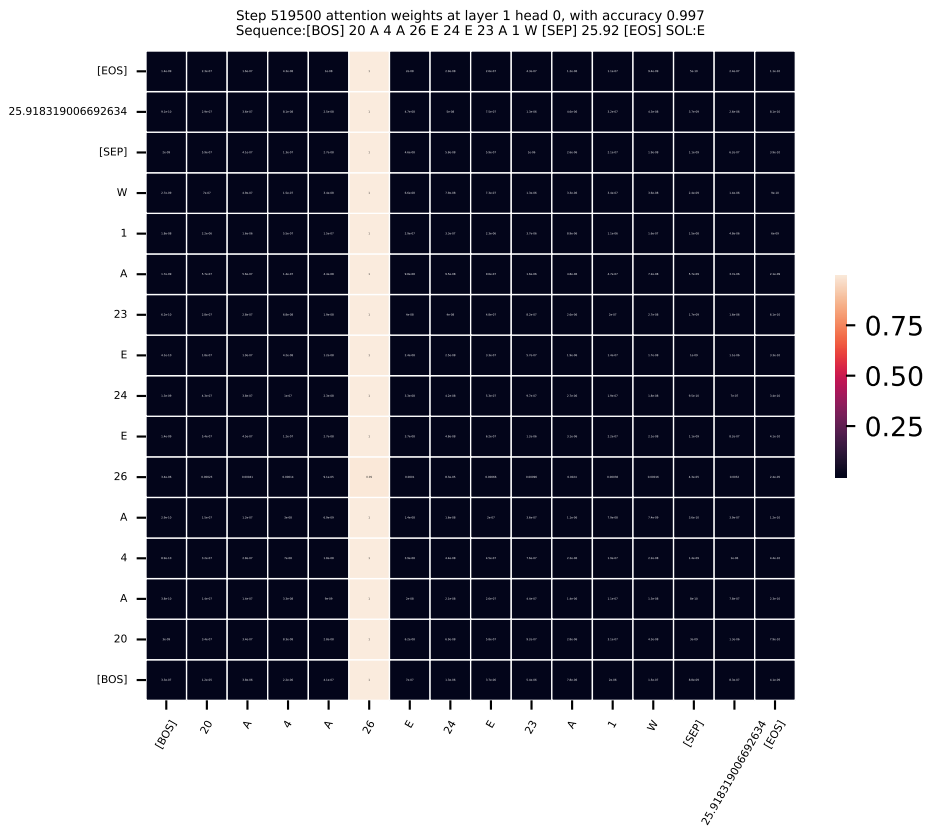

Step 510000 attention weights at layer 1 head 3, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

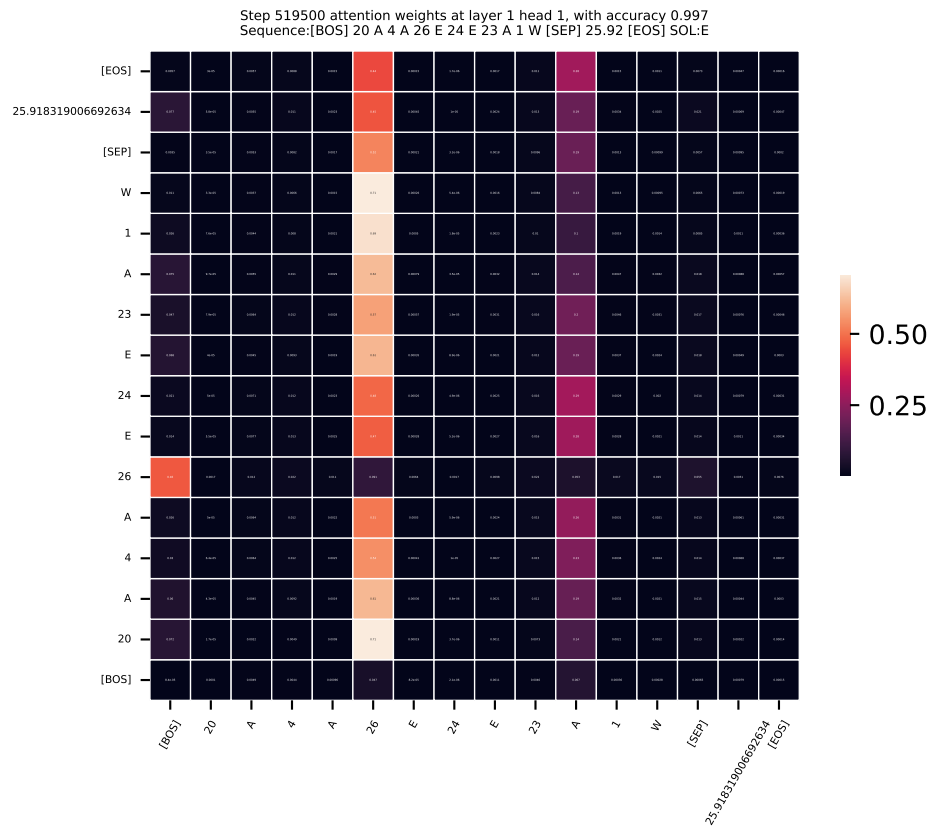

Step 519500 attention weights at layer 0 head 0, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

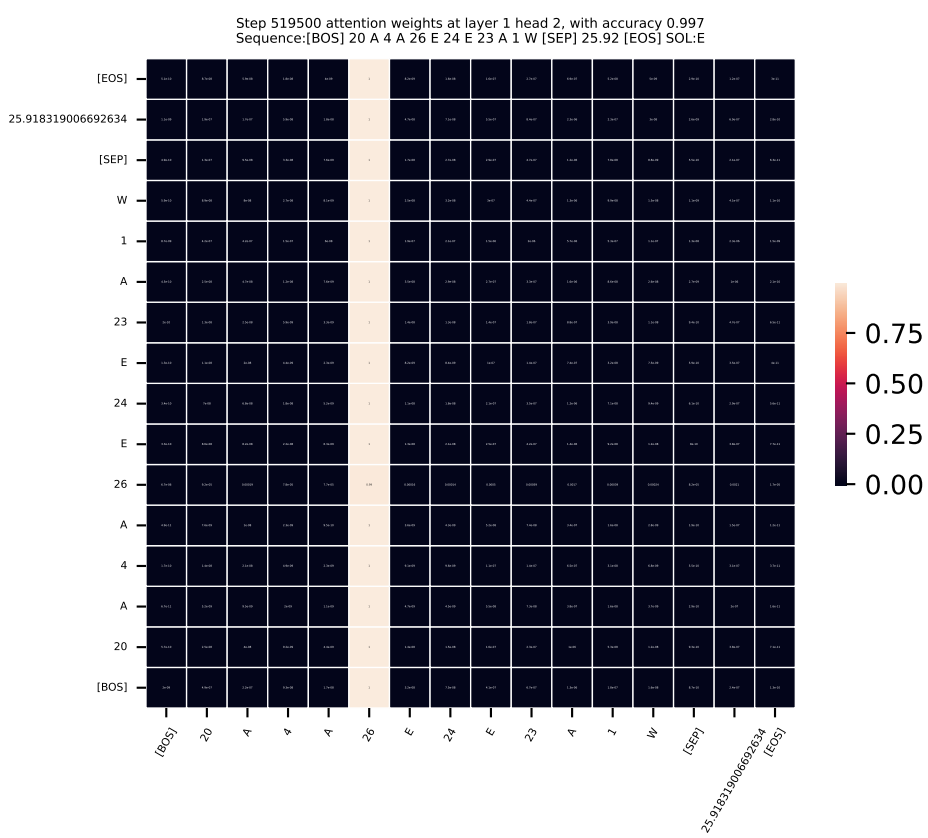

Step 519500 attention weights at layer 0 head 1, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

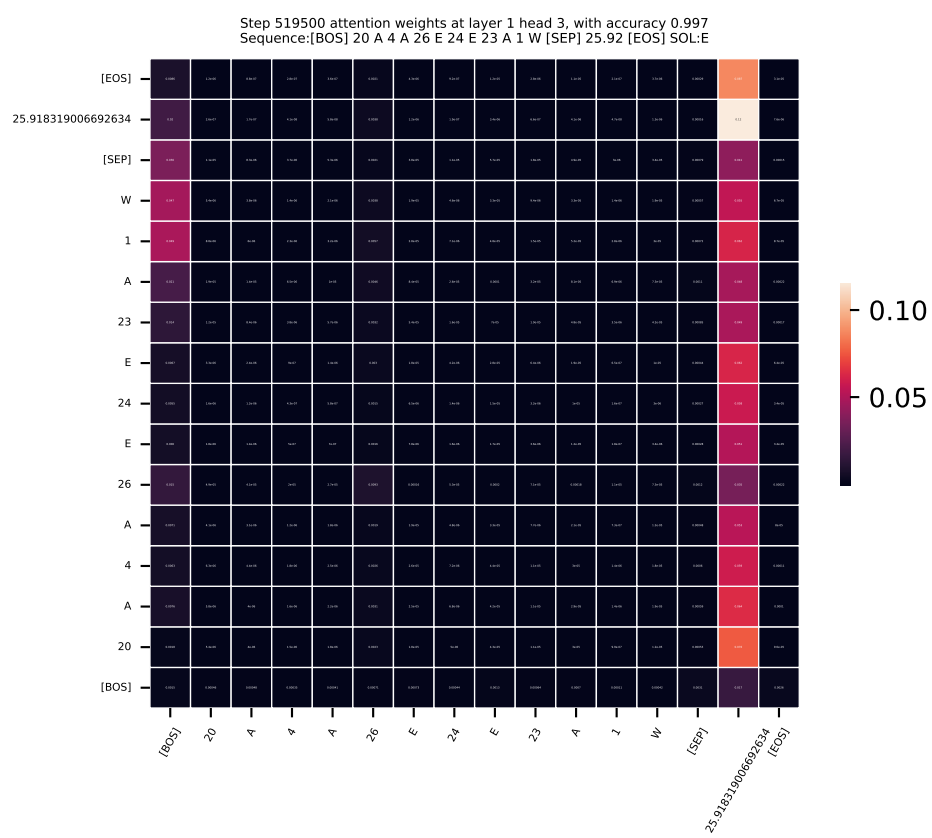

Step 519500 attention weights at layer 0 head 2, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

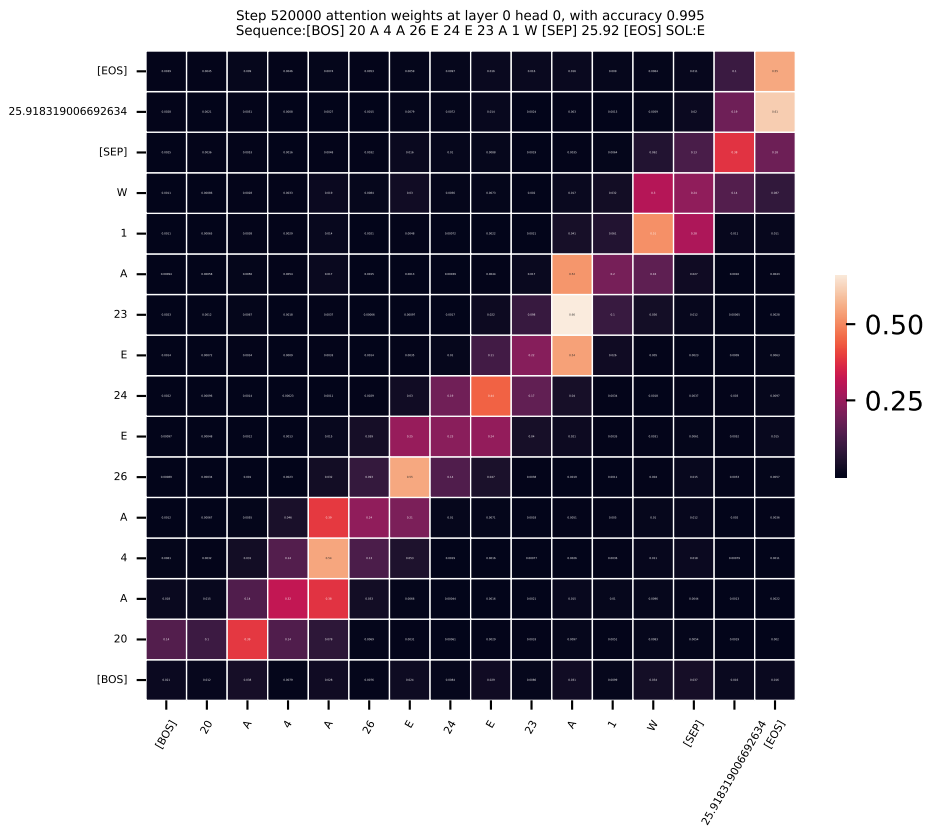

Step 519500 attention weights at layer 0 head 3, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

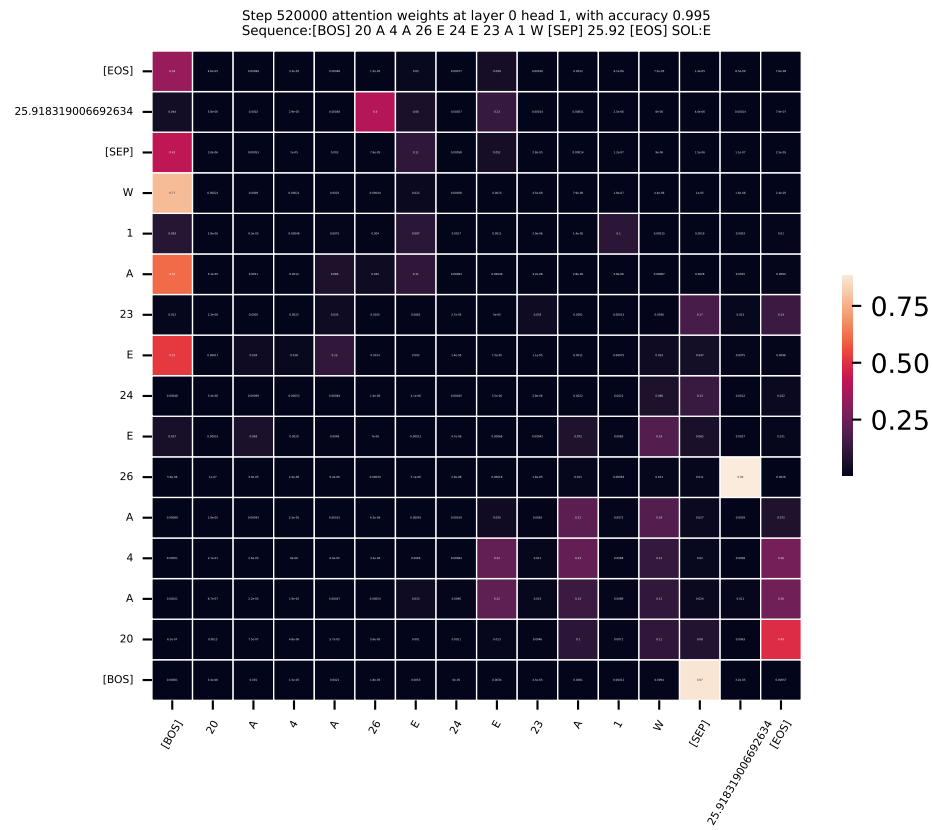

Step 519500 attention weights at layer 1 head 0, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

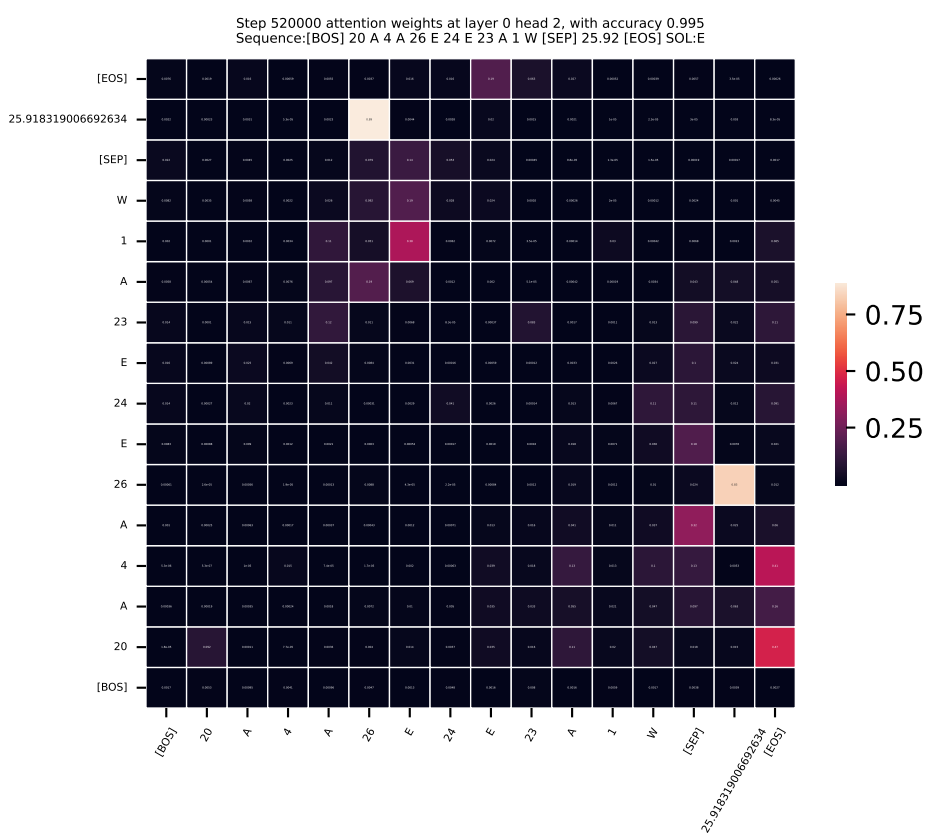

Step 519500 attention weights at layer 1 head 1, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

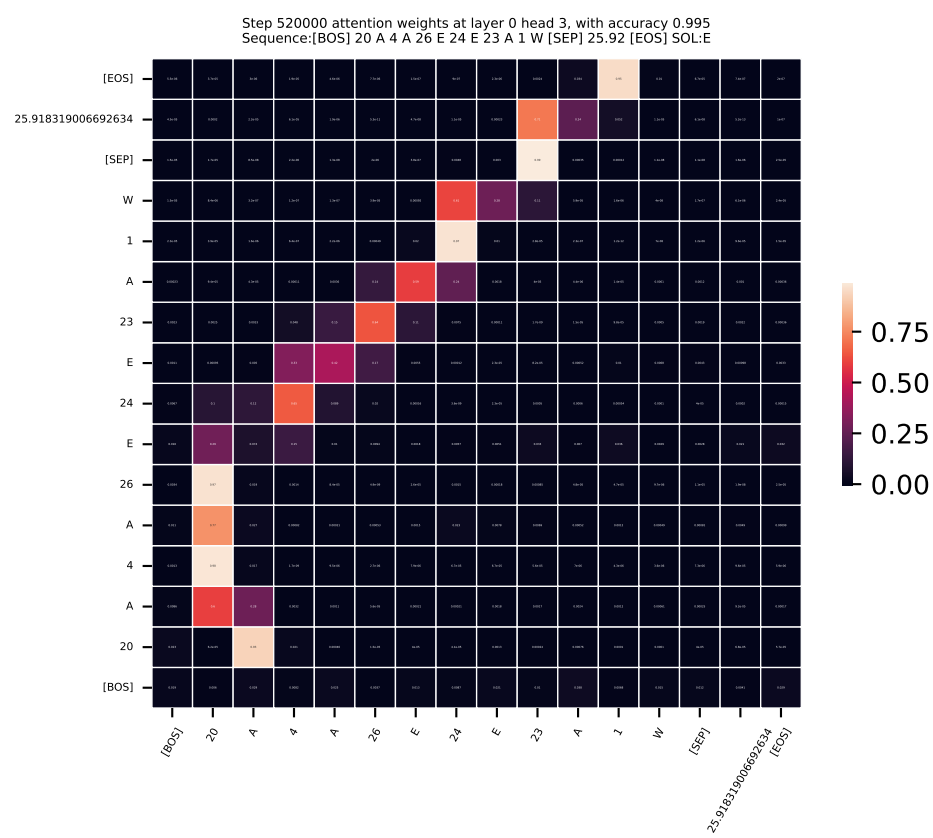

Step 519500 attention weights at layer 1 head 2, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

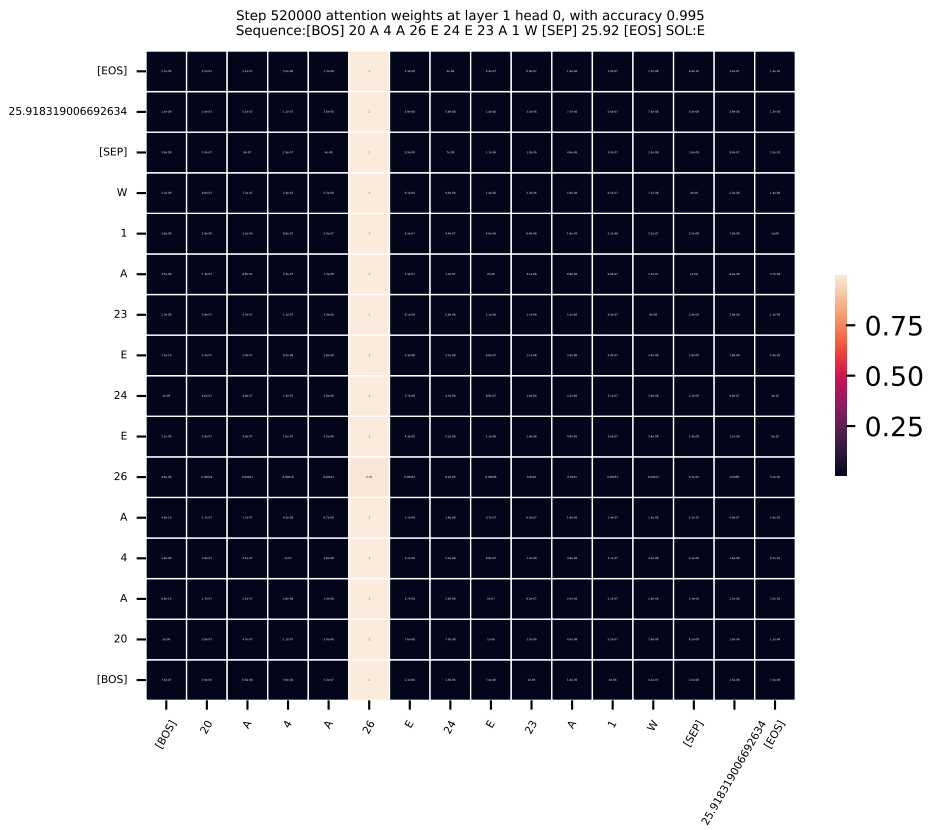

Step 519500 attention weights at layer 1 head 3, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.10
0.05

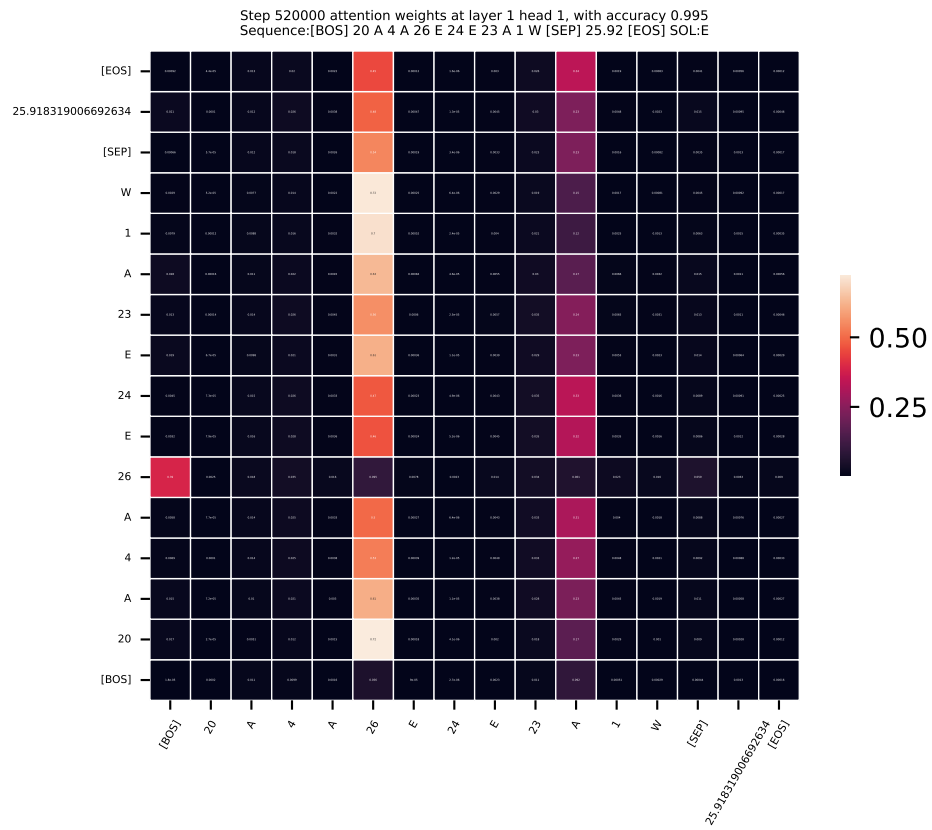

Step 520000 attention weights at layer 0 head 0, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

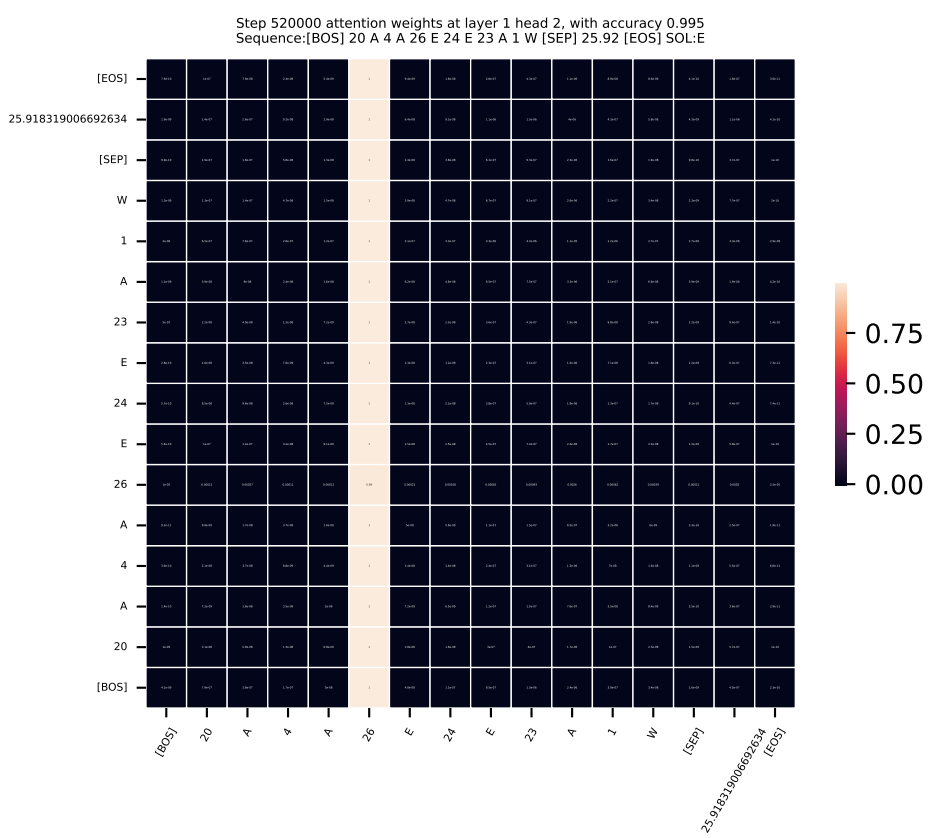

Step 520000 attention weights at layer 0 head 1, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

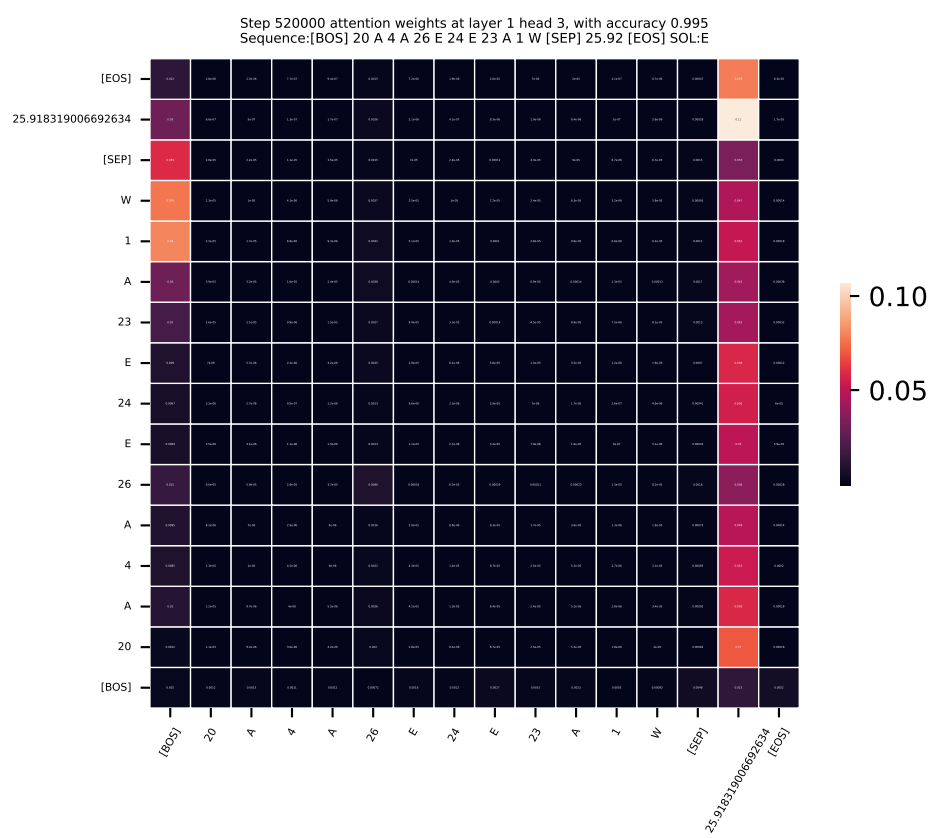

Step 520000 attention weights at layer 0 head 2, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

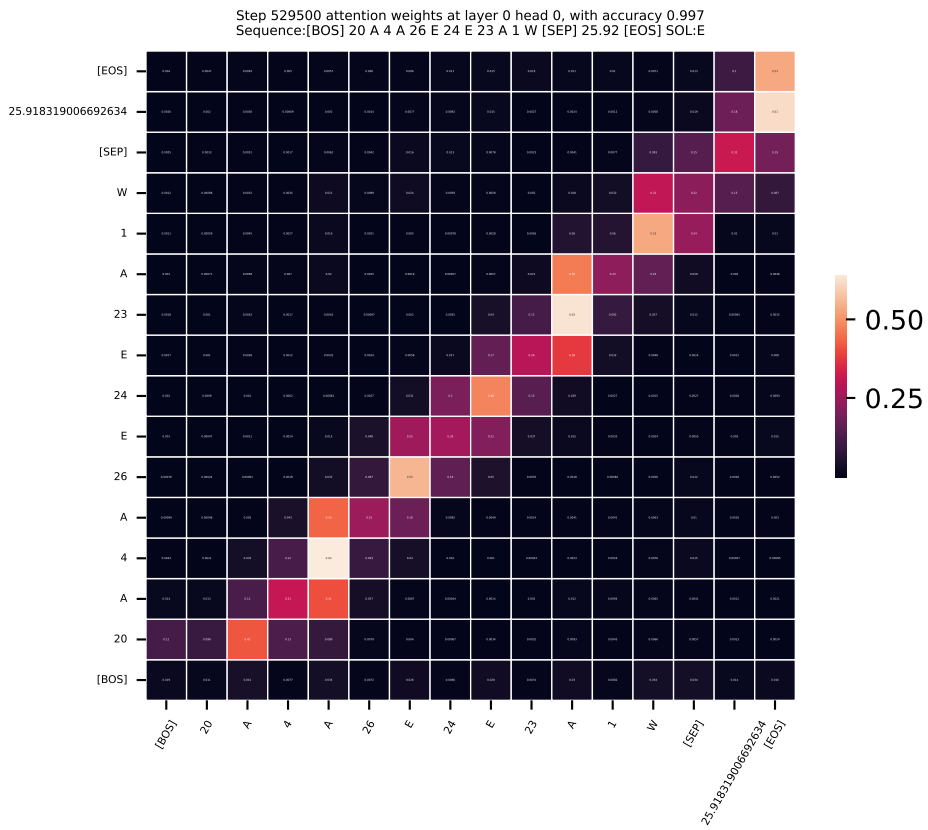

Step 520000 attention weights at layer 0 head 3, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

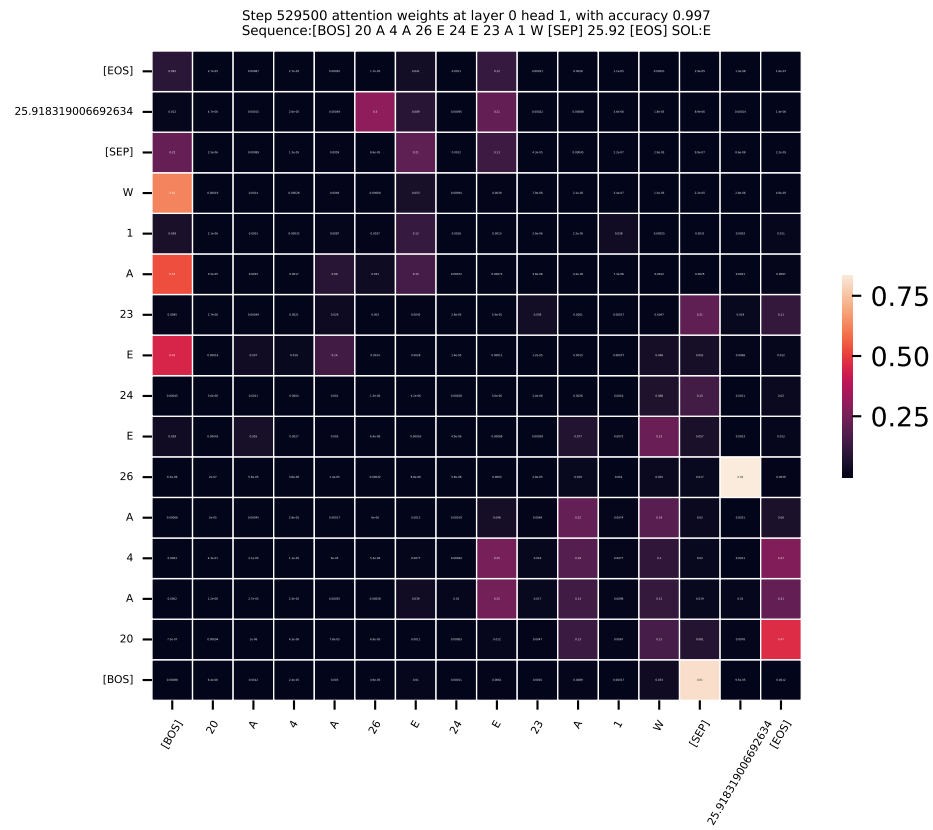

Step 520000 attention weights at layer 1 head 0, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

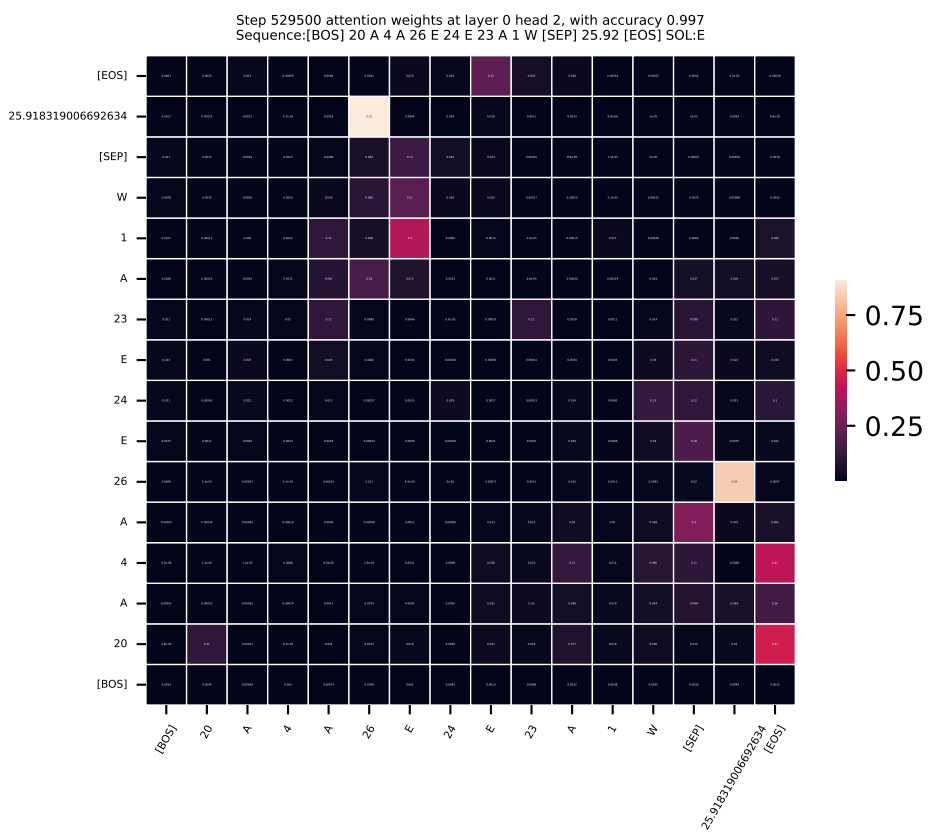

Step 520000 attention weights at layer 1 head 1, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

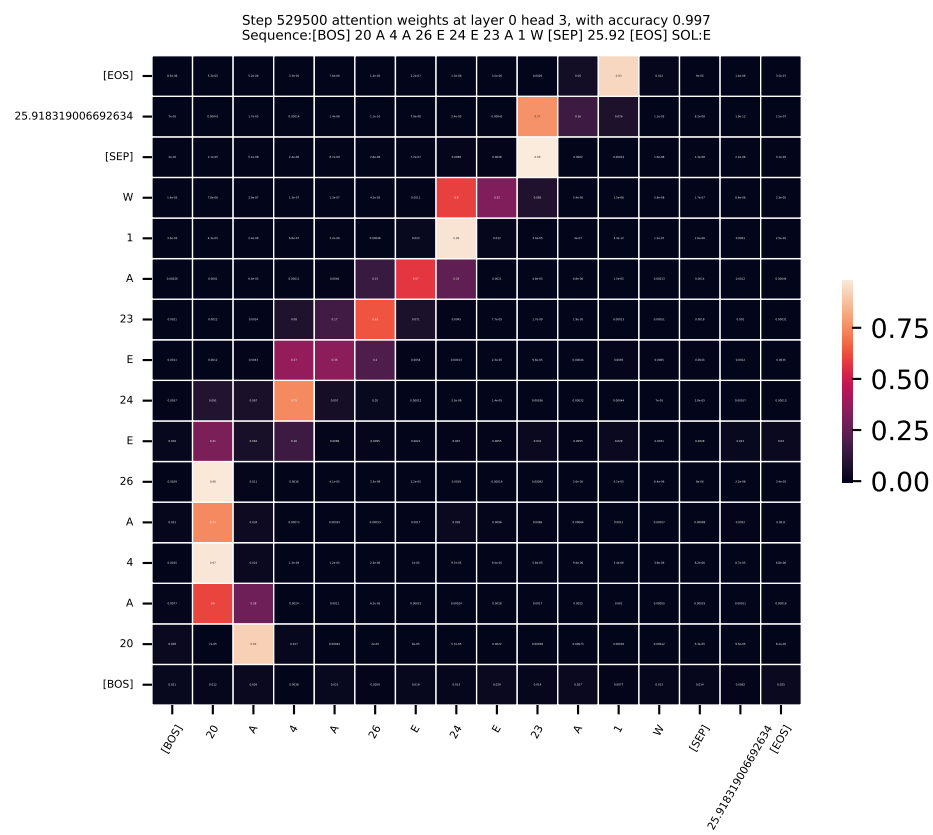

Step 520000 attention weights at layer 1 head 2, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

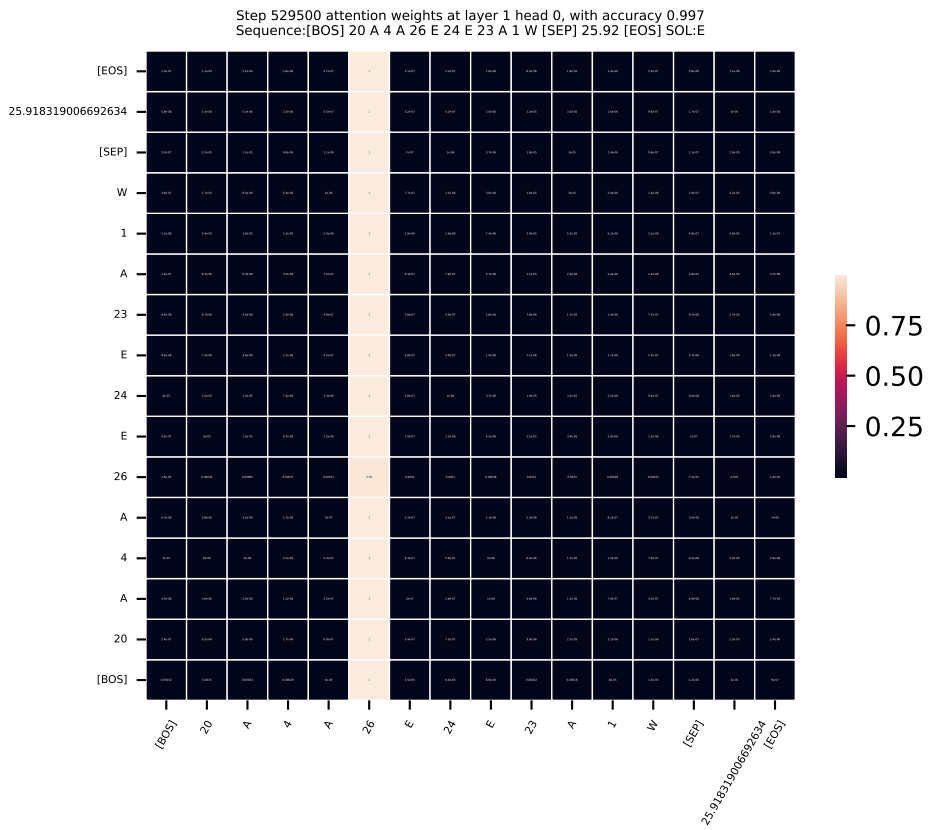

Step 520000 attention weights at layer 1 head 3, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.10
0.05

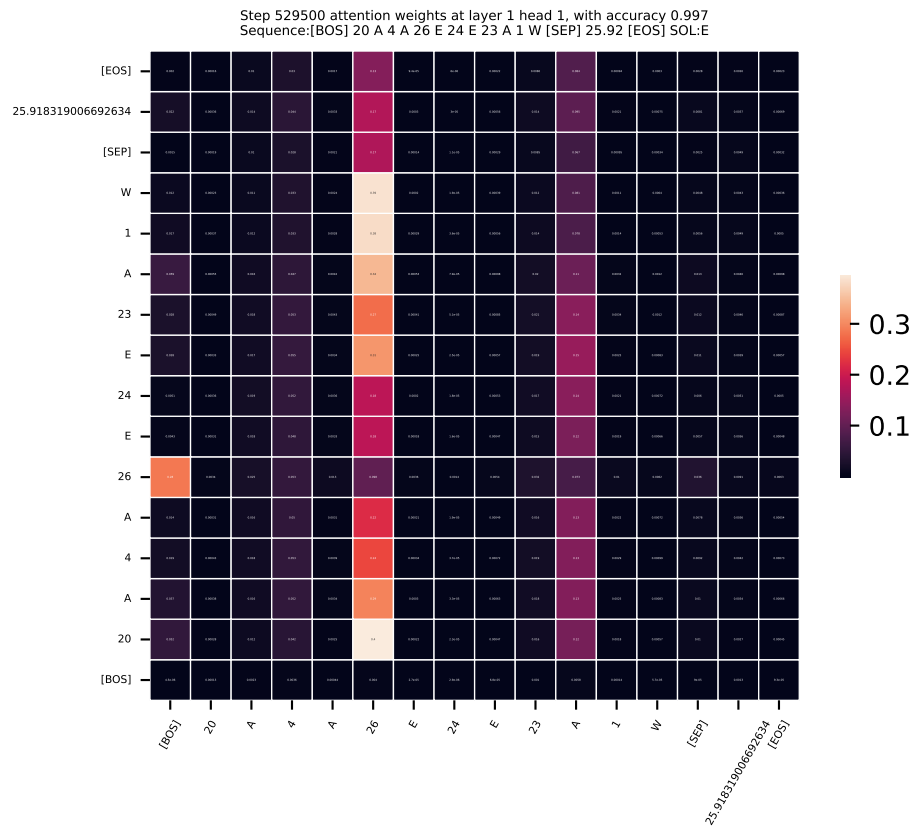

Step 529500 attention weights at layer 0 head 0, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

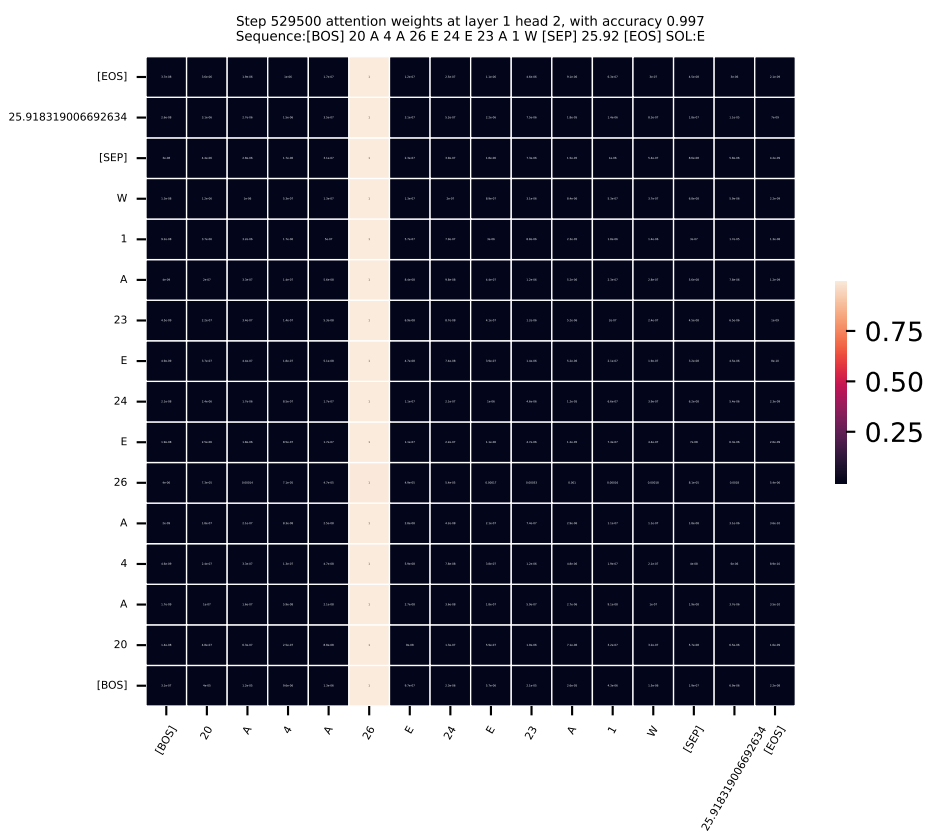

Step 529500 attention weights at layer 0 head 1, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

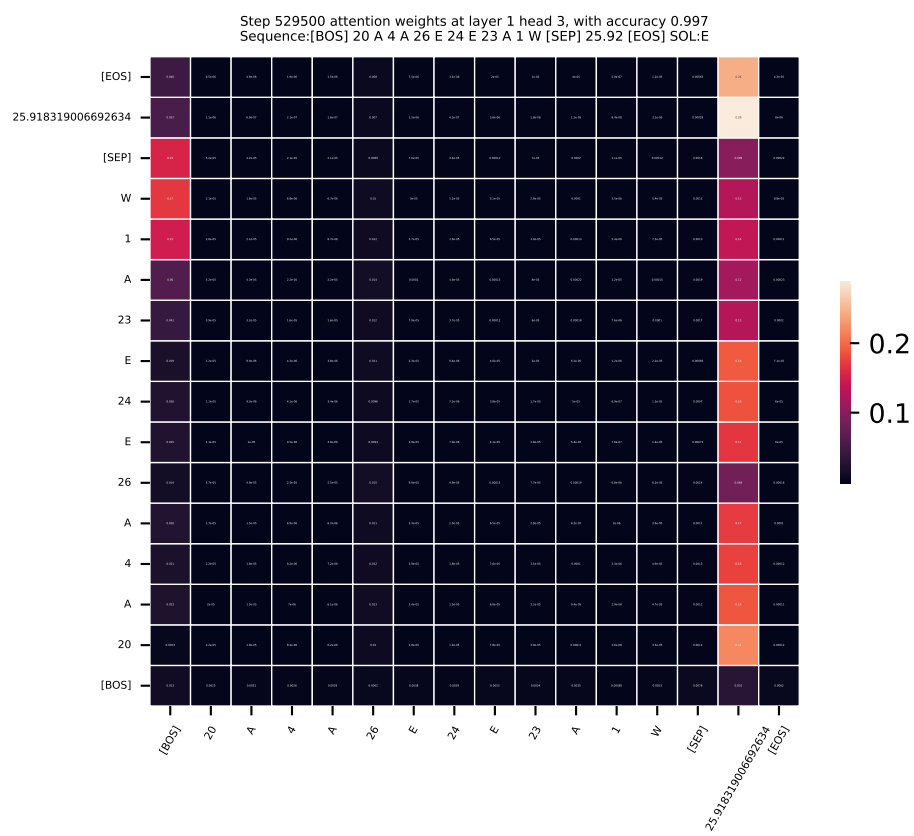

Step 529500 attention weights at layer 0 head 2, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

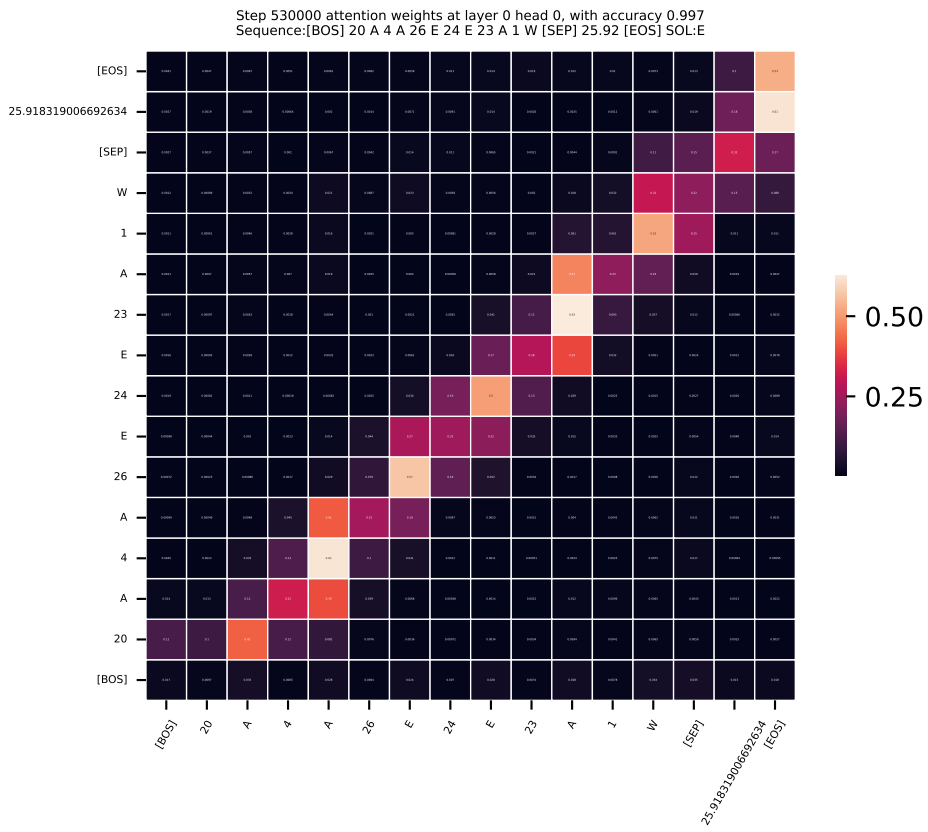

Step 529500 attention weights at layer 0 head 3, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

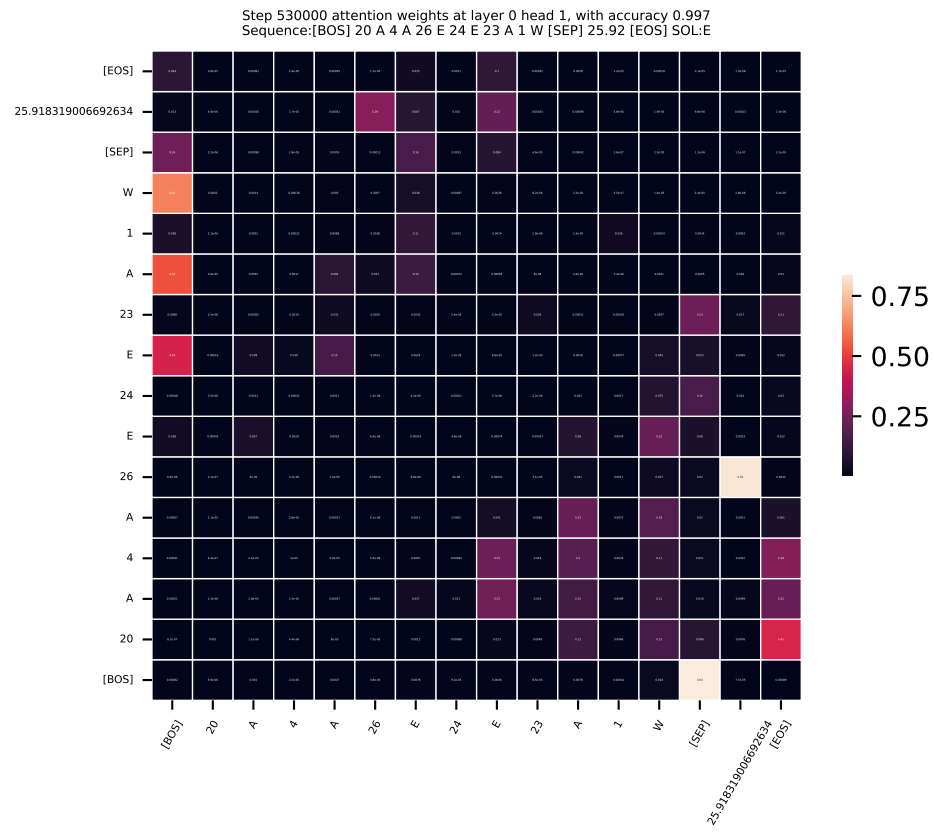

Step 529500 attention weights at layer 1 head 0, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

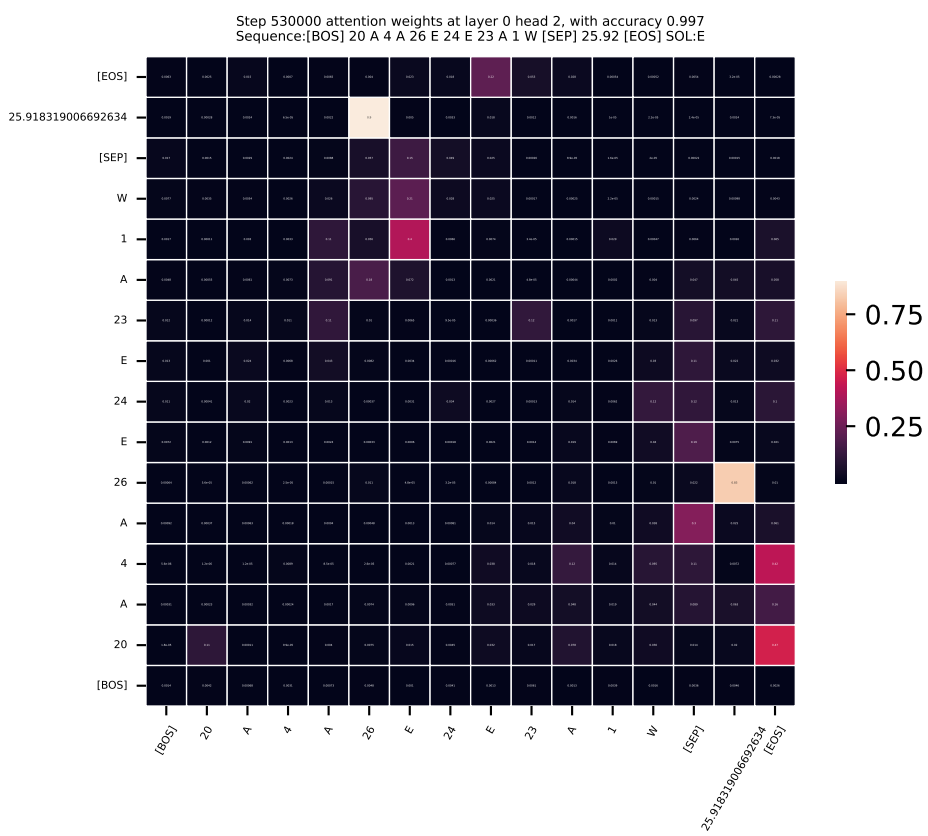

Step 529500 attention weights at layer 1 head 1, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

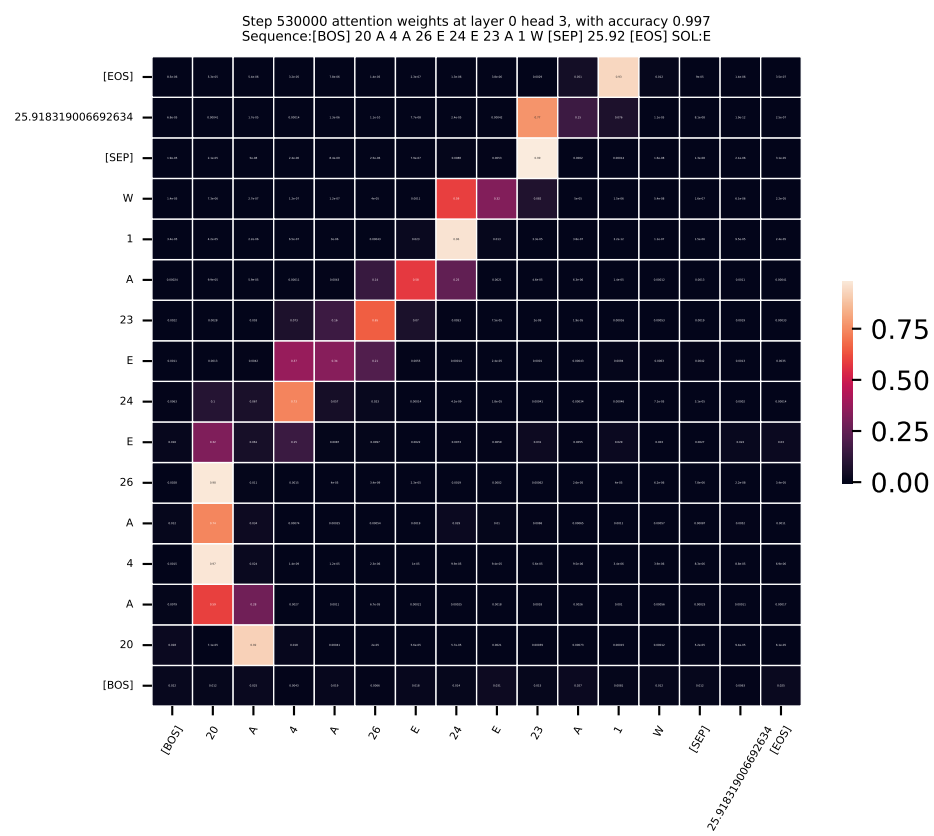

Step 529500 attention weights at layer 1 head 2, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

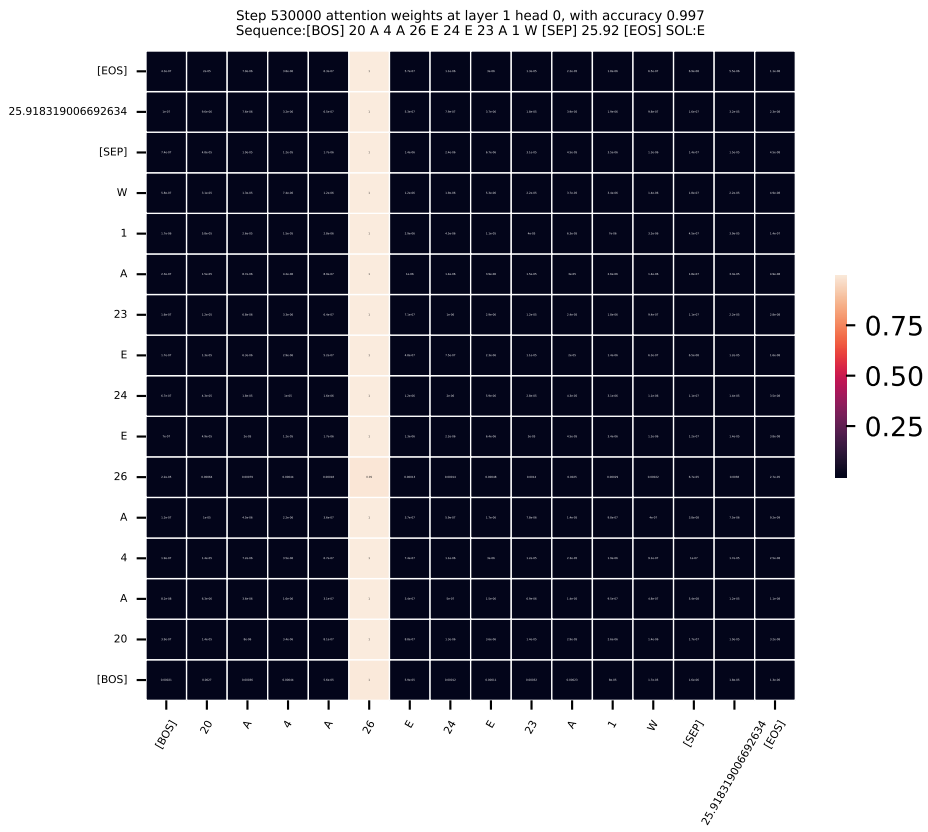

Step 529500 attention weights at layer 1 head 3, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.2
0.1

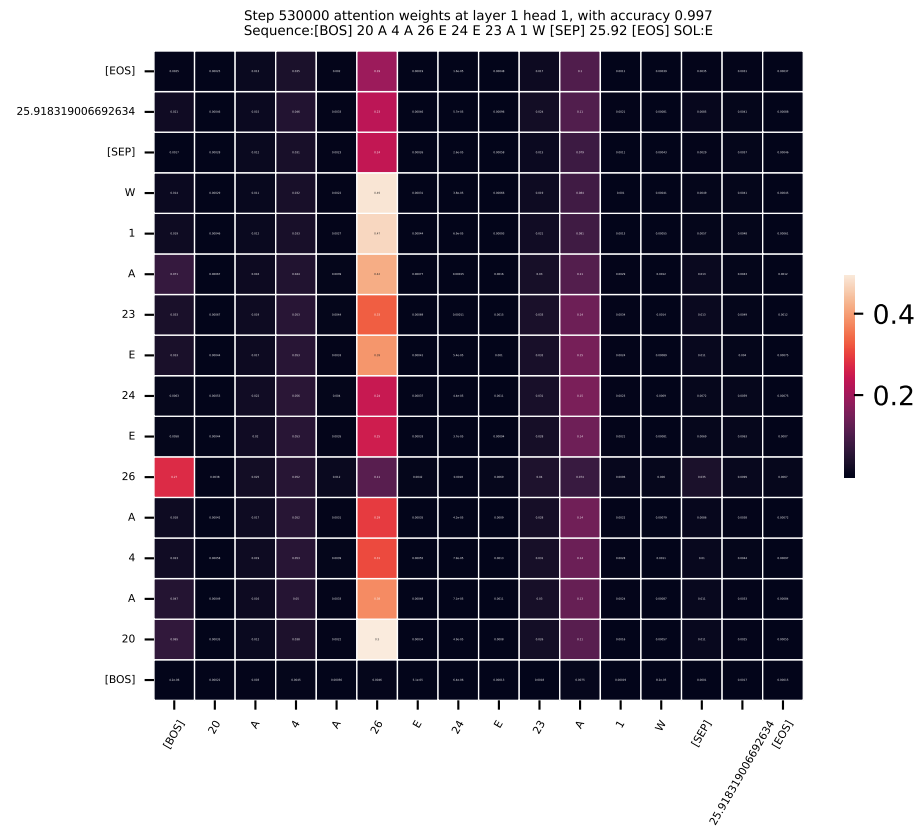

Step 530000 attention weights at layer 0 head 0, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.50
0.25

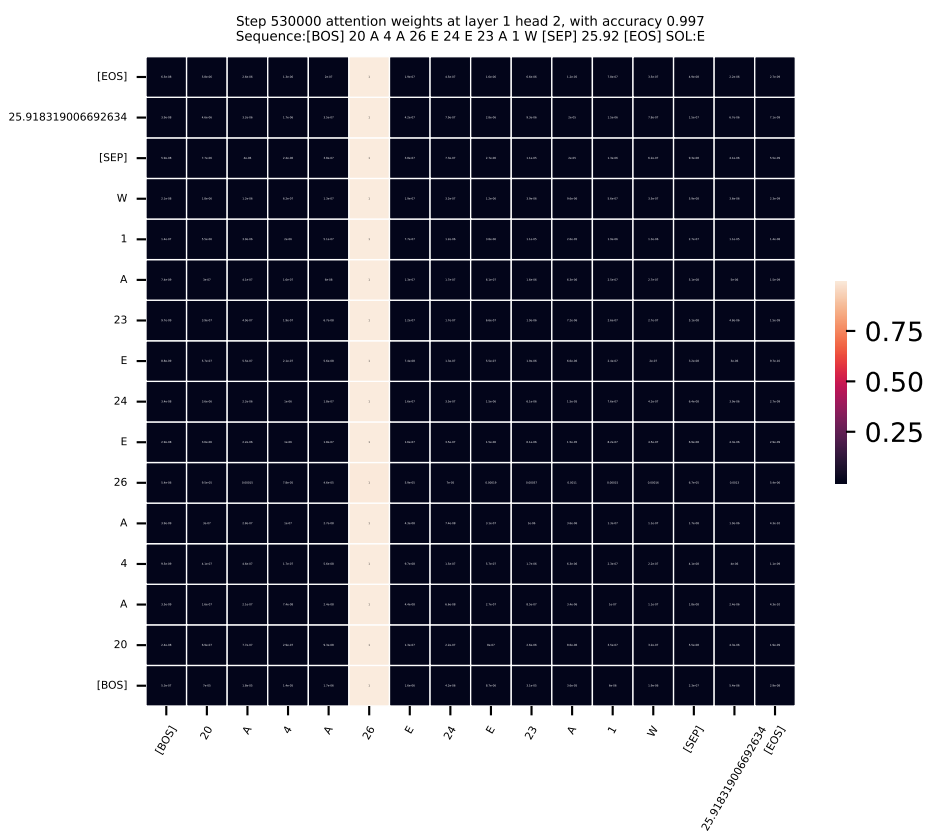

Step 530000 attention weights at layer 0 head 1, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

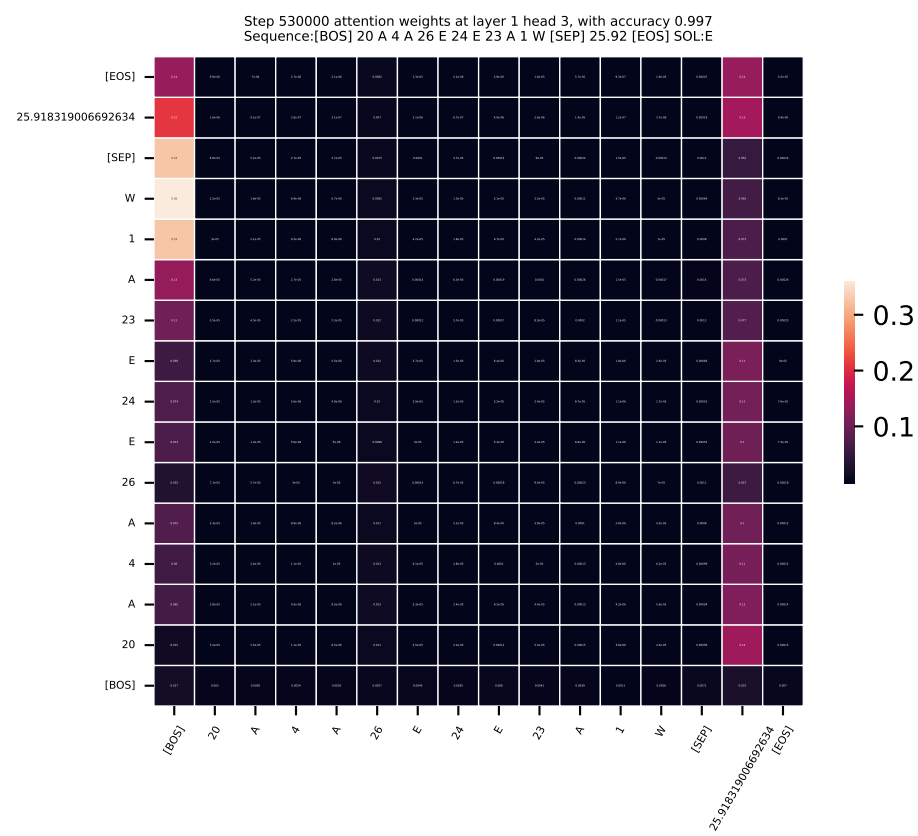

Step 530000 attention weights at layer 0 head 2, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

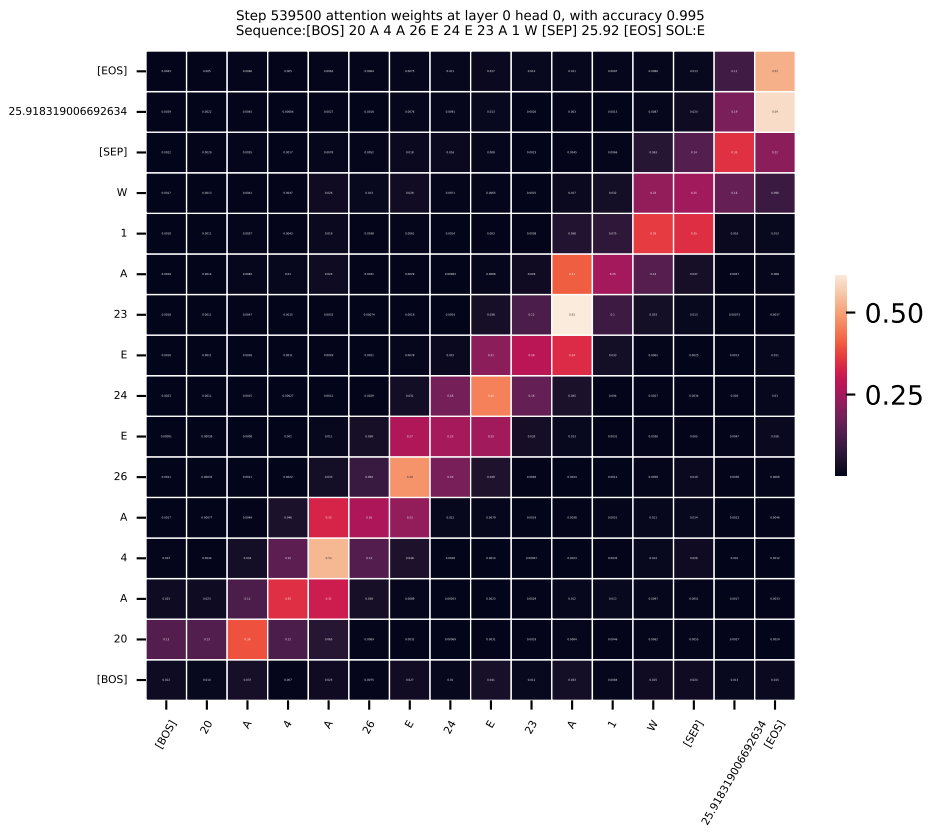

Step 530000 attention weights at layer 0 head 3, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25
0.00

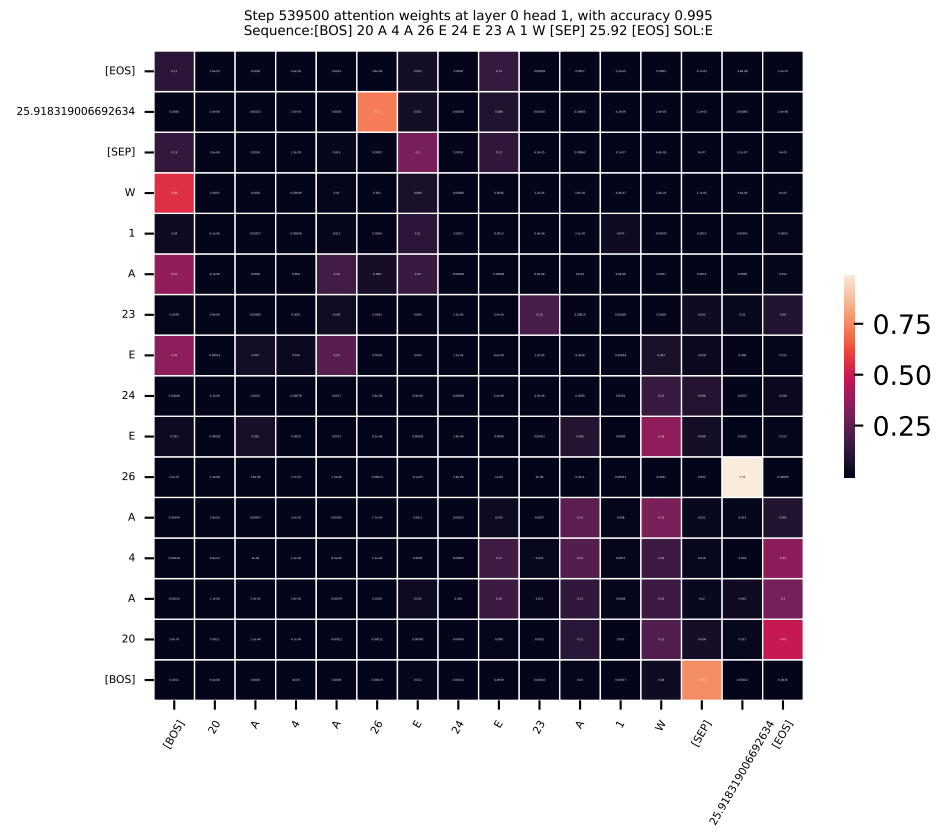

Step 530000 attention weights at layer 1 head 0, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

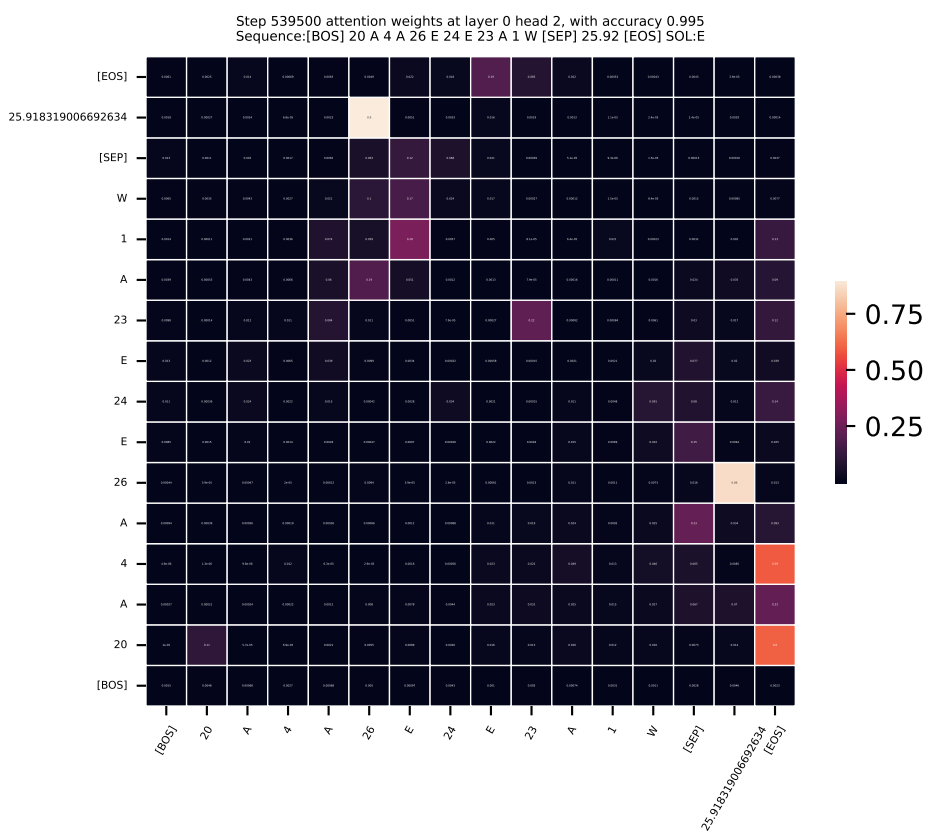

Step 530000 attention weights at layer 1 head 1, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

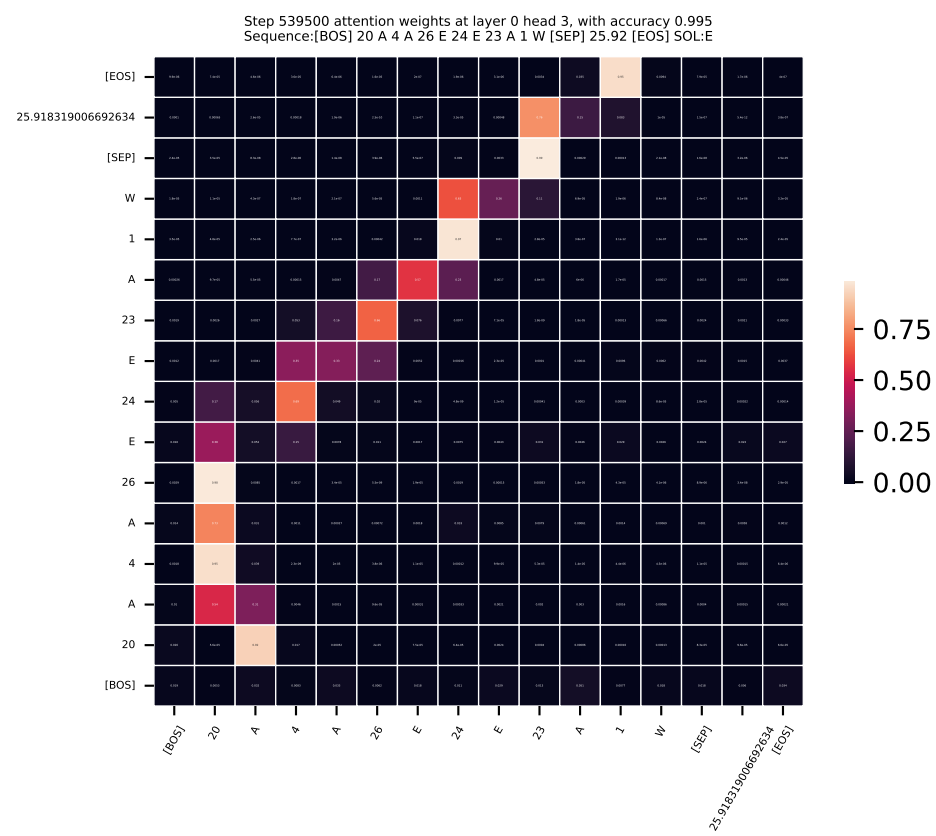

Step 530000 attention weights at layer 1 head 2, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

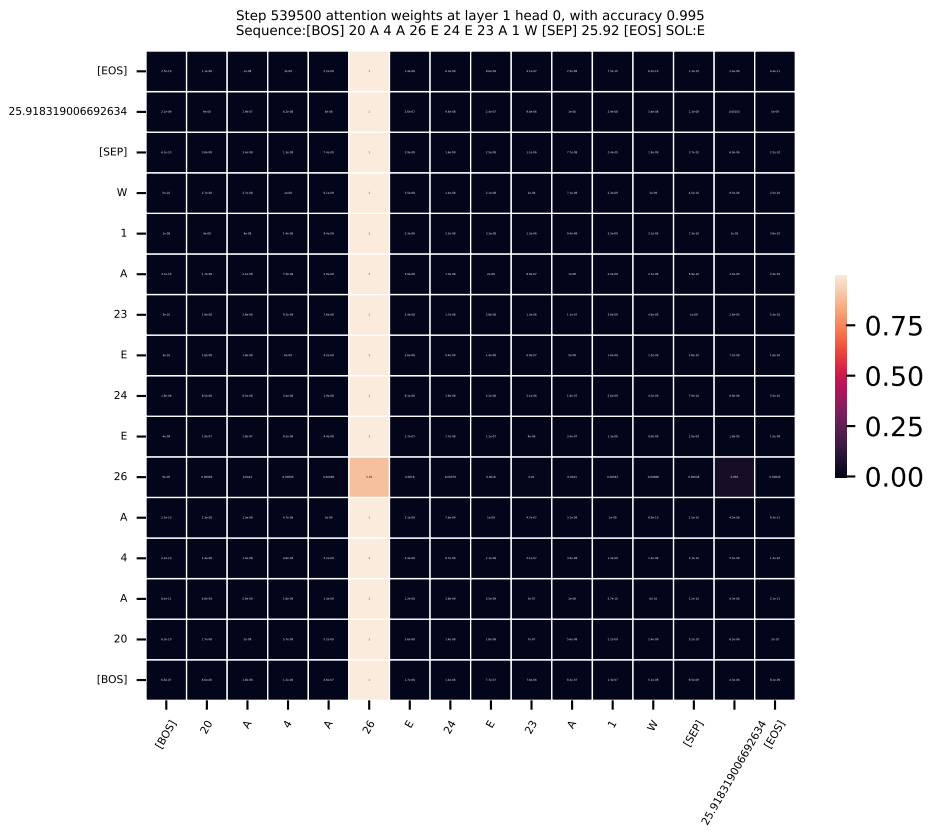

Step 530000 attention weights at layer 1 head 3, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

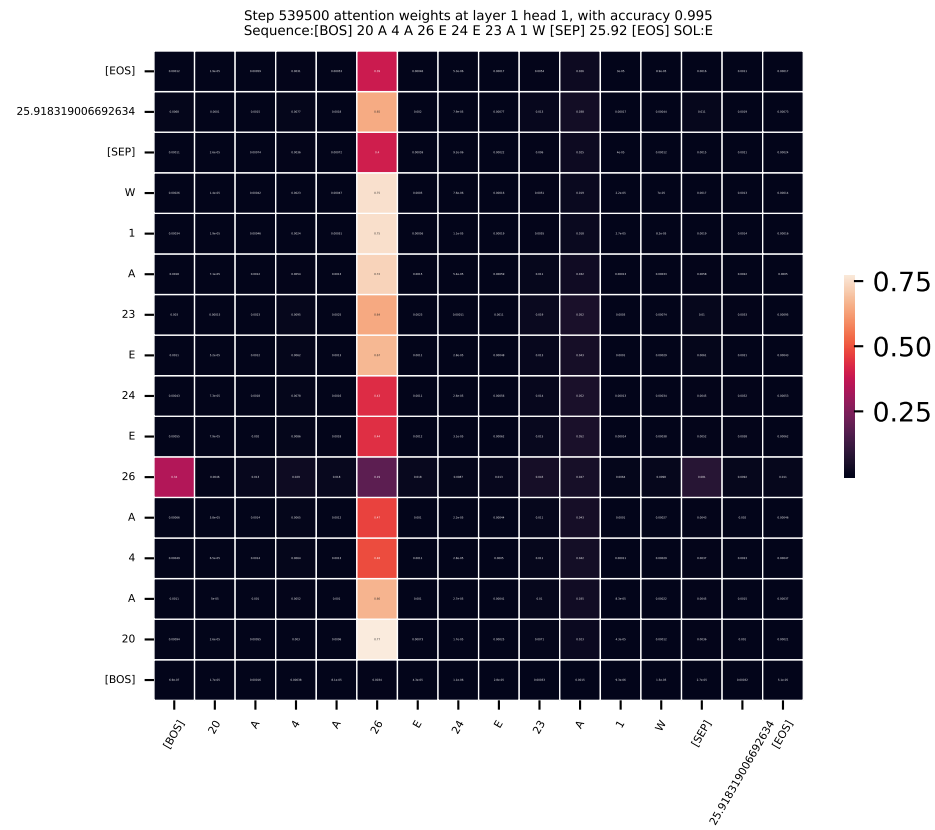

Step 539500 attention weights at layer 0 head 0, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

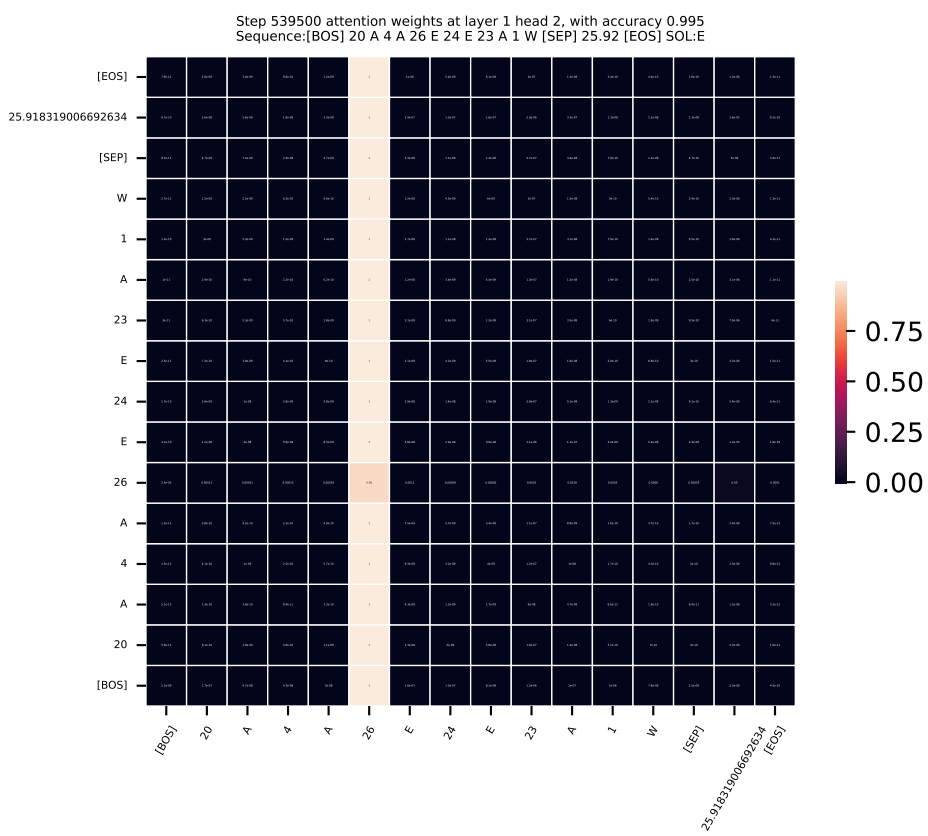

Step 539500 attention weights at layer 0 head 1, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

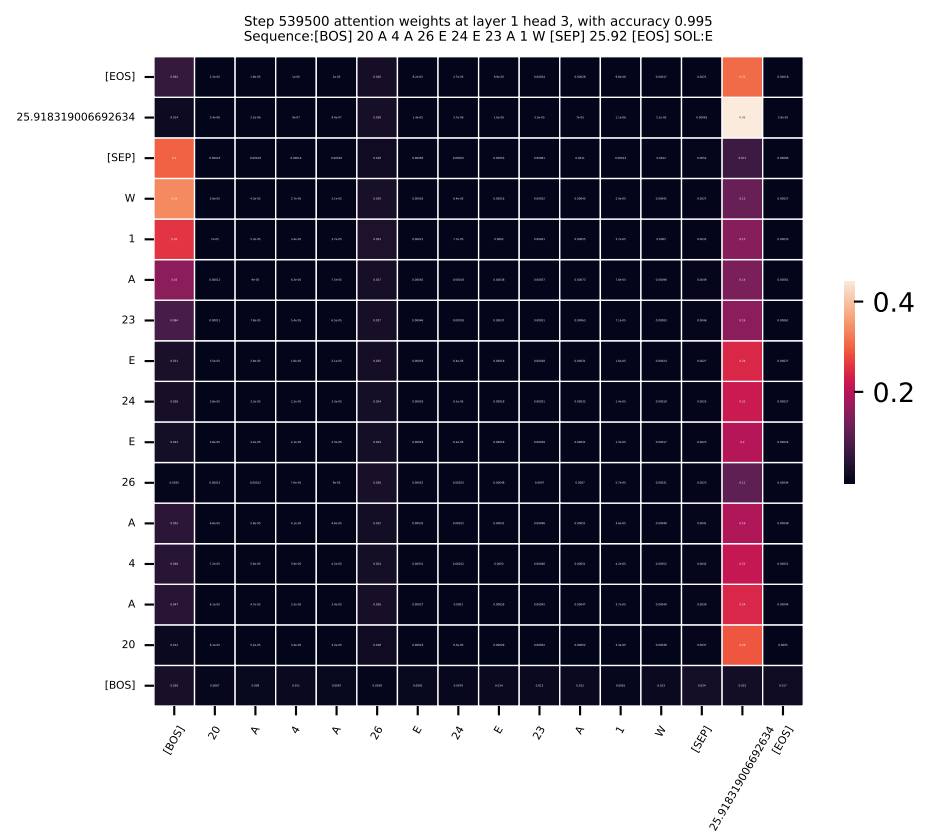

Step 539500 attention weights at layer 0 head 2, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

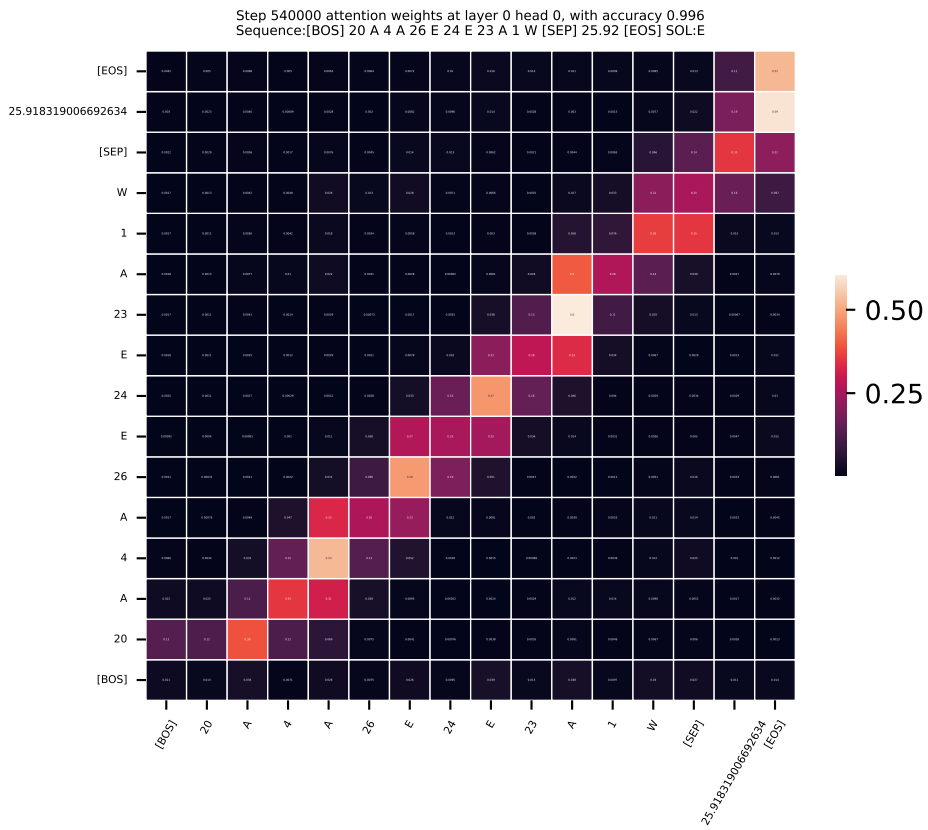

Step 539500 attention weights at layer 0 head 3, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

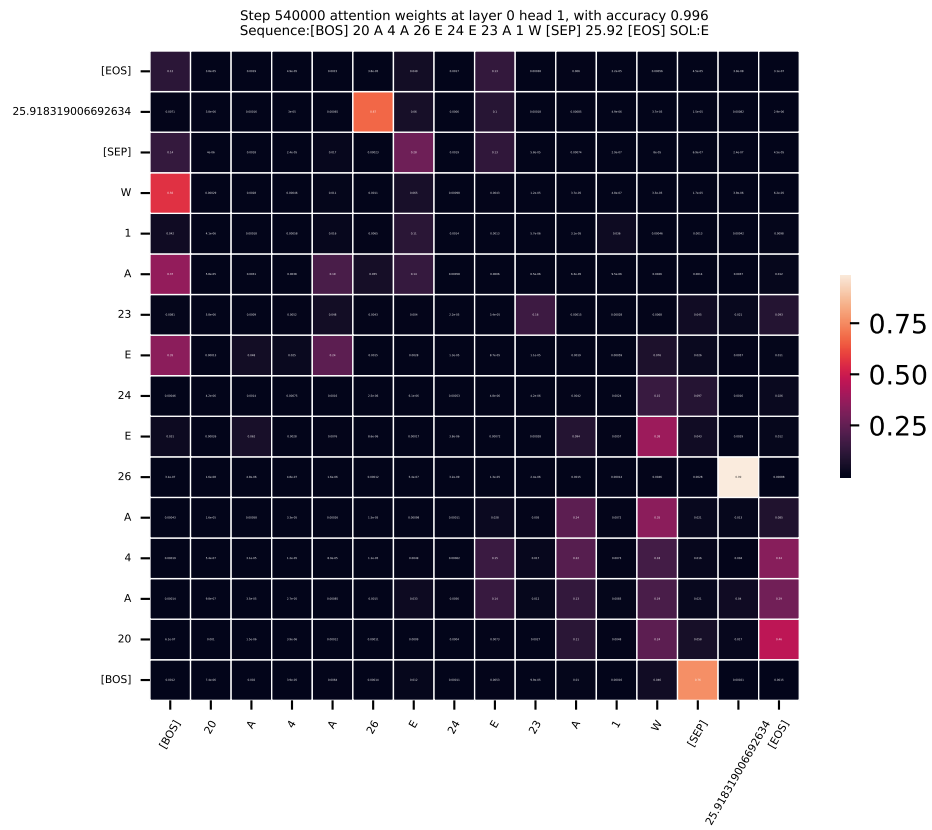

Step 539500 attention weights at layer 1 head 0, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

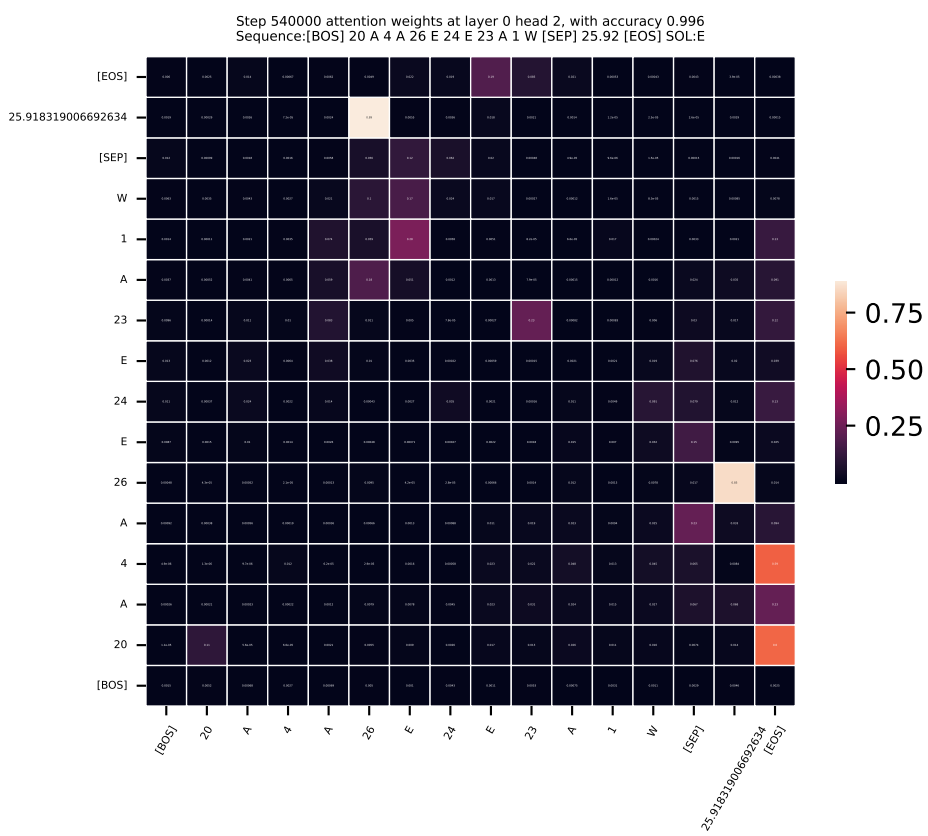

Step 539500 attention weights at layer 1 head 1, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

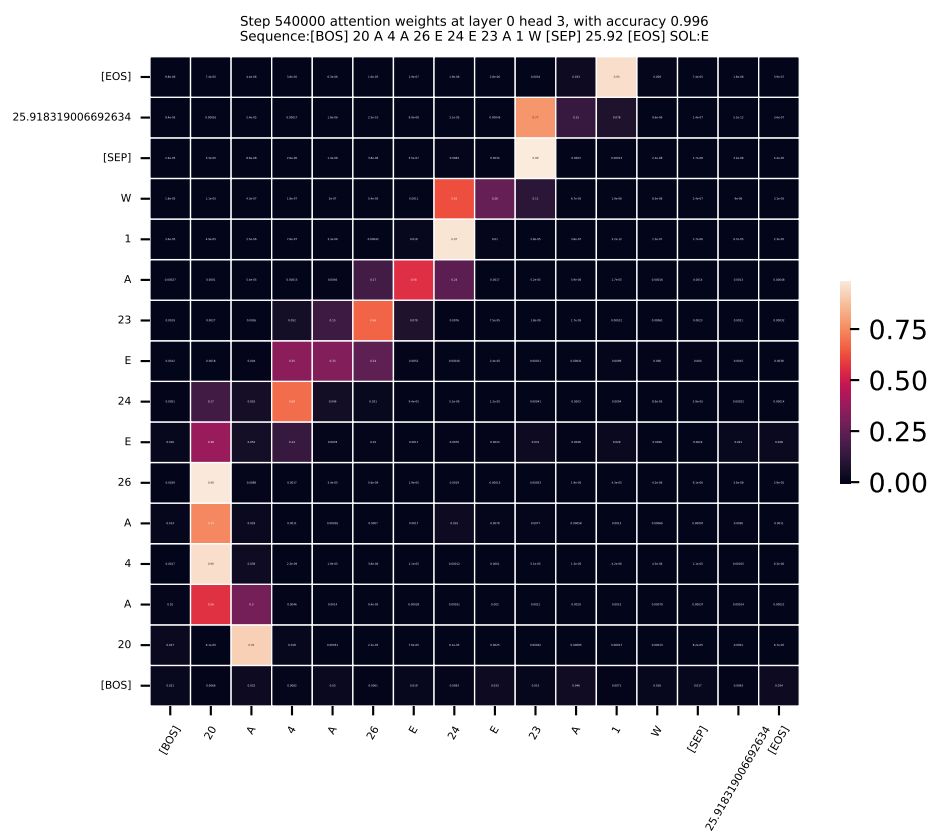

Step 539500 attention weights at layer 1 head 2, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

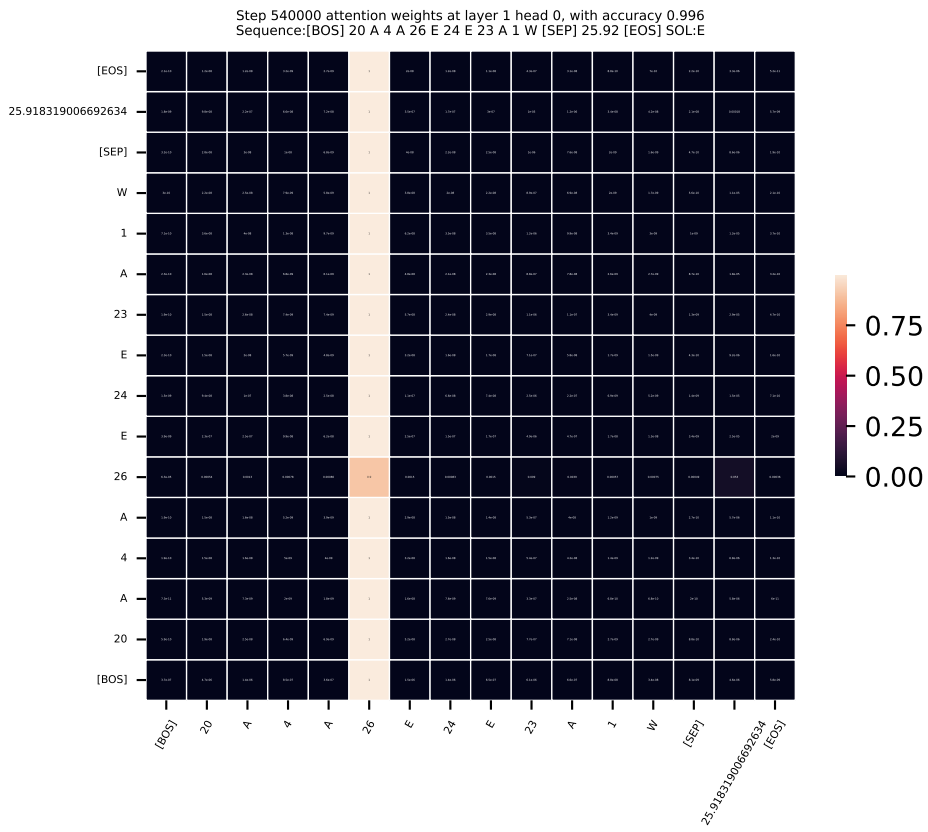

Step 539500 attention weights at layer 1 head 3, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

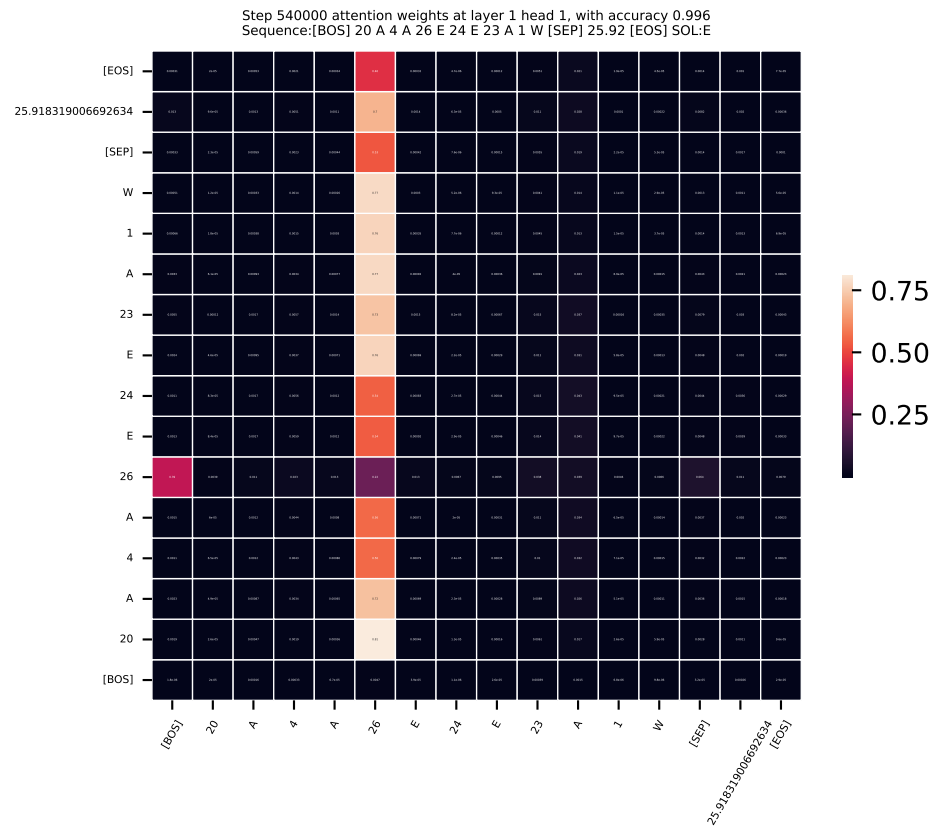

Step 540000 attention weights at layer 0 head 0, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

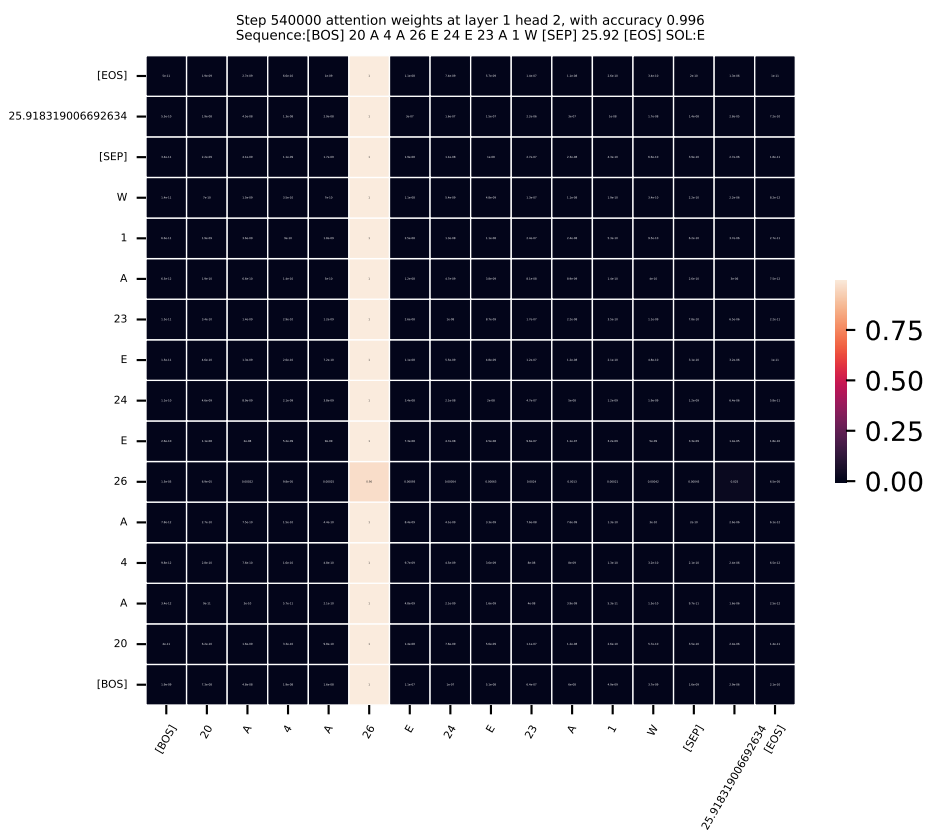

Step 540000 attention weights at layer 0 head 1, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

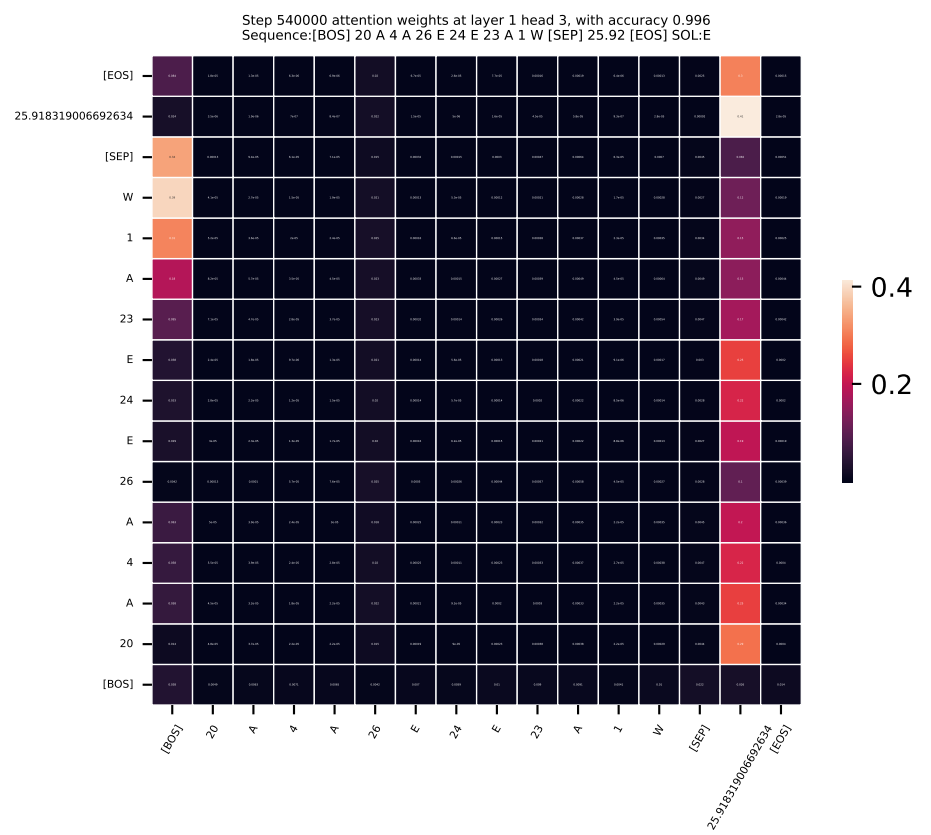

Step 540000 attention weights at layer 0 head 2, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

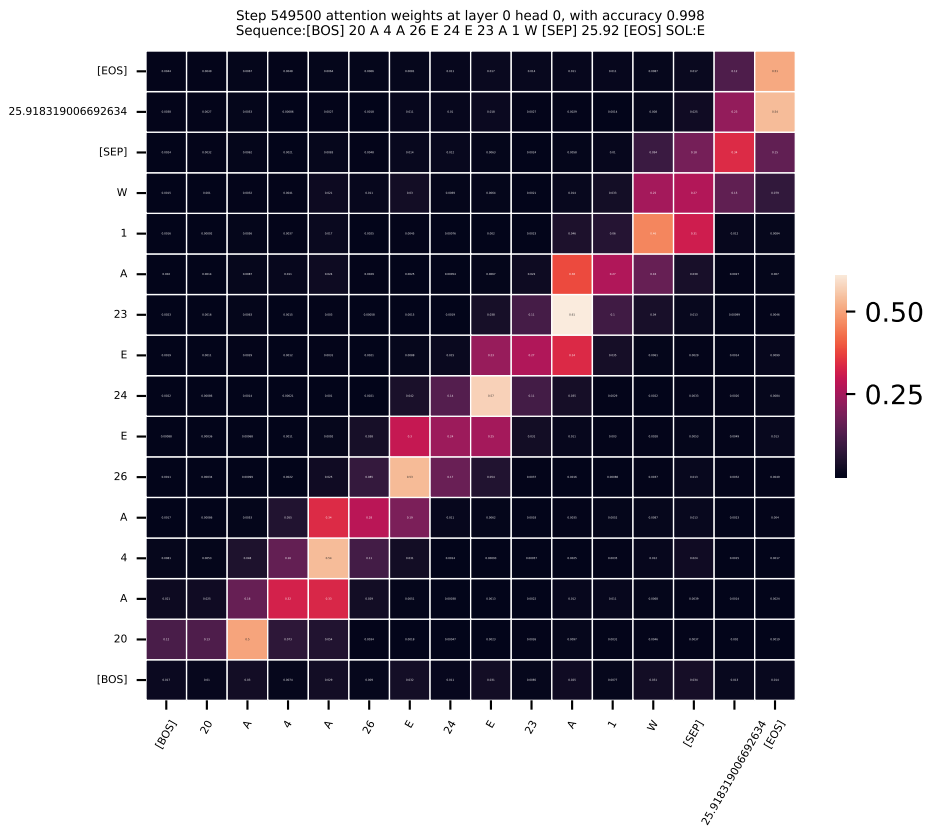

Step 540000 attention weights at layer 0 head 3, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

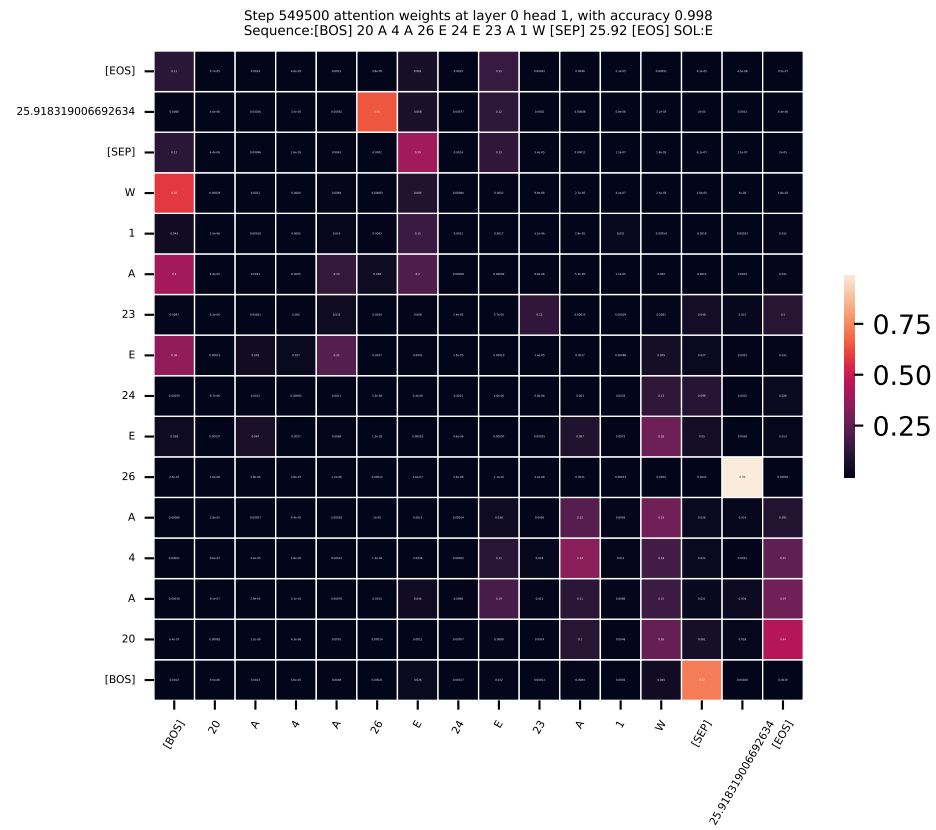

Step 540000 attention weights at layer 1 head 0, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

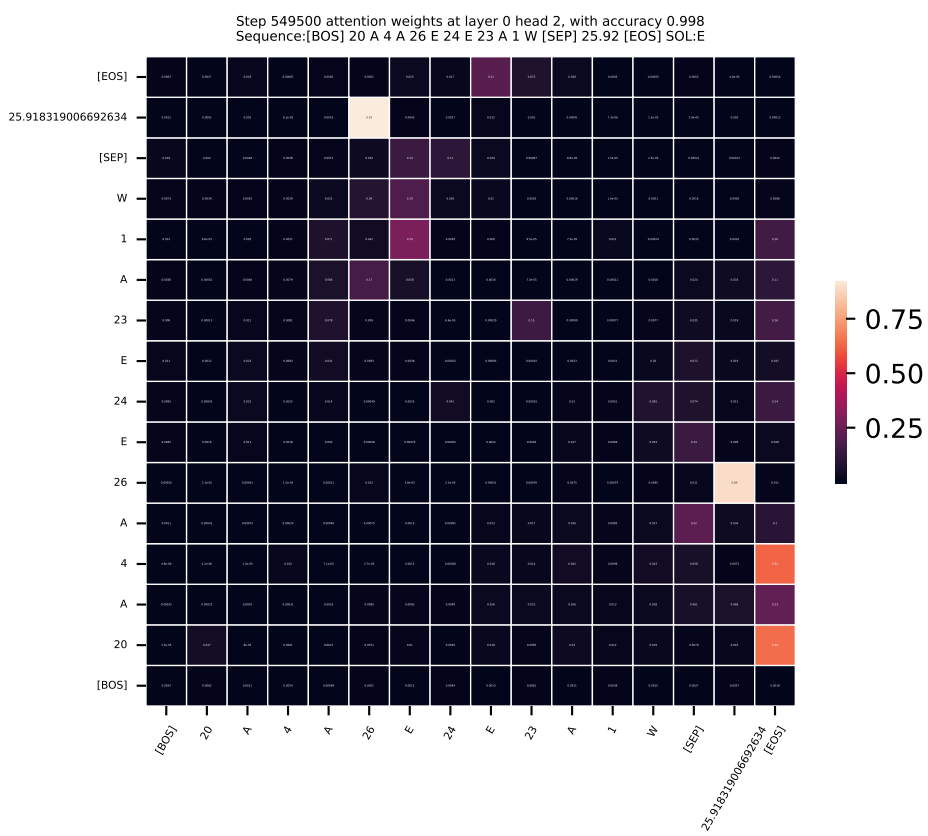

Step 540000 attention weights at layer 1 head 1, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

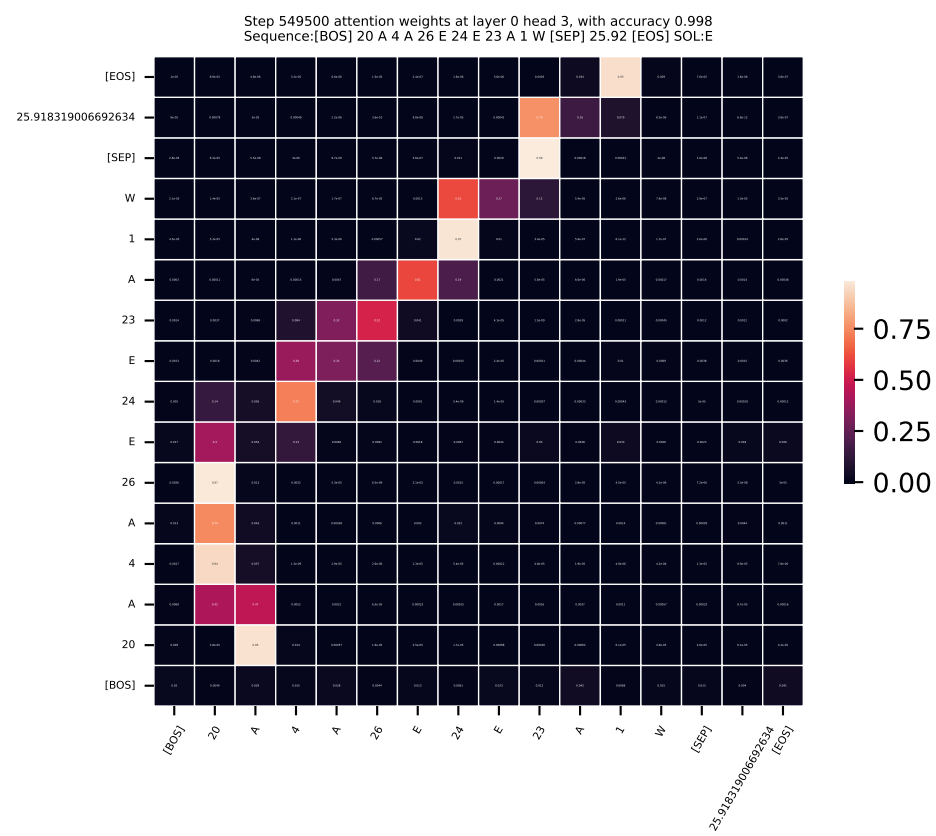

Step 540000 attention weights at layer 1 head 2, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

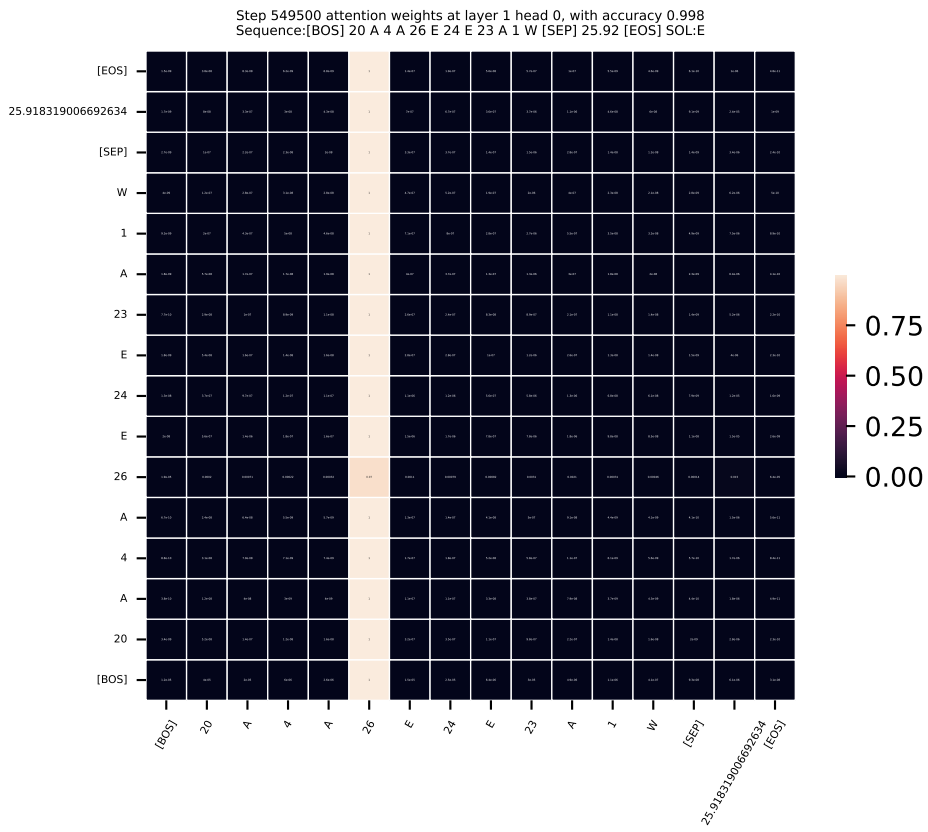

Step 540000 attention weights at layer 1 head 3, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

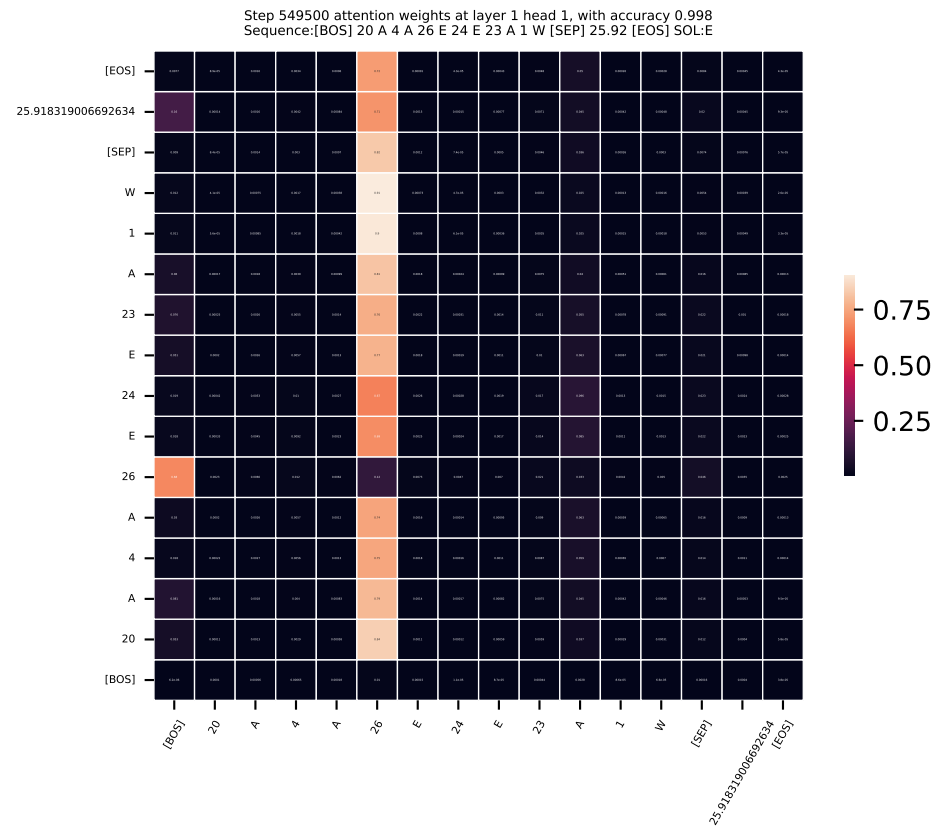

Step 549500 attention weights at layer 0 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

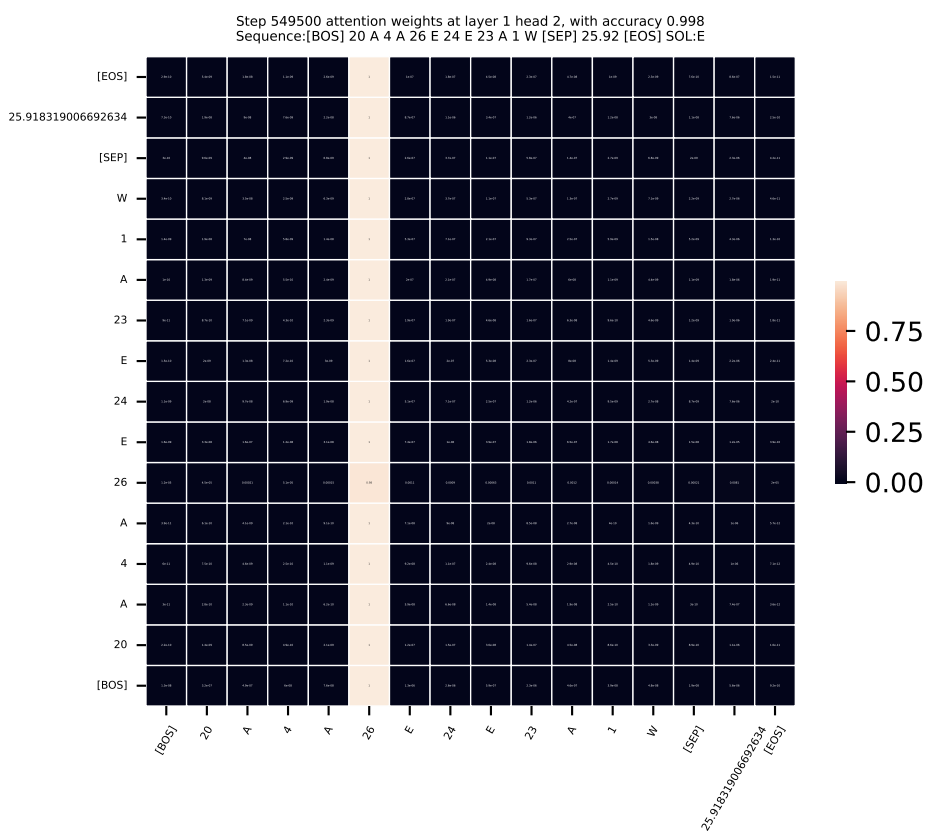

Step 549500 attention weights at layer 0 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

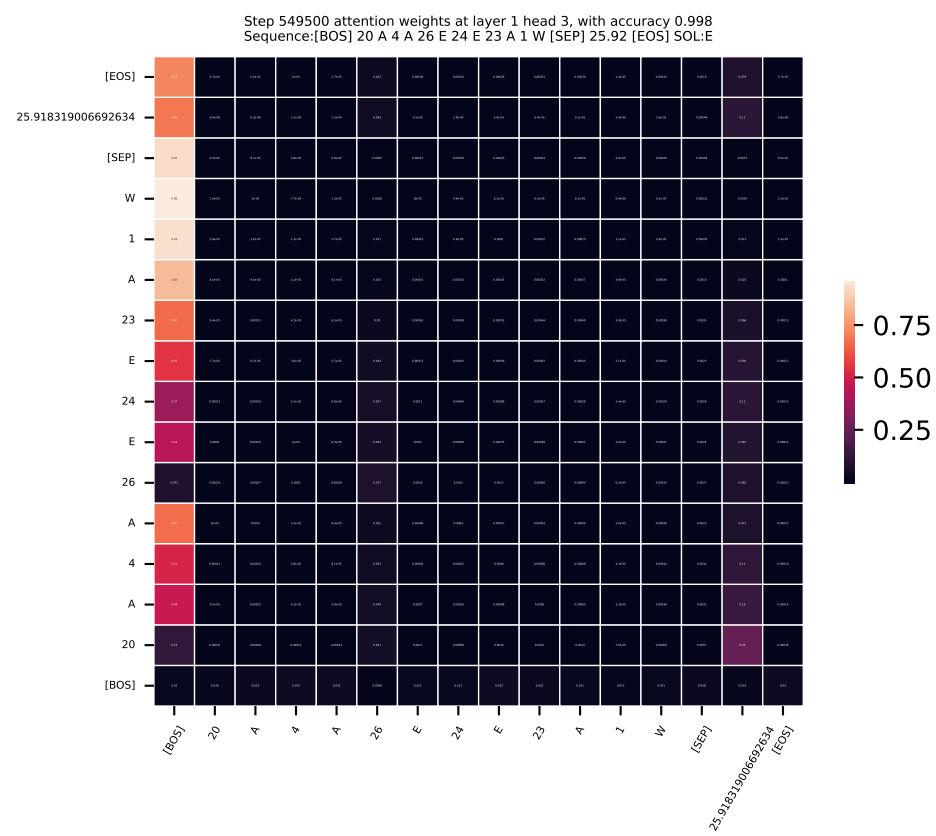

Step 549500 attention weights at layer 0 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

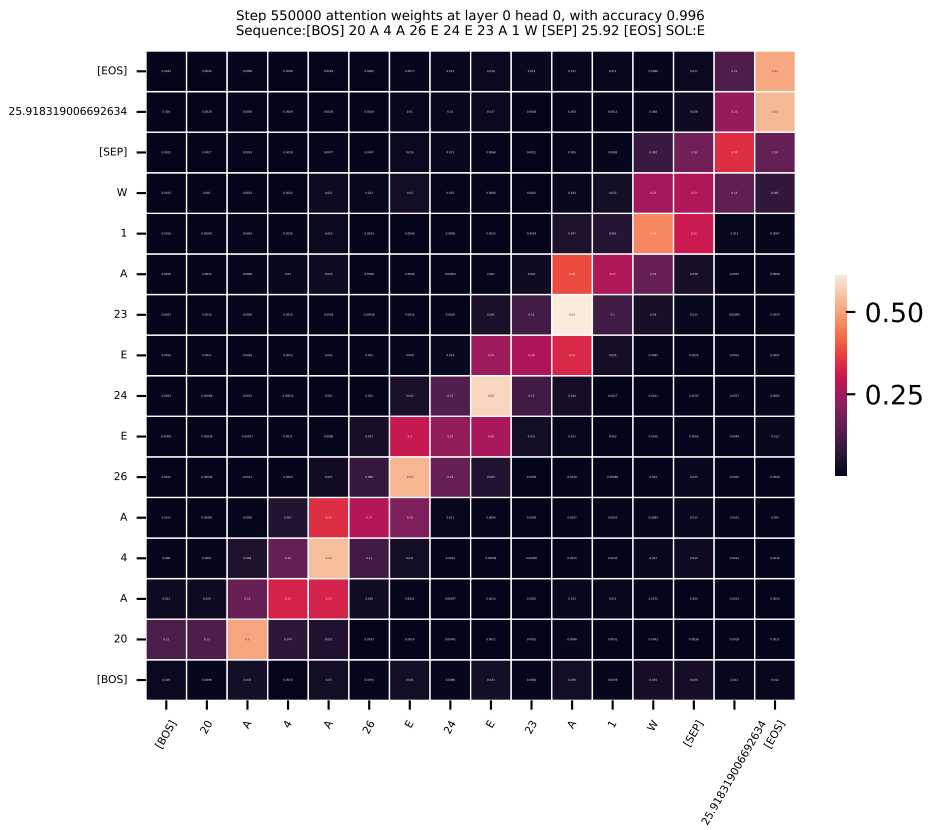

Step 549500 attention weights at layer 0 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

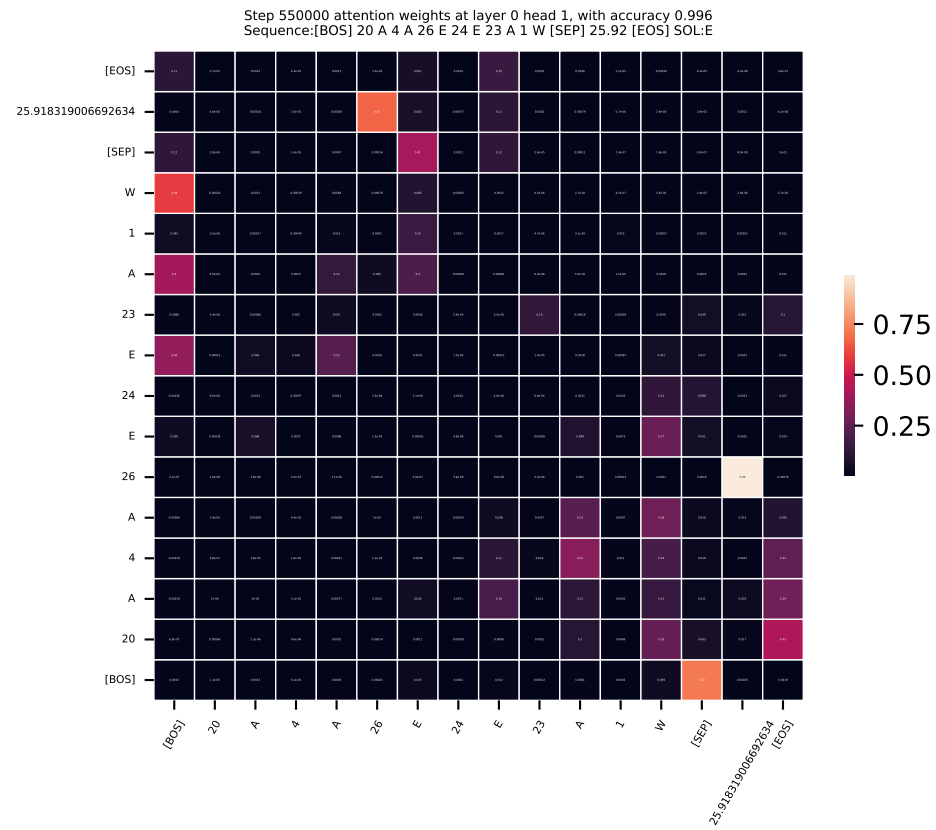

Step 549500 attention weights at layer 1 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

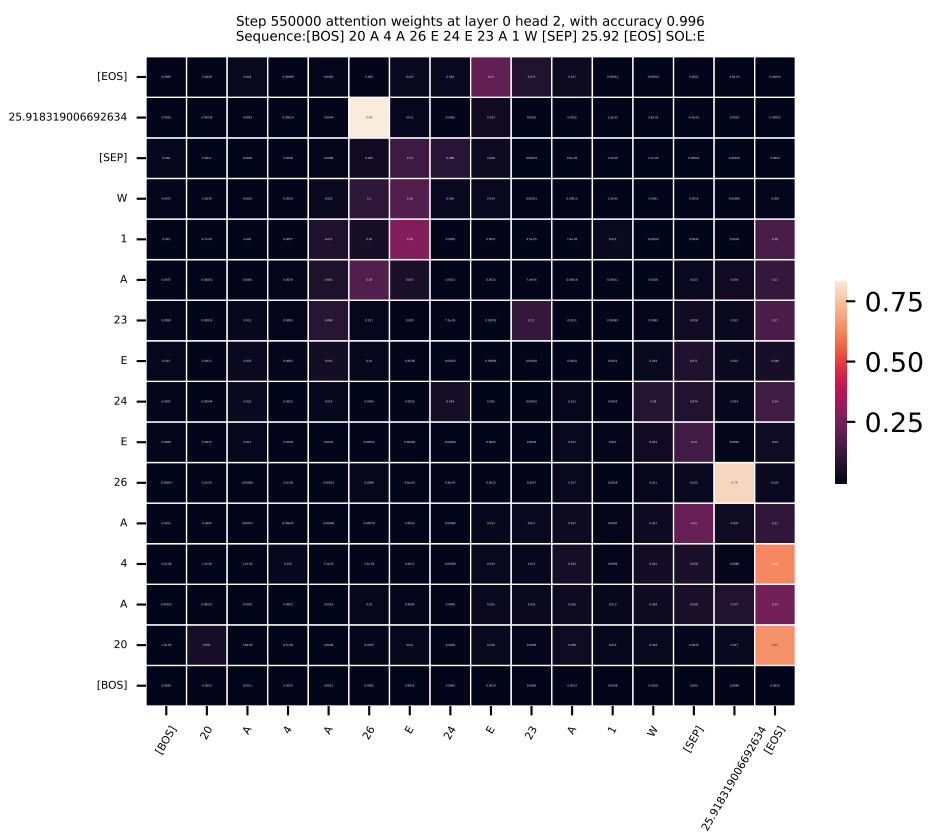

Step 549500 attention weights at layer 1 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

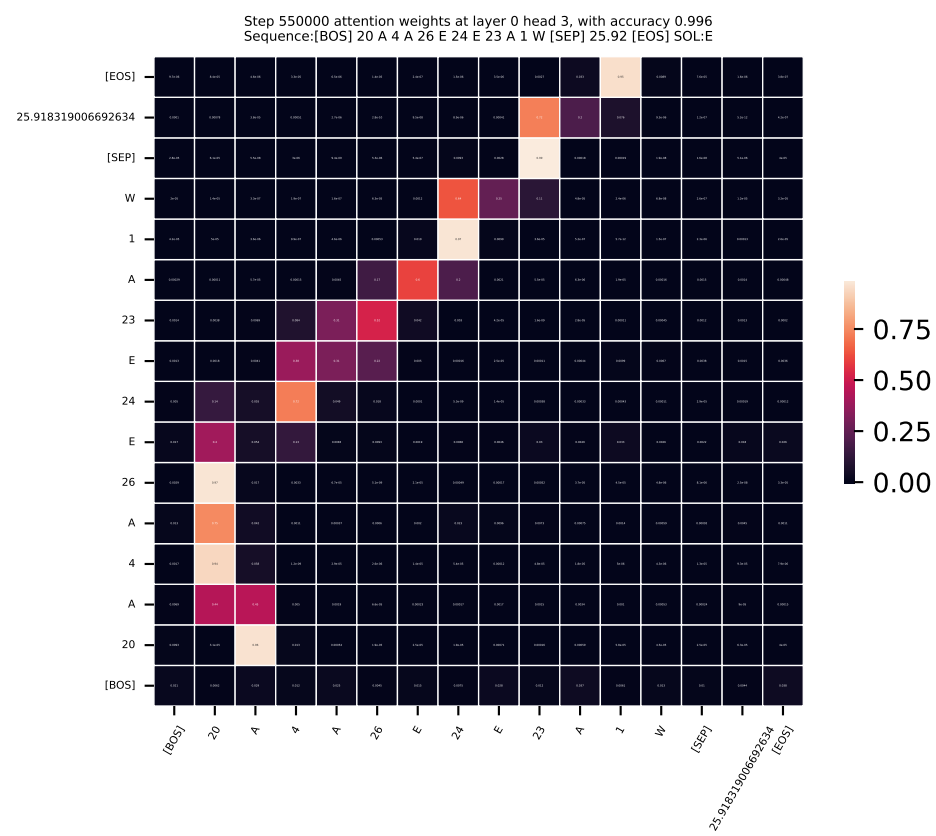

Step 549500 attention weights at layer 1 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

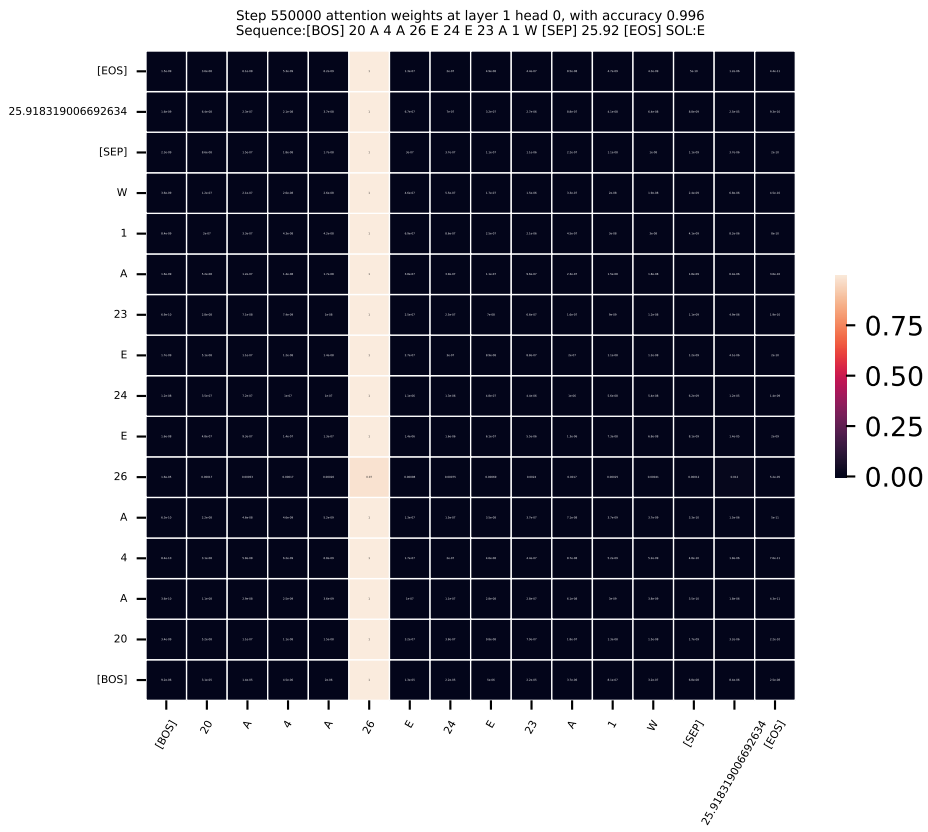

Step 549500 attention weights at layer 1 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

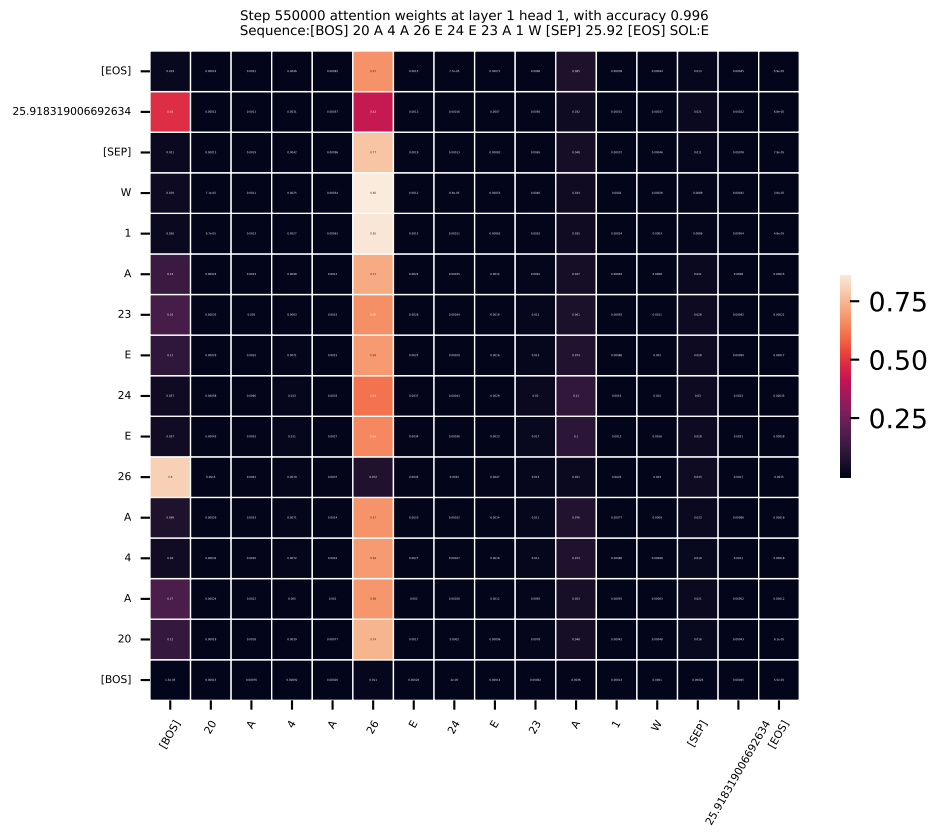

Step 550000 attention weights at layer 0 head 0, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

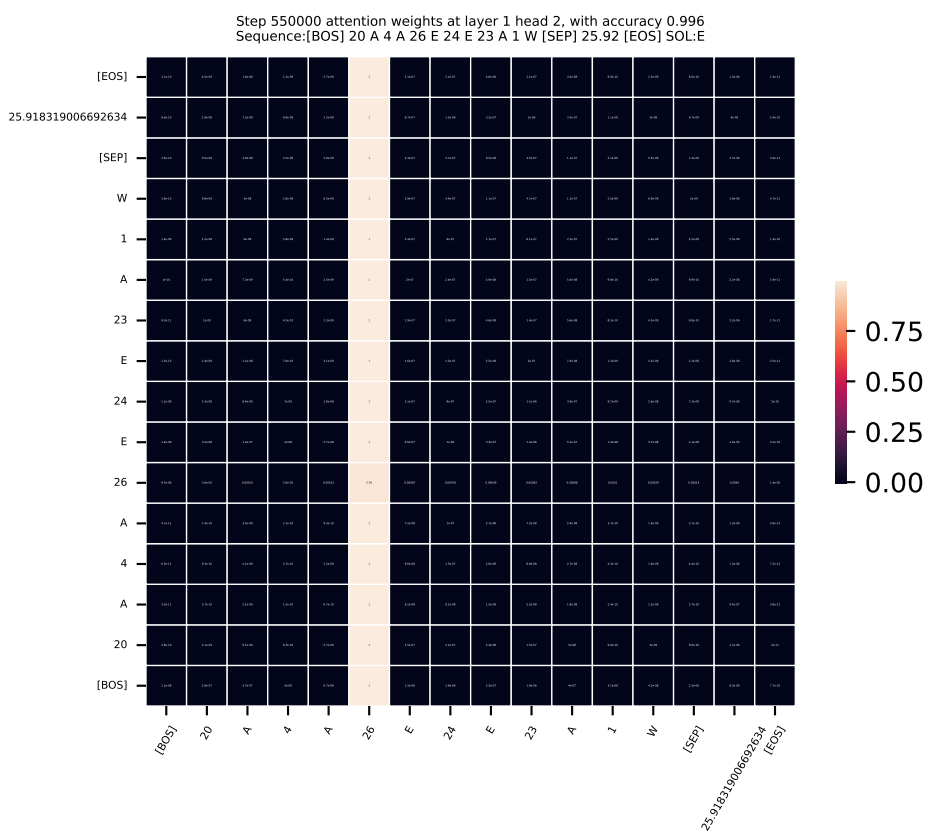

Step 550000 attention weights at layer 0 head 1, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

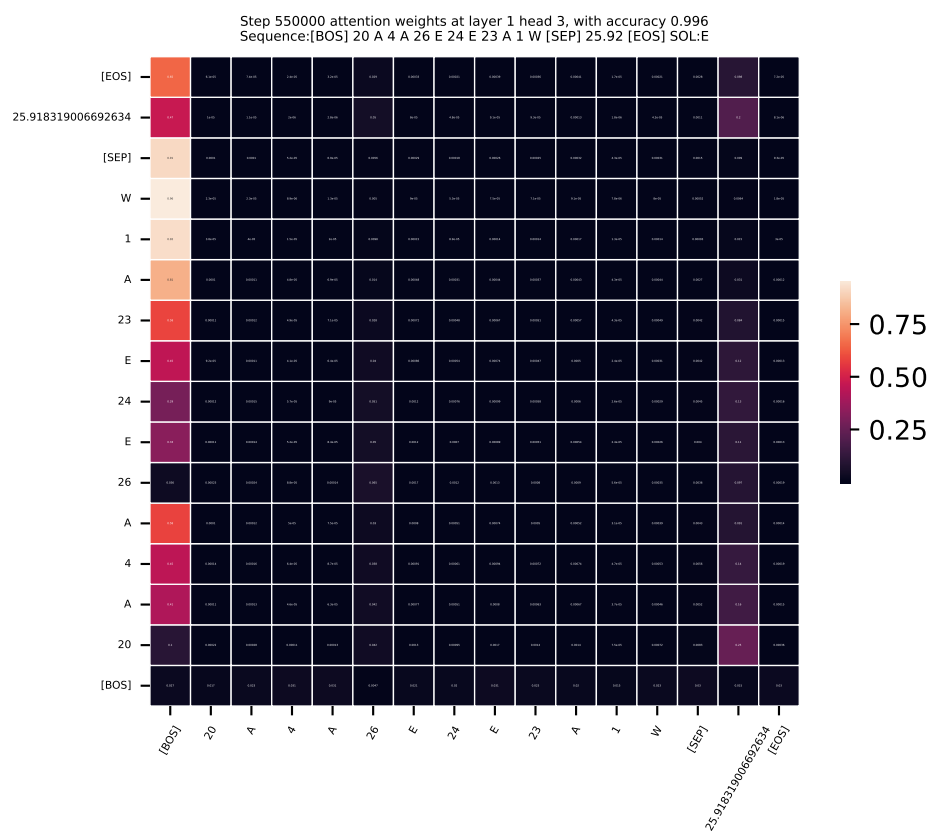

Step 550000 attention weights at layer 0 head 2, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

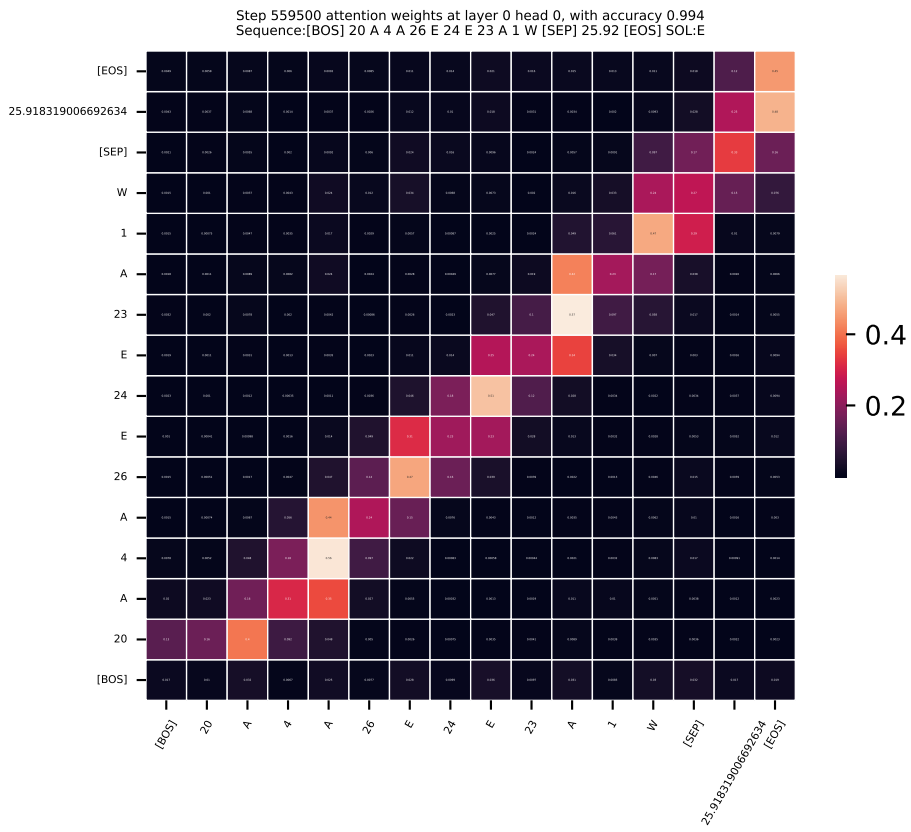

Step 550000 attention weights at layer 0 head 3, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

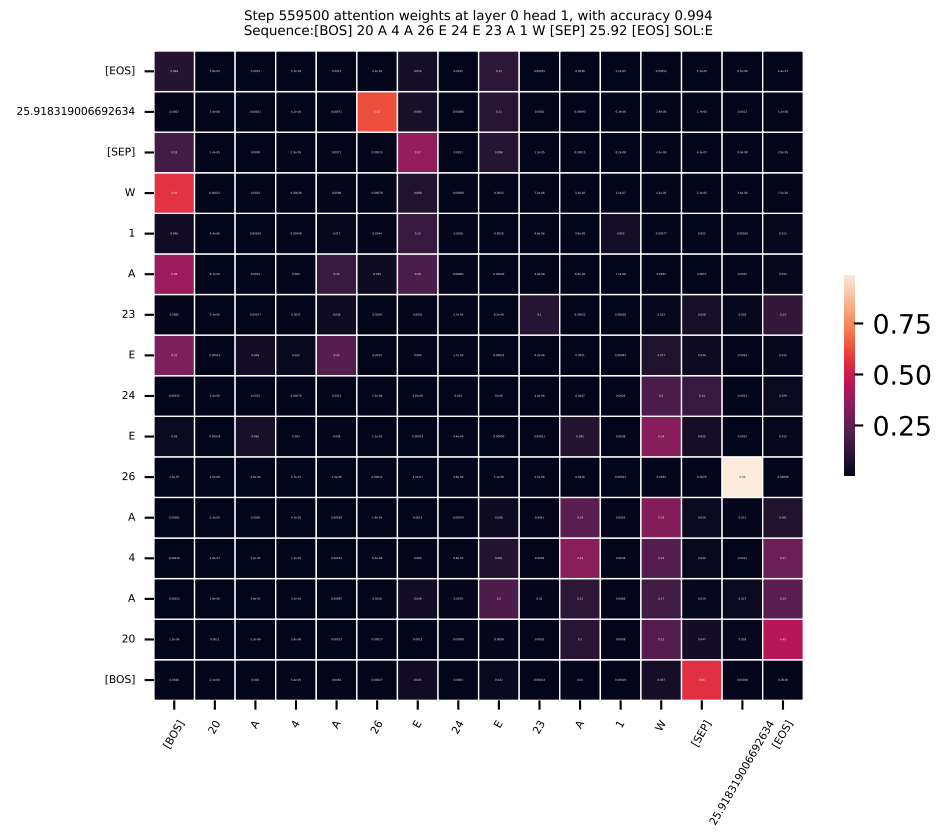

Step 550000 attention weights at layer 1 head 0, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

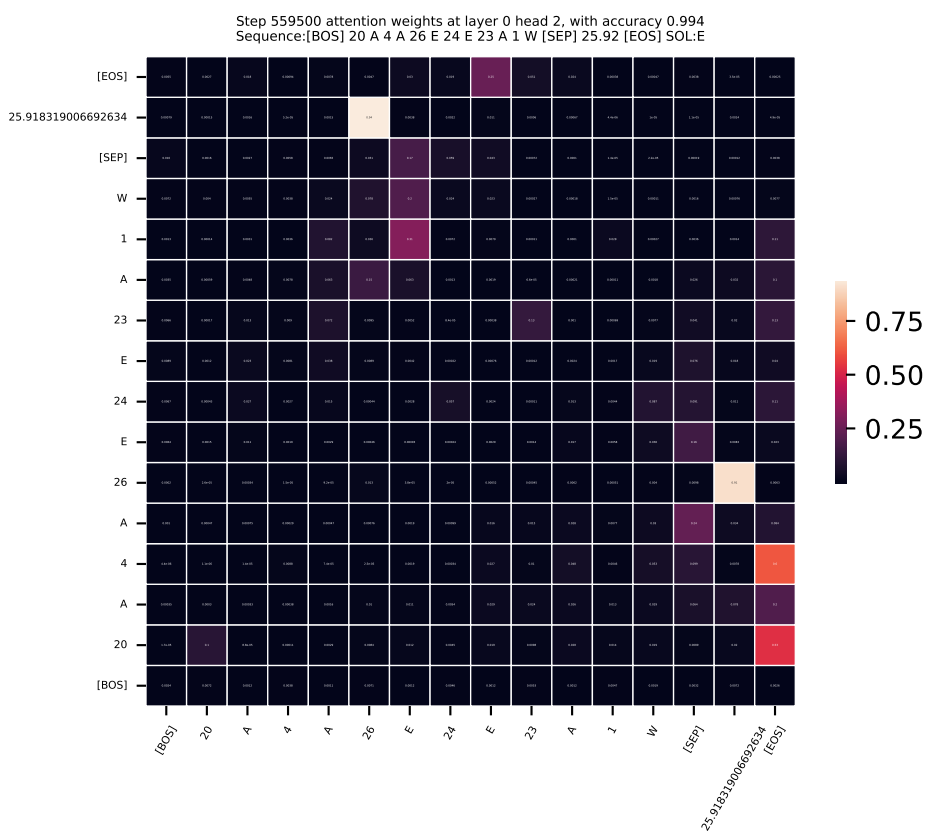

Step 550000 attention weights at layer 1 head 1, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

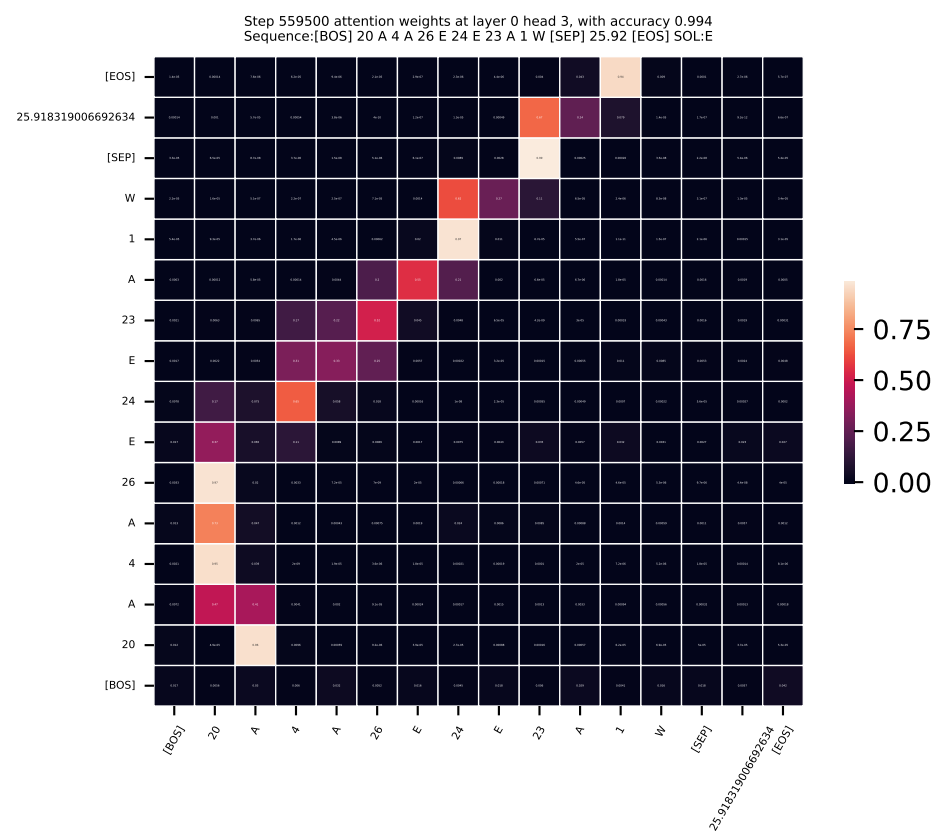

Step 550000 attention weights at layer 1 head 2, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

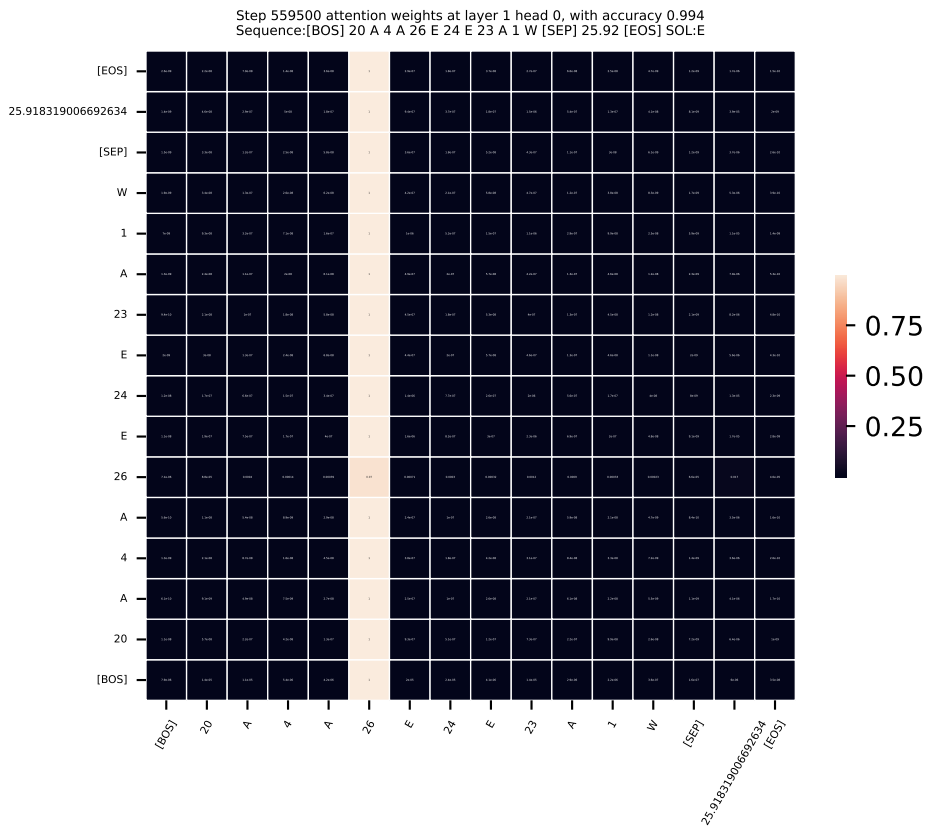

Step 550000 attention weights at layer 1 head 3, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

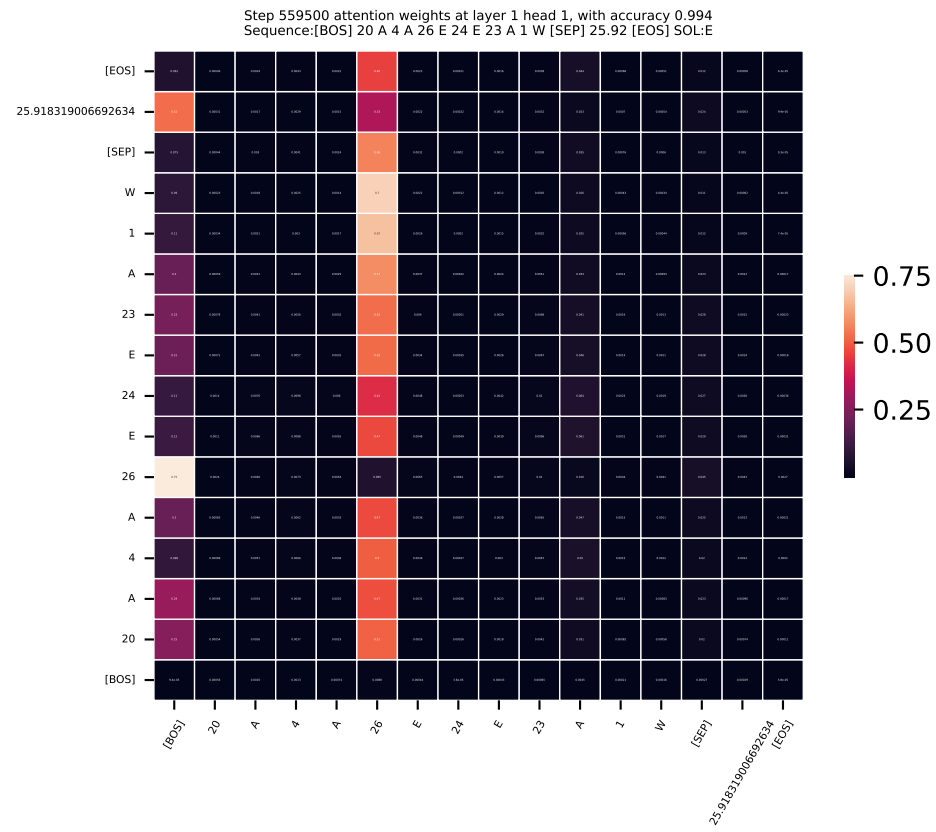

Step 559500 attention weights at layer 0 head 0, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

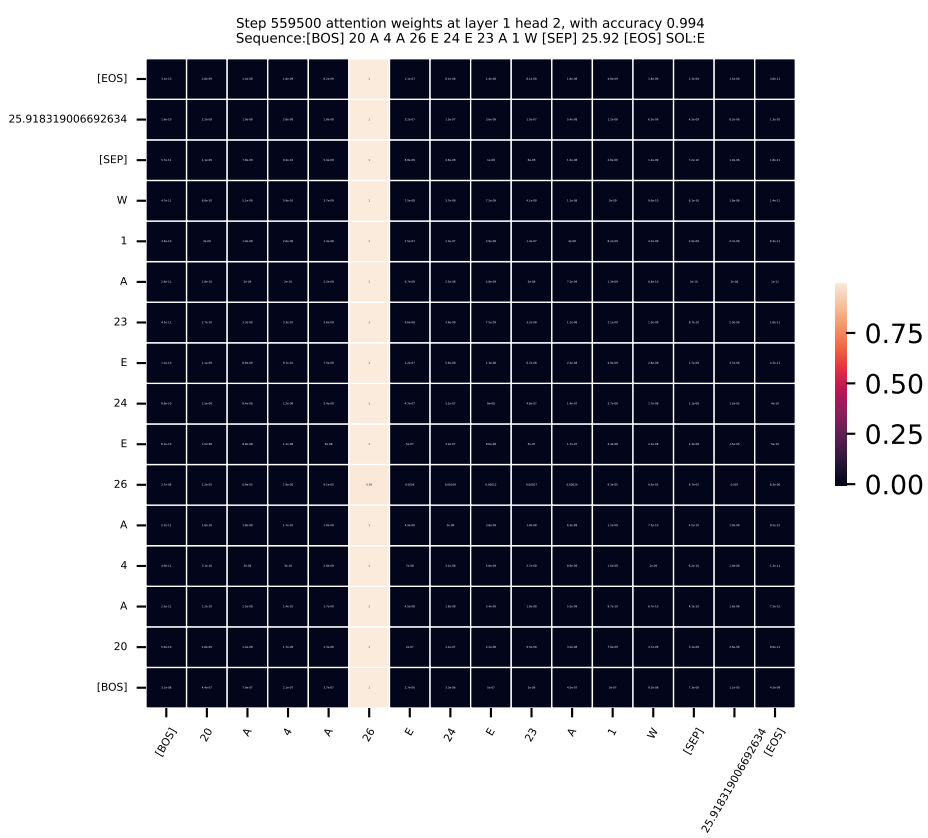

Step 559500 attention weights at layer 0 head 1, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

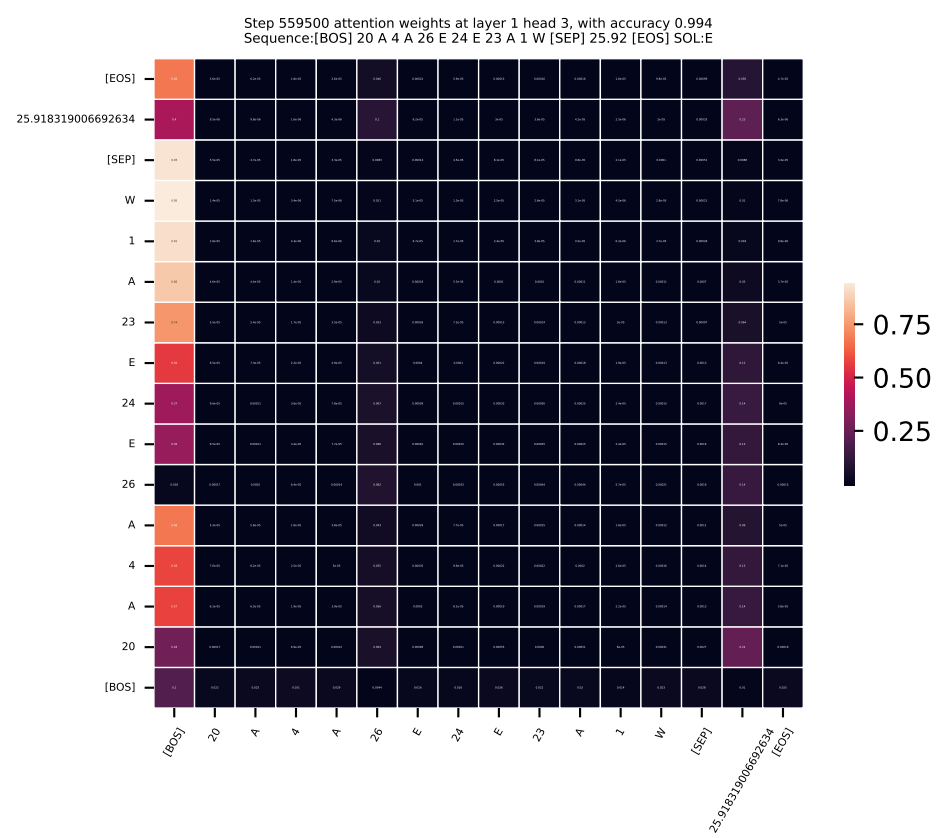

Step 559500 attention weights at layer 0 head 2, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

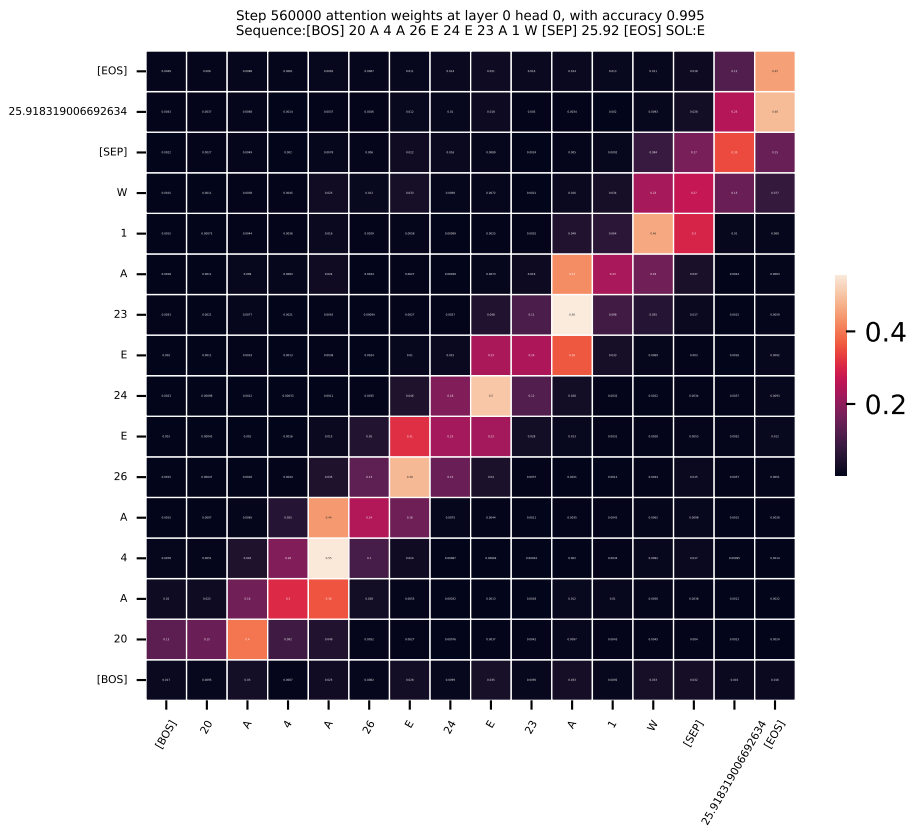

Step 559500 attention weights at layer 0 head 3, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

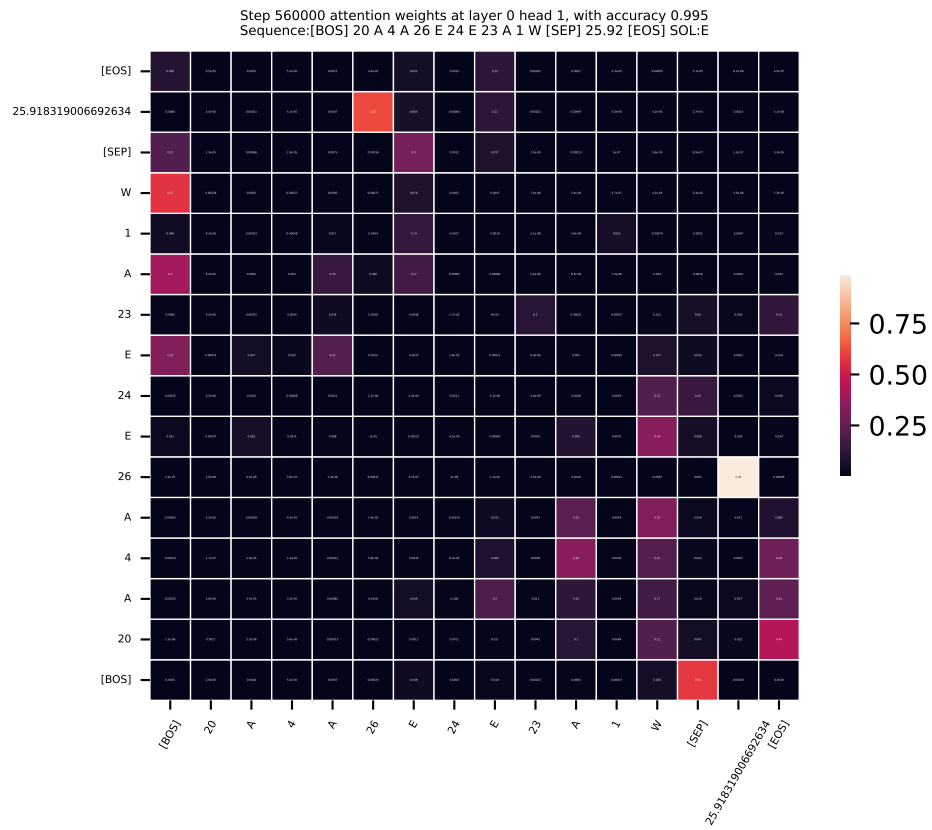

Step 559500 attention weights at layer 1 head 0, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

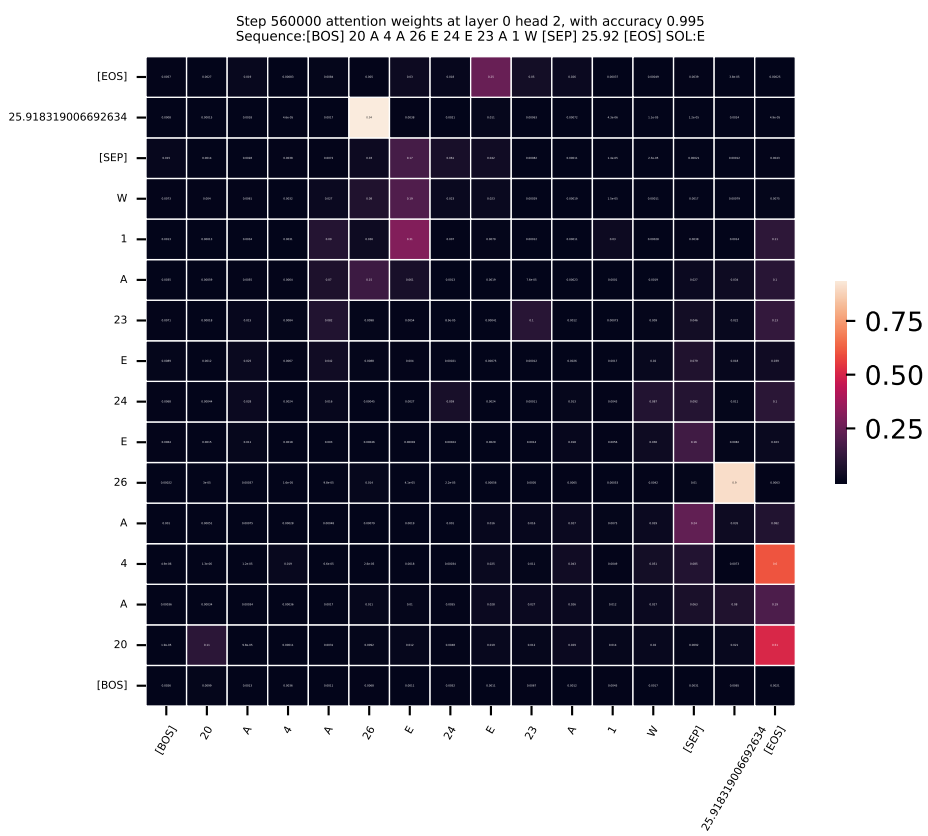

Step 559500 attention weights at layer 1 head 1, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

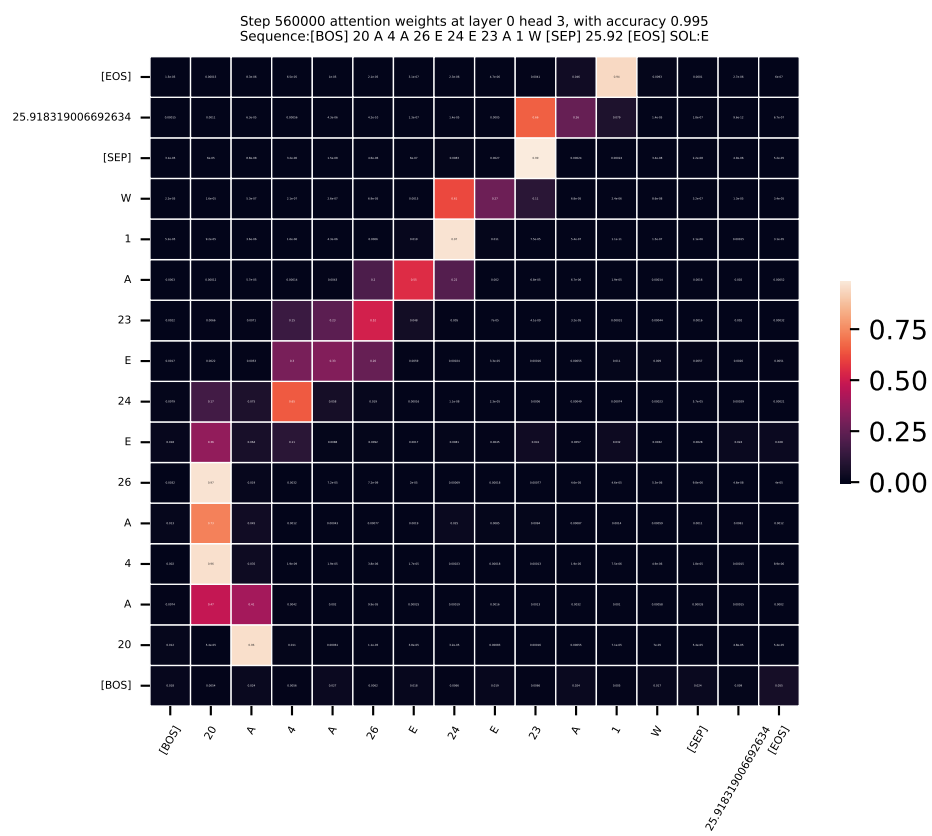

Step 559500 attention weights at layer 1 head 2, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

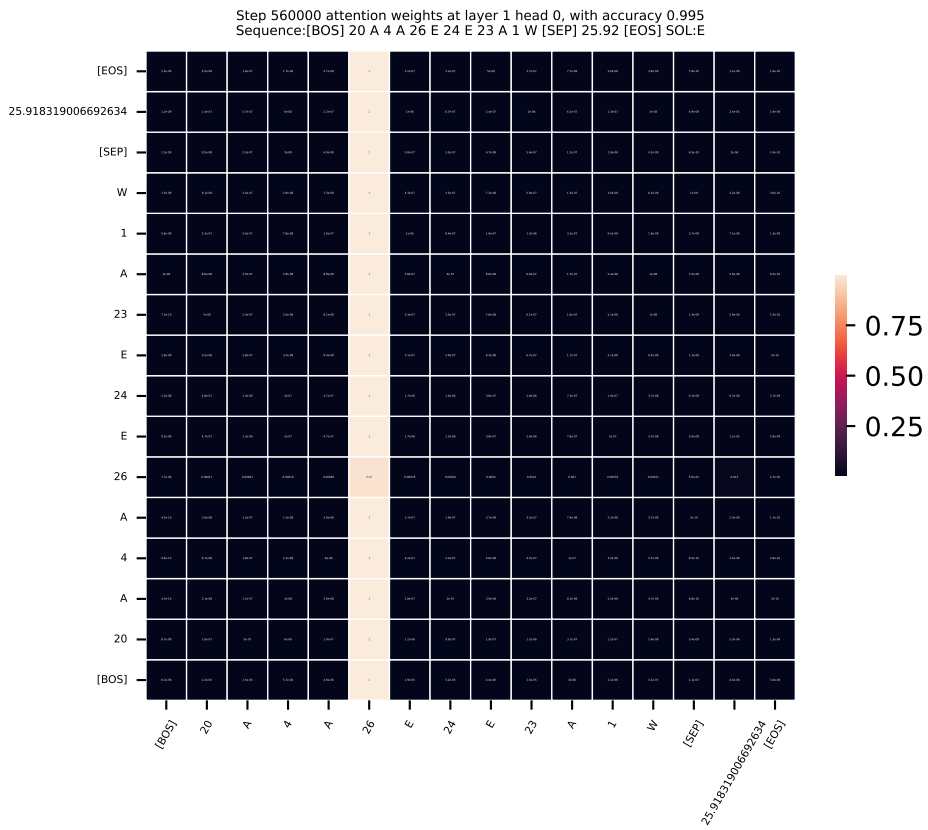

Step 559500 attention weights at layer 1 head 3, with accuracy 0.994
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

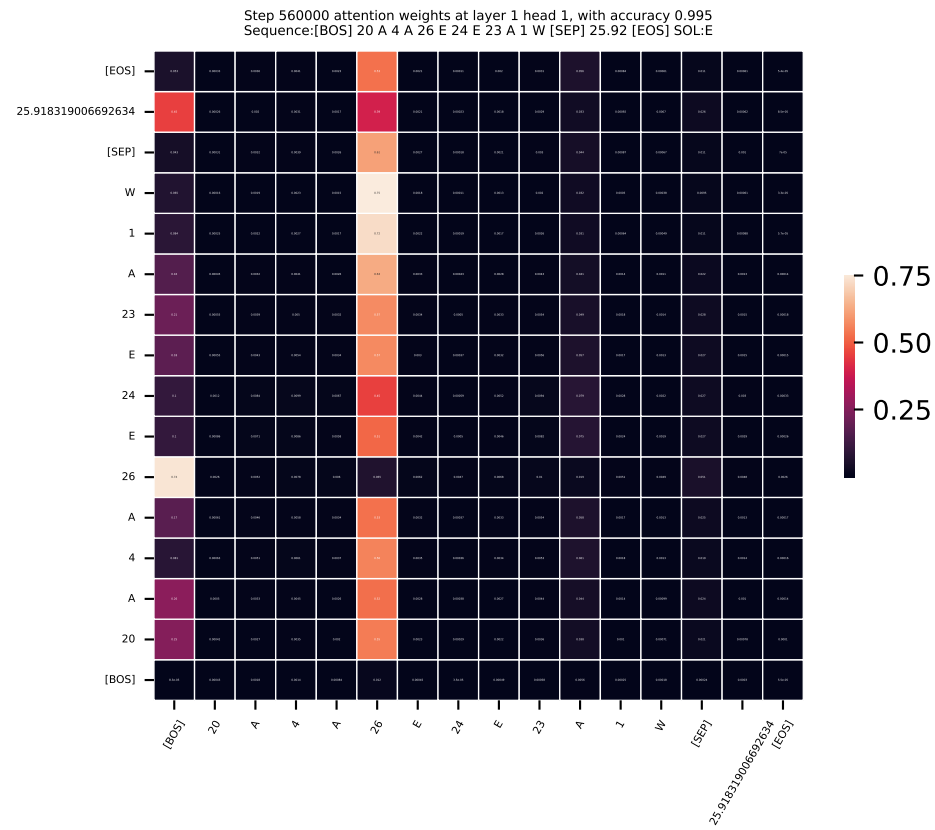

Step 560000 attention weights at layer 0 head 0, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.4
0.2

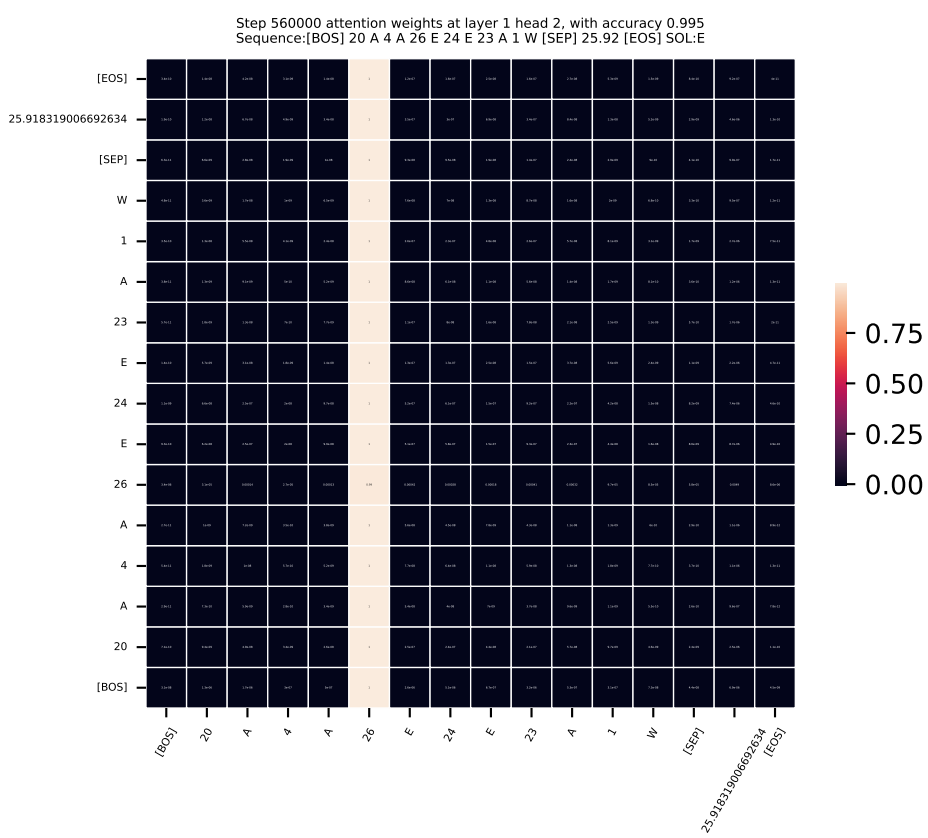

Step 560000 attention weights at layer 0 head 1, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

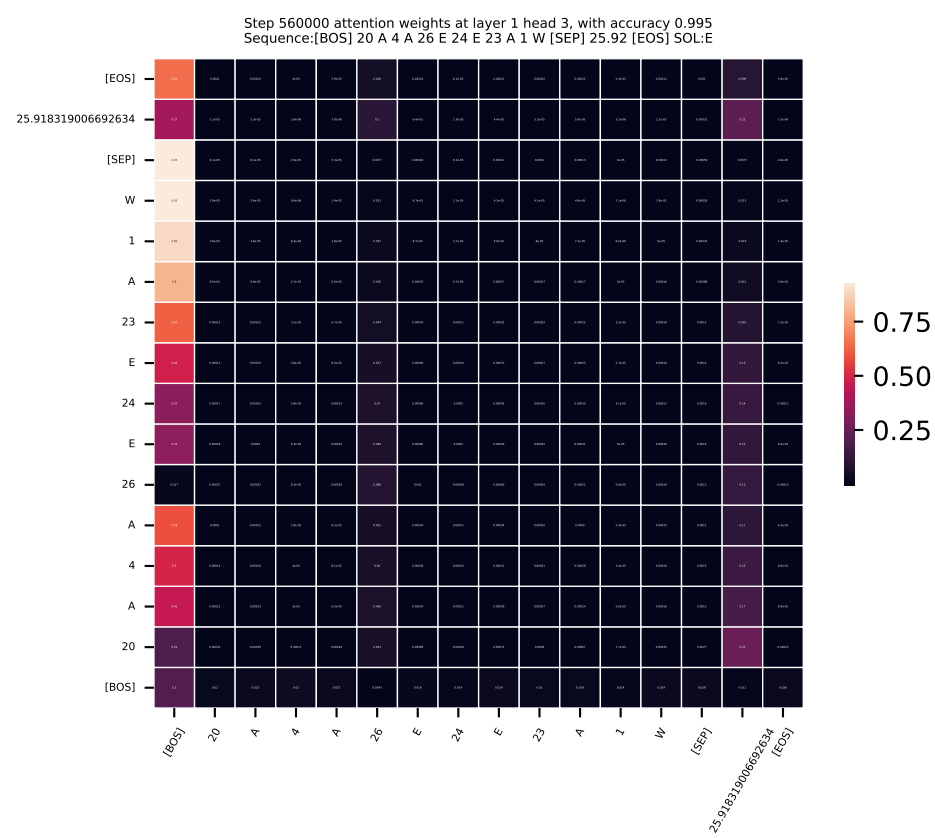

Step 560000 attention weights at layer 0 head 2, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

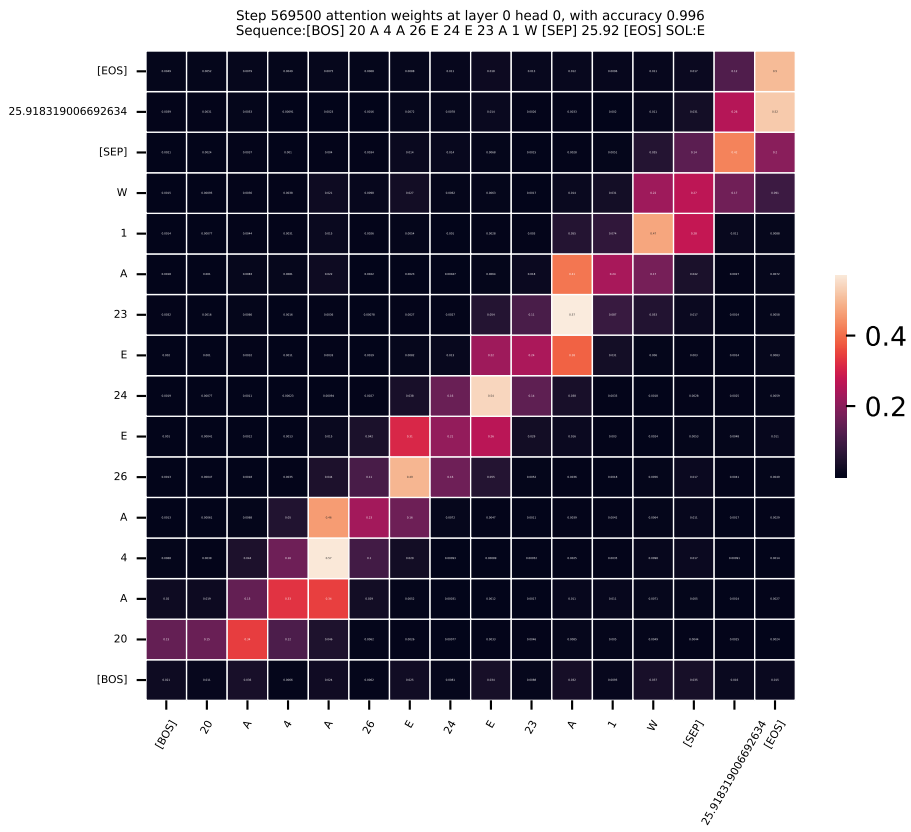

Step 560000 attention weights at layer 0 head 3, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25
0.00

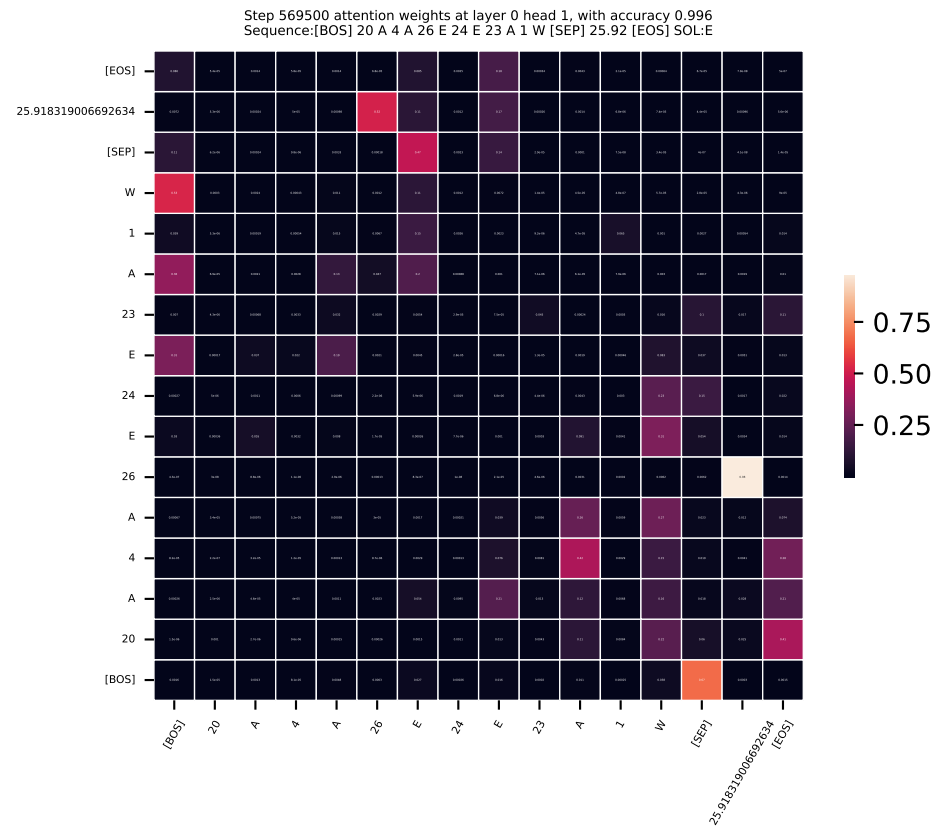

Step 560000 attention weights at layer 1 head 0, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

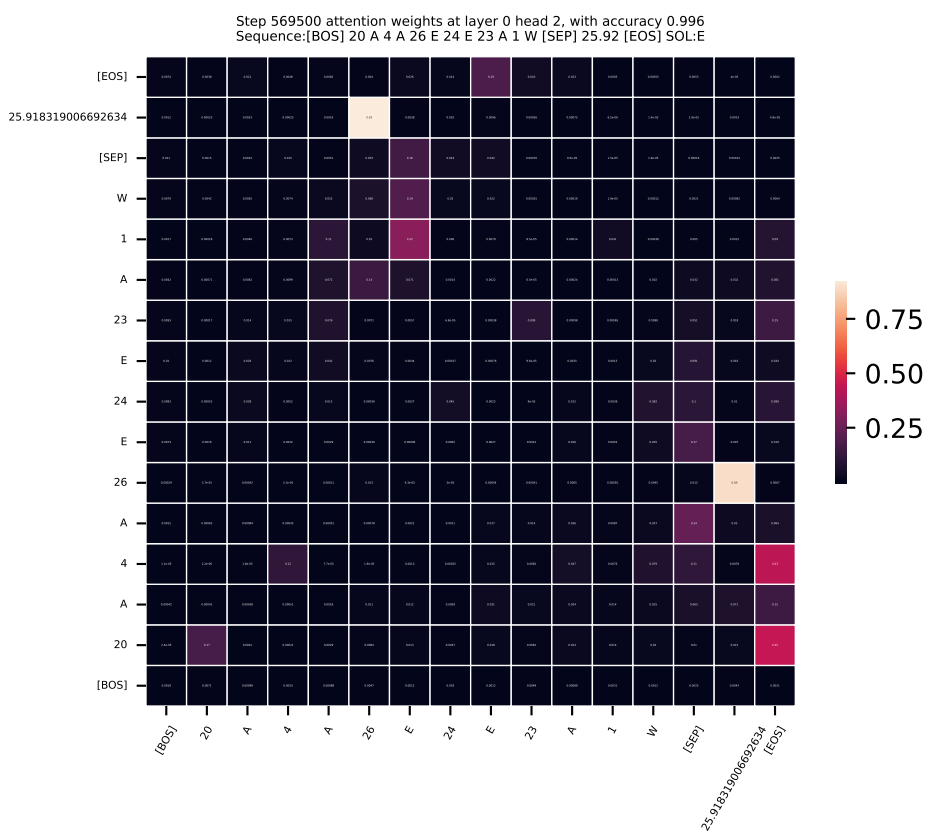

Step 560000 attention weights at layer 1 head 1, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

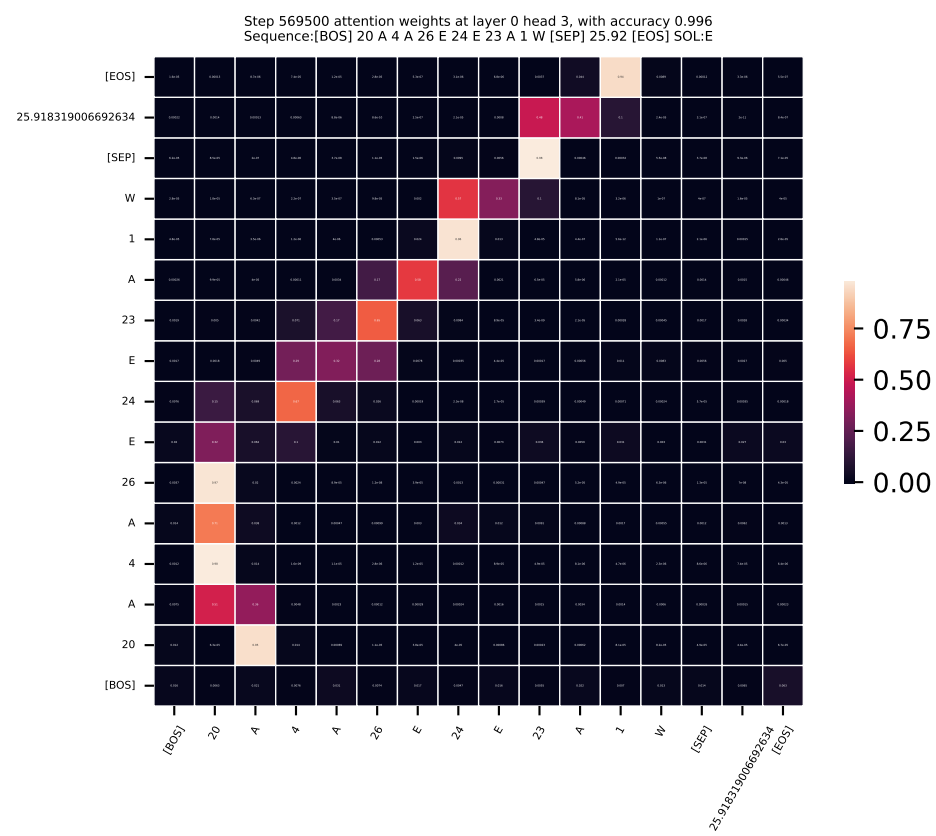

Step 560000 attention weights at layer 1 head 2, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

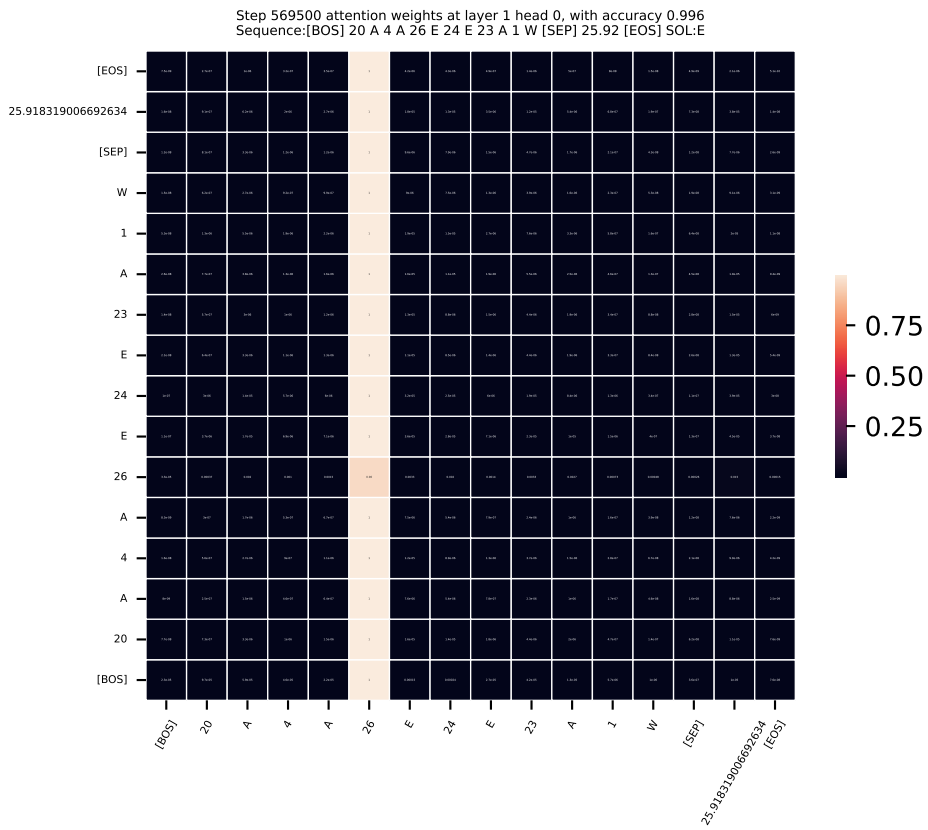

Step 560000 attention weights at layer 1 head 3, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

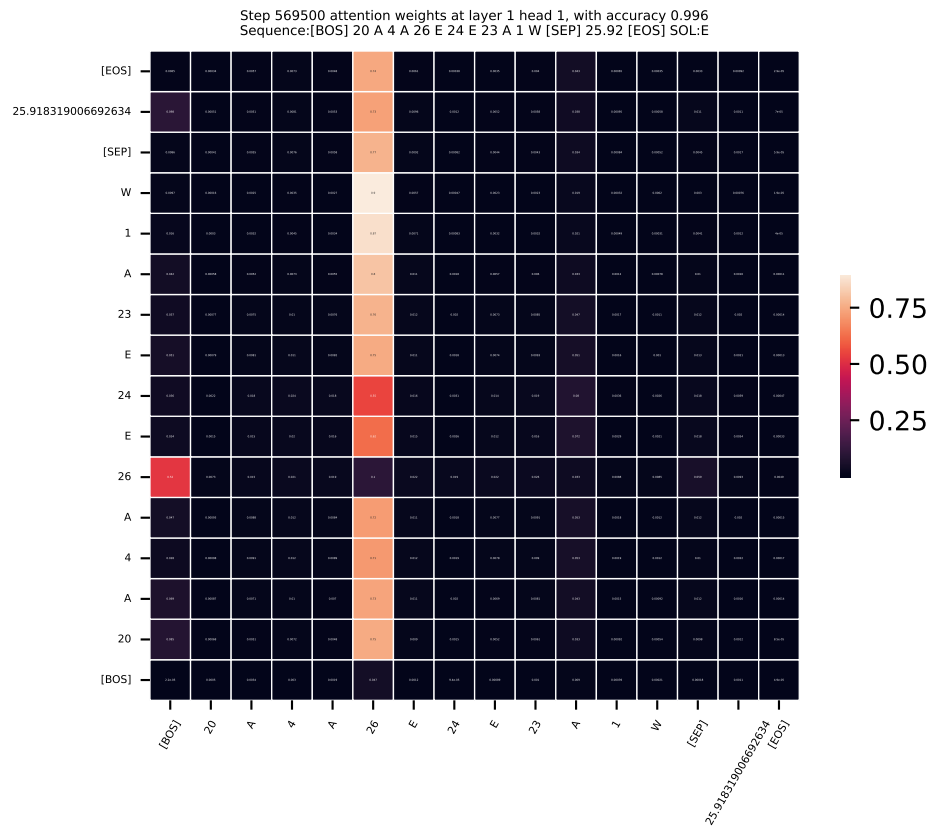

Step 569500 attention weights at layer 0 head 0, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.4
0.2

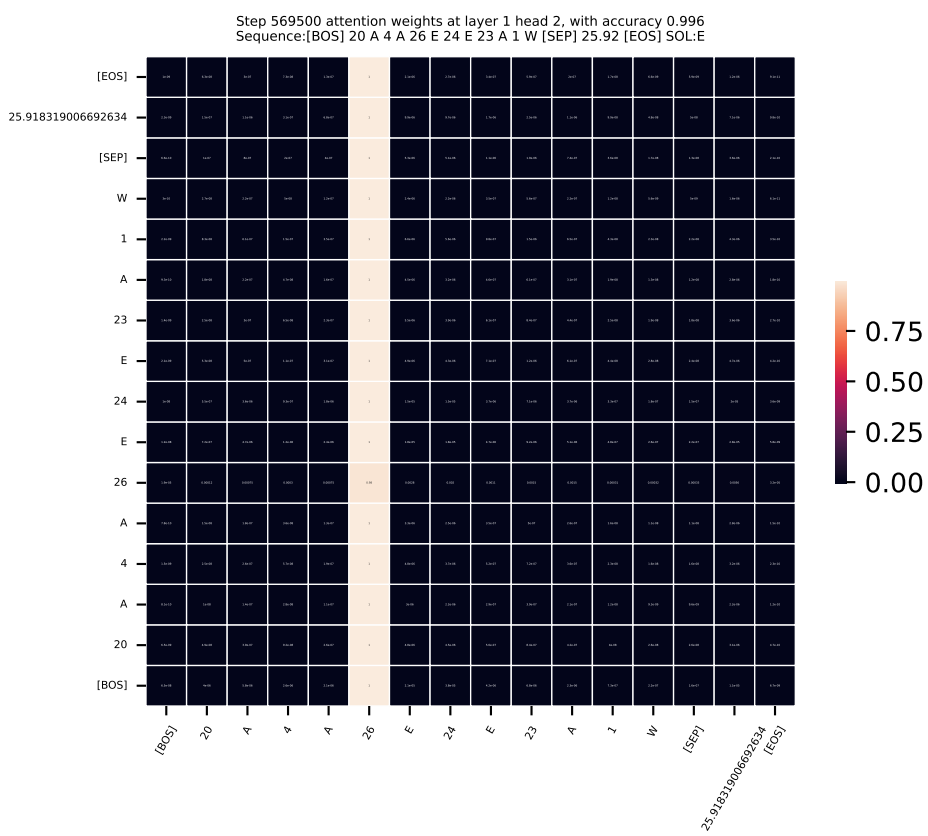

Step 569500 attention weights at layer 0 head 1, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

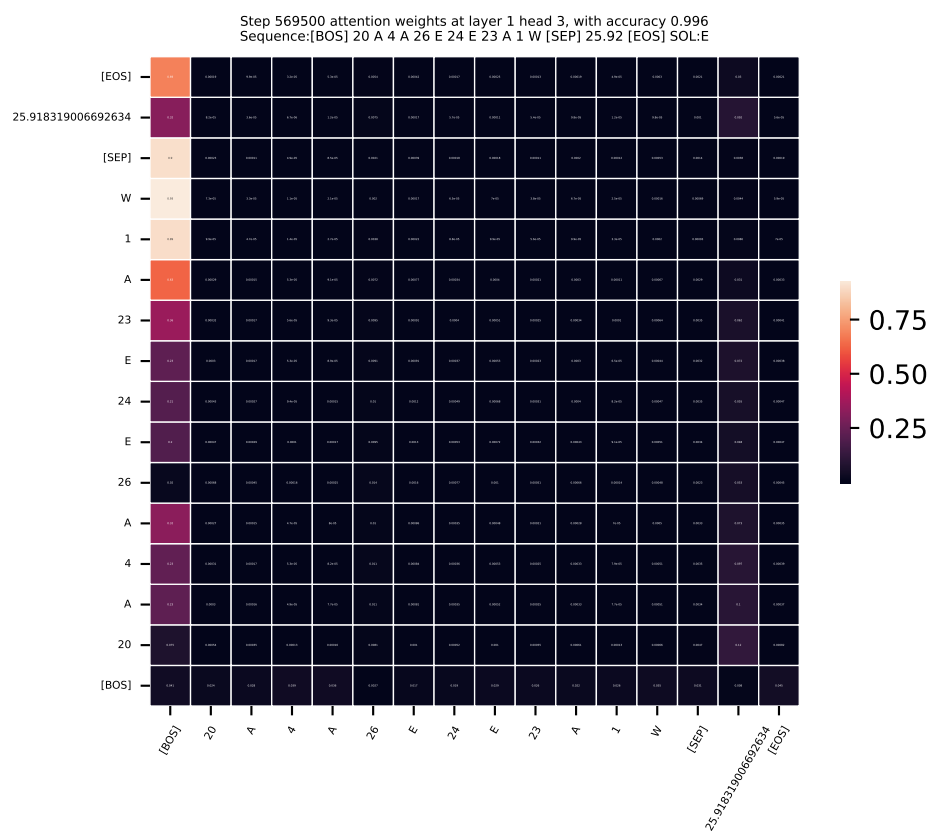

Step 569500 attention weights at layer 0 head 2, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

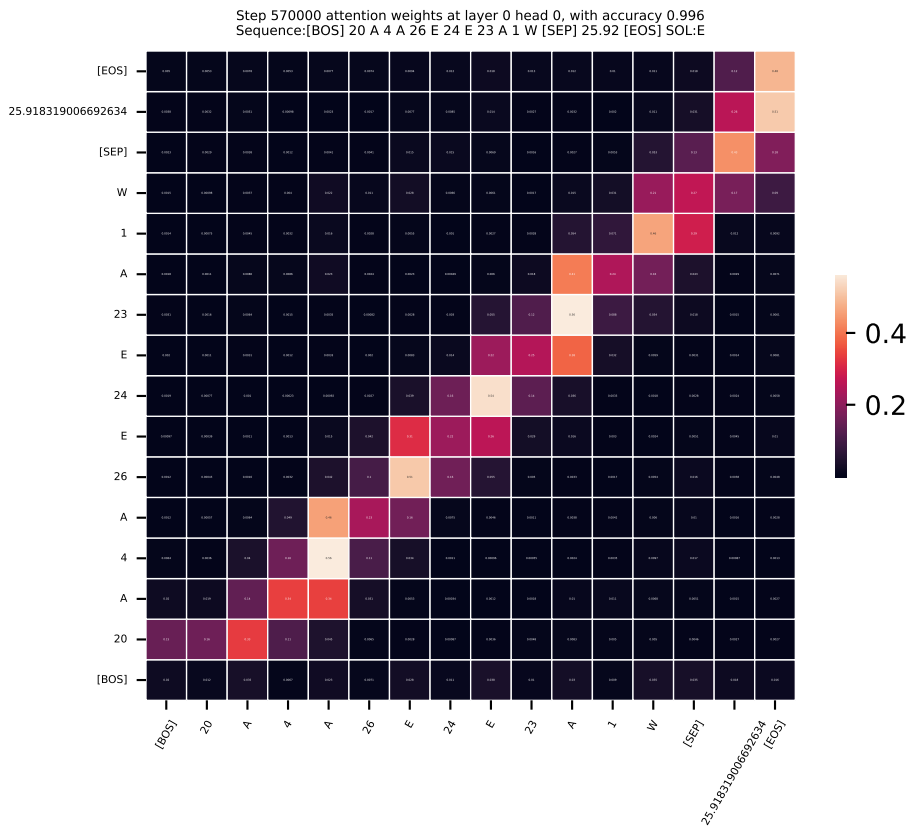

Step 569500 attention weights at layer 0 head 3, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25
0.00

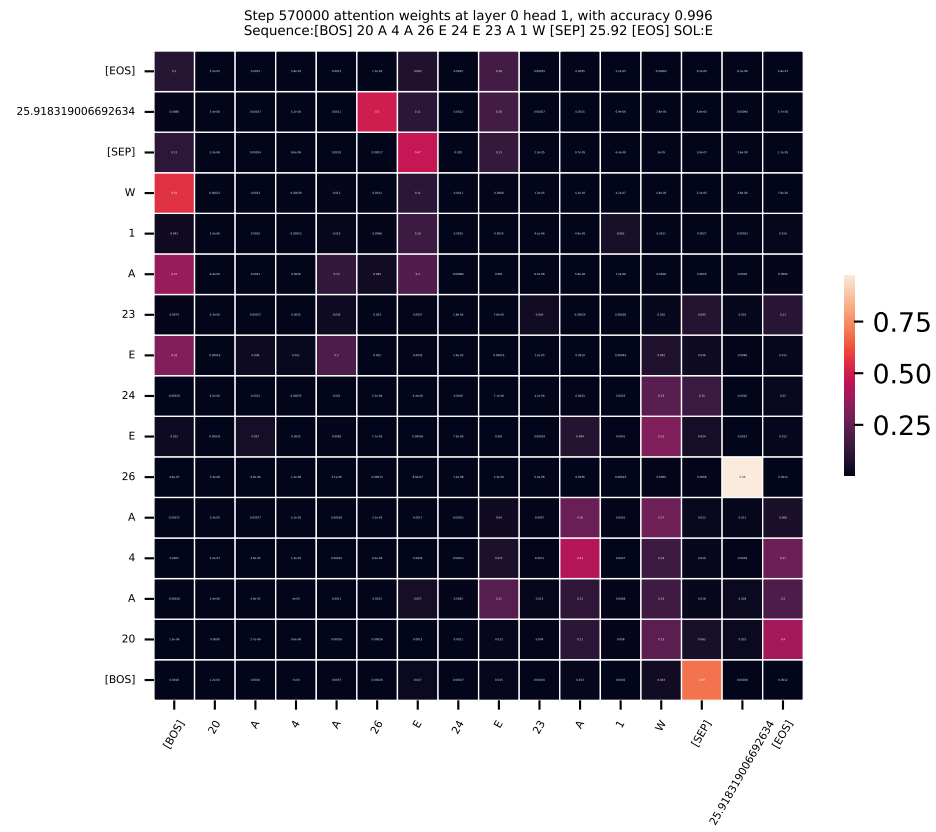

Step 569500 attention weights at layer 1 head 0, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

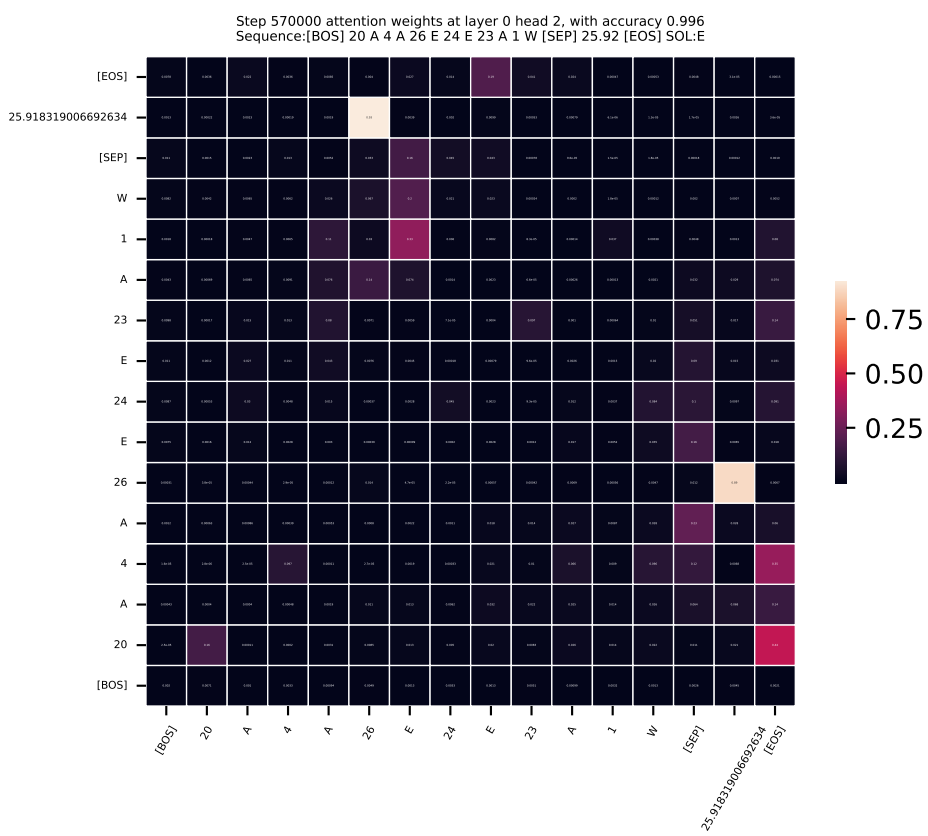

Step 569500 attention weights at layer 1 head 1, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

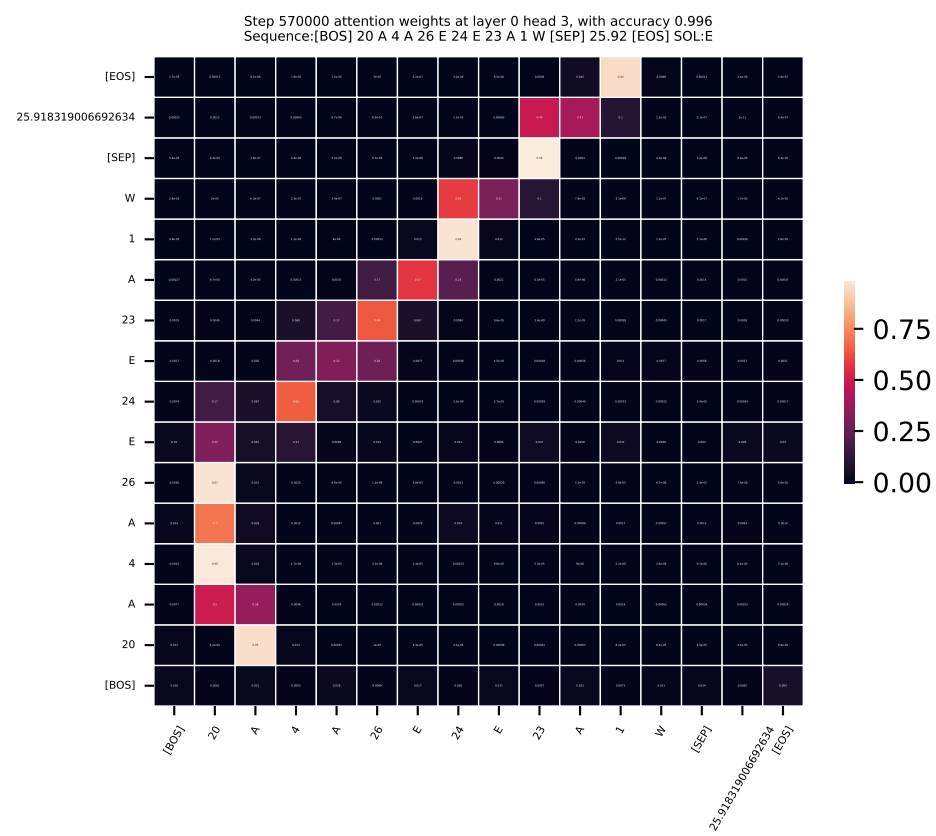

Step 569500 attention weights at layer 1 head 2, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

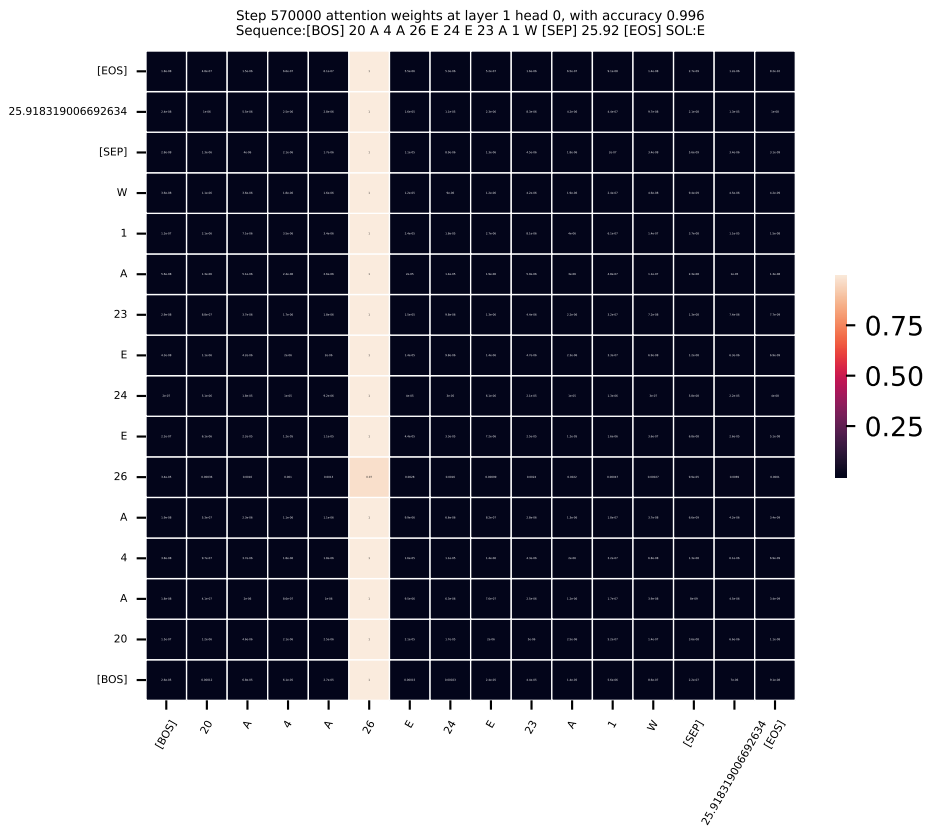

Step 569500 attention weights at layer 1 head 3, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

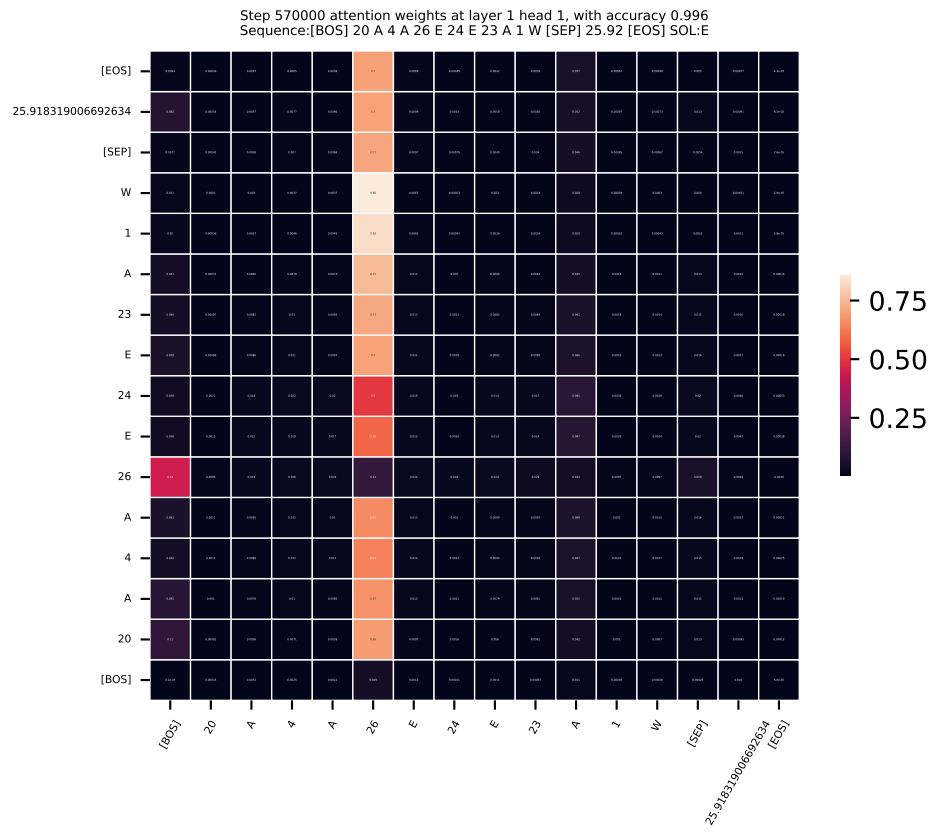

Step 570000 attention weights at layer 0 head 0, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

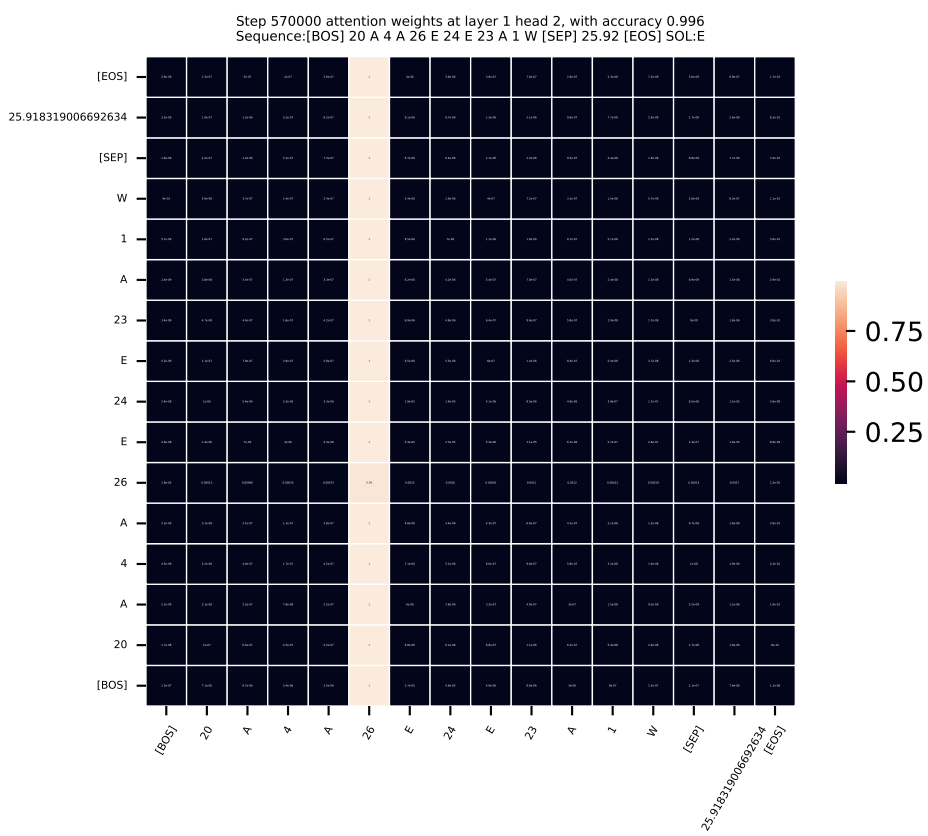

Step 570000 attention weights at layer 0 head 1, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

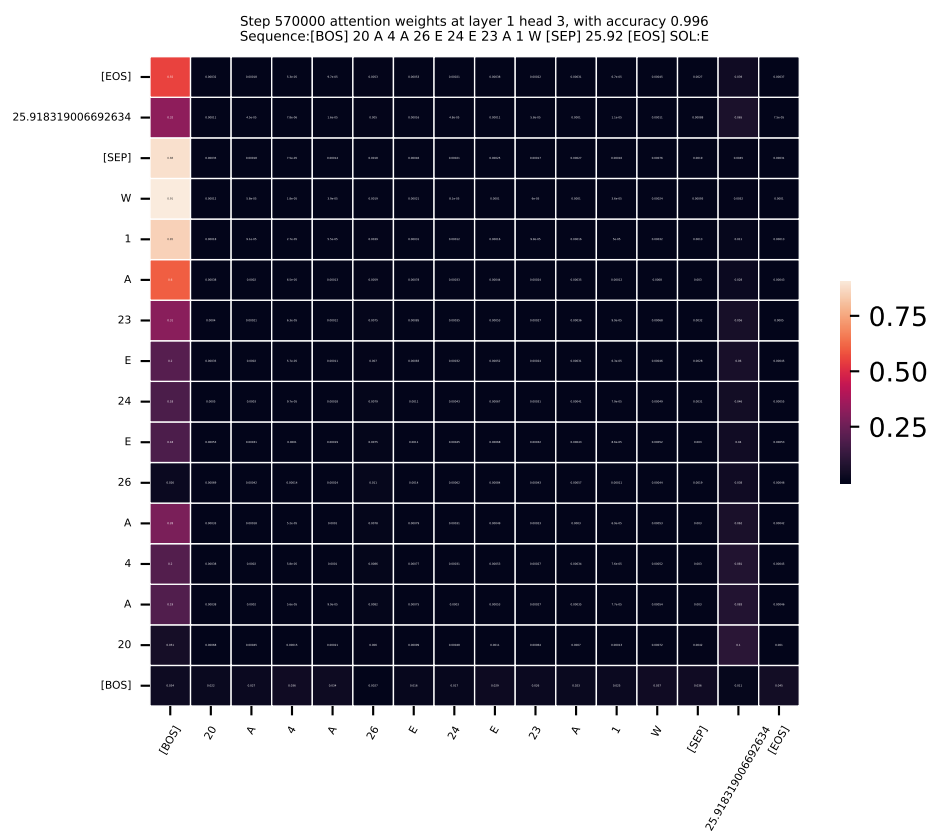

Step 570000 attention weights at layer 0 head 2, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

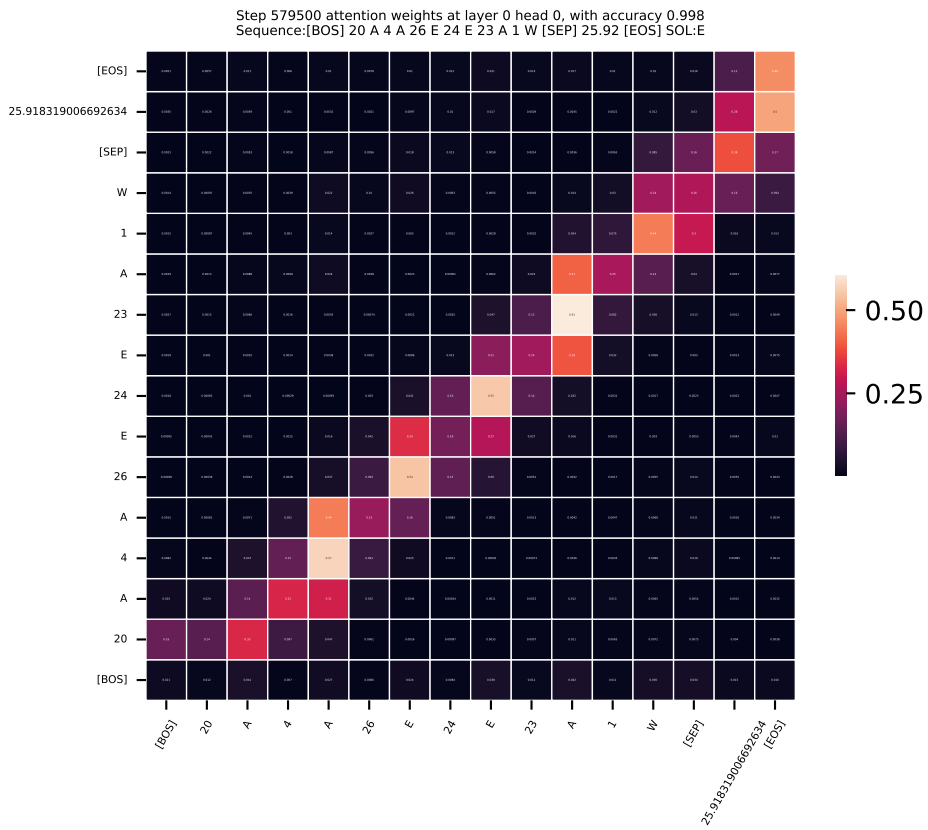

Step 570000 attention weights at layer 0 head 3, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

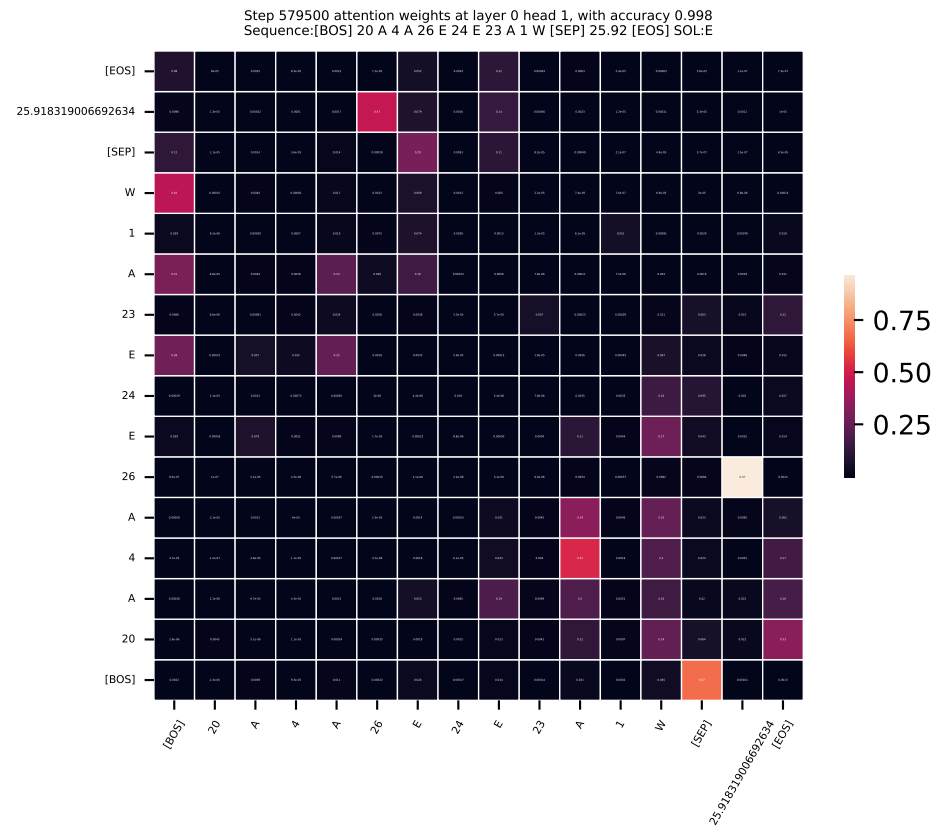

Step 570000 attention weights at layer 1 head 0, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

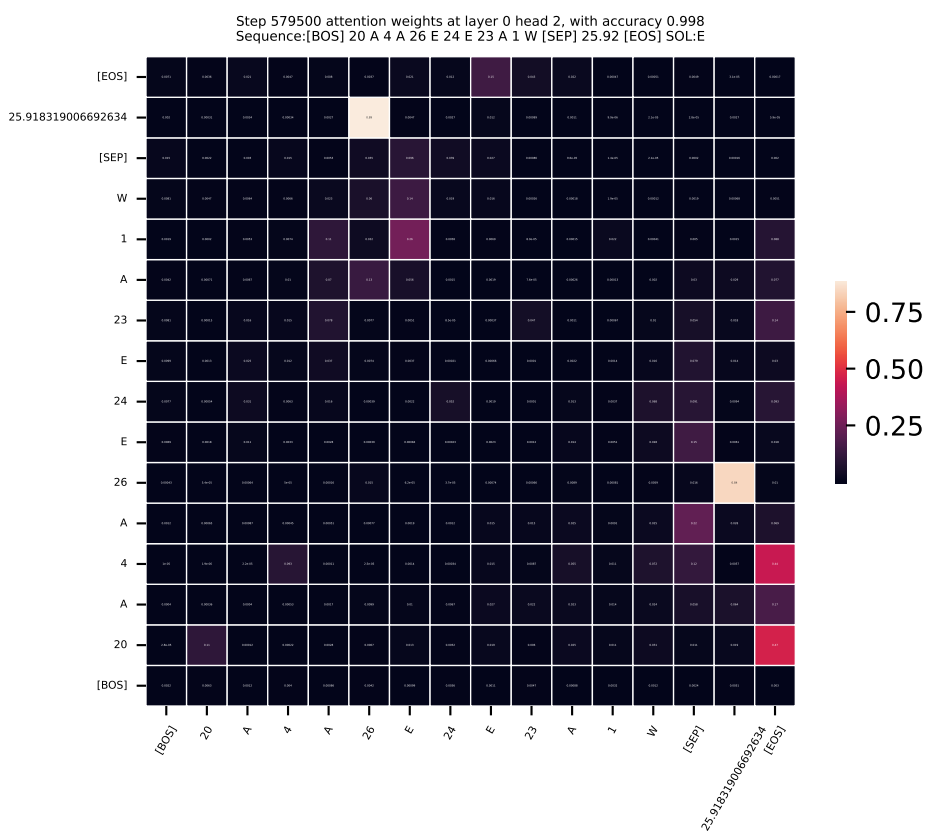

Step 570000 attention weights at layer 1 head 1, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

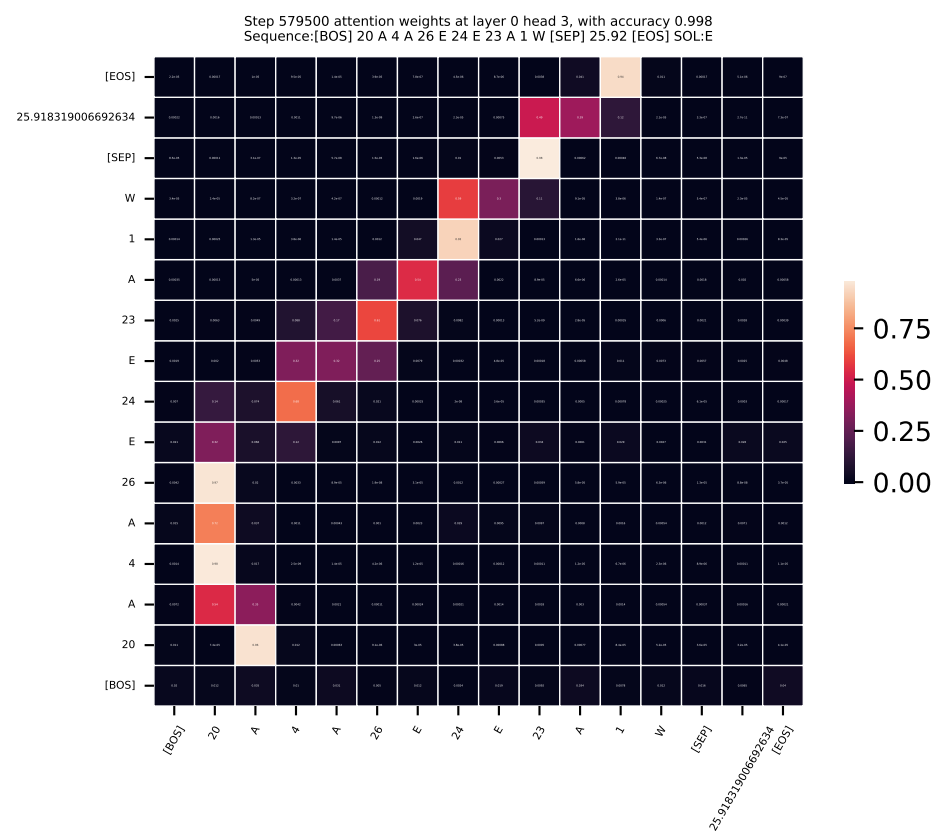

Step 570000 attention weights at layer 1 head 2, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

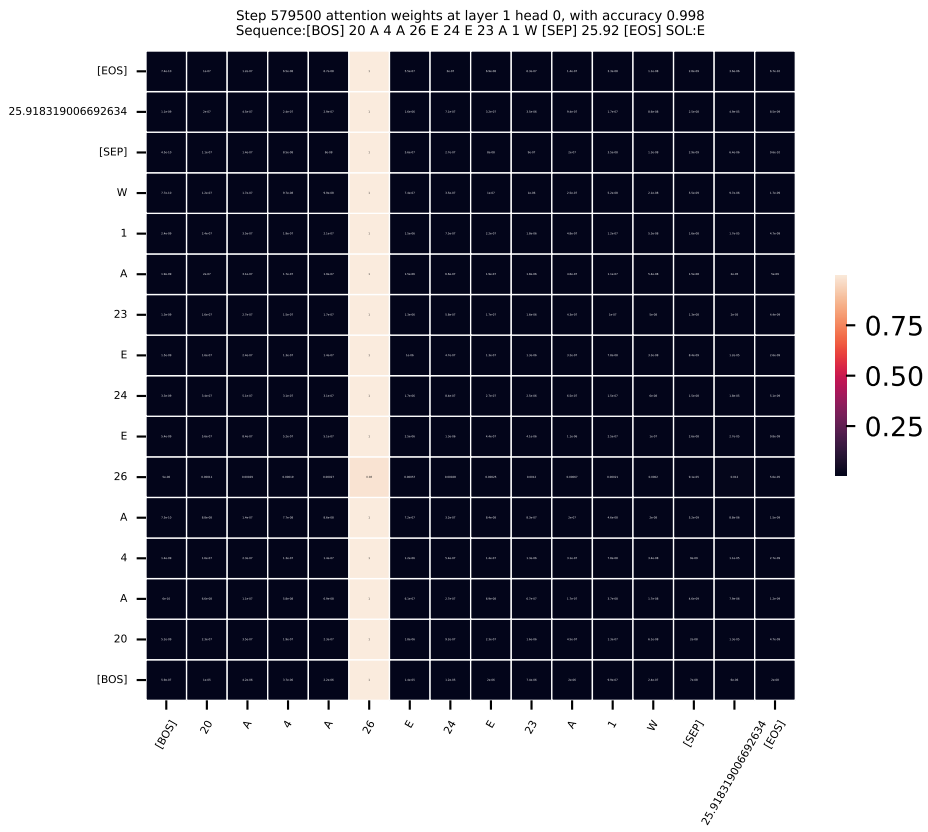

Step 570000 attention weights at layer 1 head 3, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

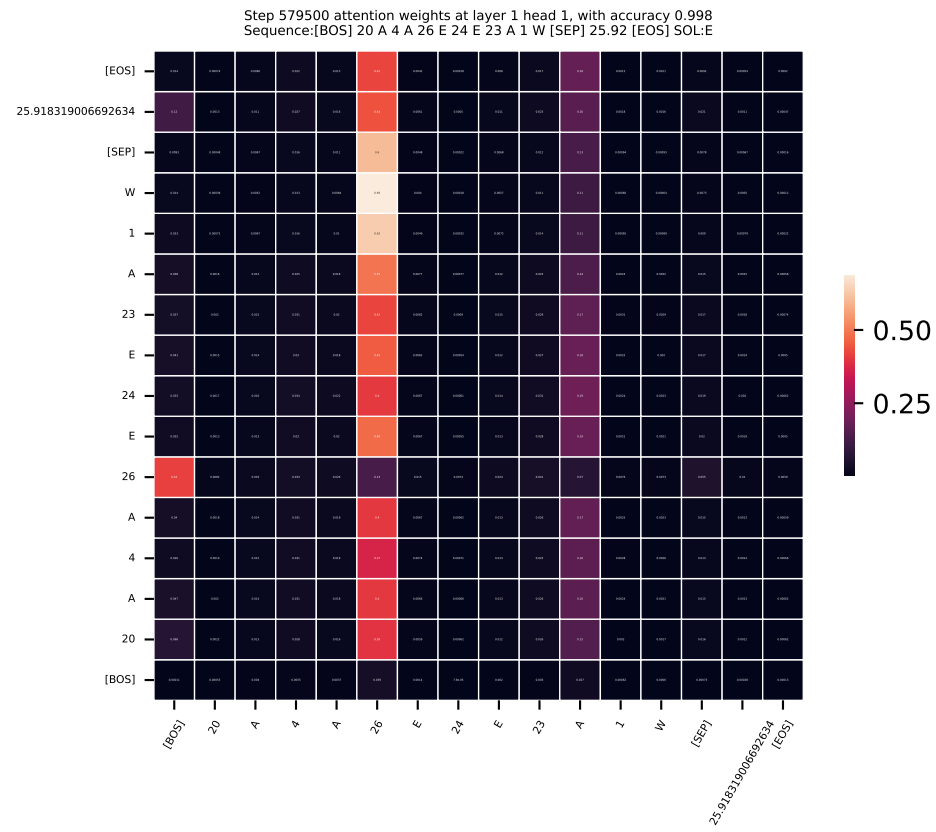

Step 579500 attention weights at layer 0 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

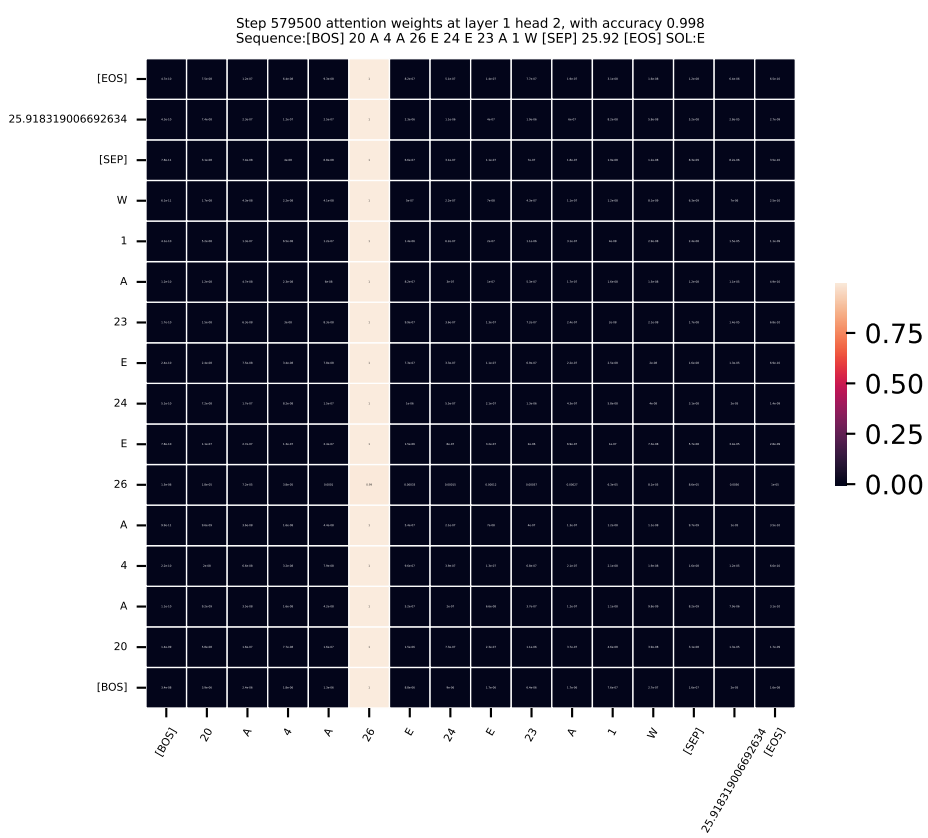

Step 579500 attention weights at layer 0 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

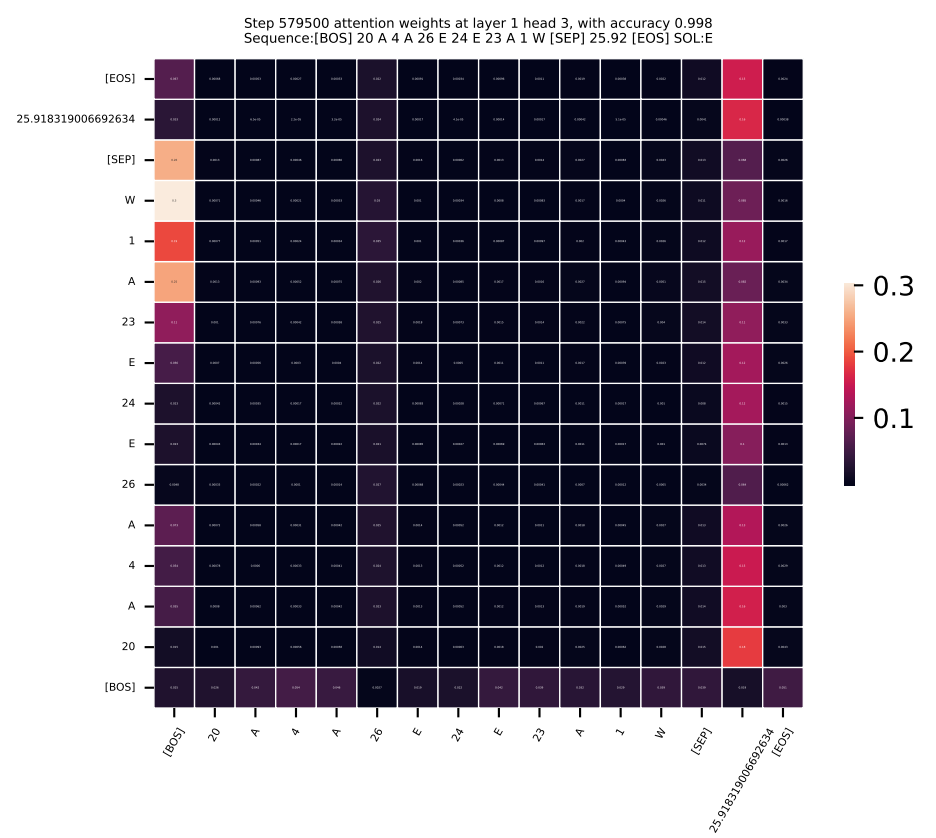

Step 579500 attention weights at layer 0 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

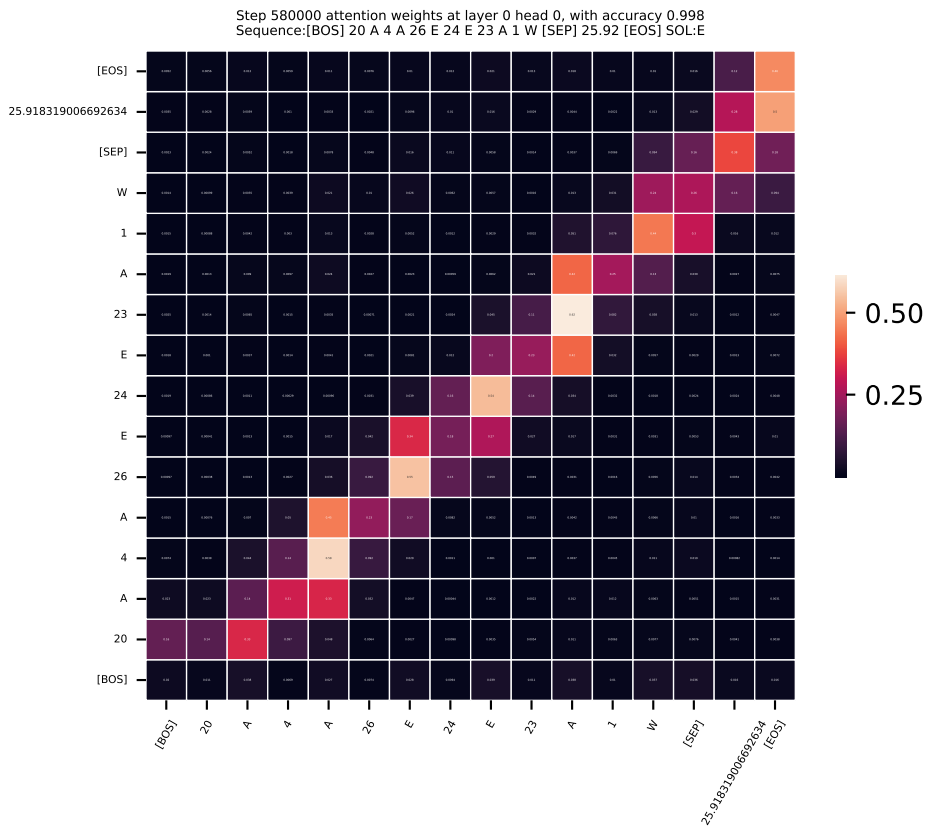

Step 579500 attention weights at layer 0 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

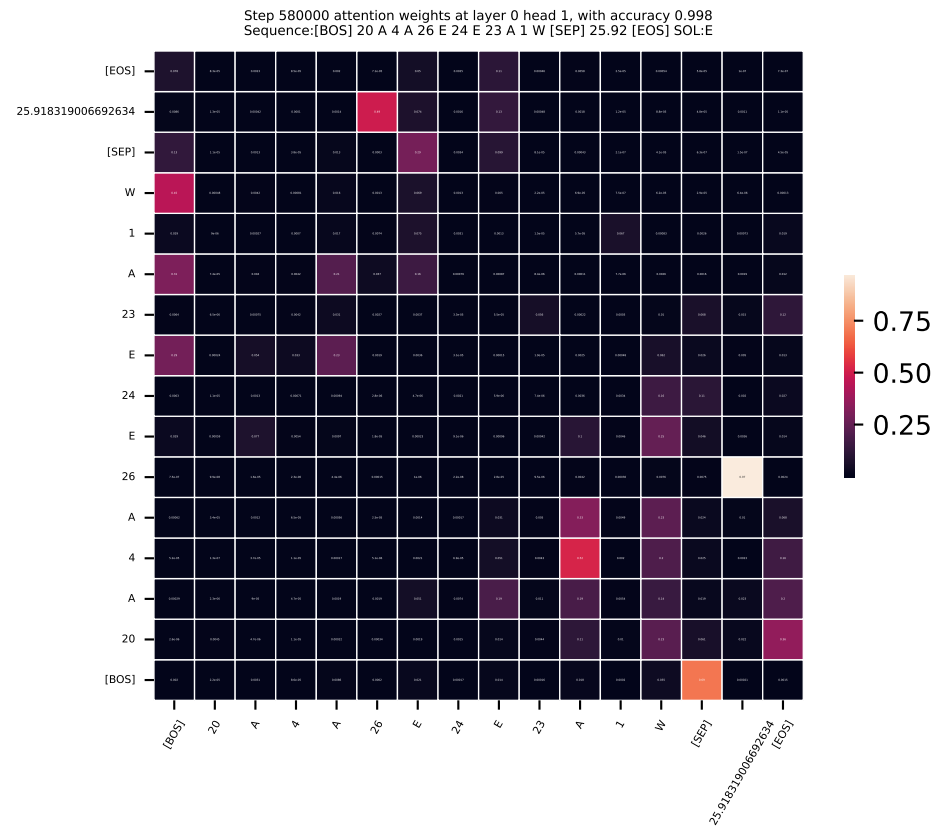

Step 579500 attention weights at layer 1 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

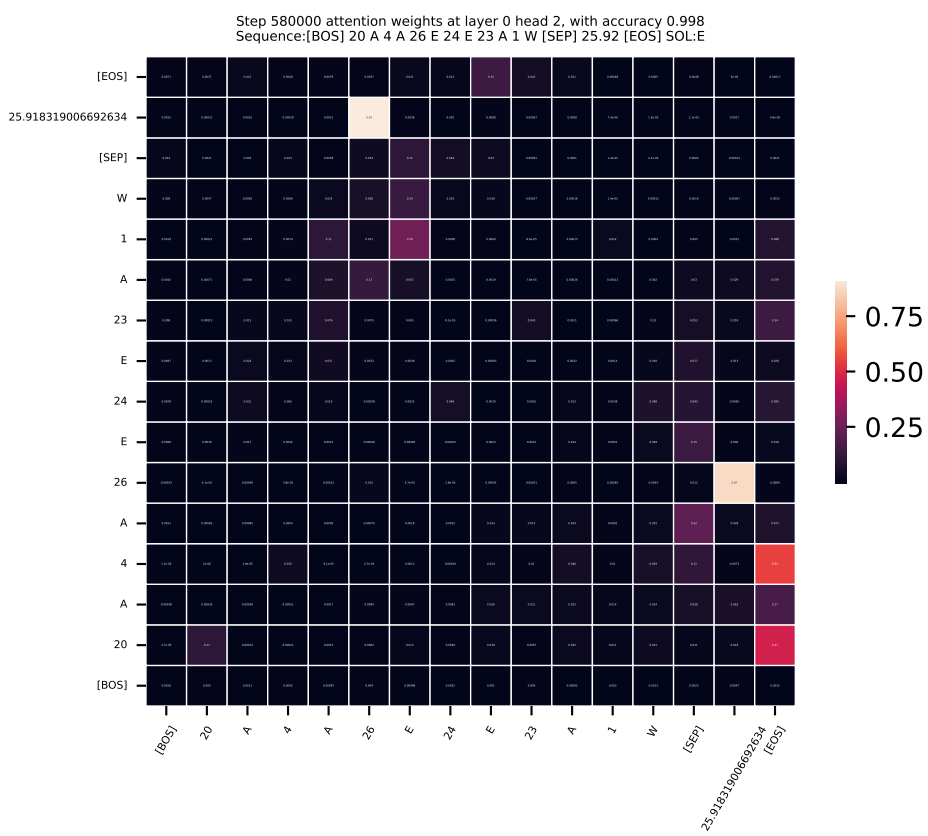

Step 579500 attention weights at layer 1 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

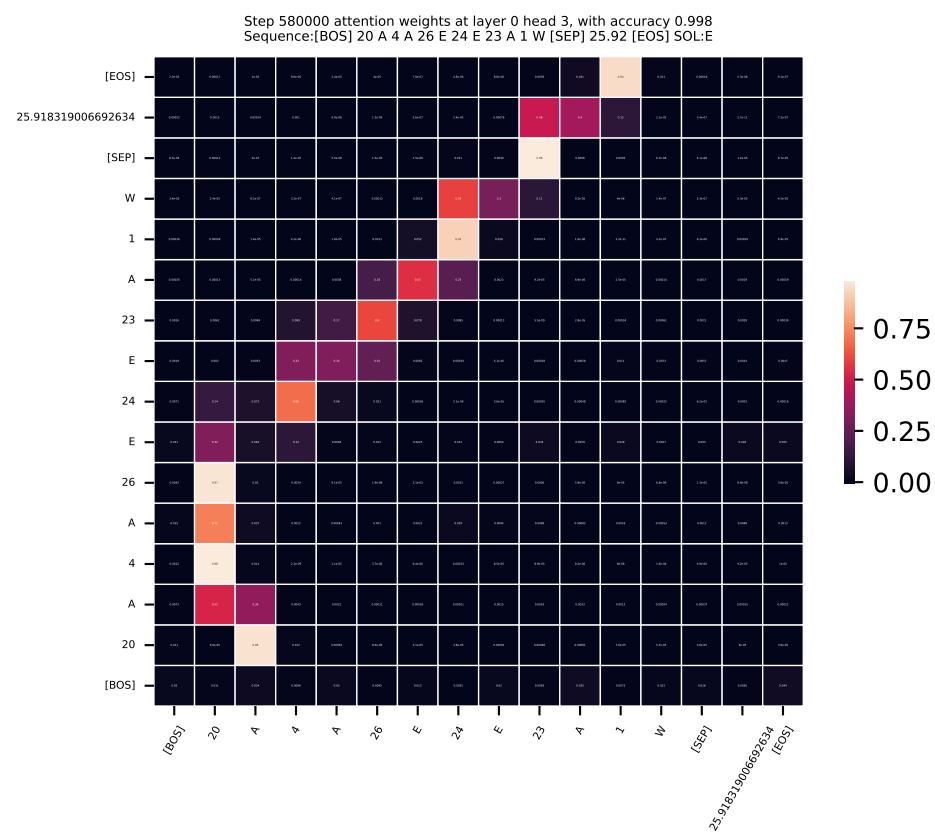

Step 579500 attention weights at layer 1 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

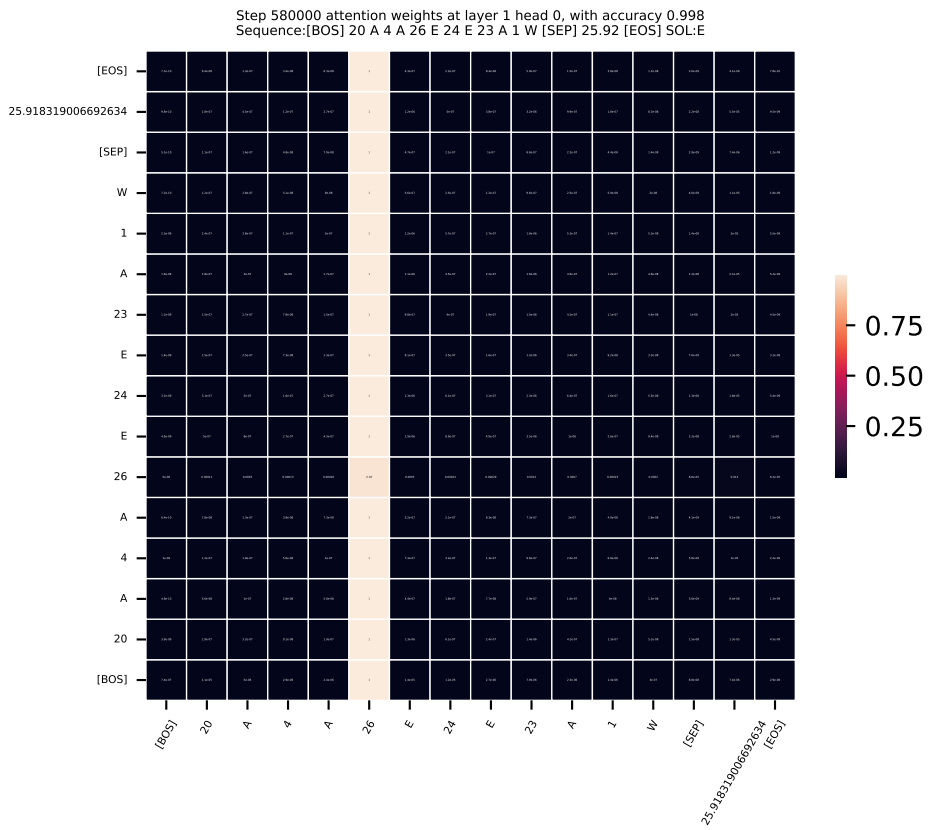

Step 579500 attention weights at layer 1 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

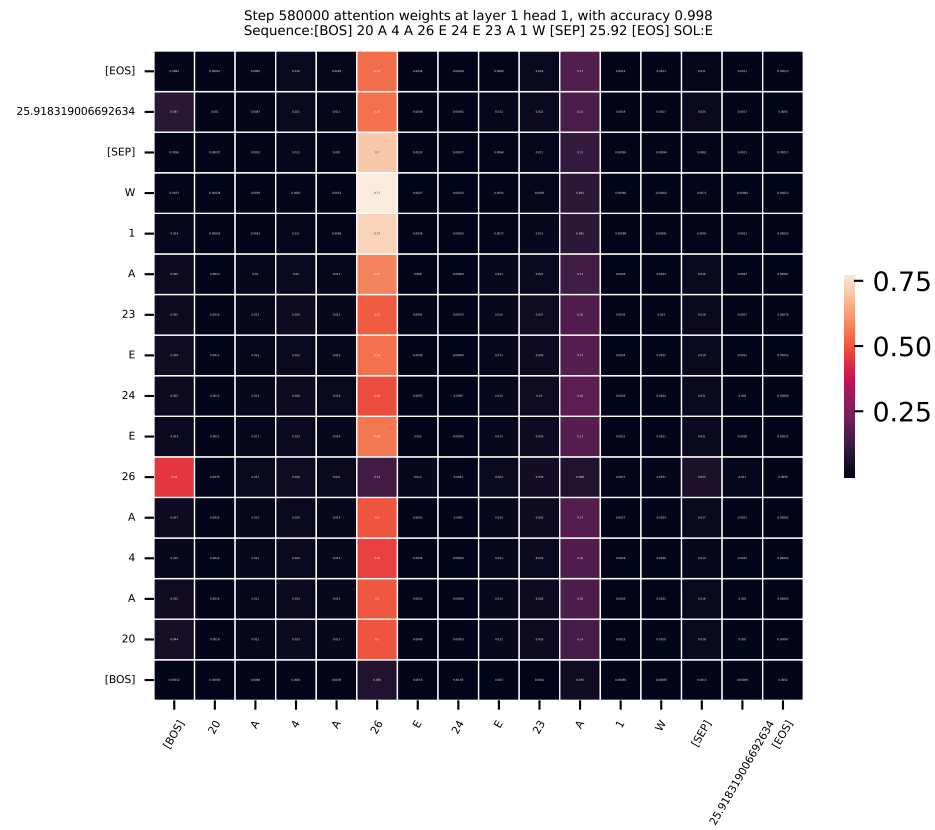

Step 580000 attention weights at layer 0 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.50
0.25

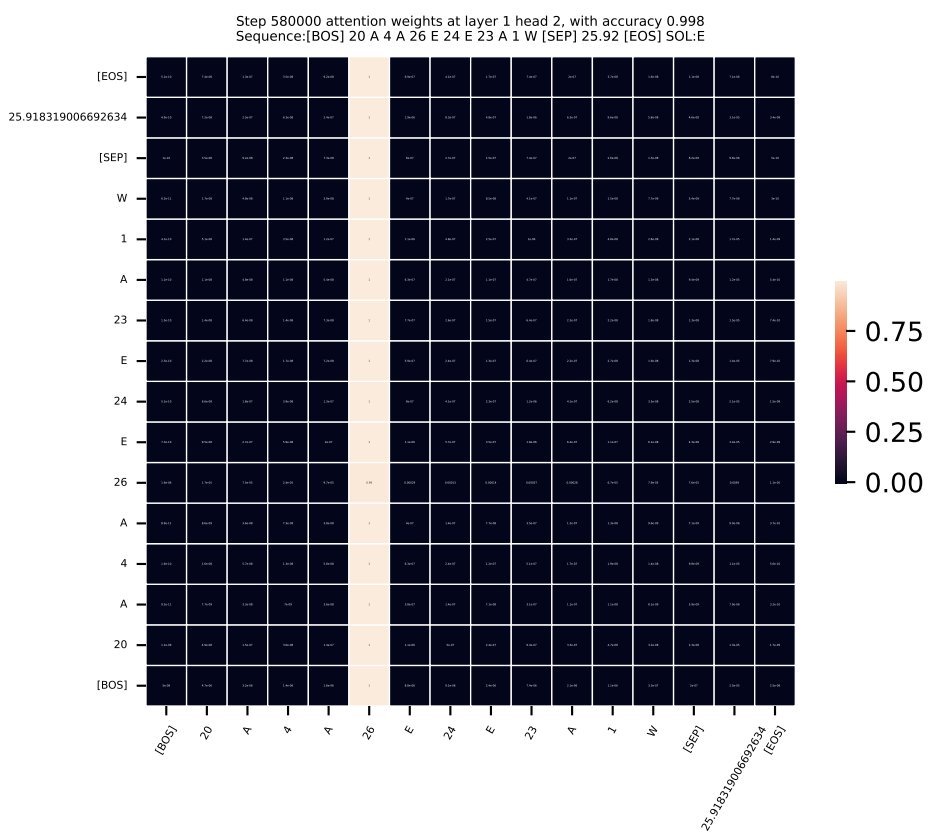

Step 580000 attention weights at layer 0 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

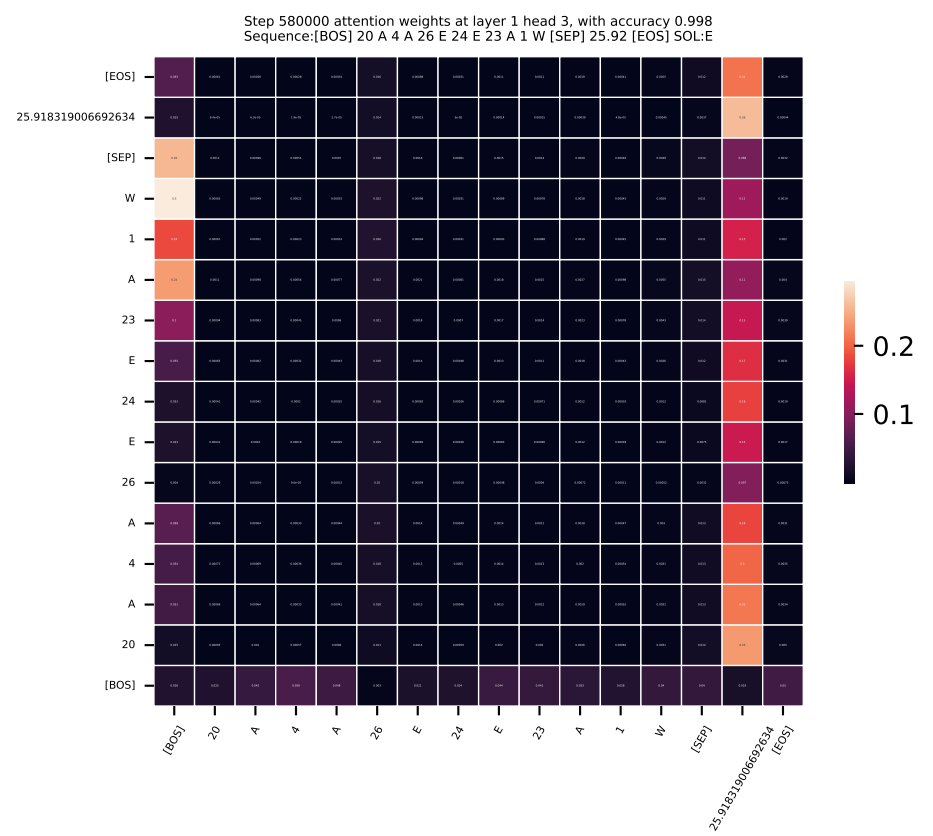

Step 580000 attention weights at layer 0 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

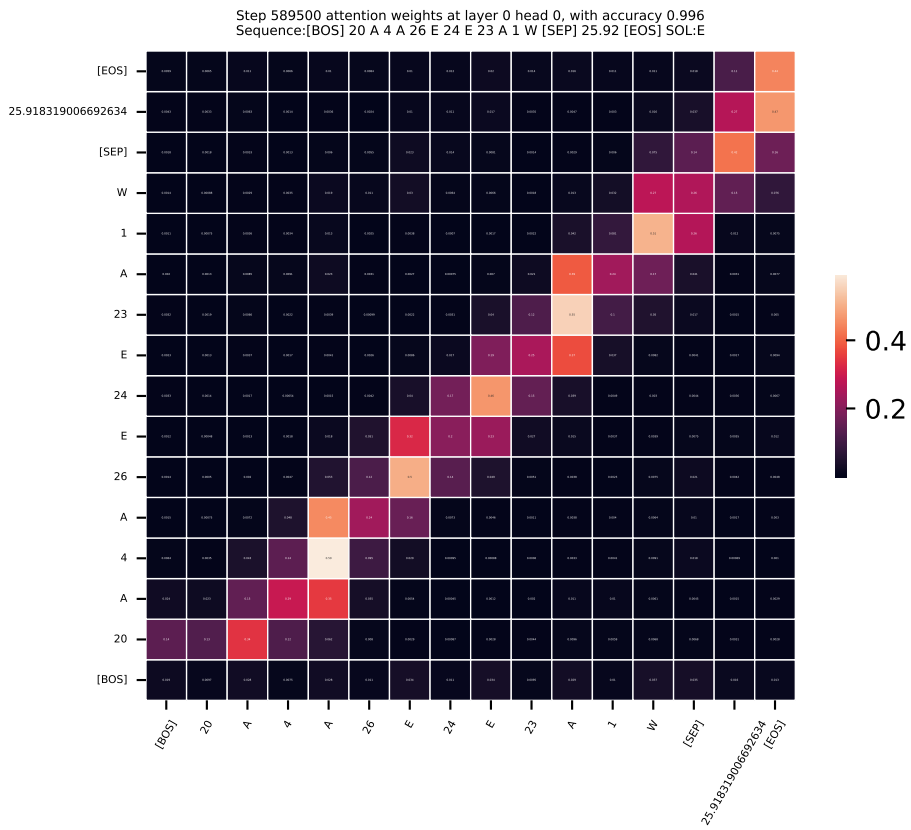

Step 580000 attention weights at layer 0 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

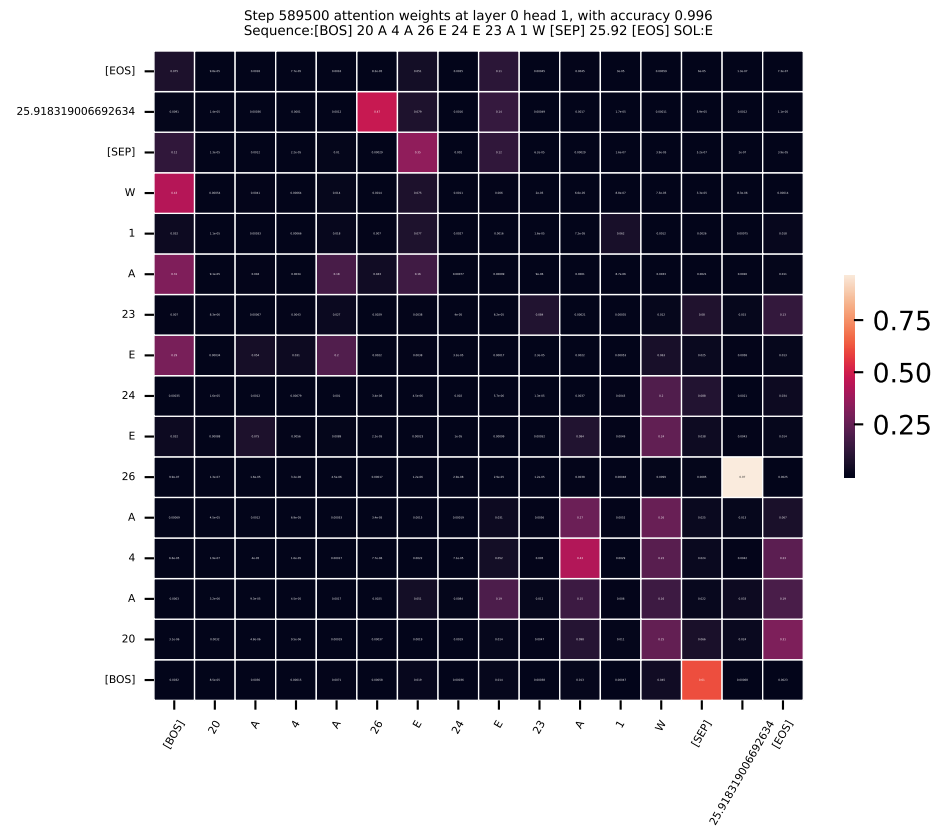

Step 580000 attention weights at layer 1 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

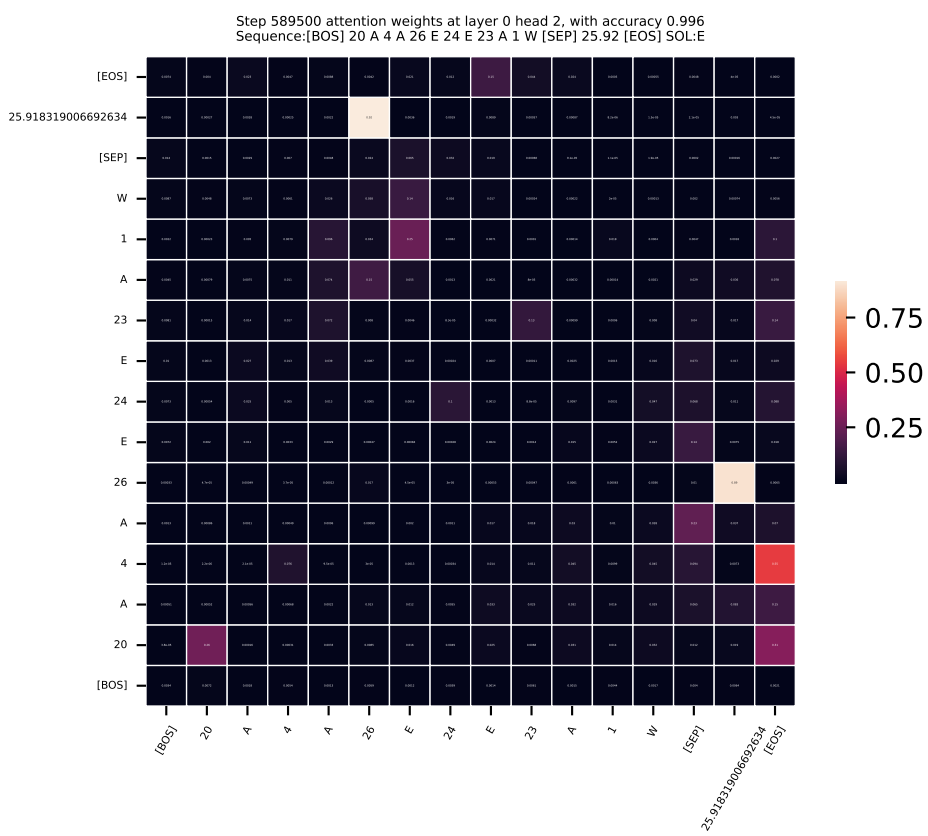

Step 580000 attention weights at layer 1 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

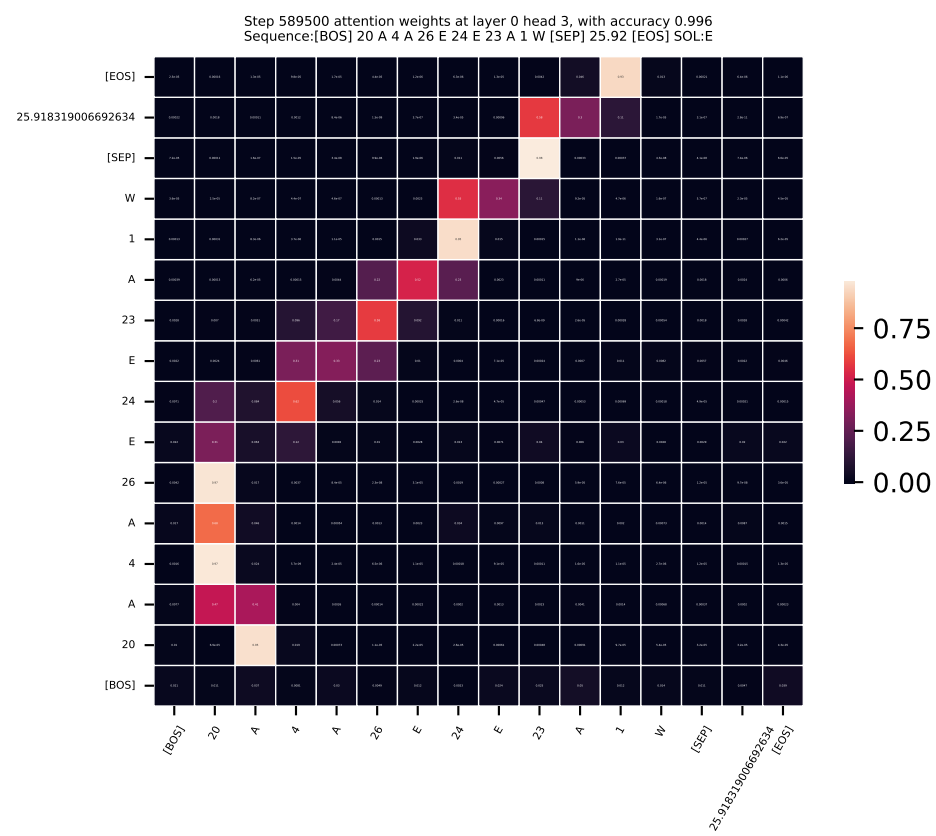

Step 580000 attention weights at layer 1 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

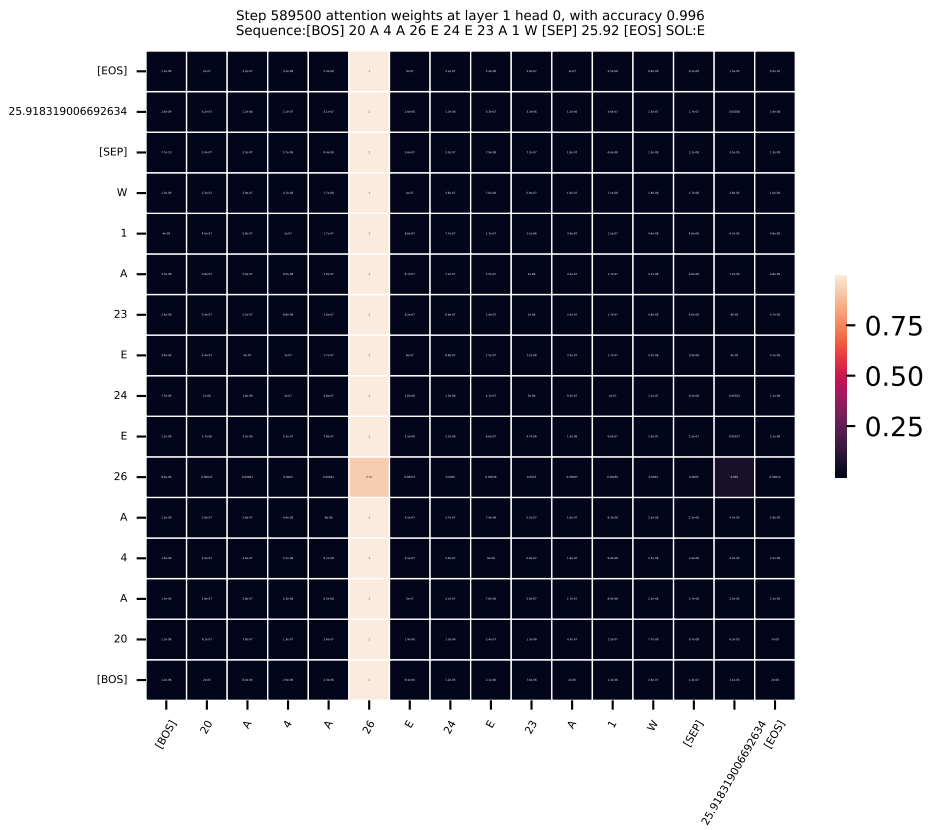

Step 580000 attention weights at layer 1 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.2
0.1

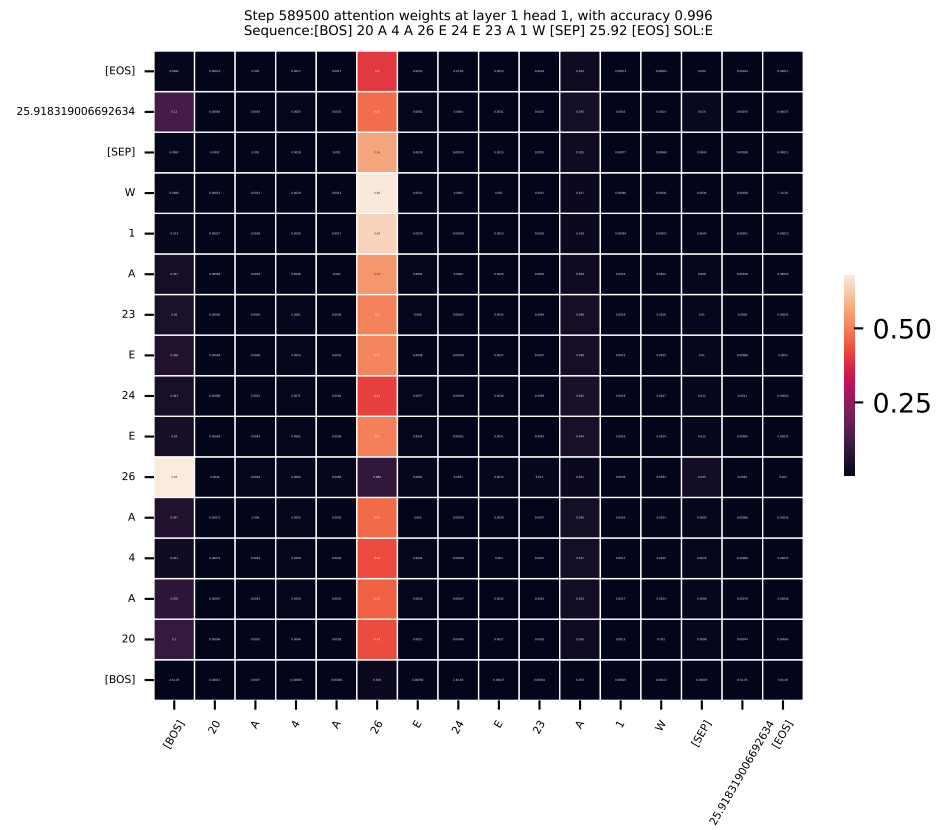

Step 589500 attention weights at layer 0 head 0, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

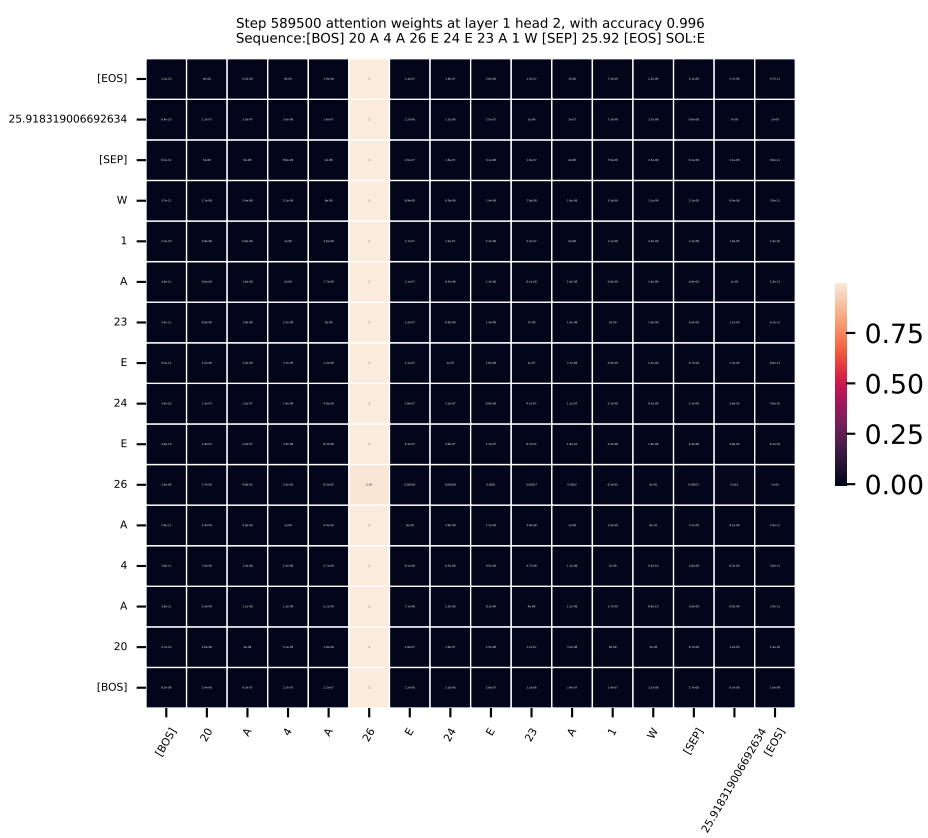

Step 589500 attention weights at layer 0 head 1, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

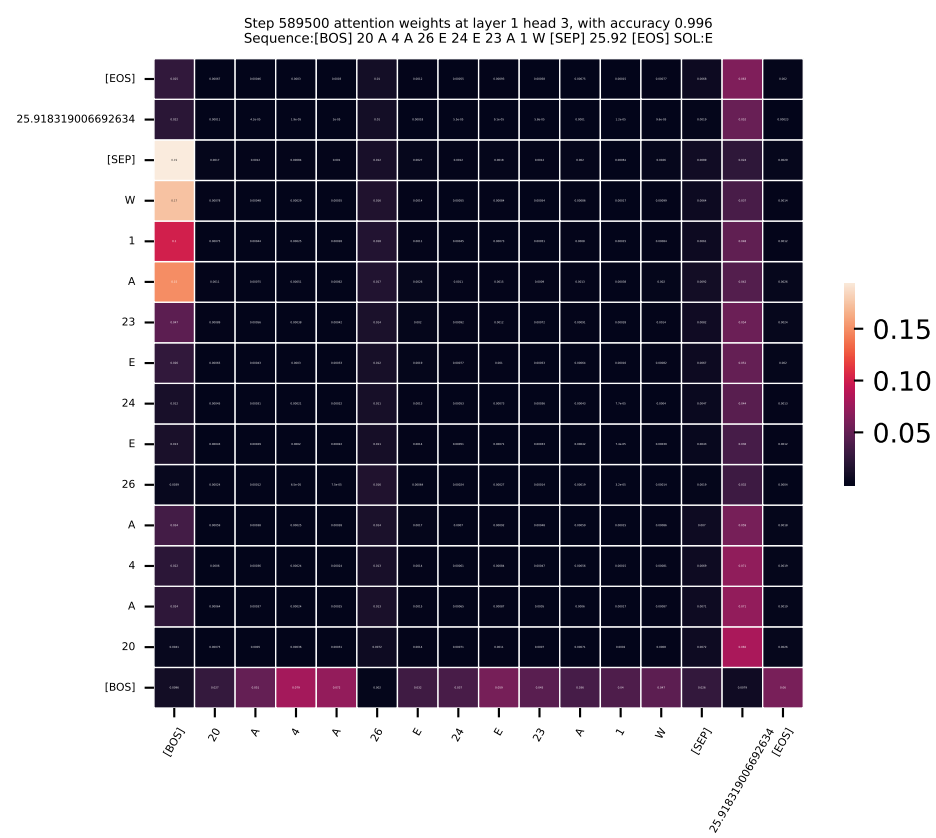

Step 589500 attention weights at layer 0 head 2, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

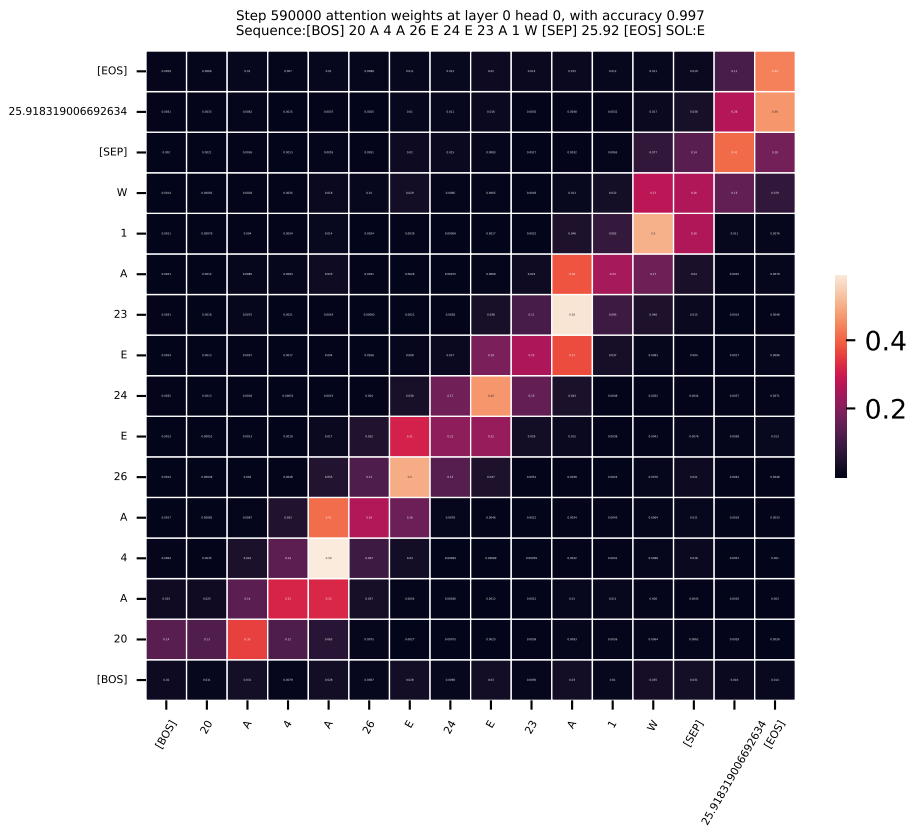

Step 589500 attention weights at layer 0 head 3, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

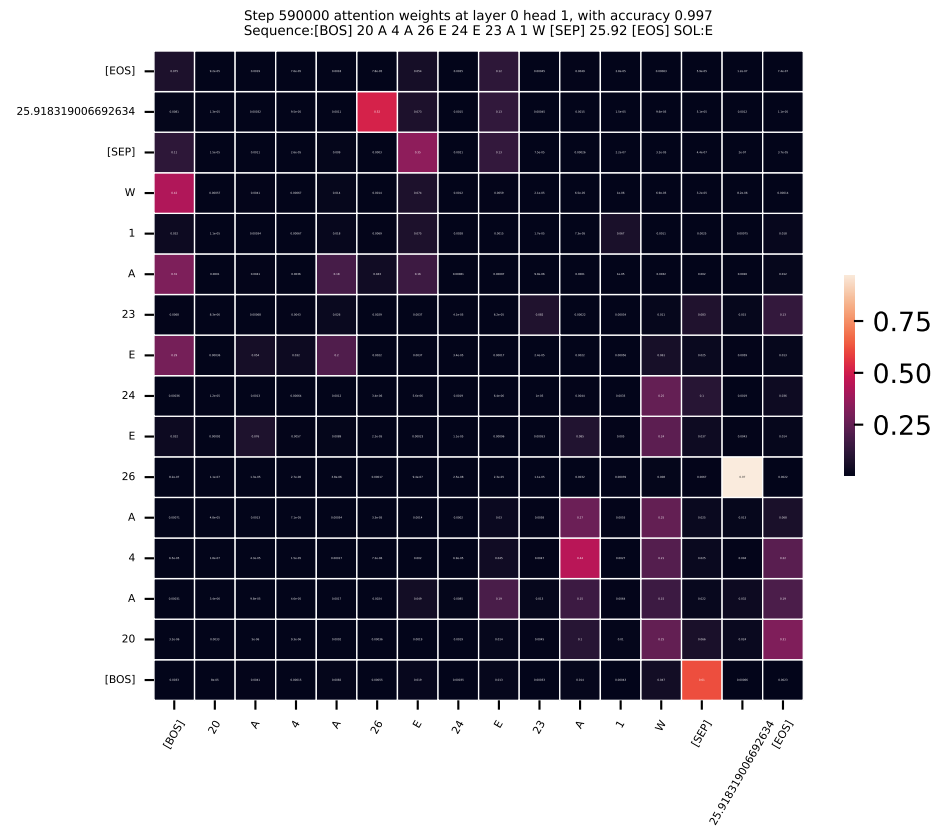

Step 589500 attention weights at layer 1 head 0, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

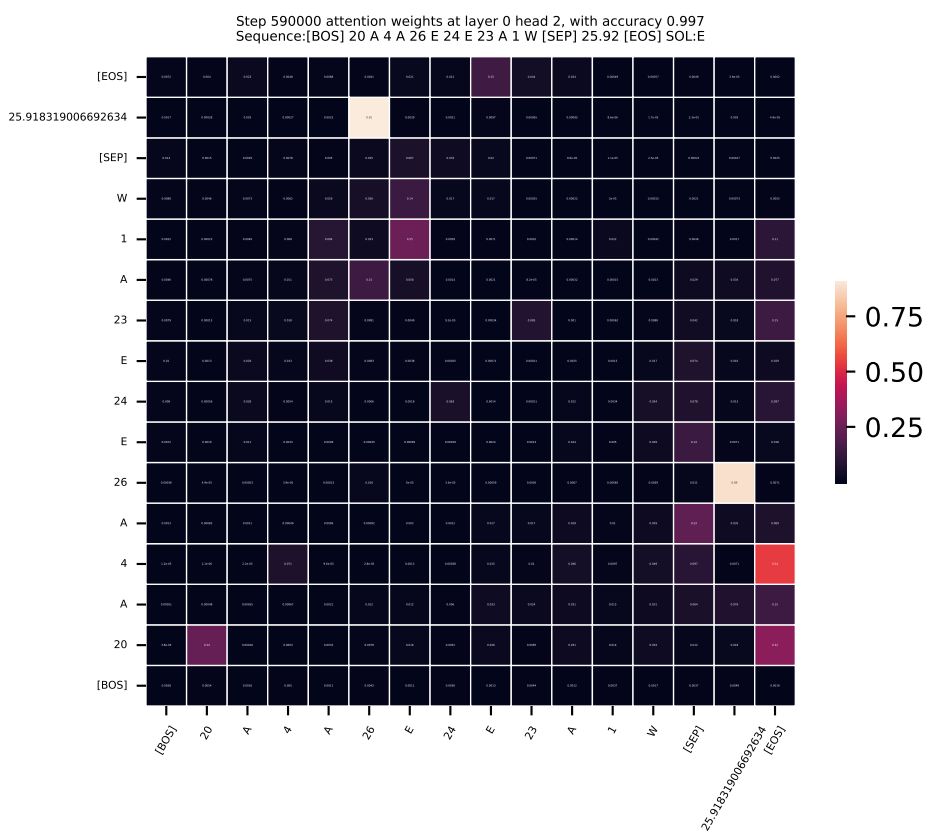

Step 589500 attention weights at layer 1 head 1, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

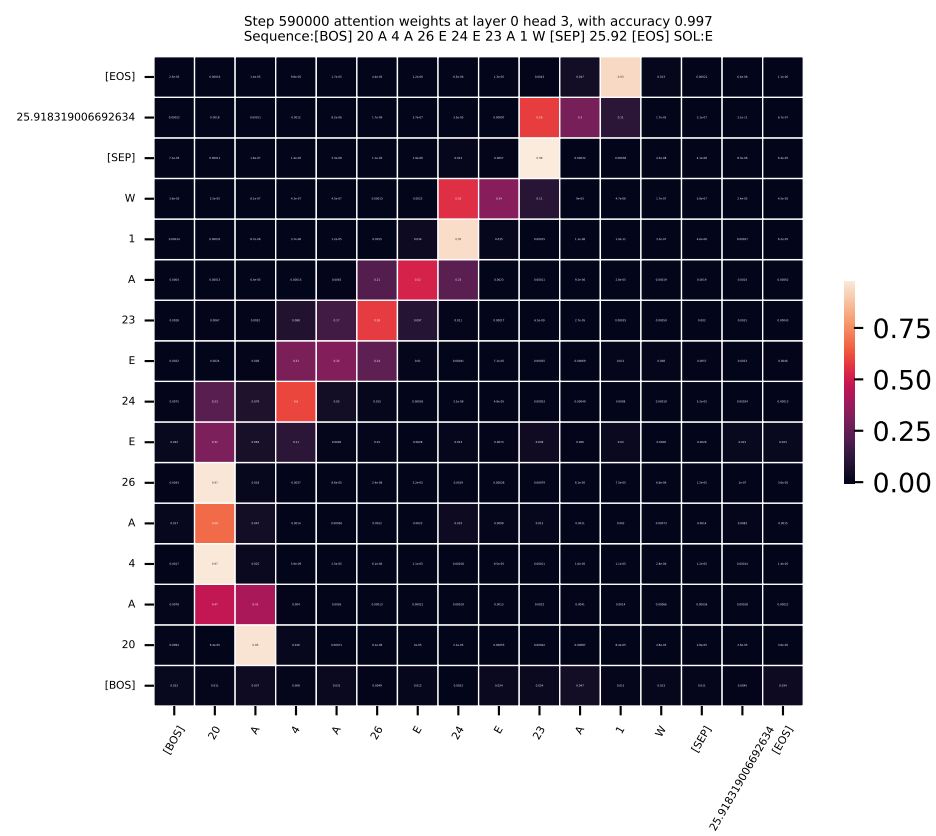

Step 589500 attention weights at layer 1 head 2, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

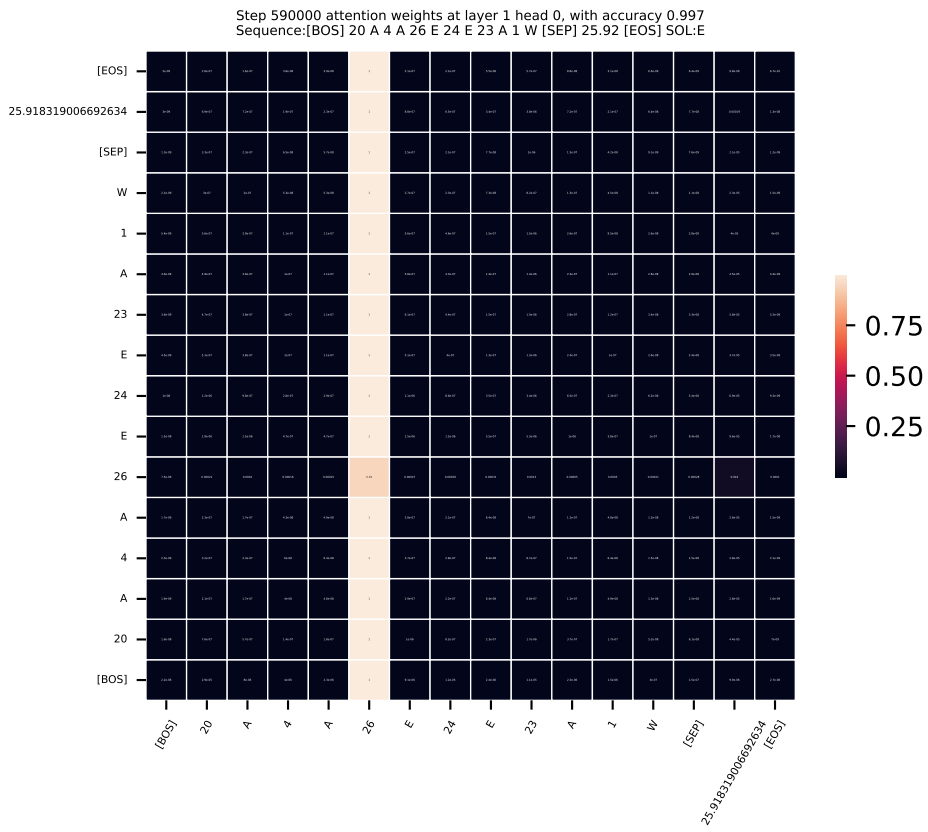

Step 589500 attention weights at layer 1 head 3, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.15
0.10
0.05

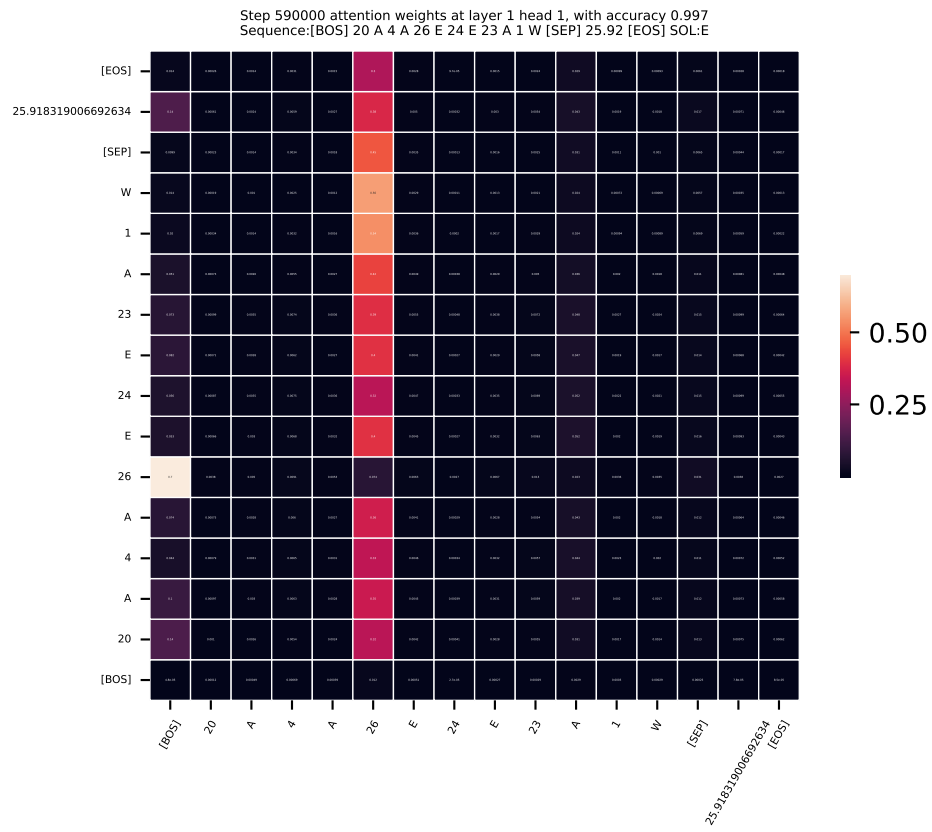

Step 590000 attention weights at layer 0 head 0, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

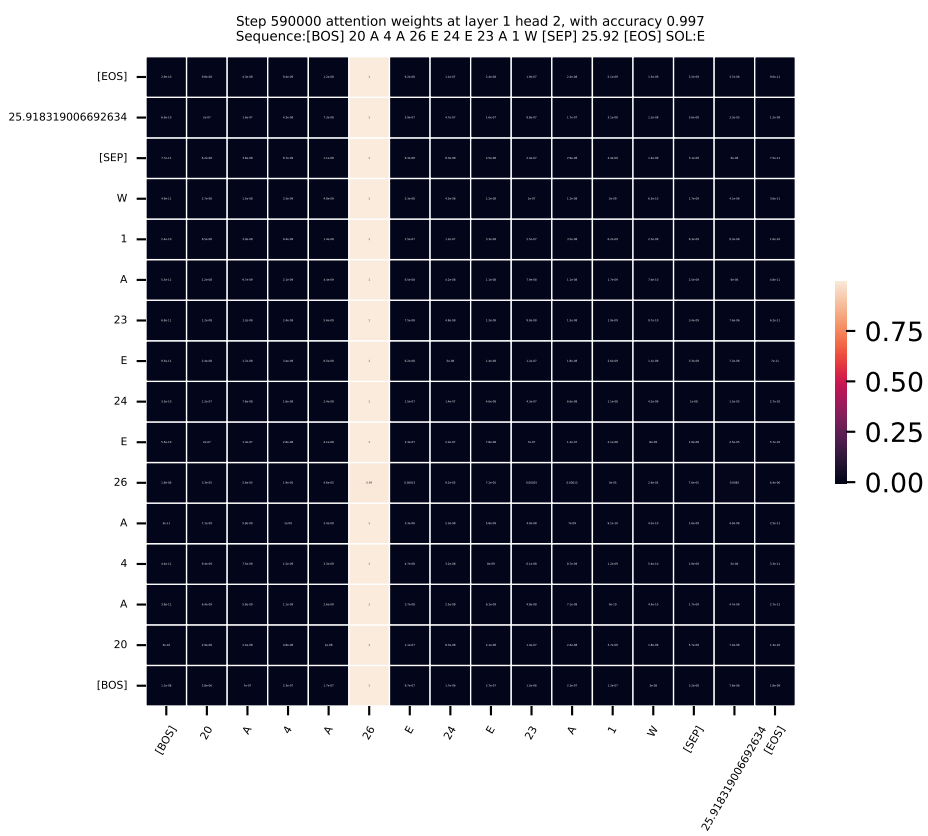

Step 590000 attention weights at layer 0 head 1, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

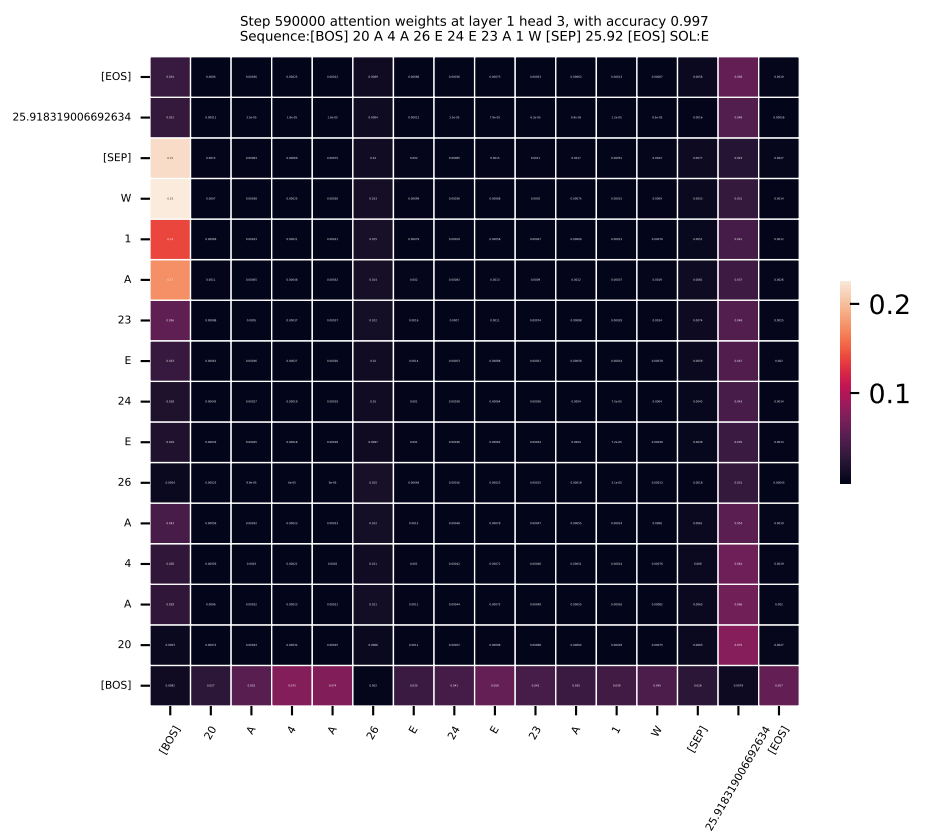

Step 590000 attention weights at layer 0 head 2, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

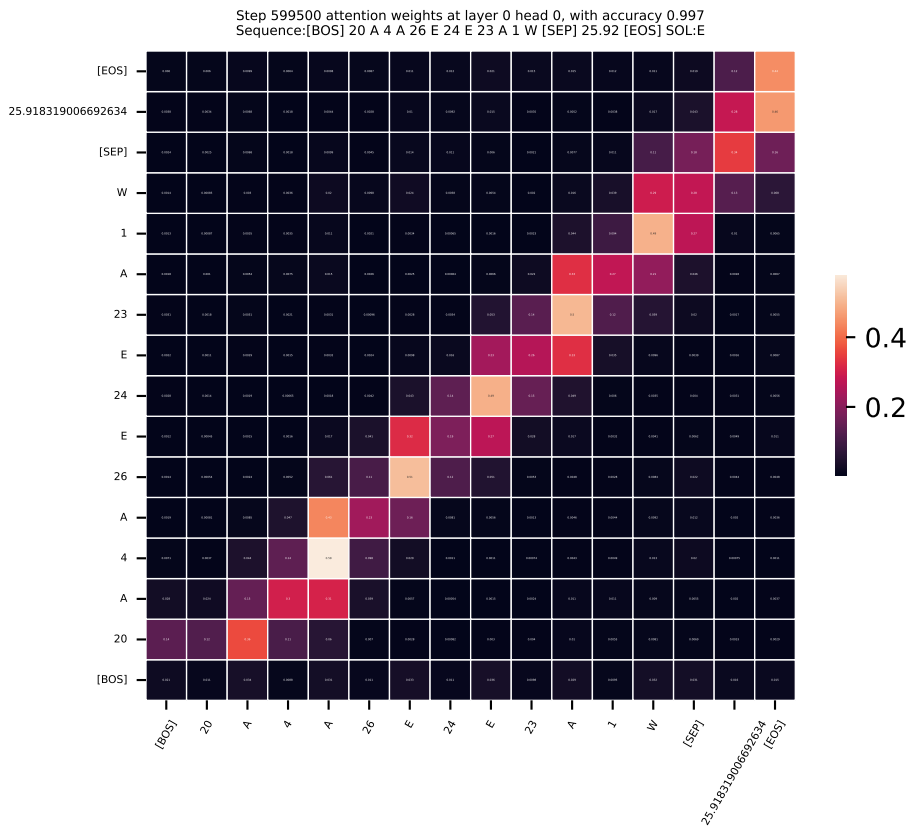

Step 590000 attention weights at layer 0 head 3, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

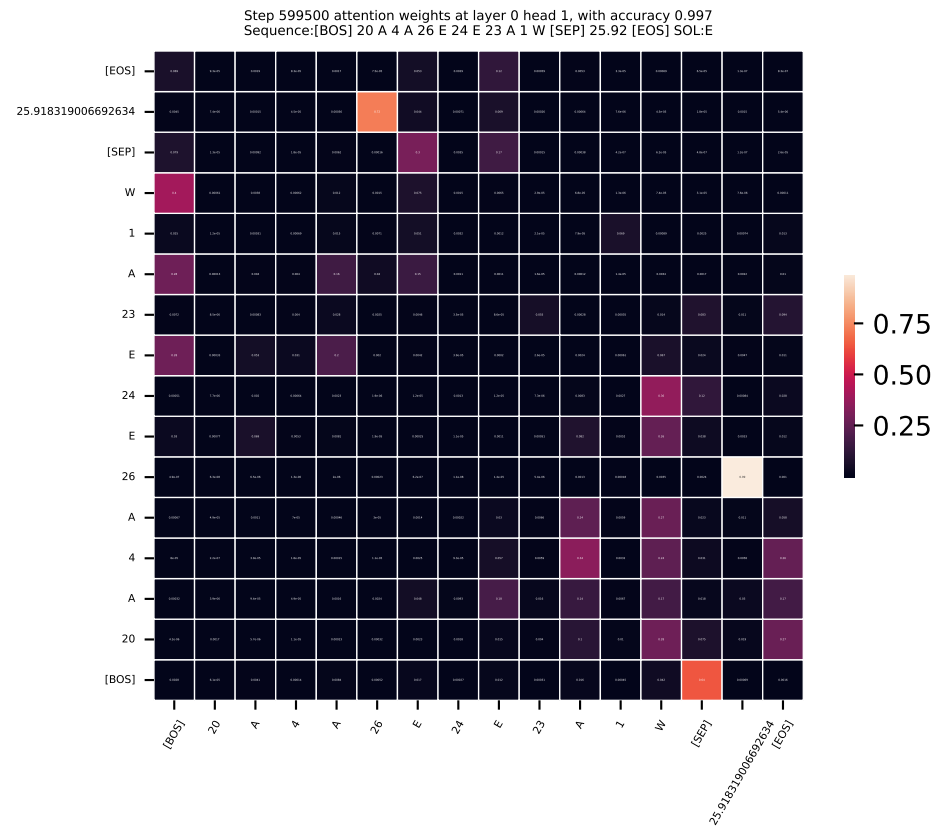

Step 590000 attention weights at layer 1 head 0, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

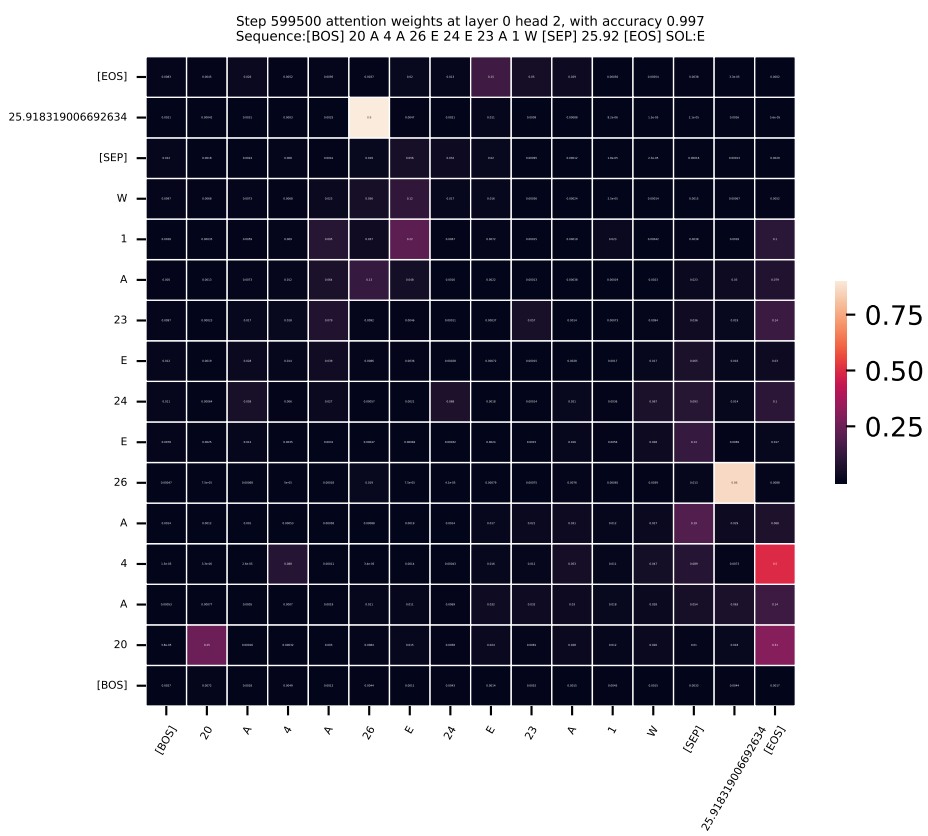

Step 590000 attention weights at layer 1 head 1, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

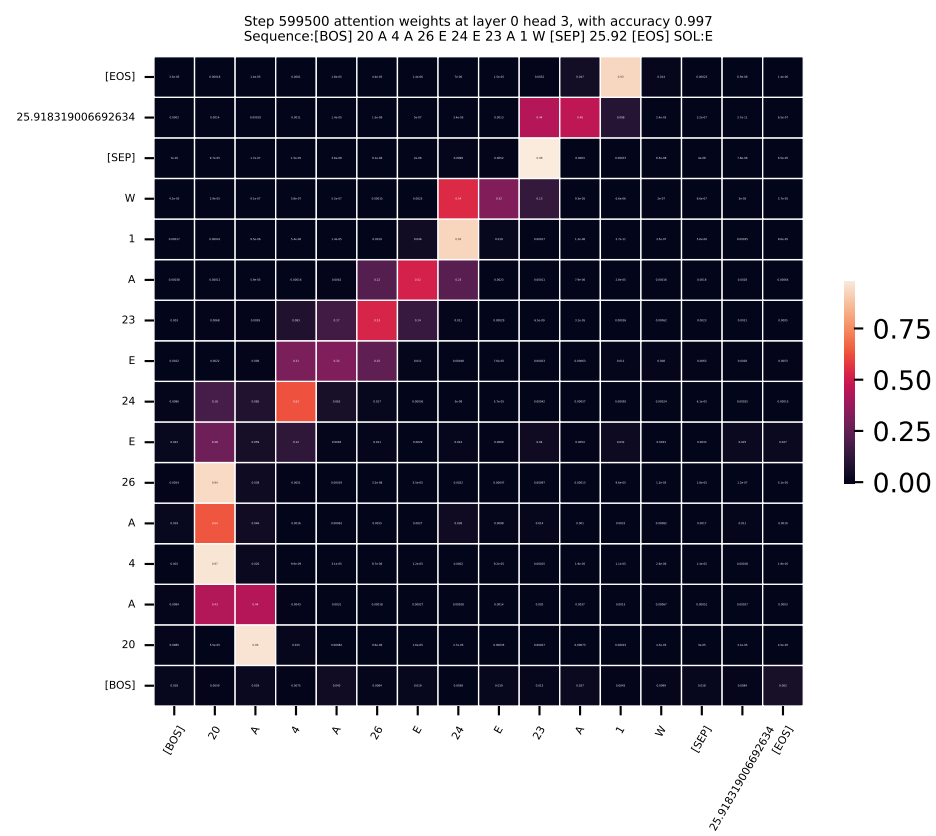

Step 590000 attention weights at layer 1 head 2, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

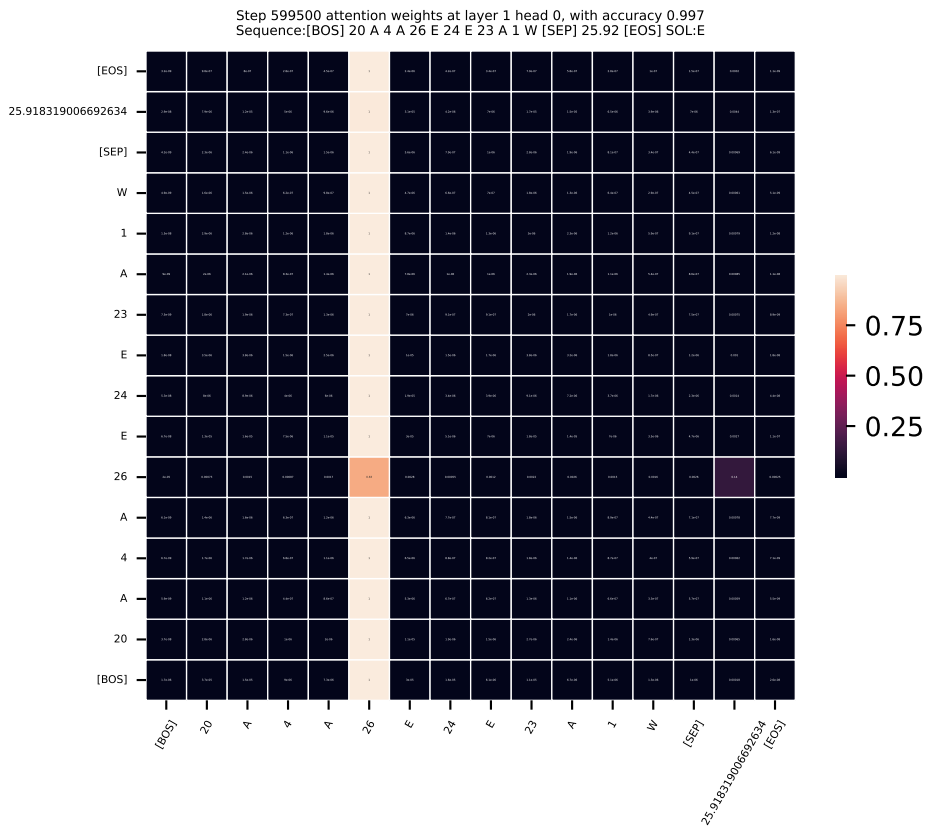

Step 590000 attention weights at layer 1 head 3, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.2
0.1

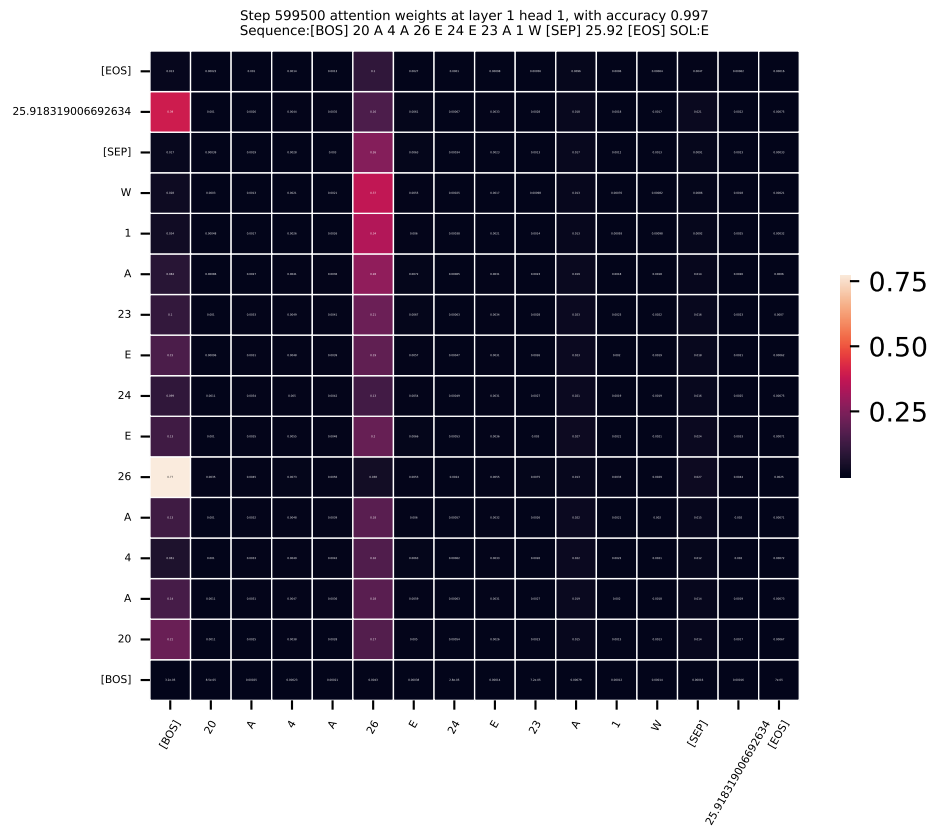

Step 599500 attention weights at layer 0 head 0, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

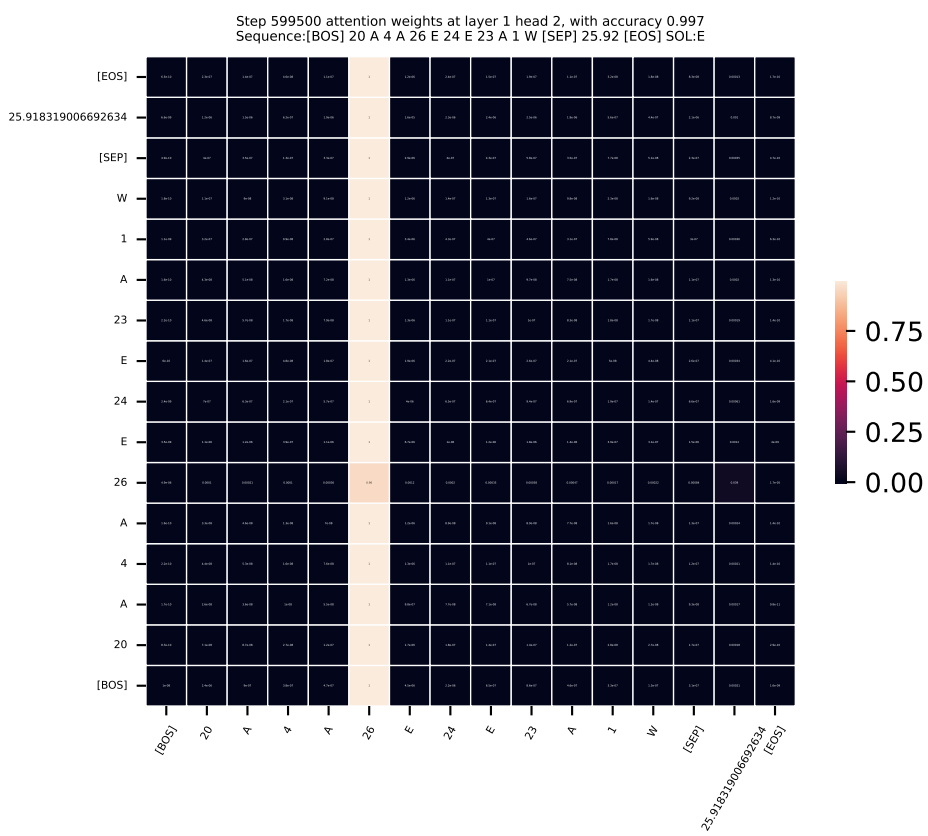

Step 599500 attention weights at layer 0 head 1, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

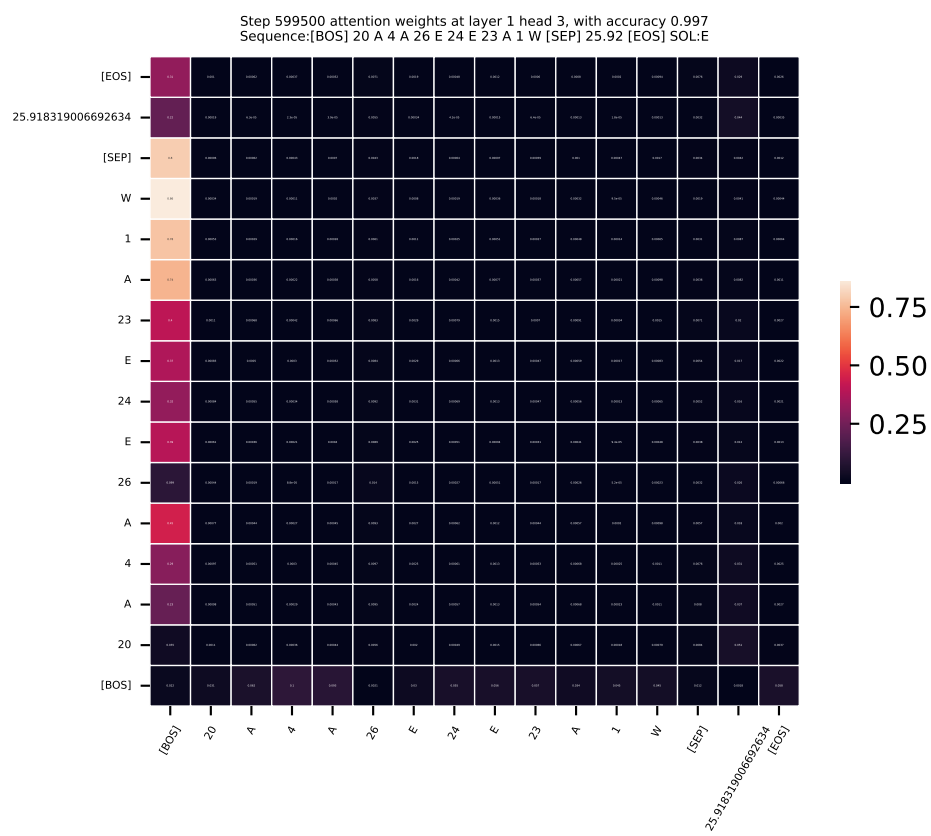

Step 599500 attention weights at layer 0 head 2, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

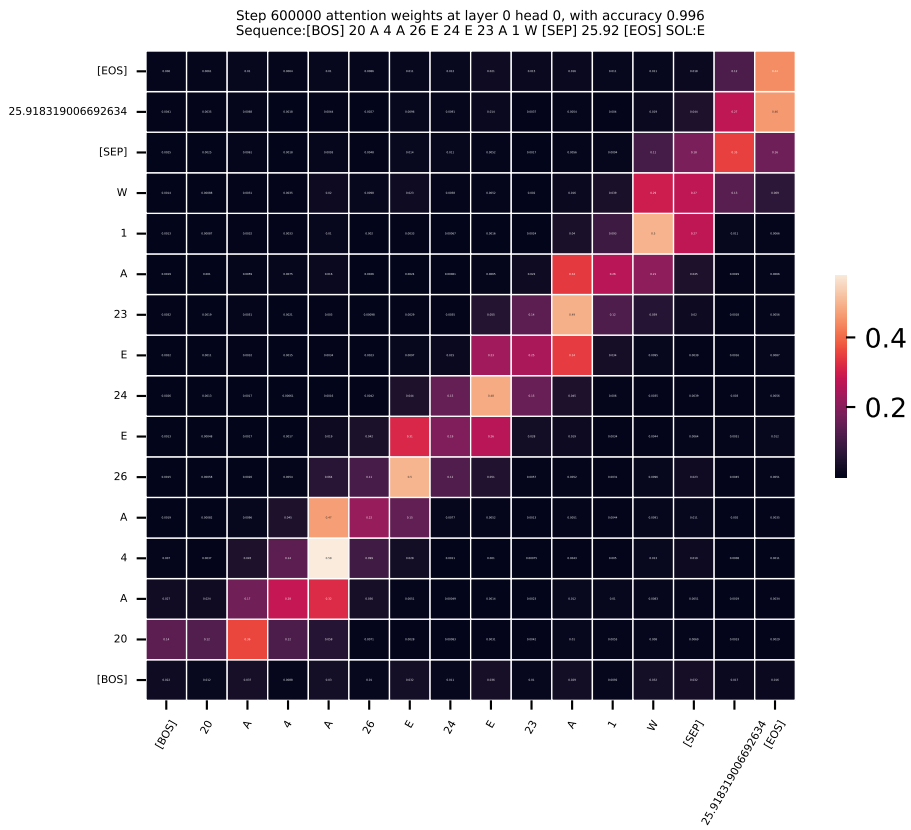

Step 599500 attention weights at layer 0 head 3, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

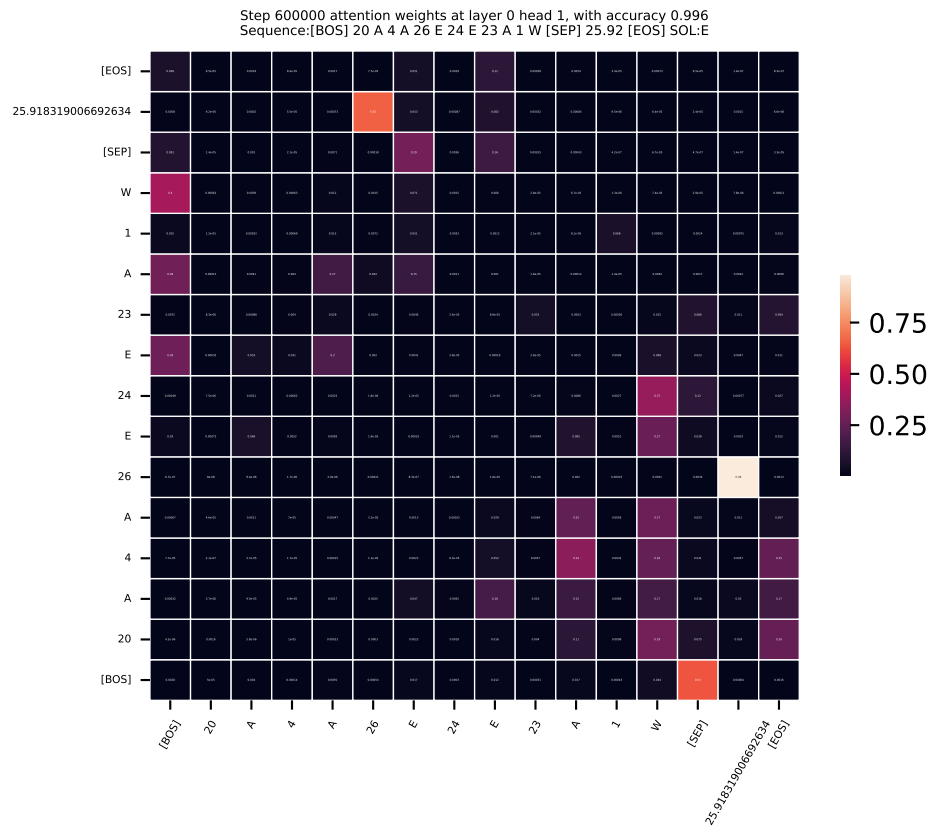

Step 599500 attention weights at layer 1 head 0, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

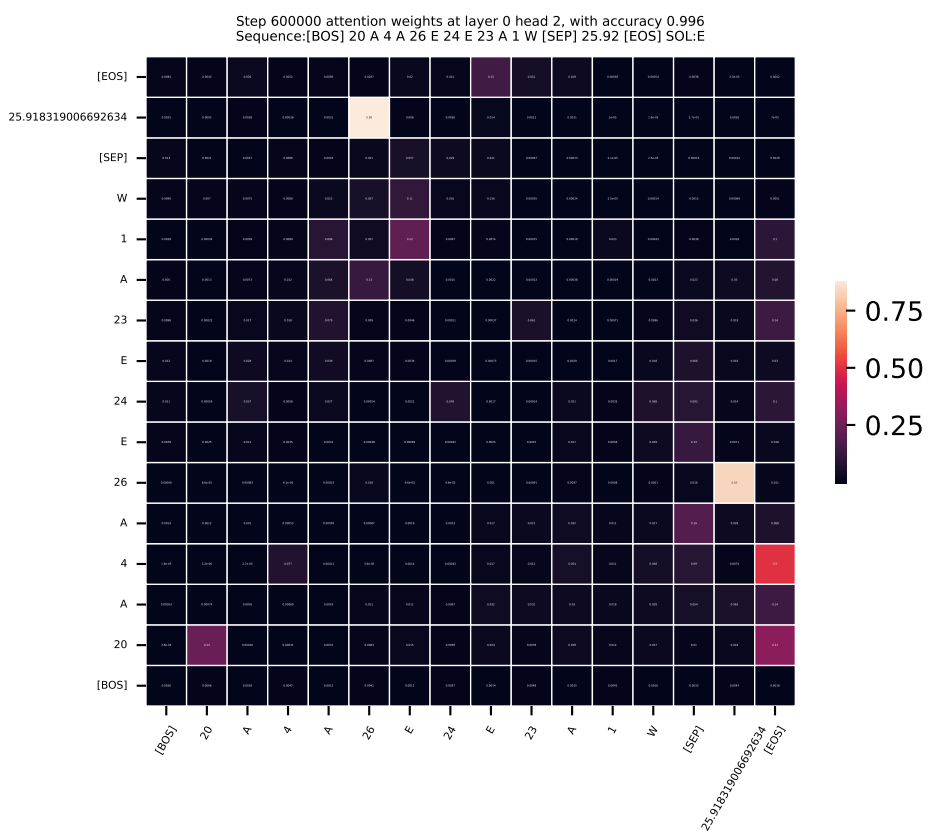

Step 599500 attention weights at layer 1 head 1, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

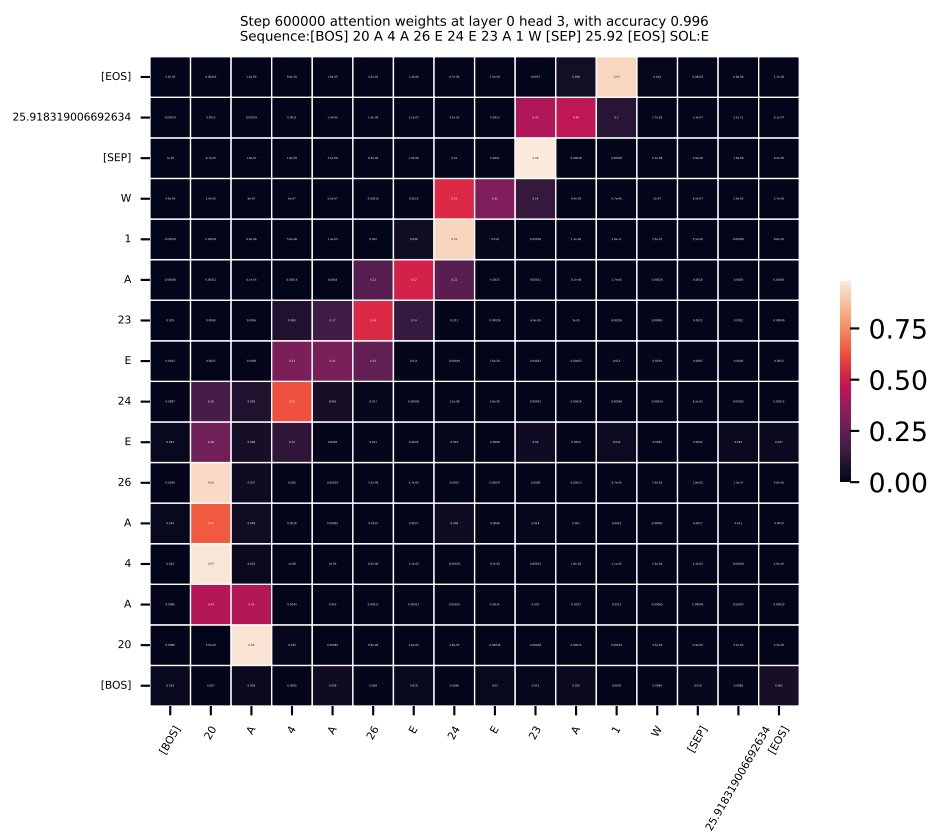

Step 599500 attention weights at layer 1 head 2, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

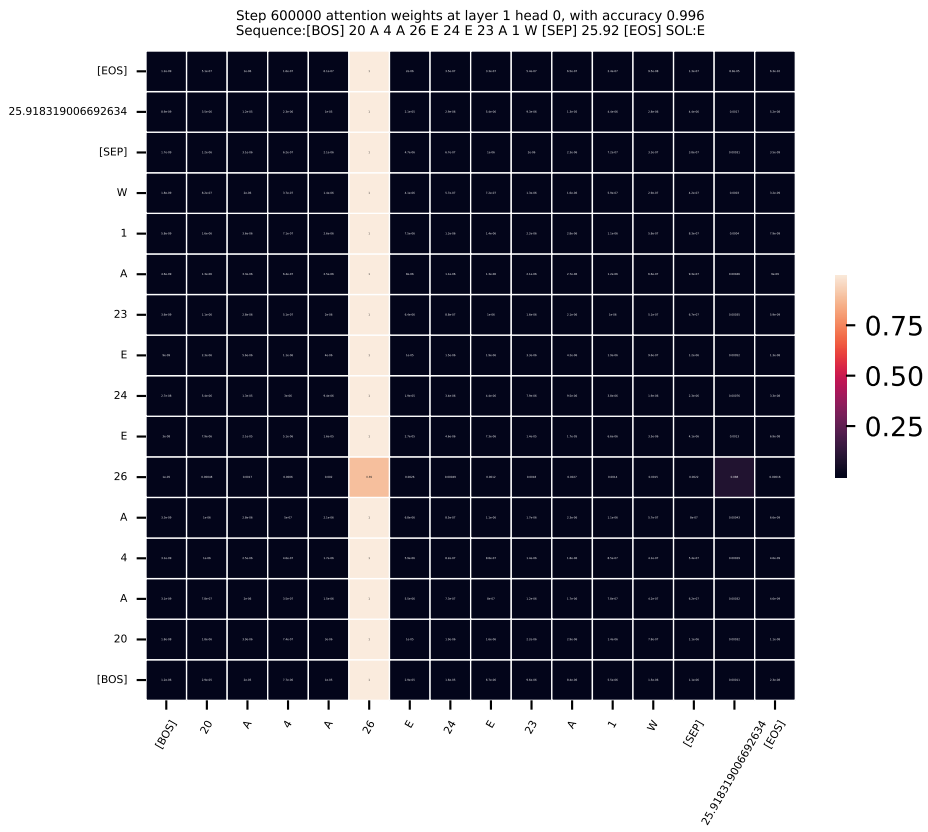

Step 599500 attention weights at layer 1 head 3, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

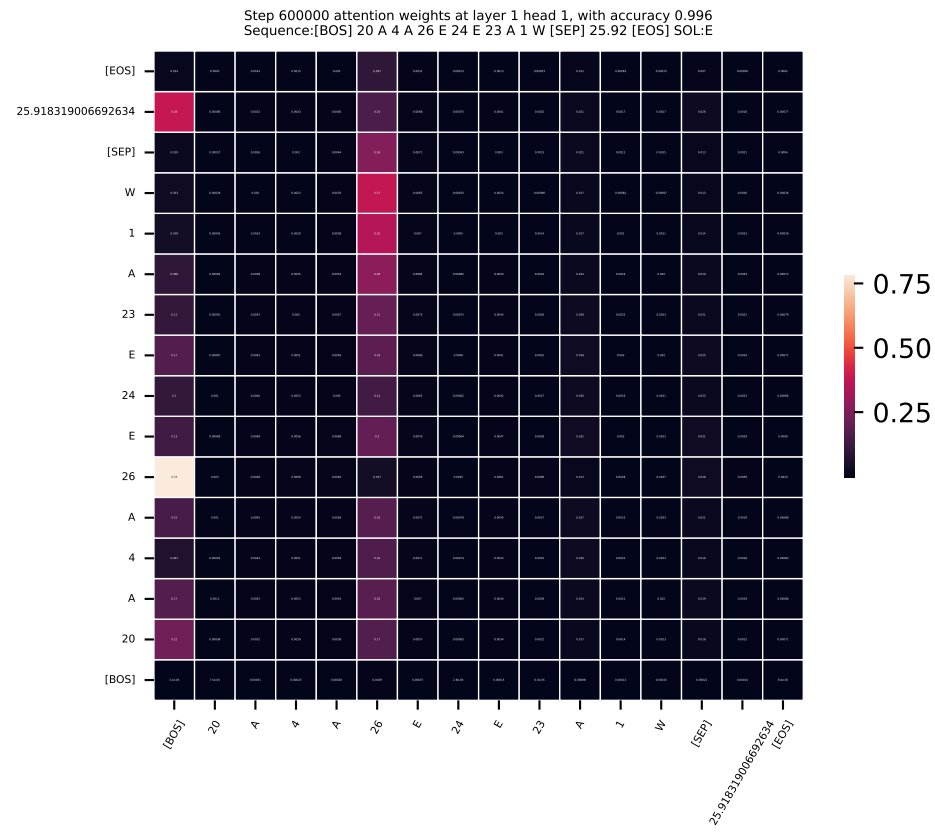

Step 600000 attention weights at layer 0 head 0, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

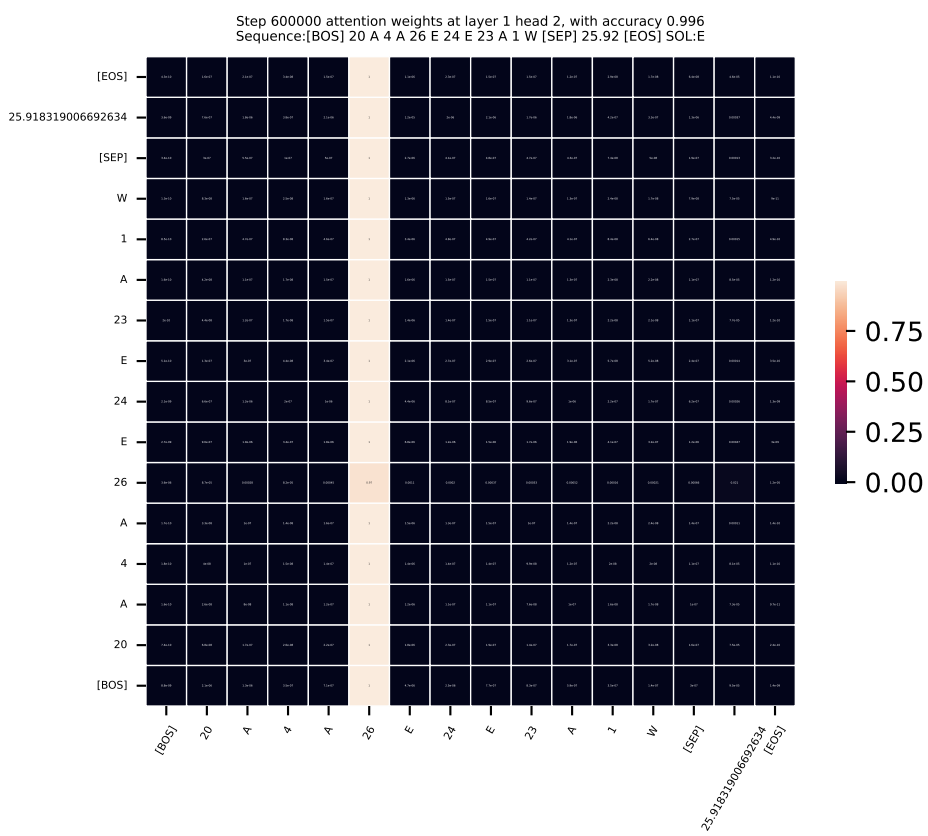

Step 600000 attention weights at layer 0 head 1, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

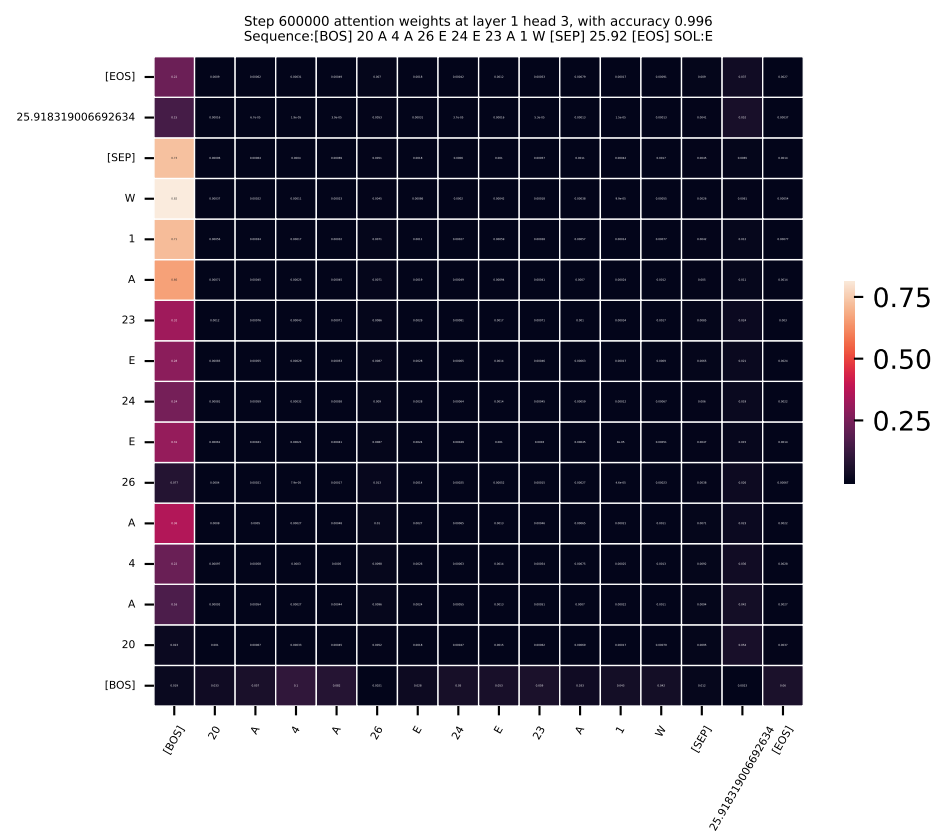

Step 600000 attention weights at layer 0 head 2, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

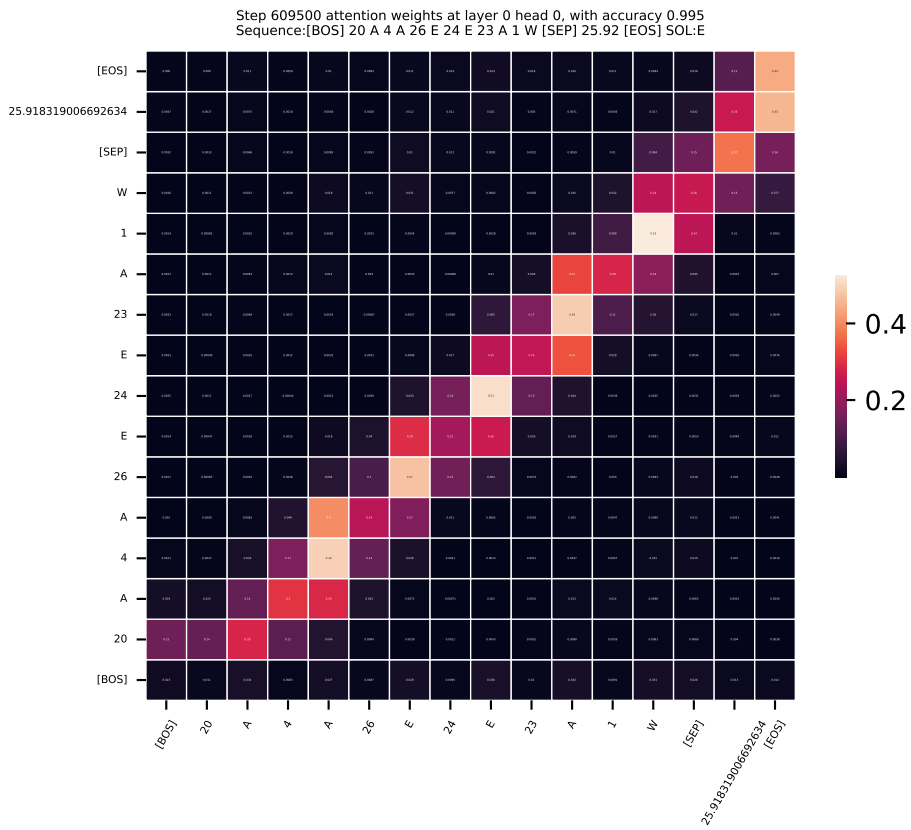

Step 600000 attention weights at layer 0 head 3, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

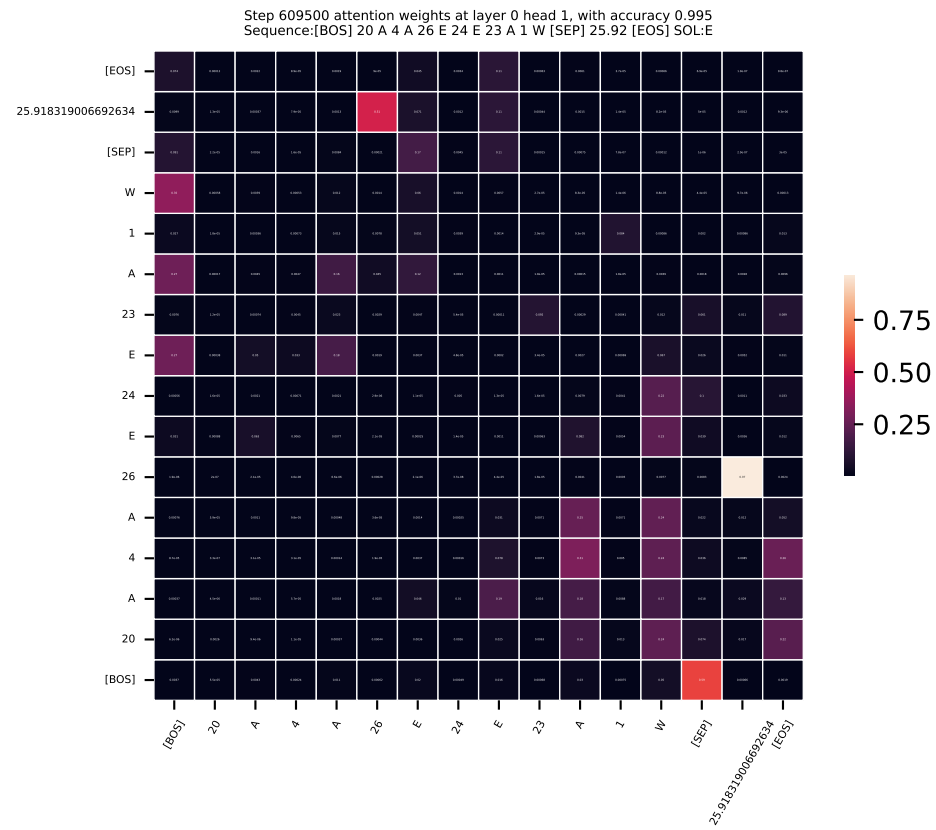

Step 600000 attention weights at layer 1 head 0, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

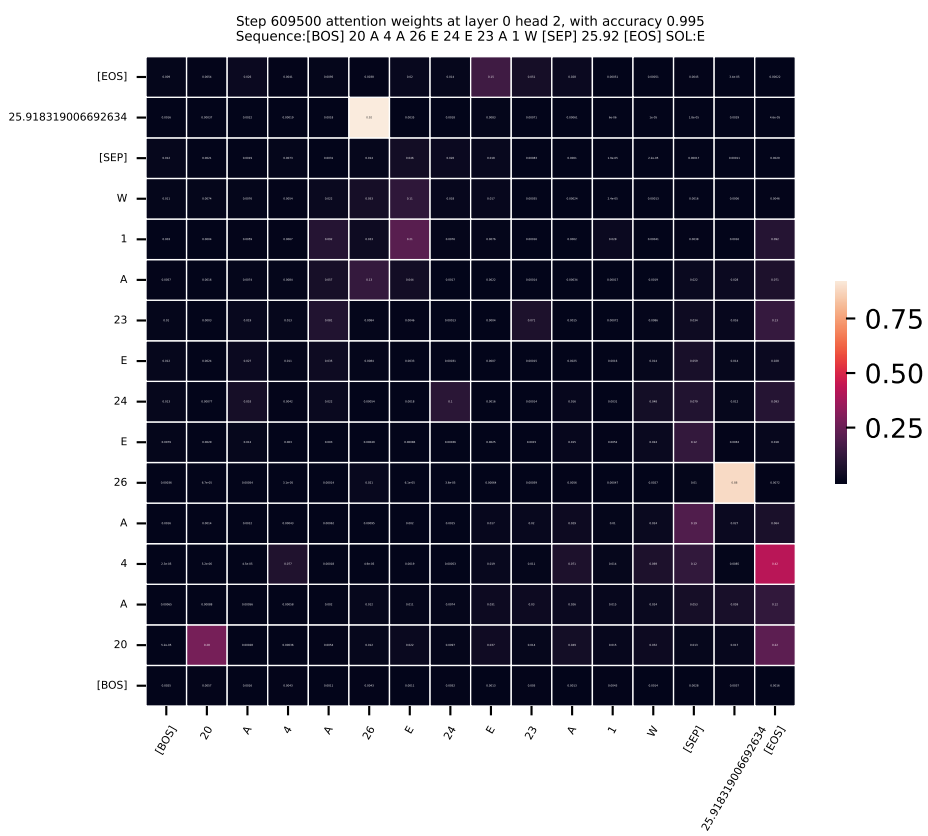

Step 600000 attention weights at layer 1 head 1, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

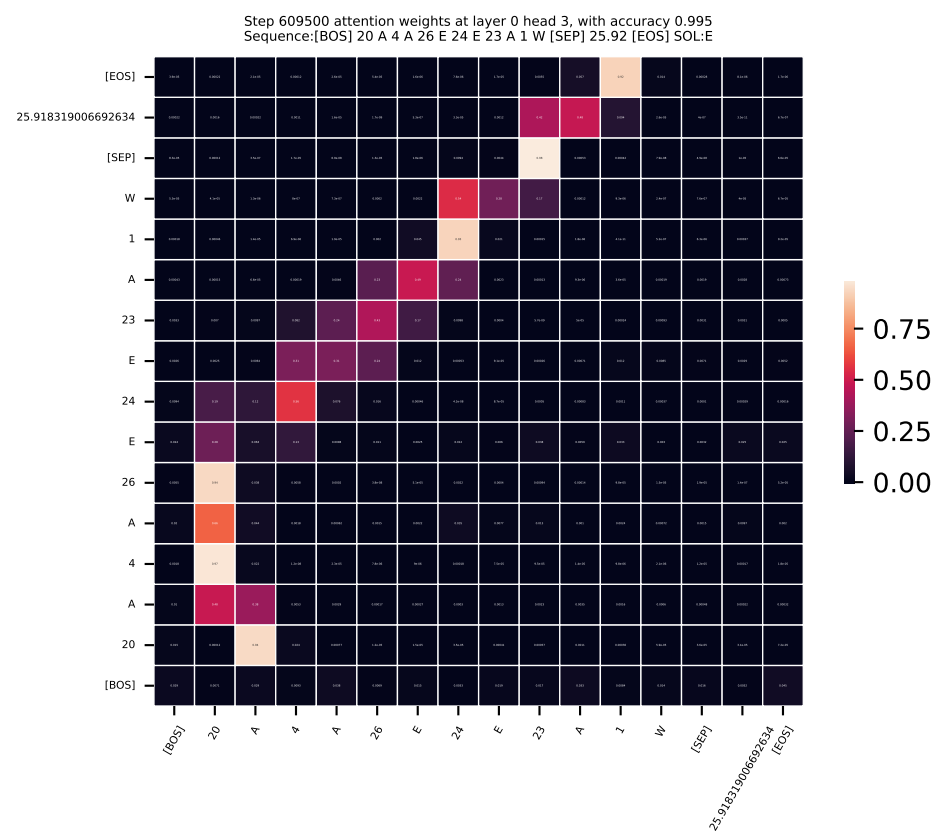

Step 600000 attention weights at layer 1 head 2, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

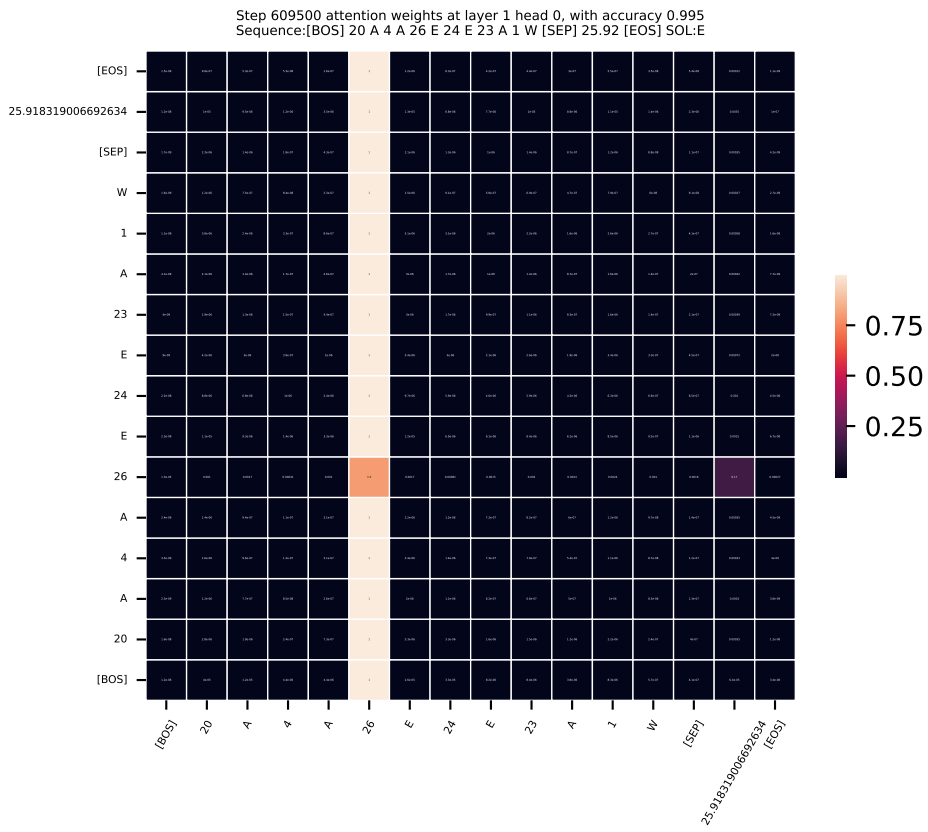

Step 600000 attention weights at layer 1 head 3, with accuracy 0.996
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

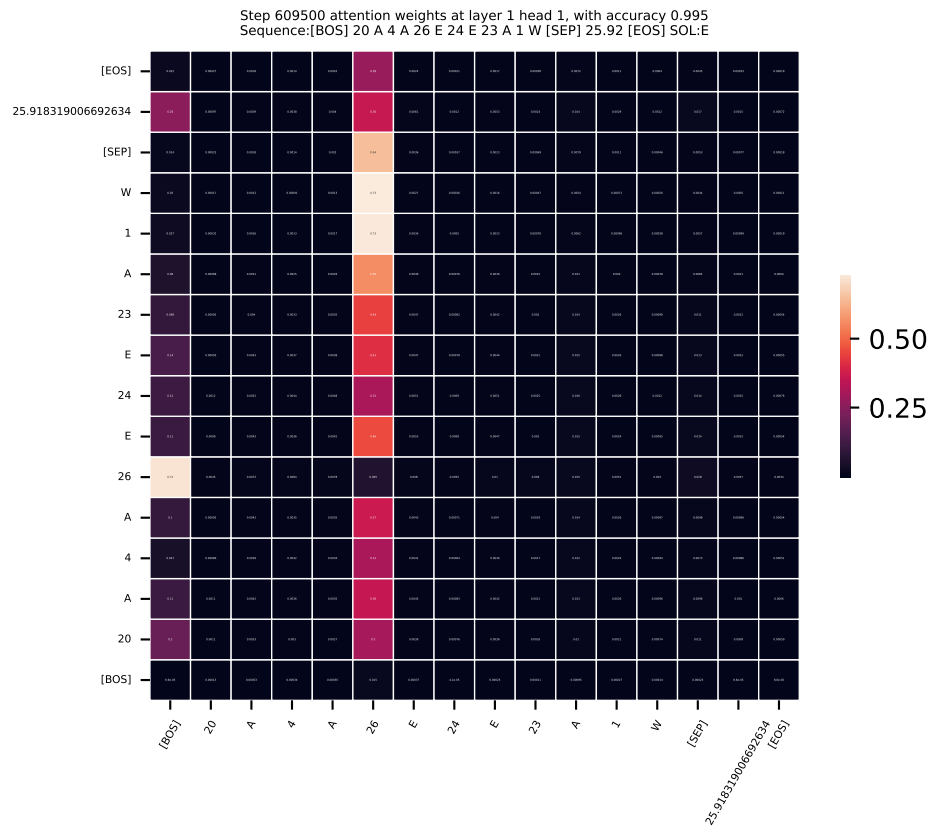

Step 609500 attention weights at layer 0 head 0, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

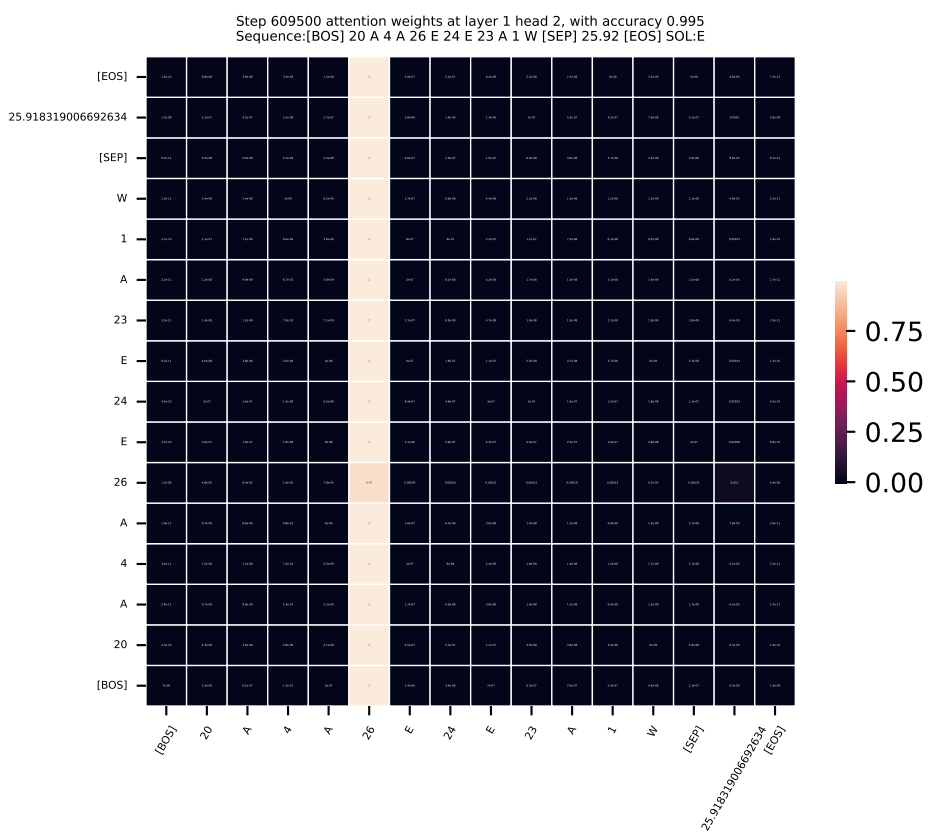

Step 609500 attention weights at layer 0 head 1, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

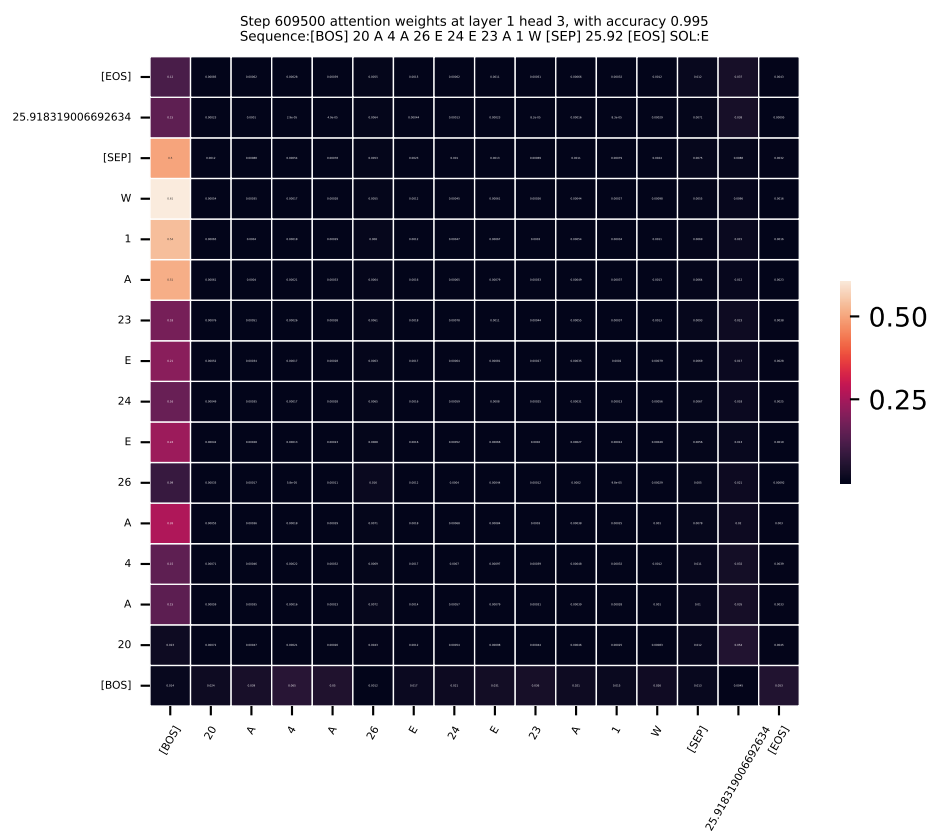

Step 609500 attention weights at layer 0 head 2, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

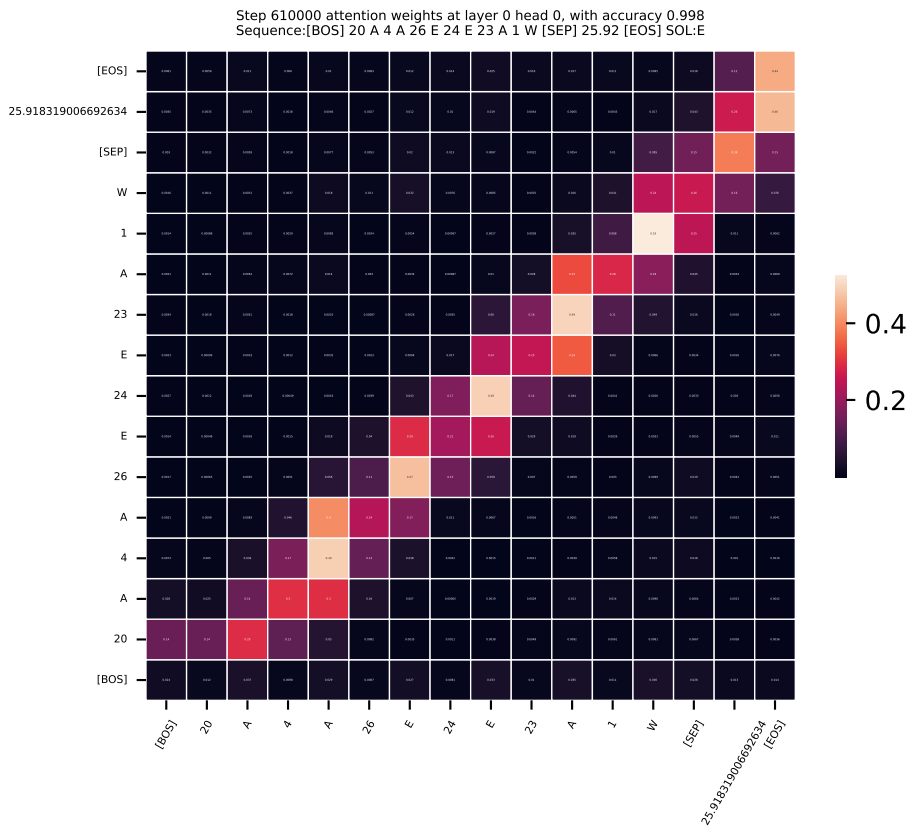

Step 609500 attention weights at layer 0 head 3, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

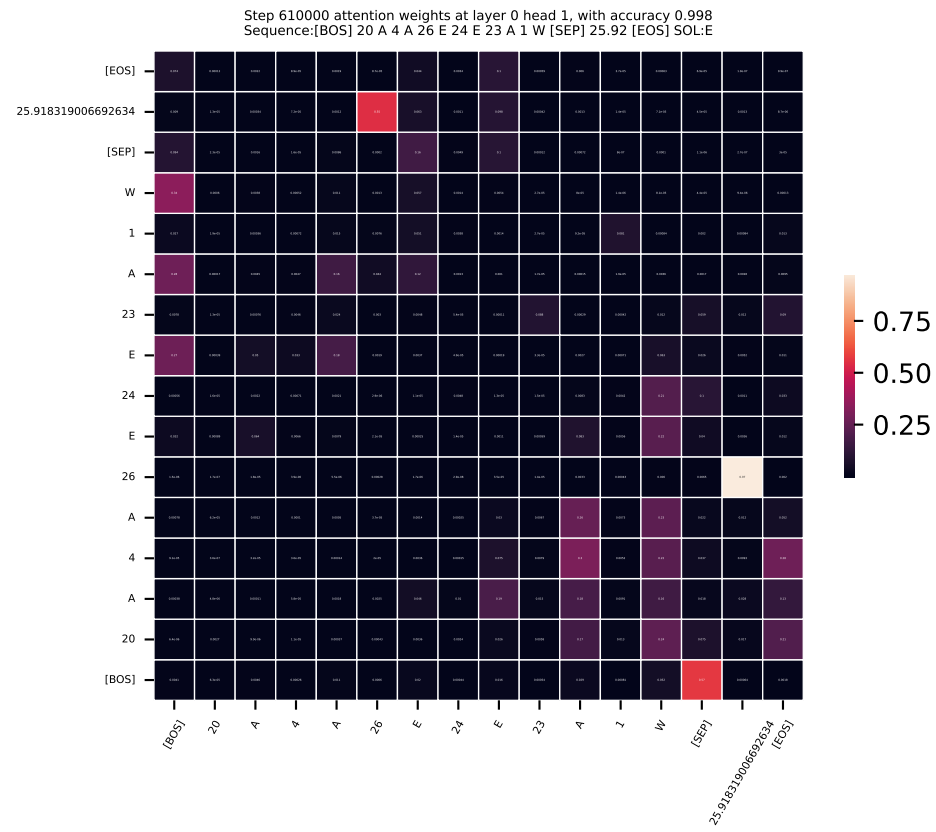

Step 609500 attention weights at layer 1 head 0, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

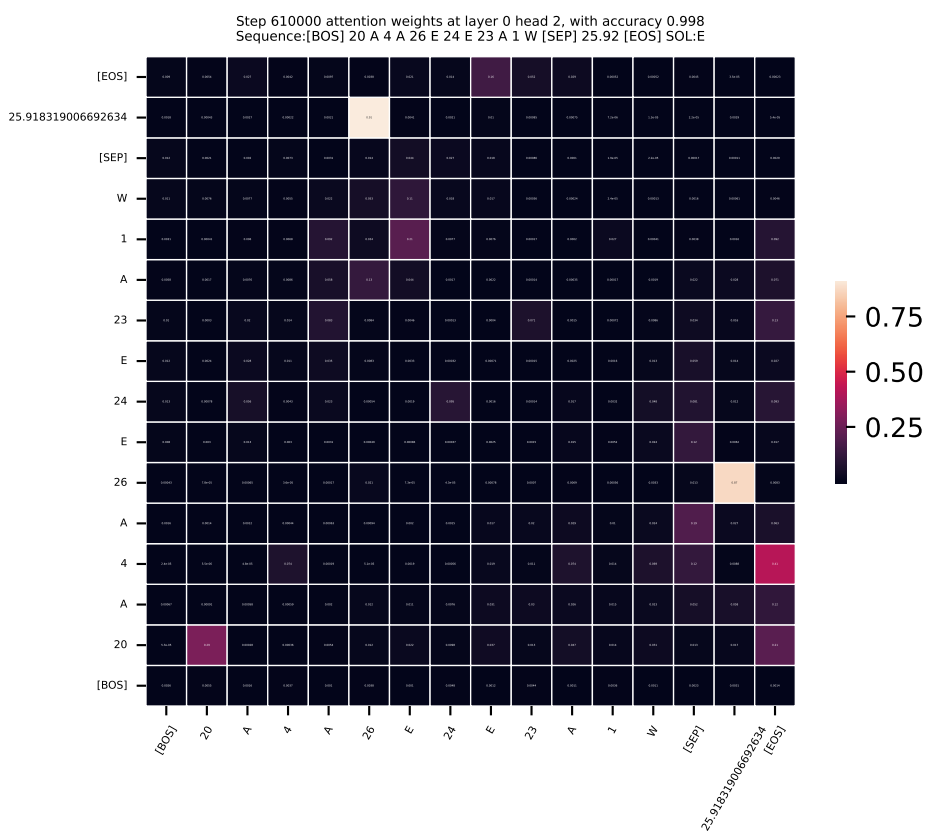

Step 609500 attention weights at layer 1 head 1, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

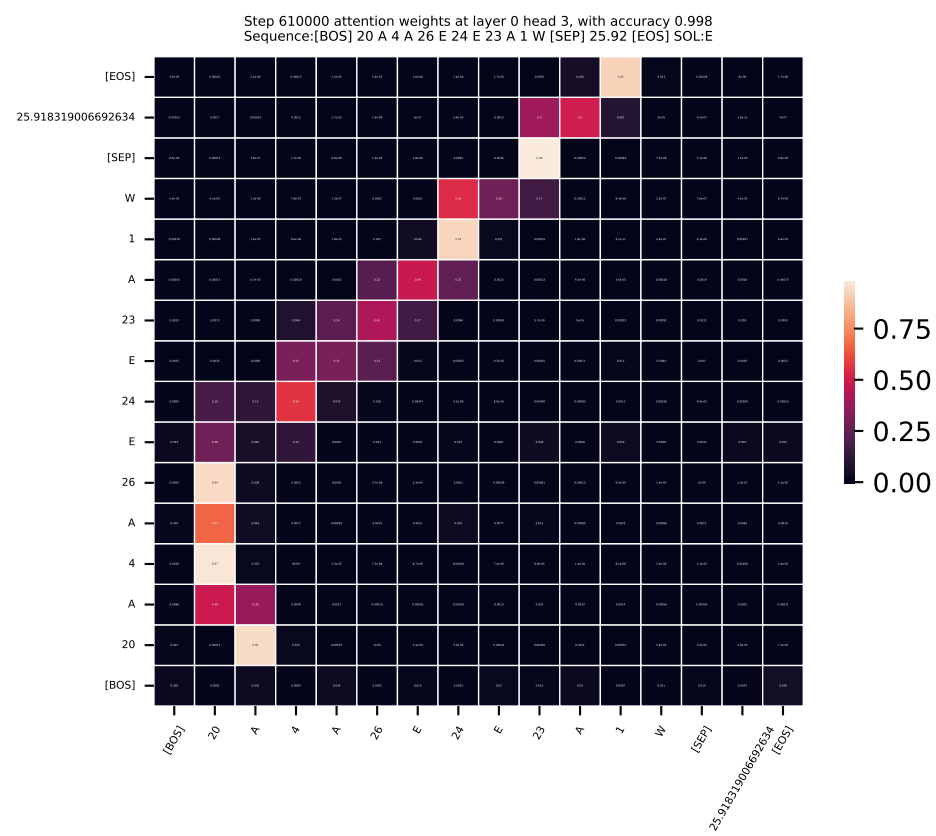

Step 609500 attention weights at layer 1 head 2, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

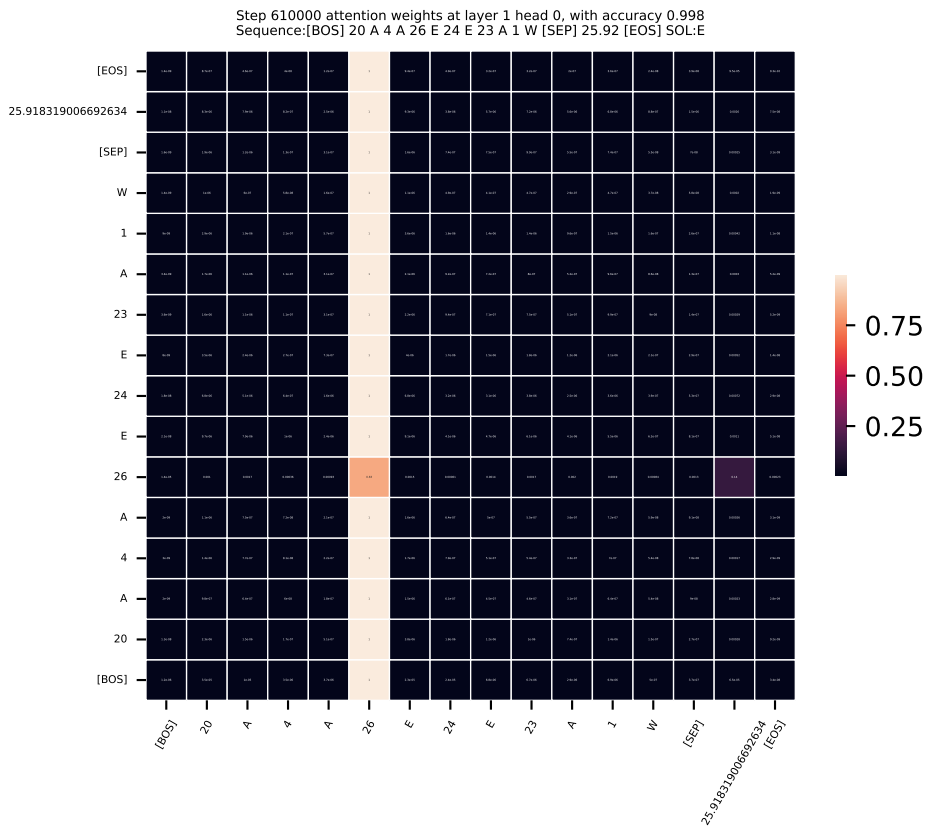

Step 609500 attention weights at layer 1 head 3, with accuracy 0.995
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

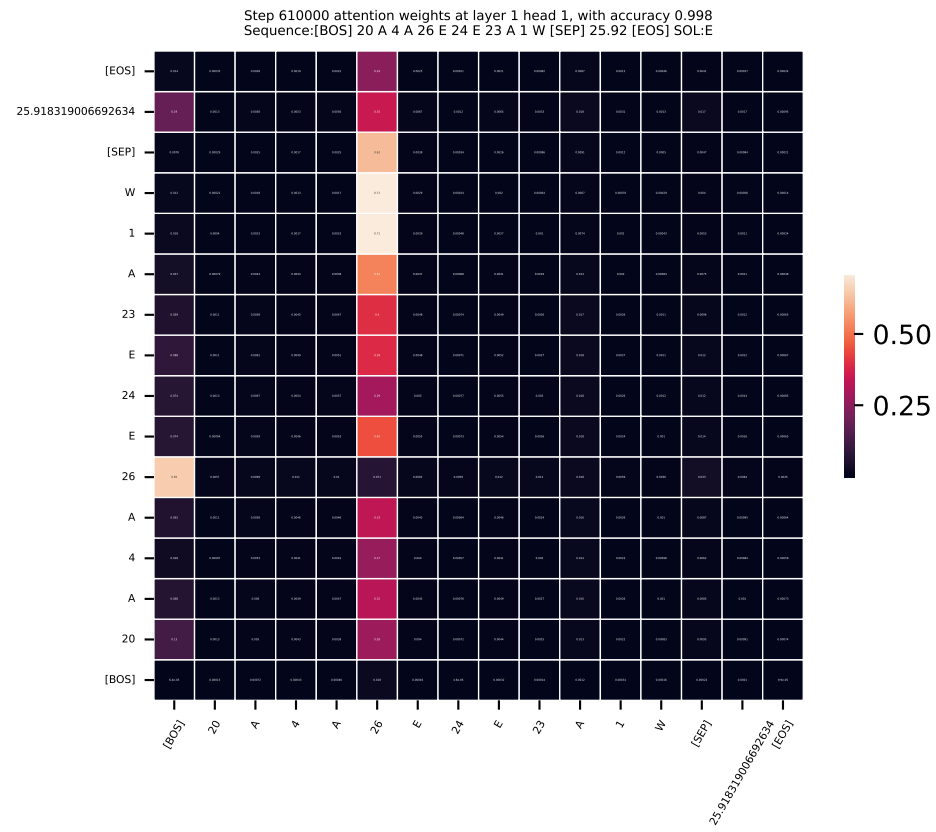

Step 610000 attention weights at layer 0 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.4
0.2

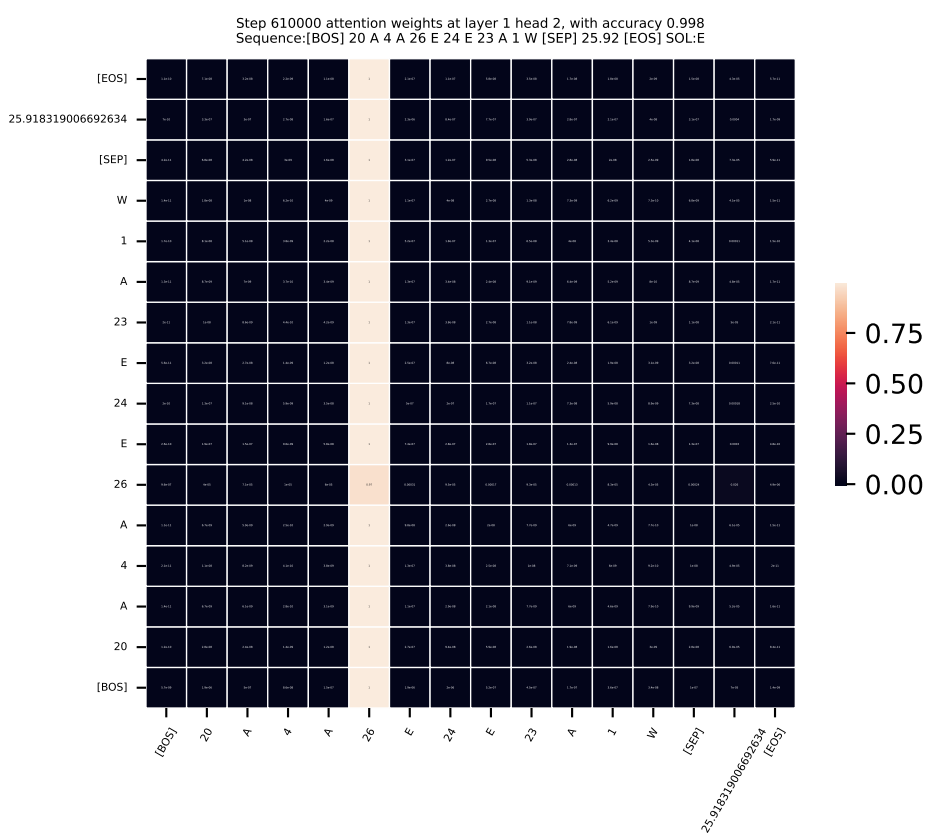

Step 610000 attention weights at layer 0 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

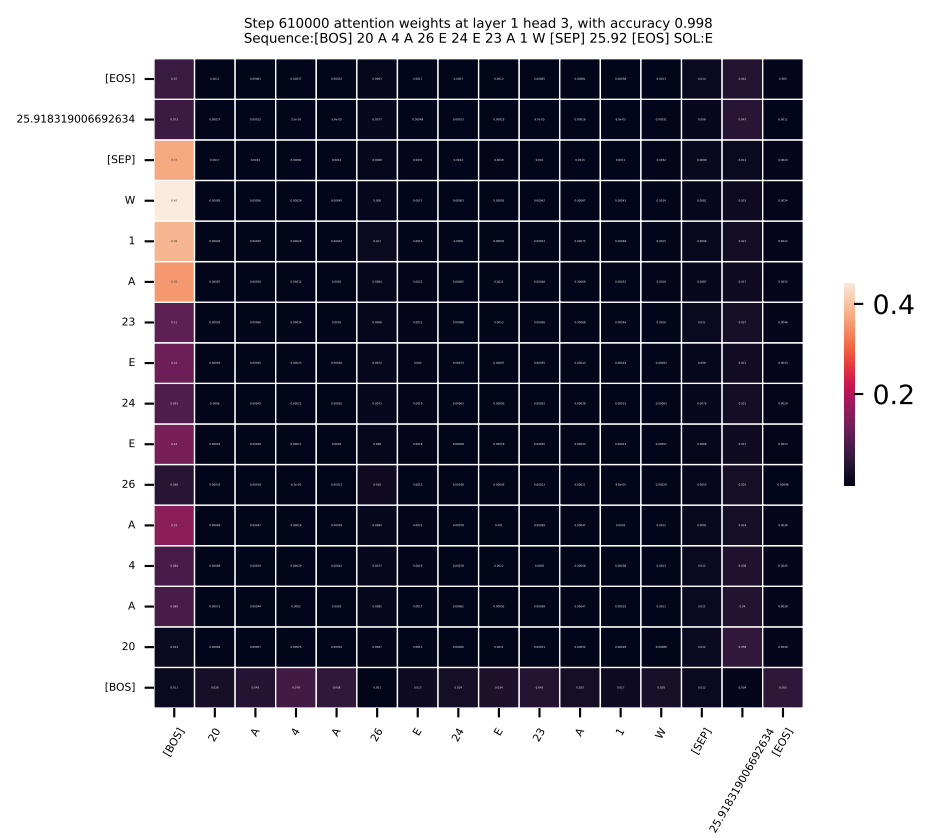

Step 610000 attention weights at layer 0 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

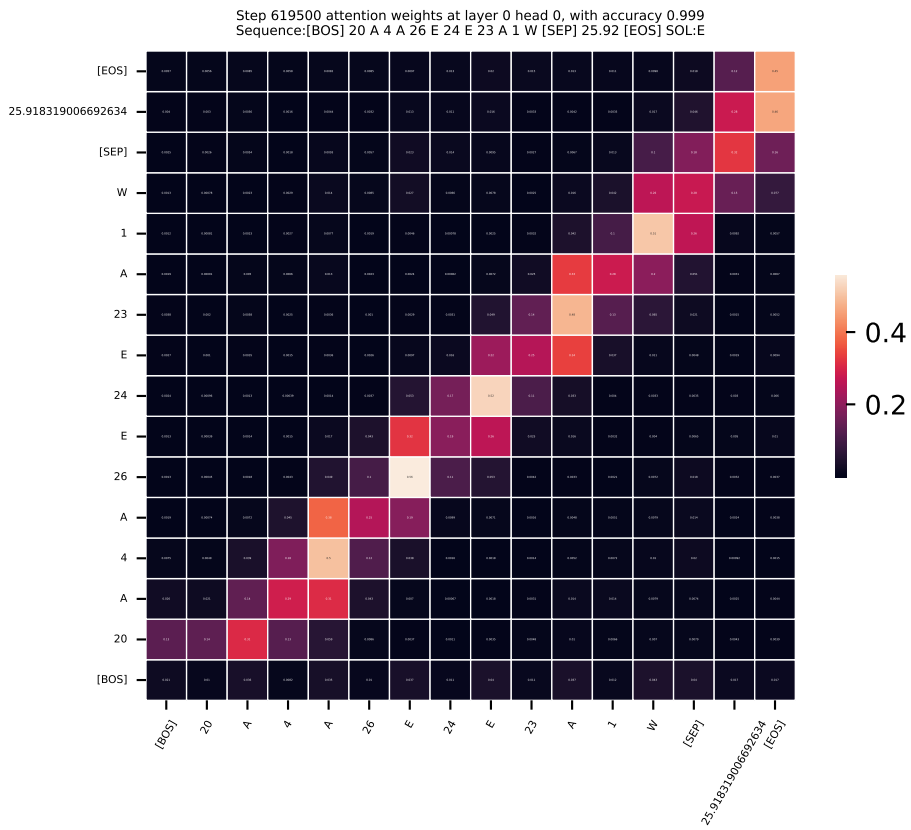

Step 610000 attention weights at layer 0 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25
0.00

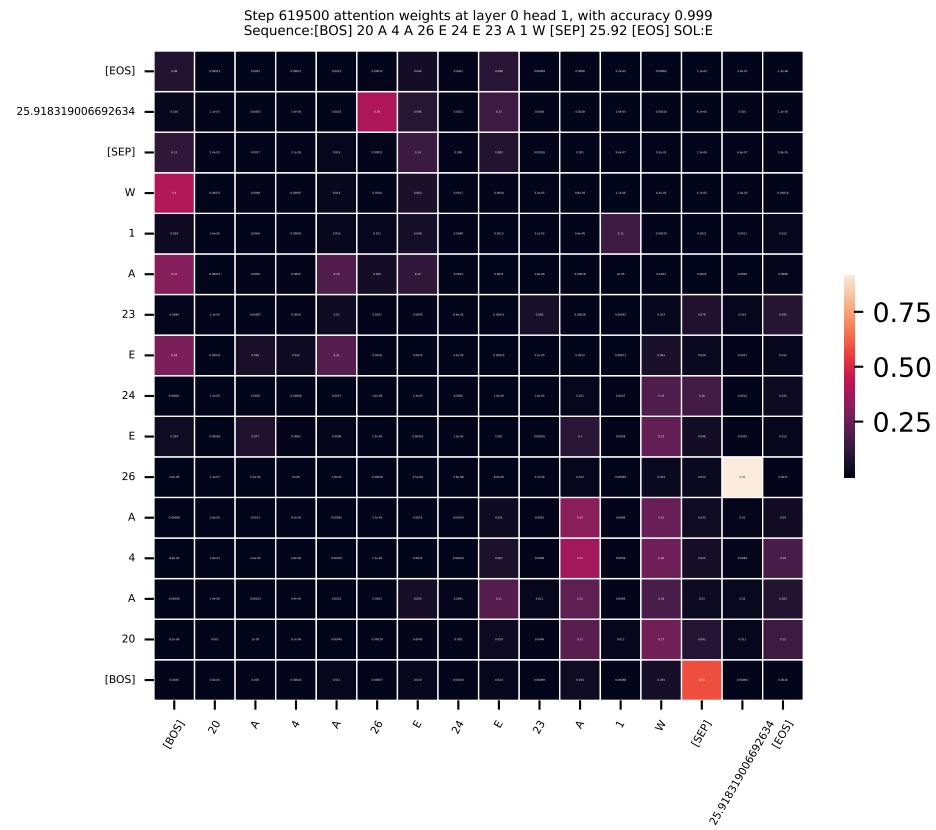

Step 610000 attention weights at layer 1 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

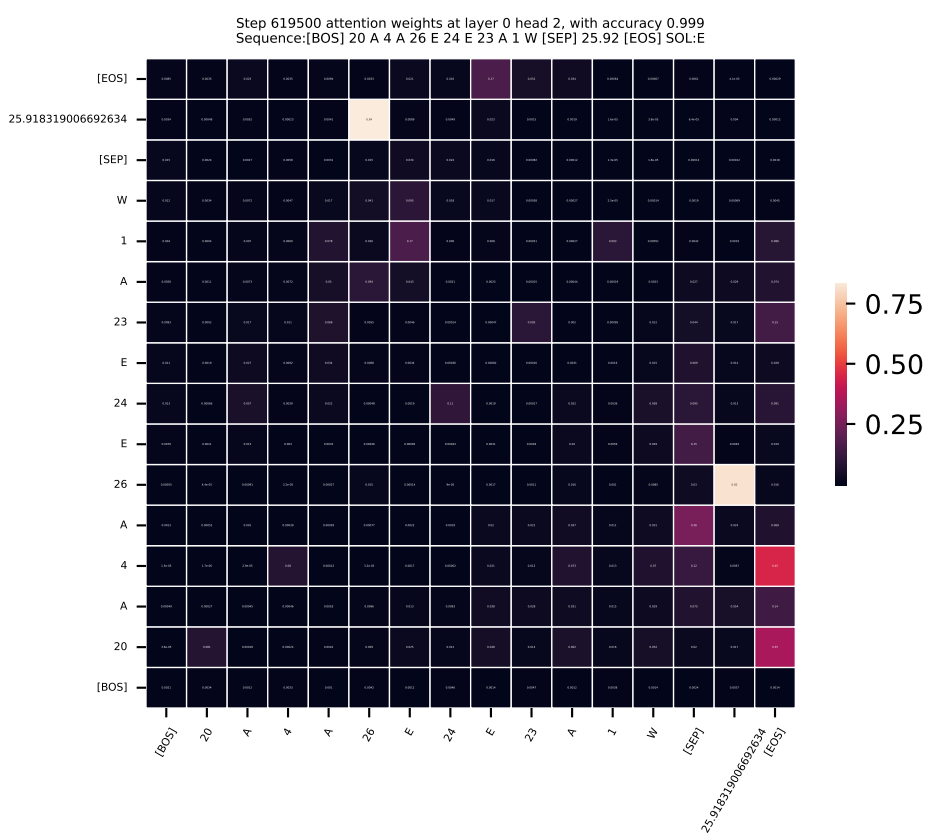

Step 610000 attention weights at layer 1 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

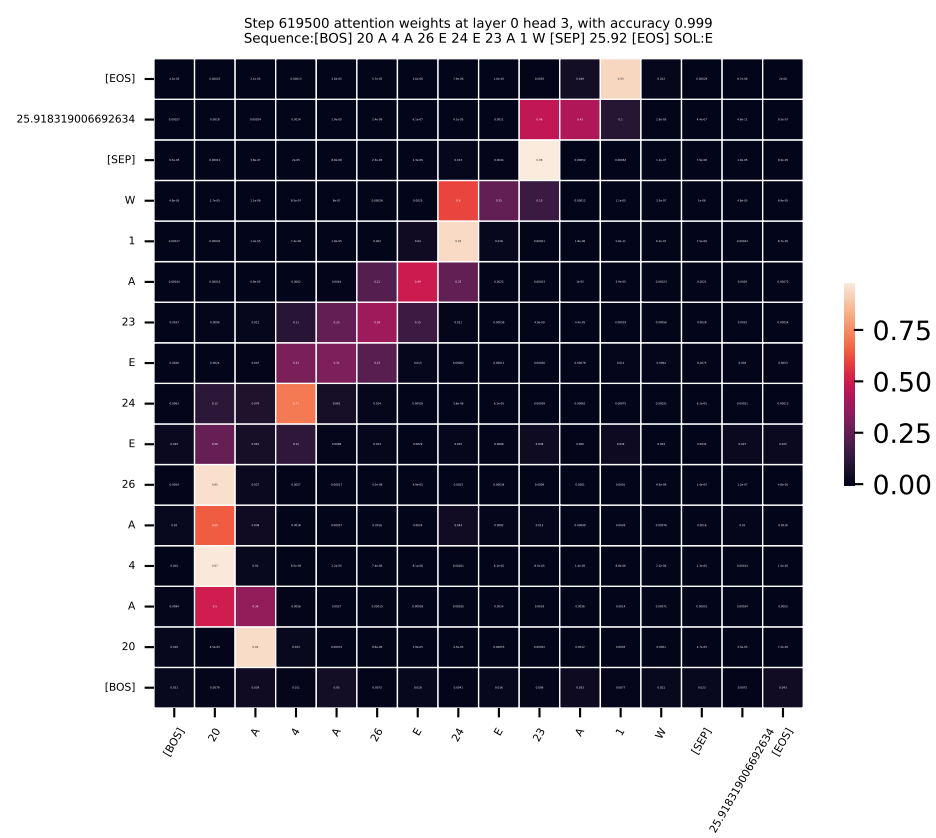

Step 610000 attention weights at layer 1 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

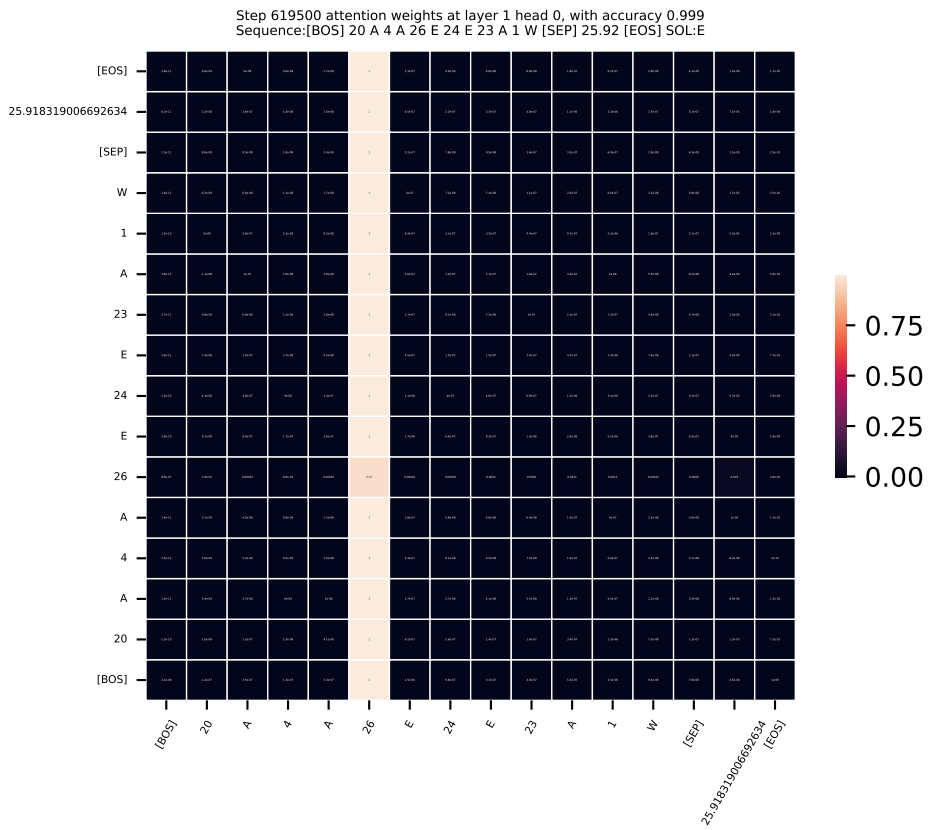

Step 610000 attention weights at layer 1 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

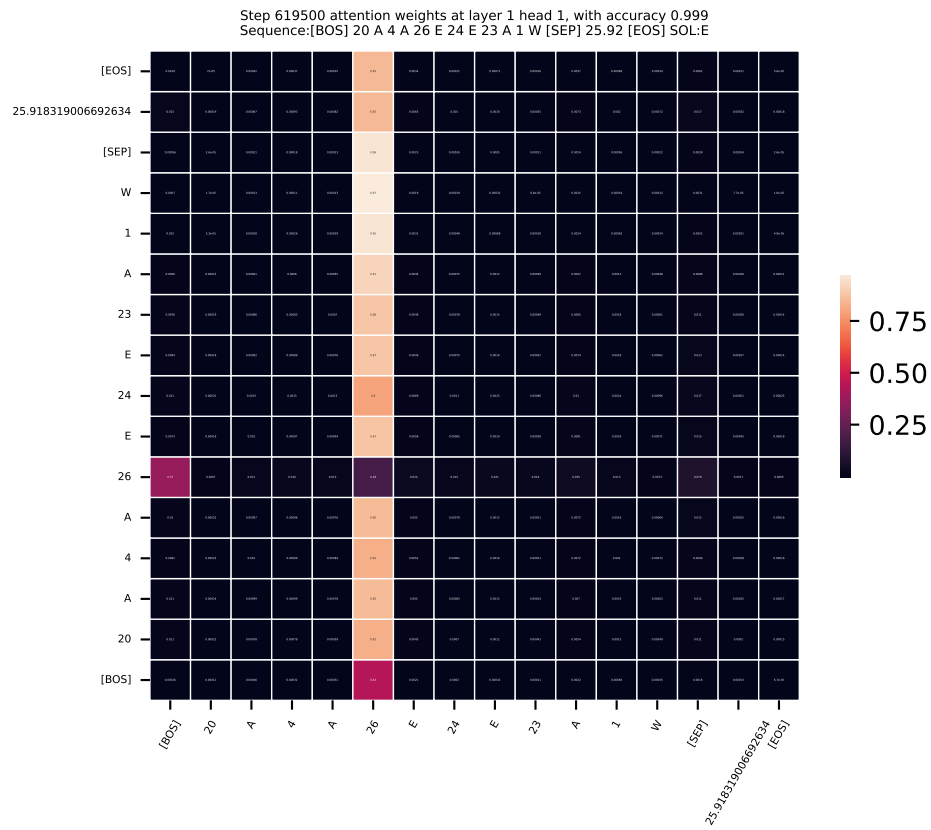

Step 619500 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

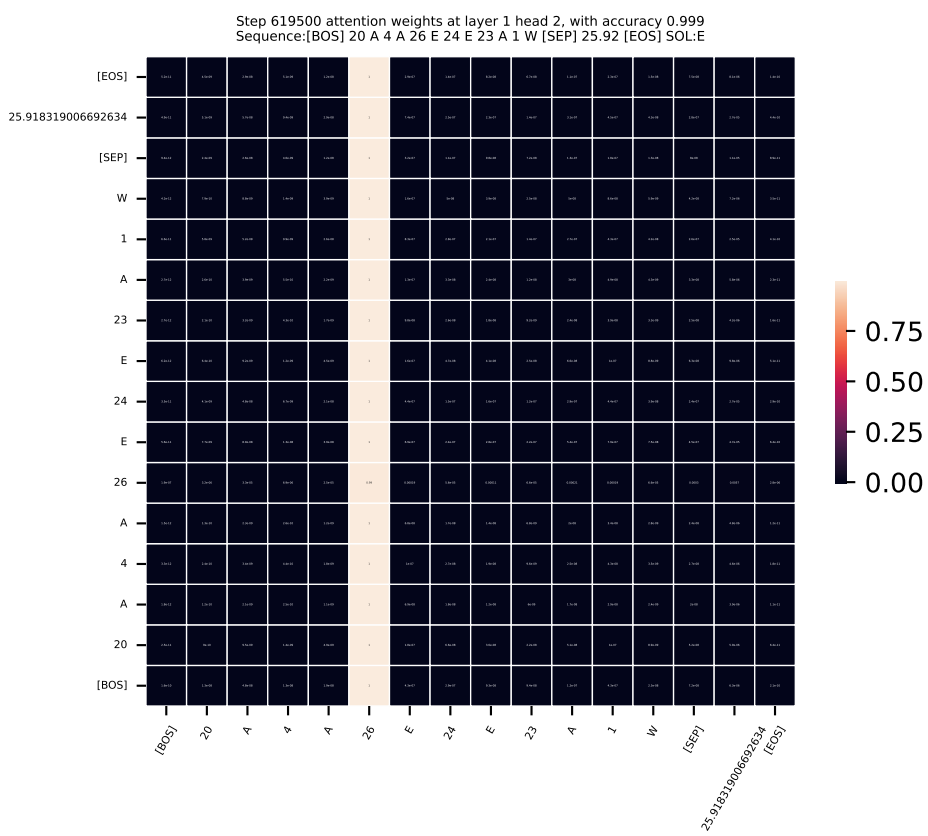

Step 619500 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

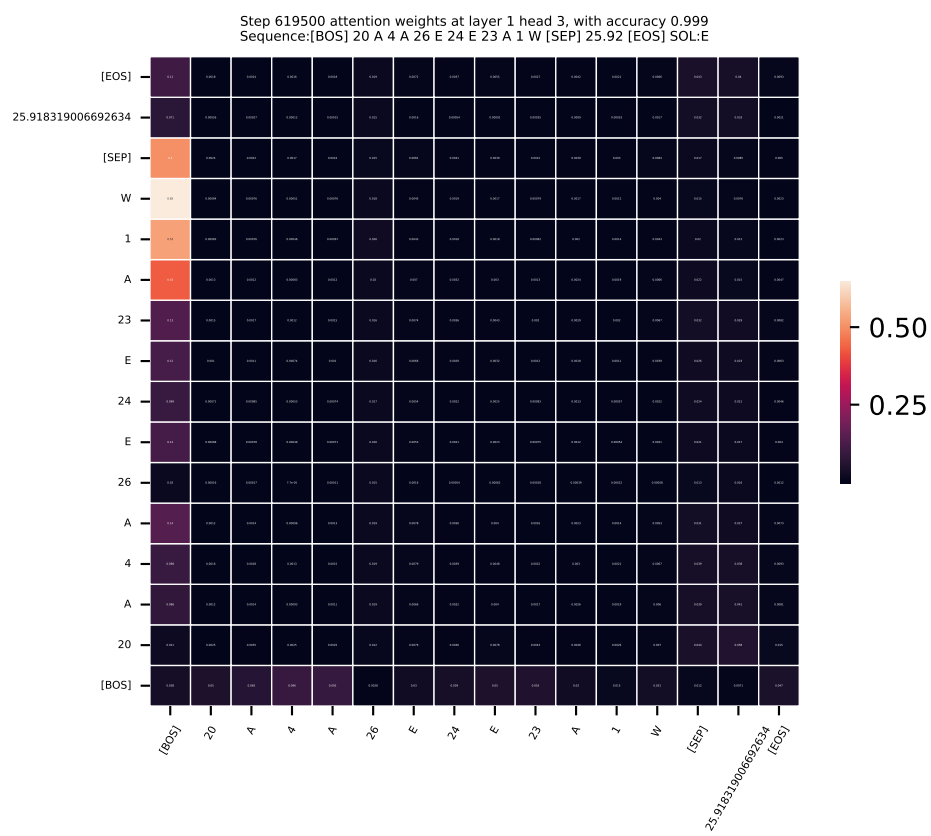

Step 619500 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

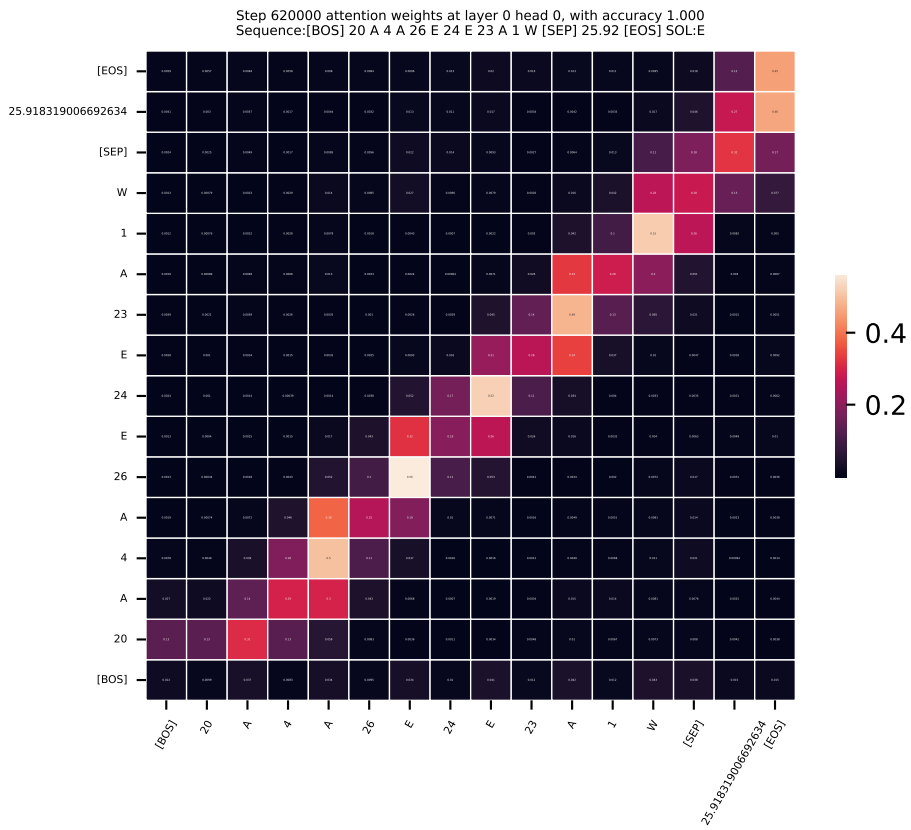

Step 619500 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

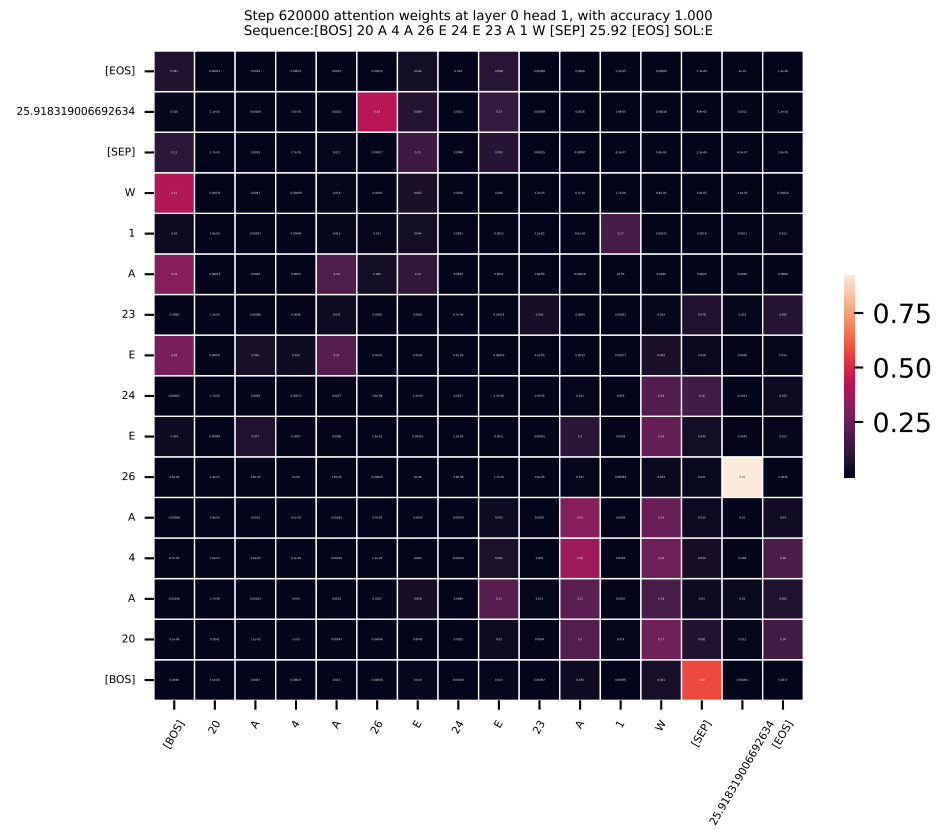

Step 619500 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25
0.00

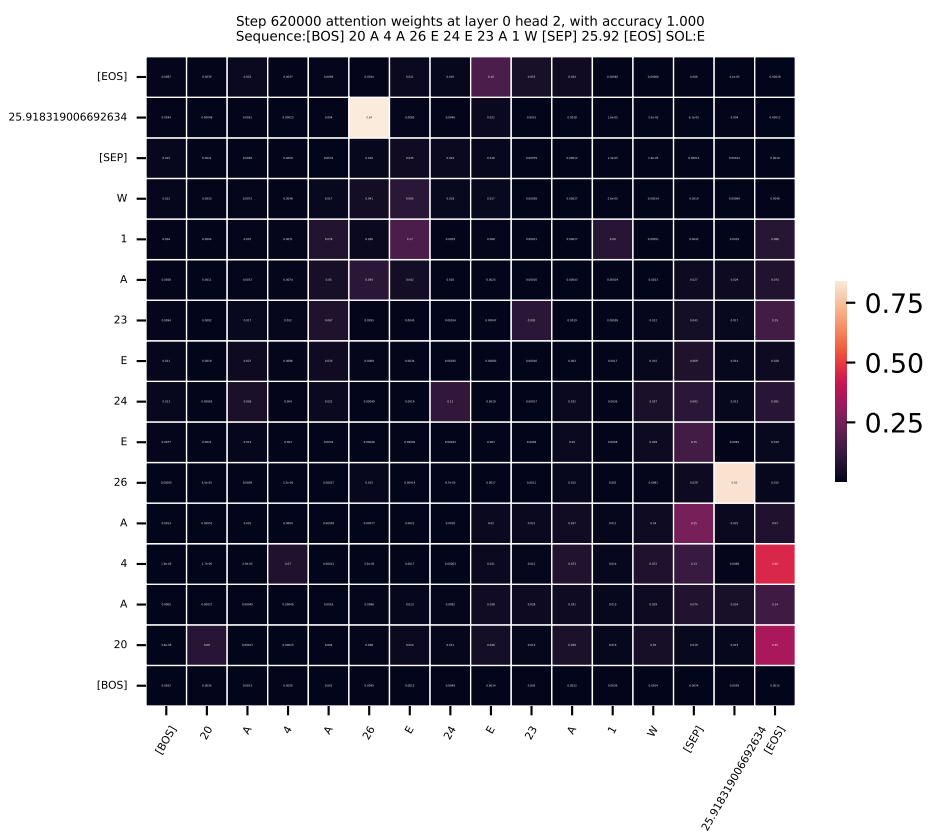

Step 619500 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

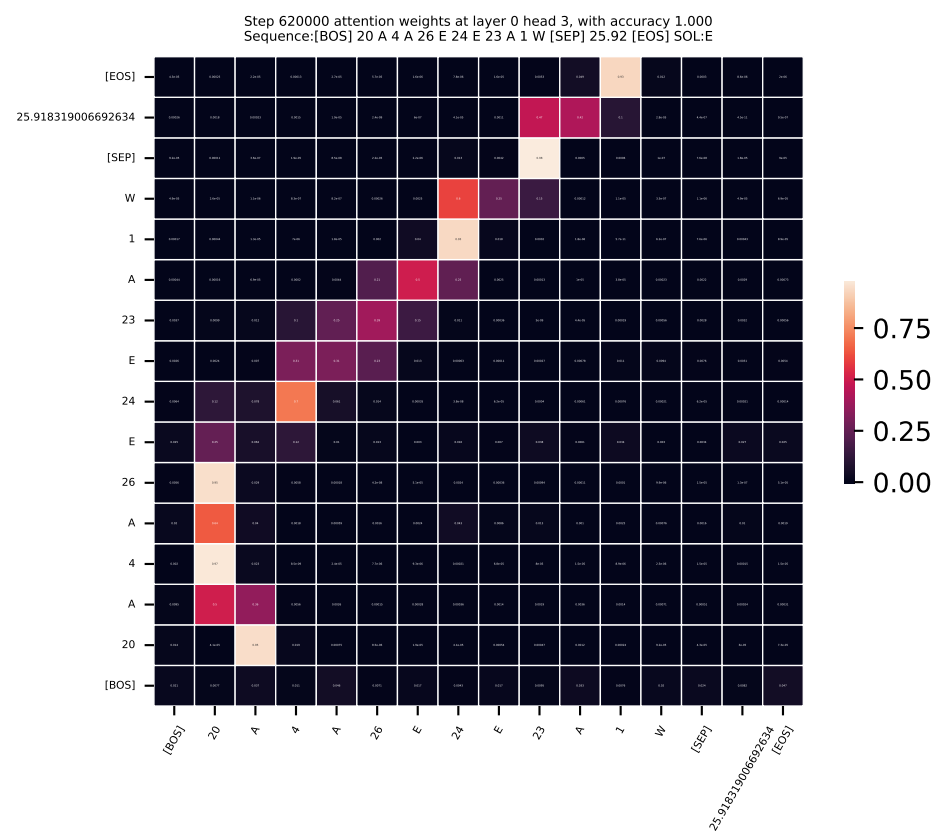

Step 619500 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25
0.00

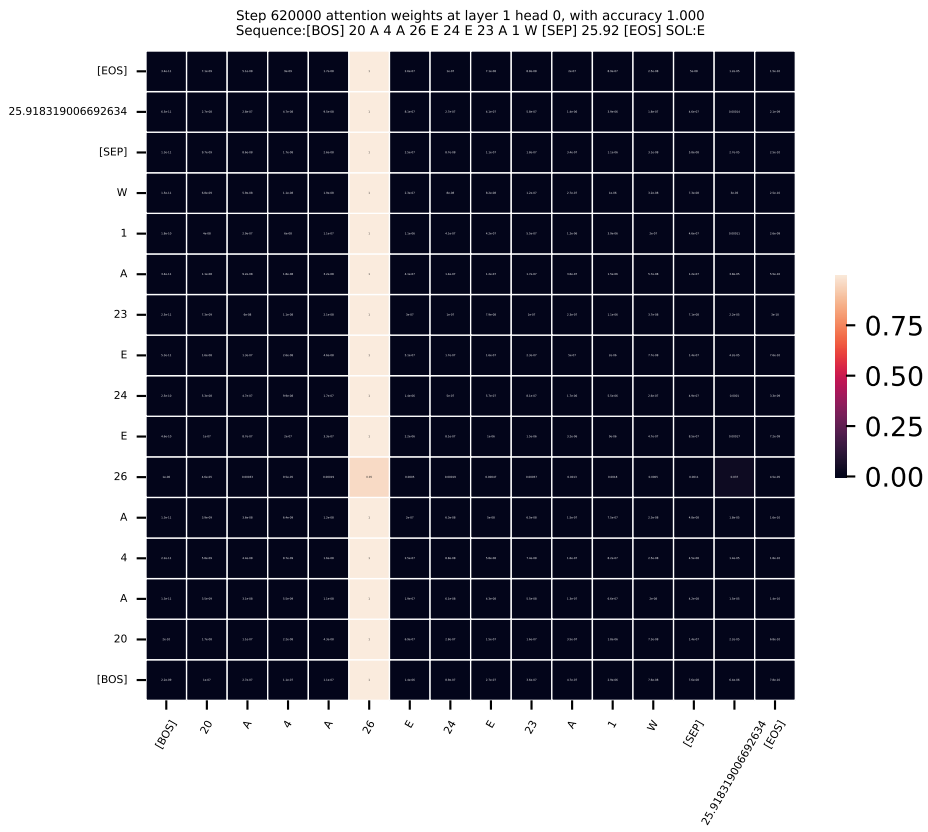

Step 619500 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.50
0.25

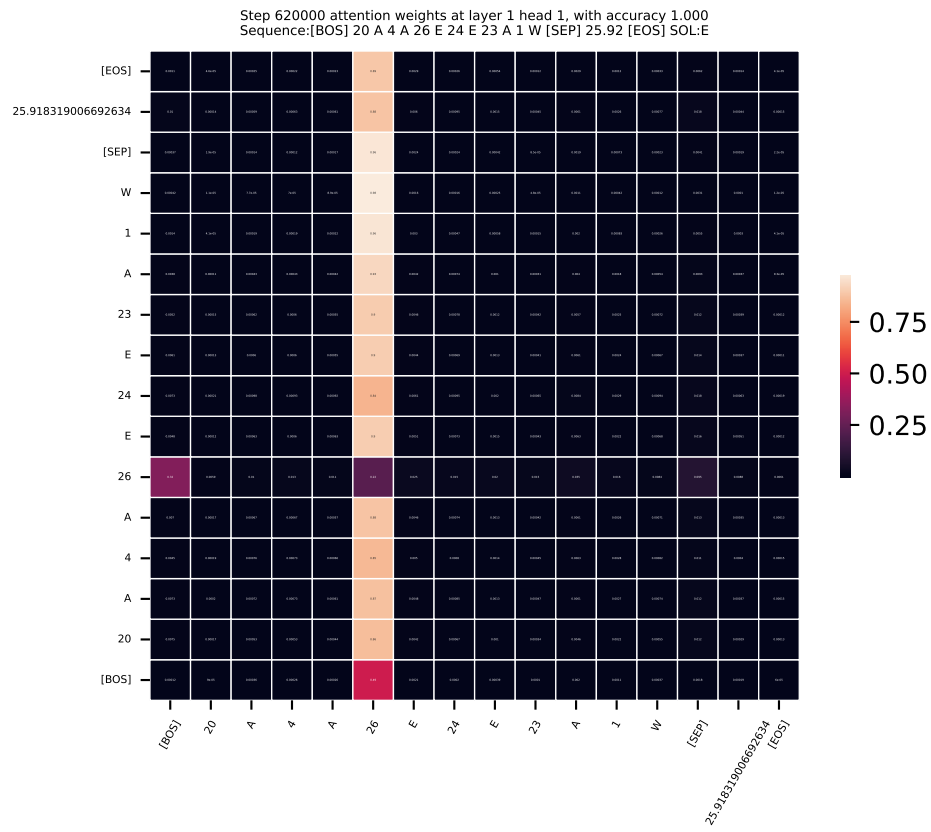

Step 620000 attention weights at layer 0 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

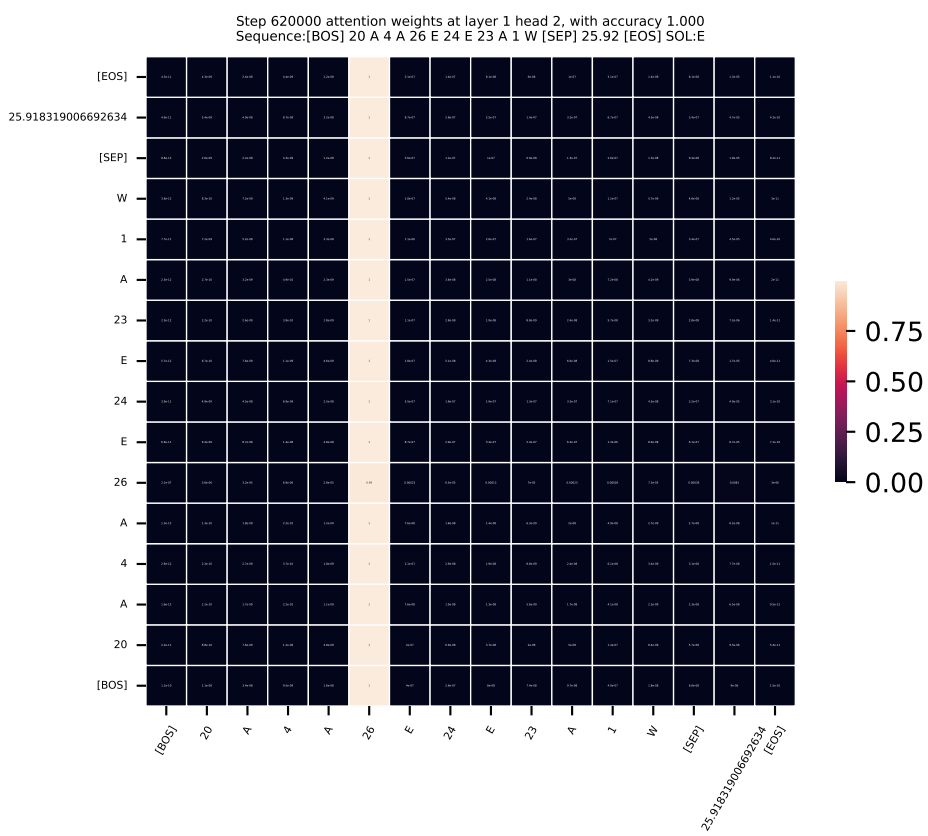

Step 620000 attention weights at layer 0 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

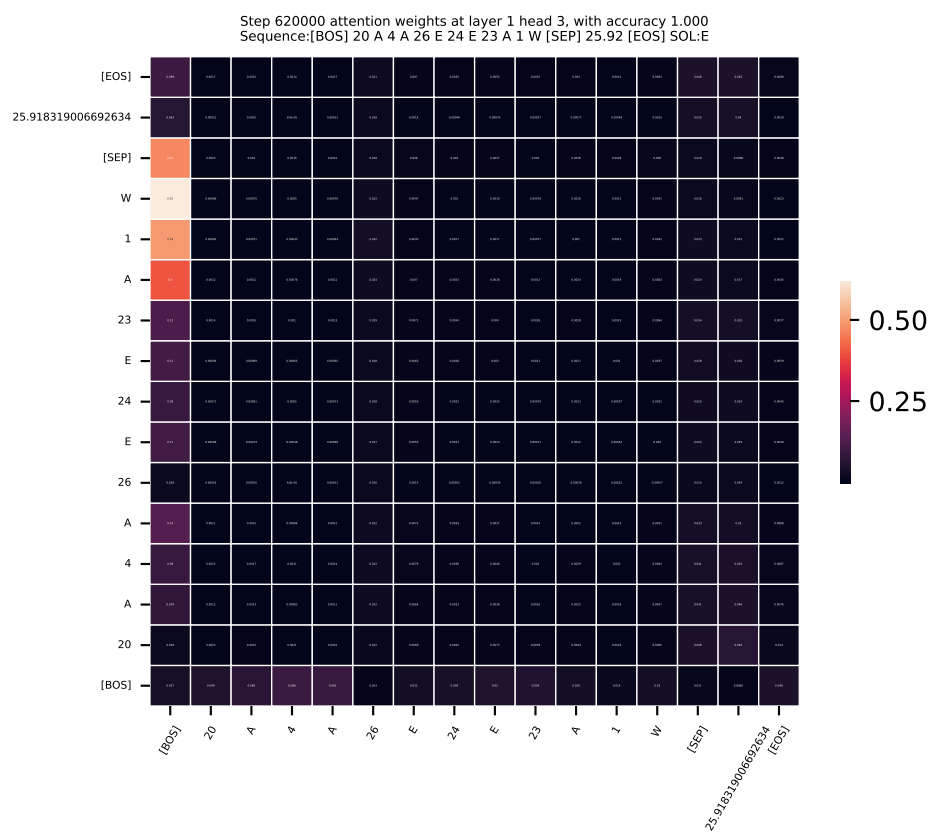

Step 620000 attention weights at layer 0 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

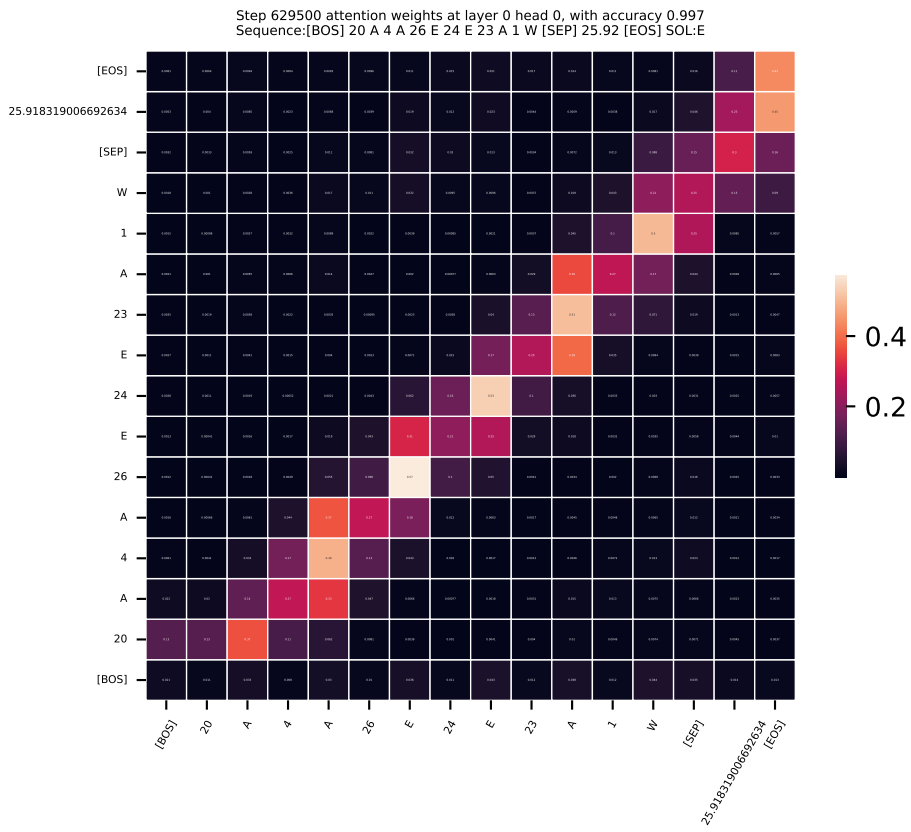

Step 620000 attention weights at layer 0 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

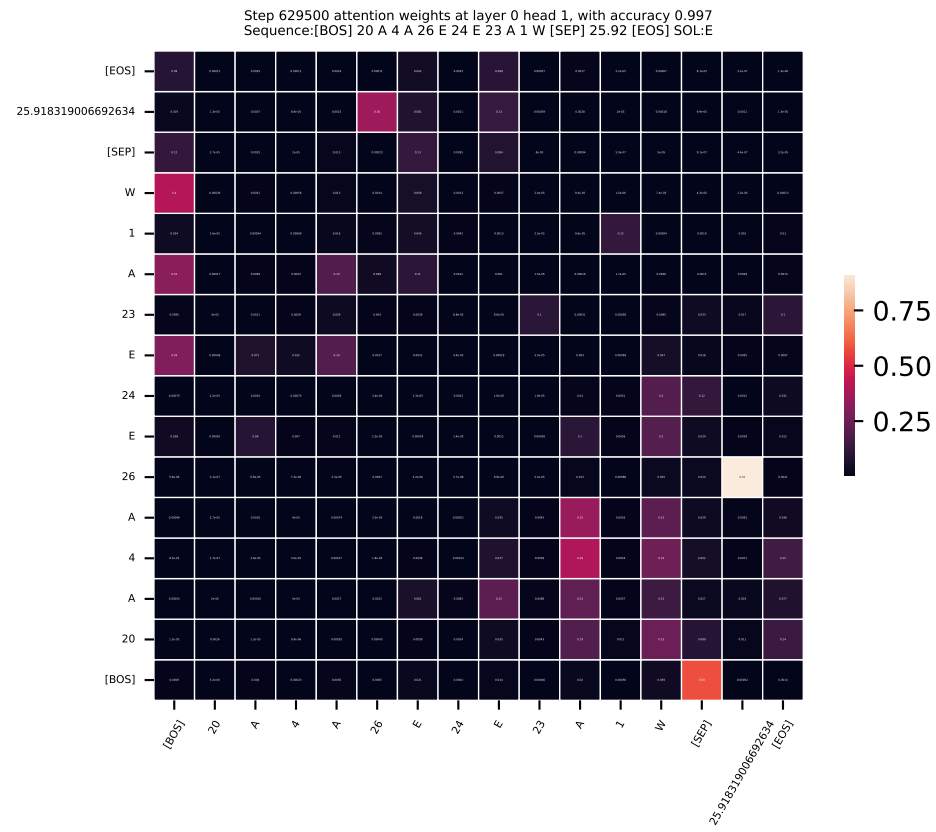

Step 620000 attention weights at layer 1 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

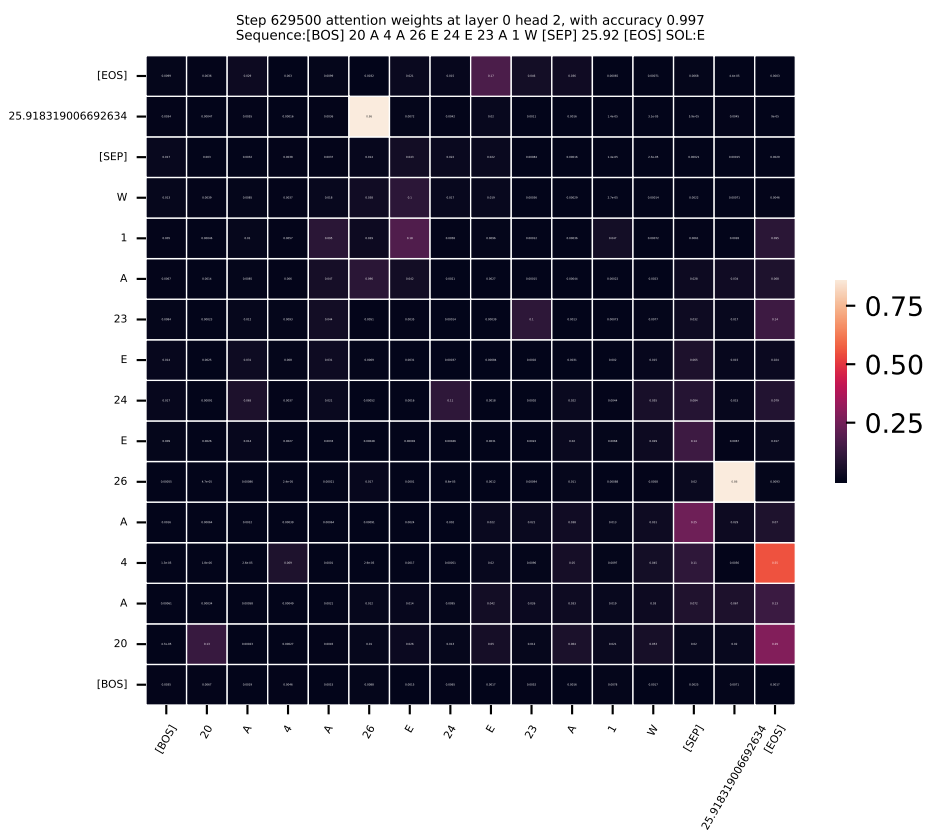

Step 620000 attention weights at layer 1 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

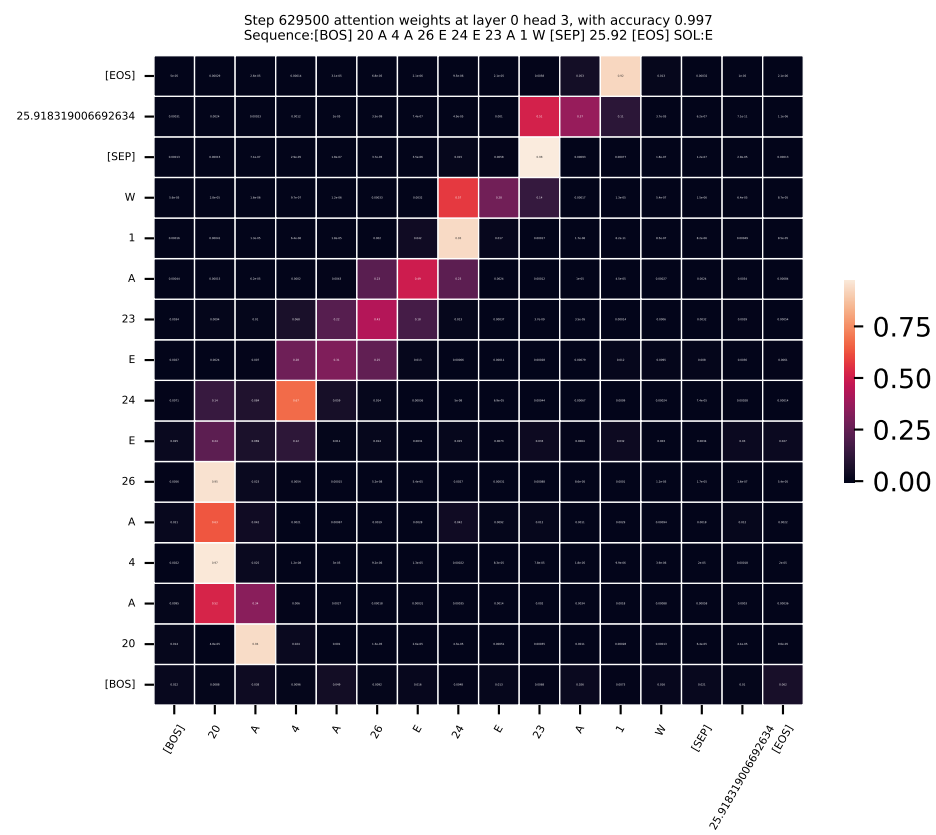

Step 620000 attention weights at layer 1 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

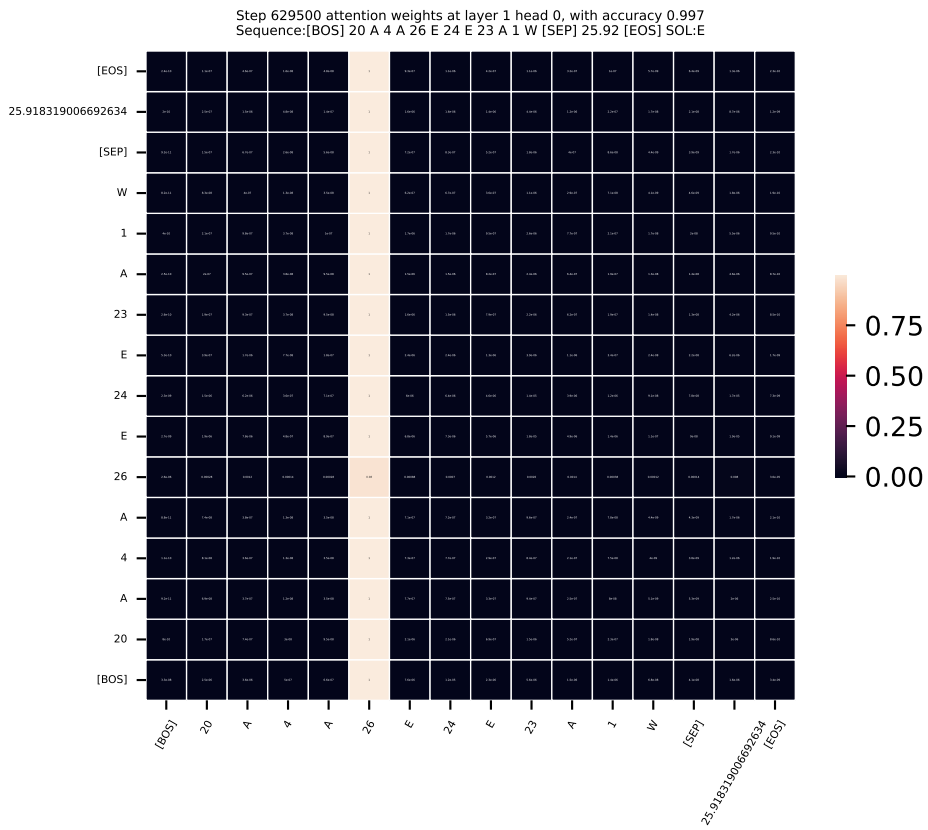

Step 620000 attention weights at layer 1 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

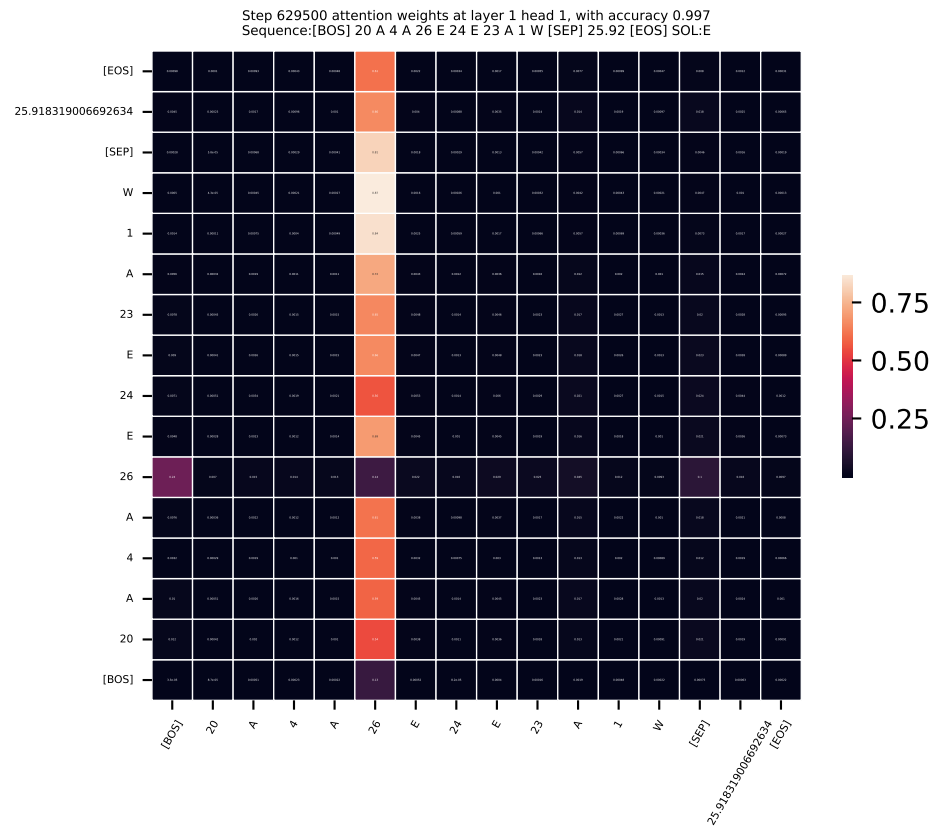

Step 629500 attention weights at layer 0 head 0, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

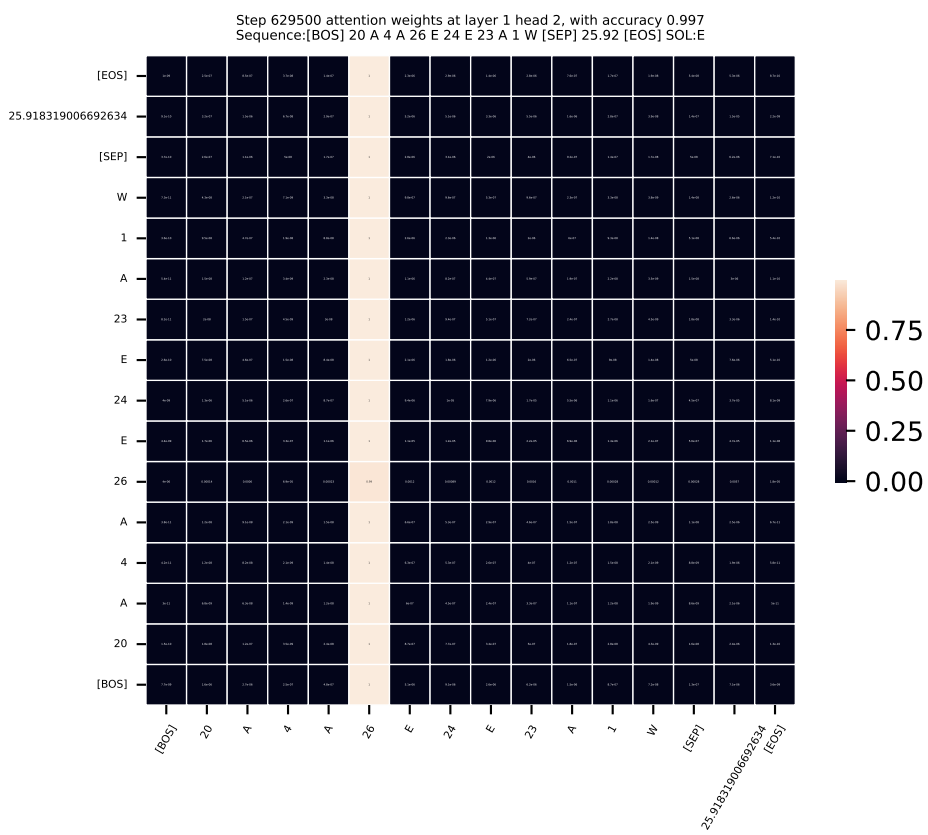

Step 629500 attention weights at layer 0 head 1, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

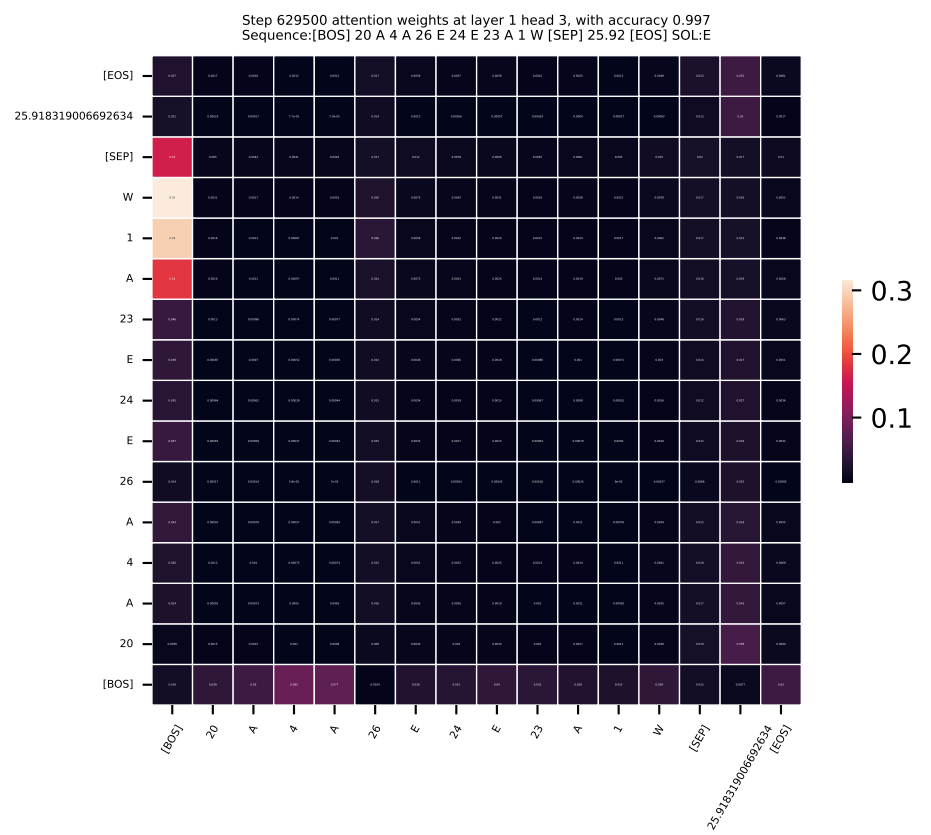

Step 629500 attention weights at layer 0 head 2, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

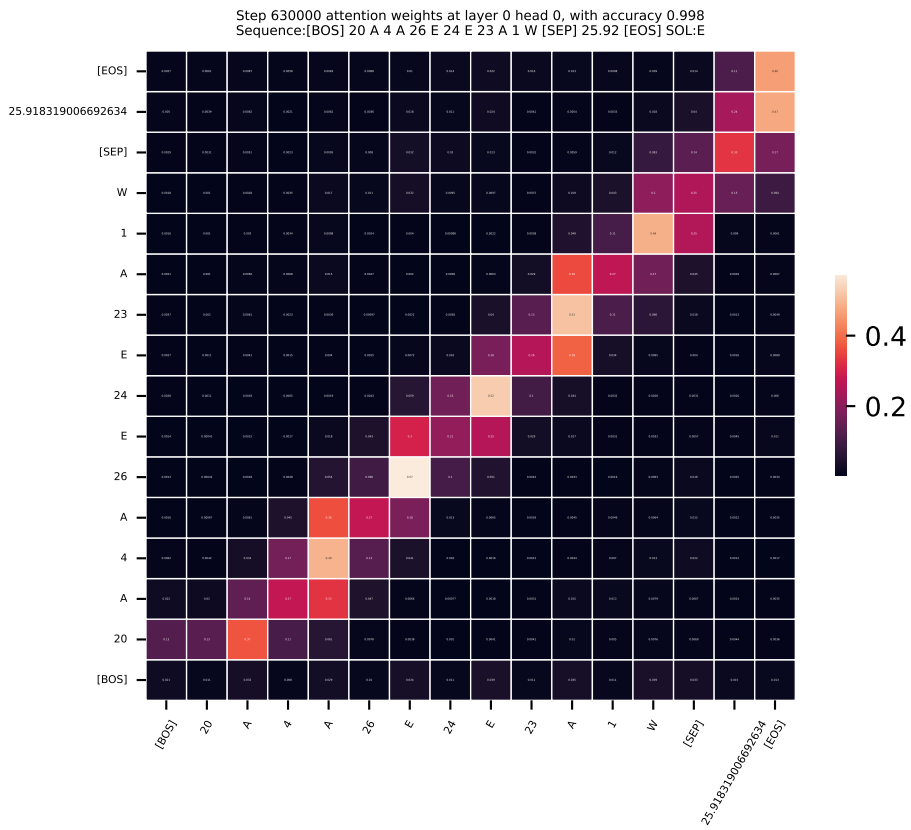

Step 629500 attention weights at layer 0 head 3, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

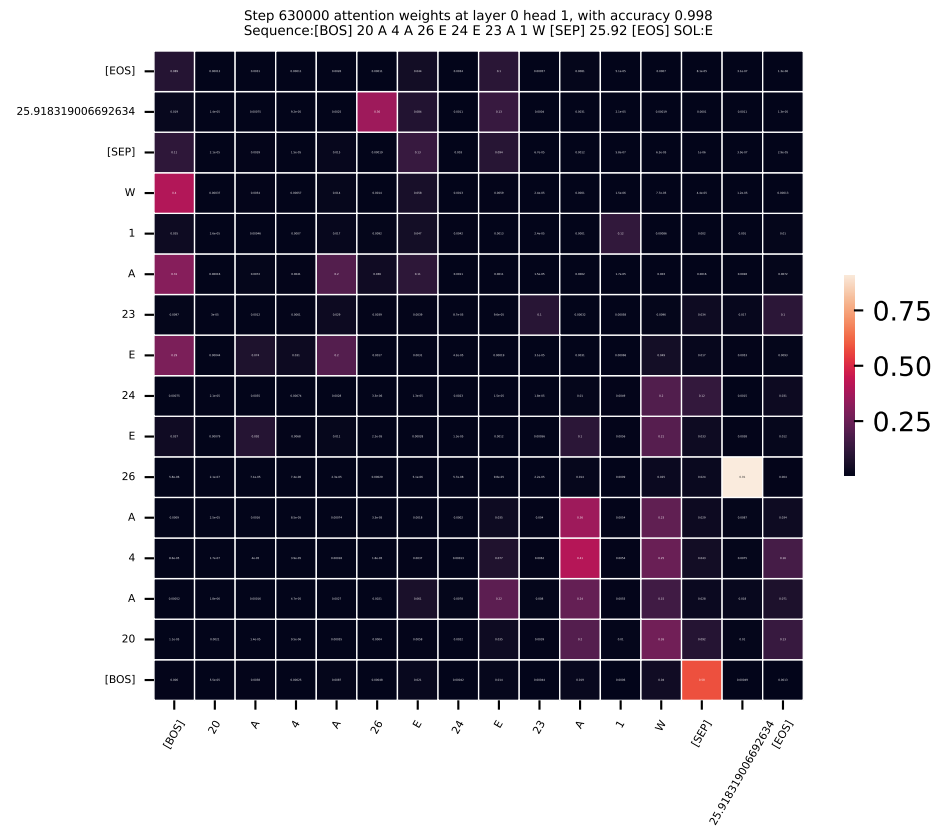

Step 629500 attention weights at layer 1 head 0, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

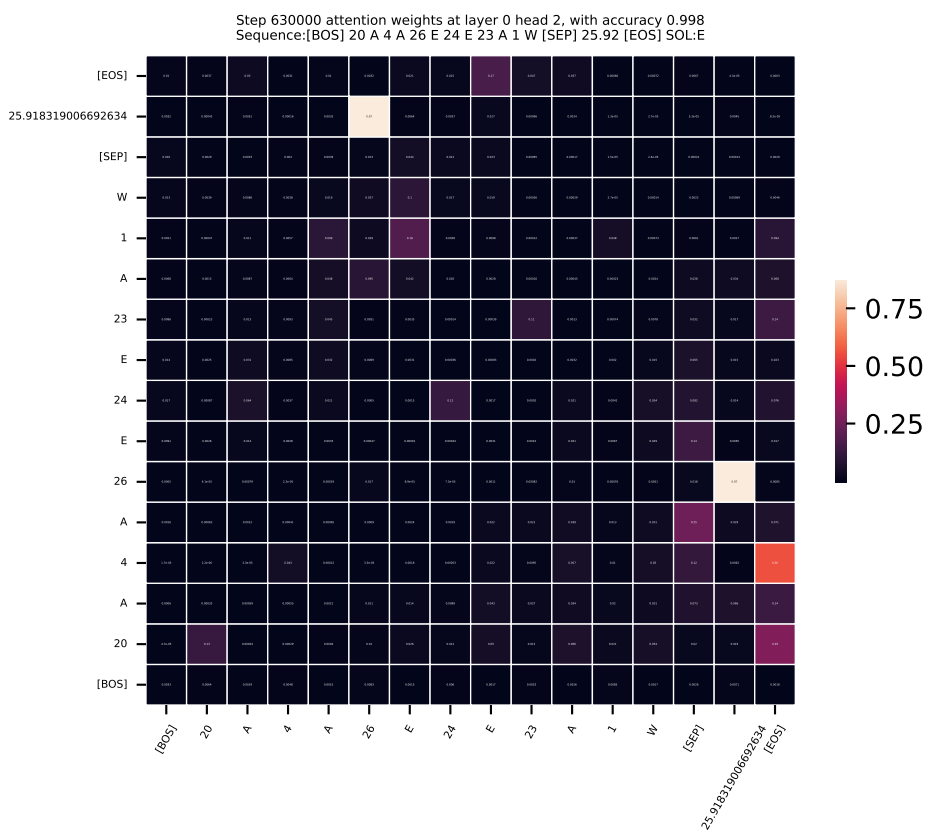

Step 629500 attention weights at layer 1 head 1, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

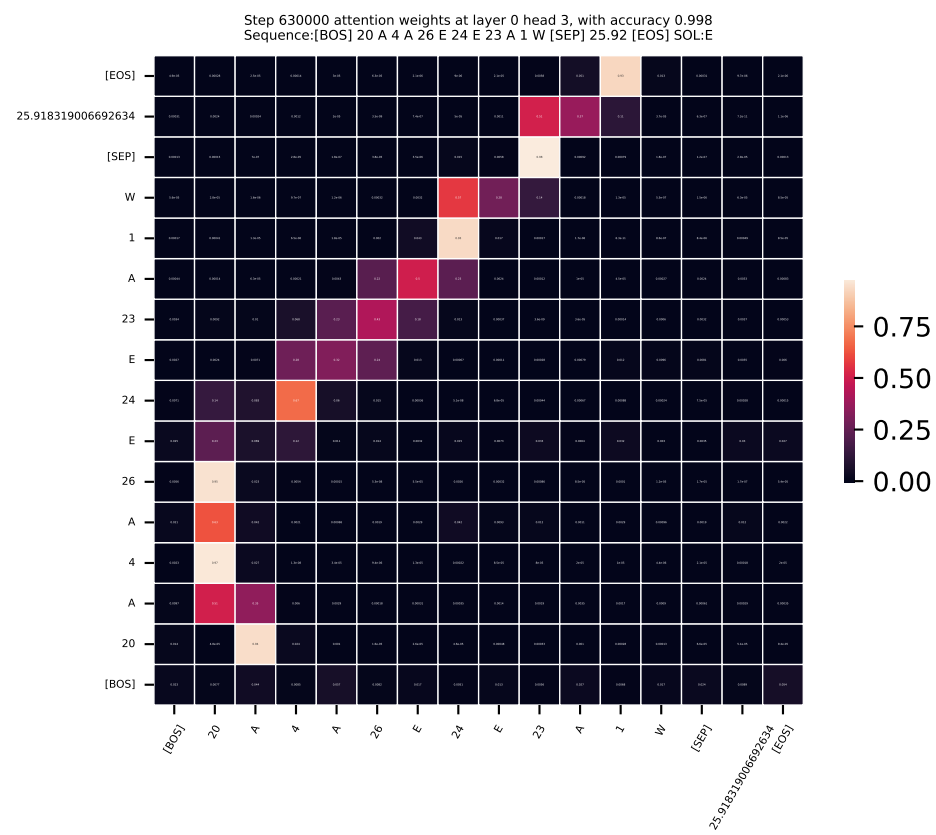

Step 629500 attention weights at layer 1 head 2, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

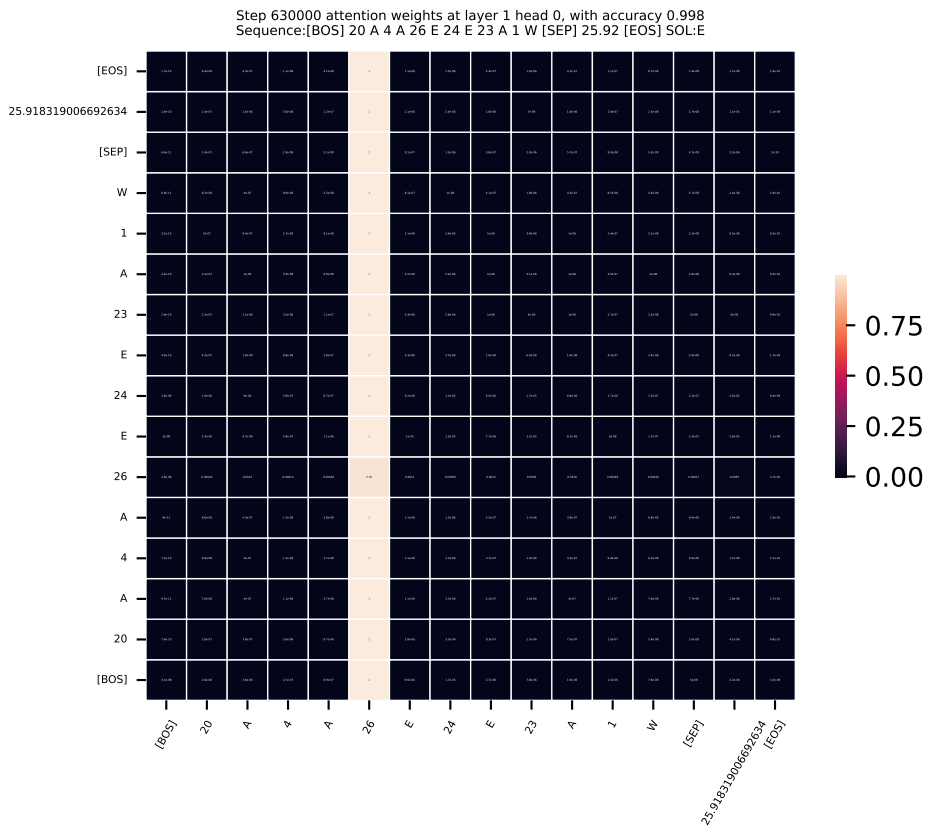

Step 629500 attention weights at layer 1 head 3, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

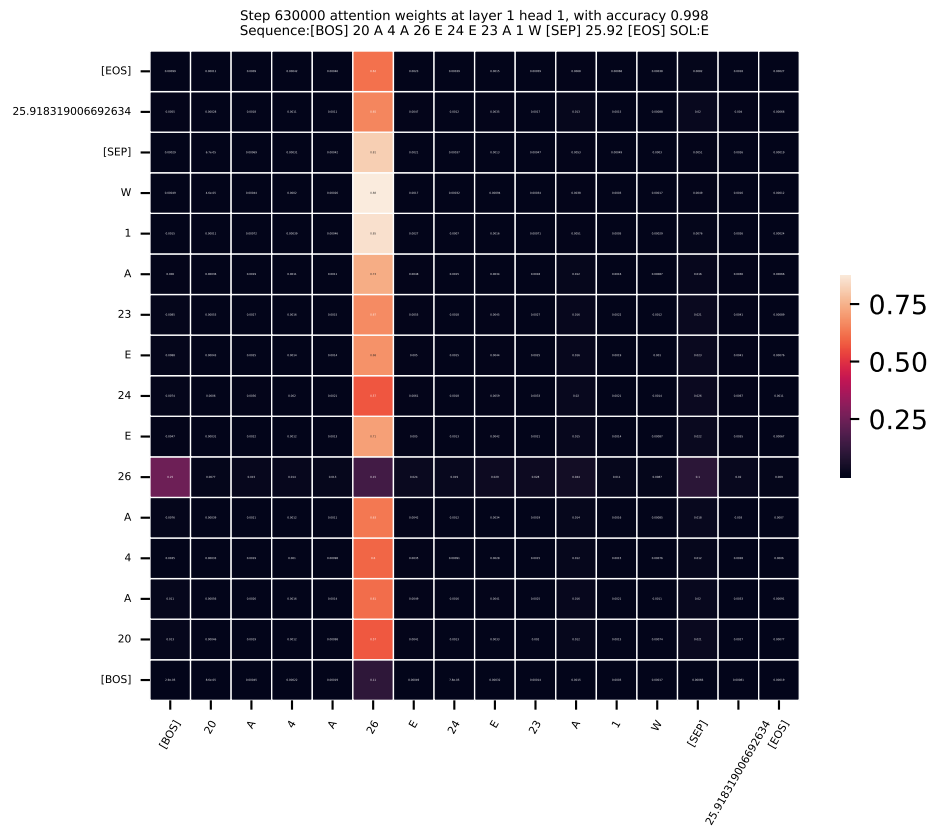

Step 630000 attention weights at layer 0 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

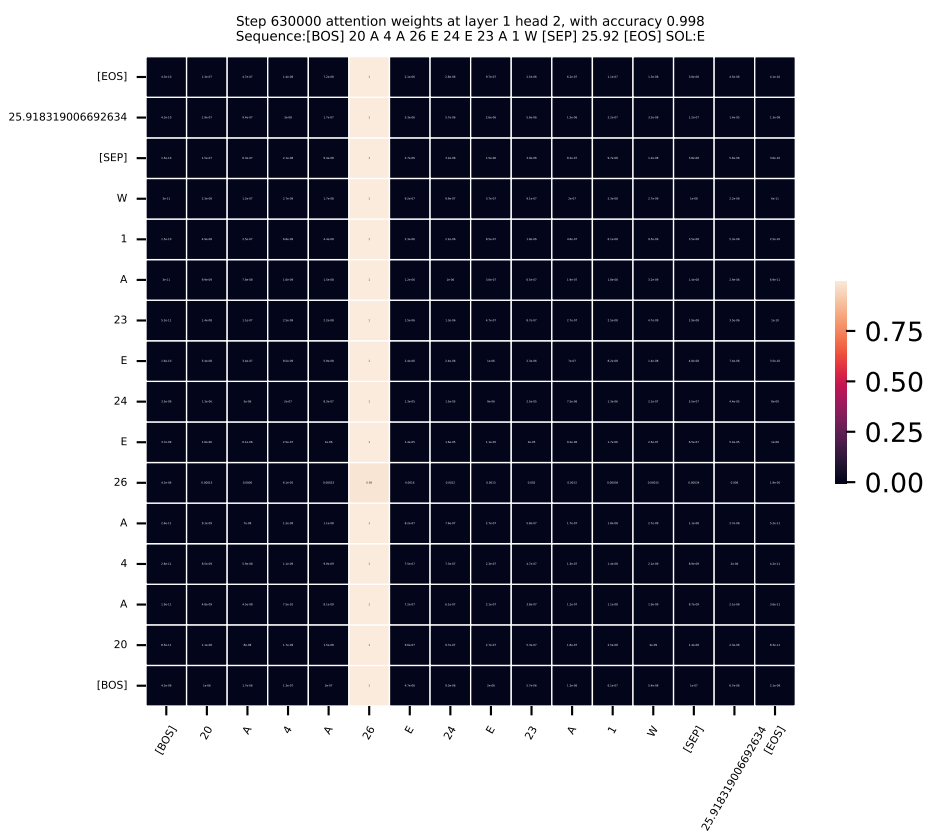

Step 630000 attention weights at layer 0 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

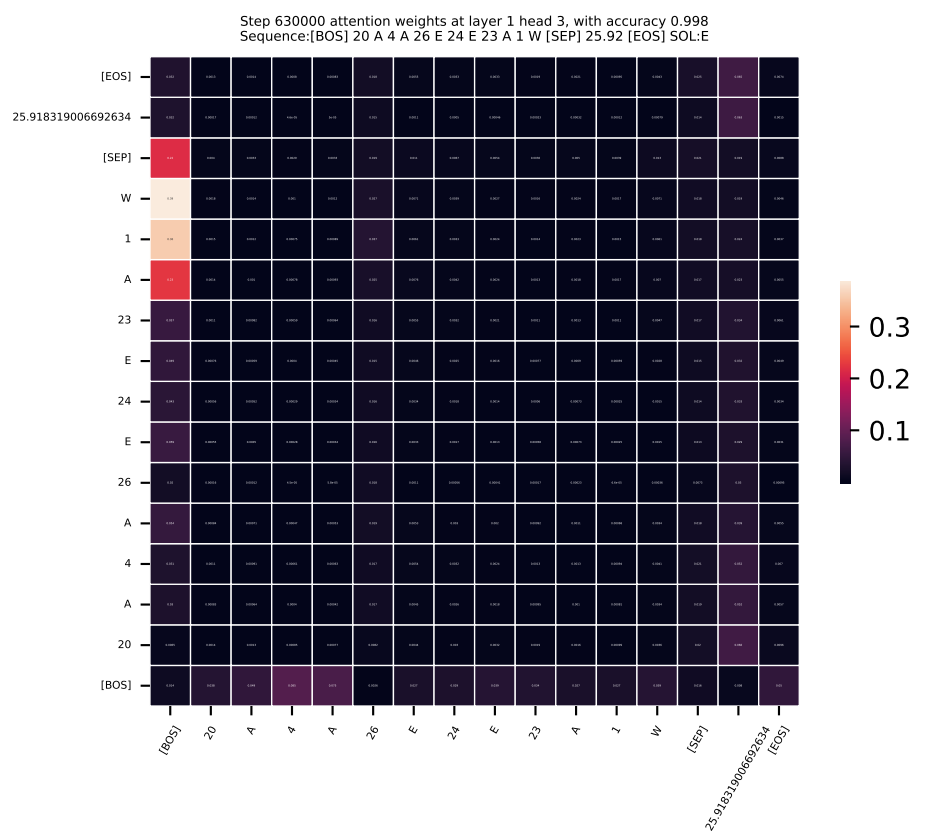

Step 630000 attention weights at layer 0 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

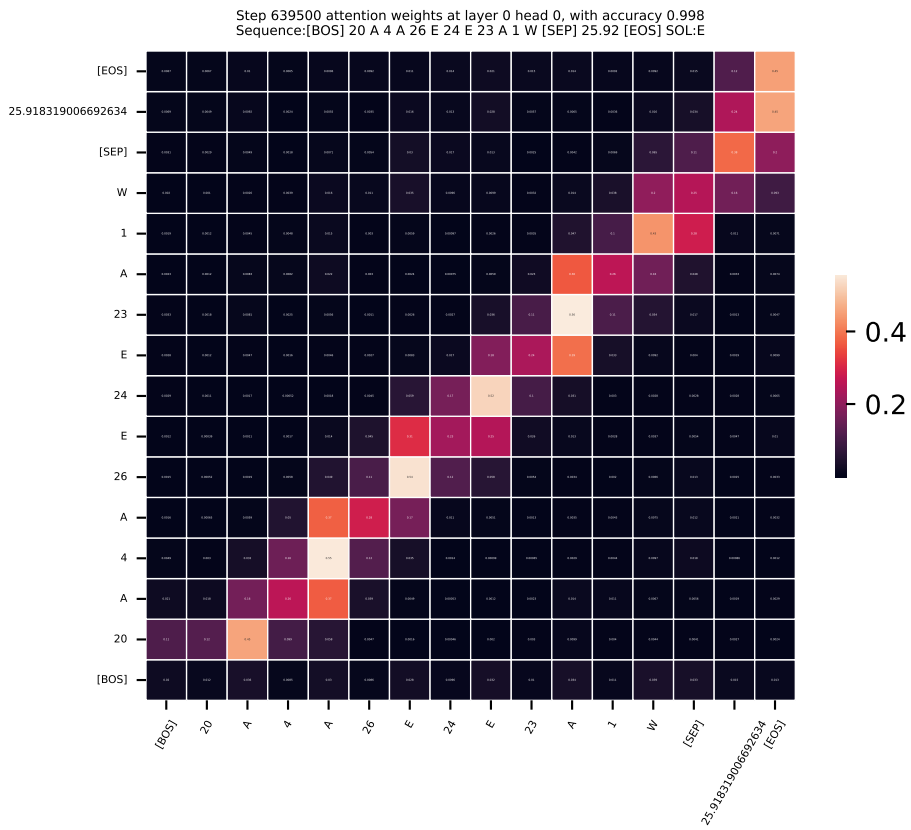

Step 630000 attention weights at layer 0 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

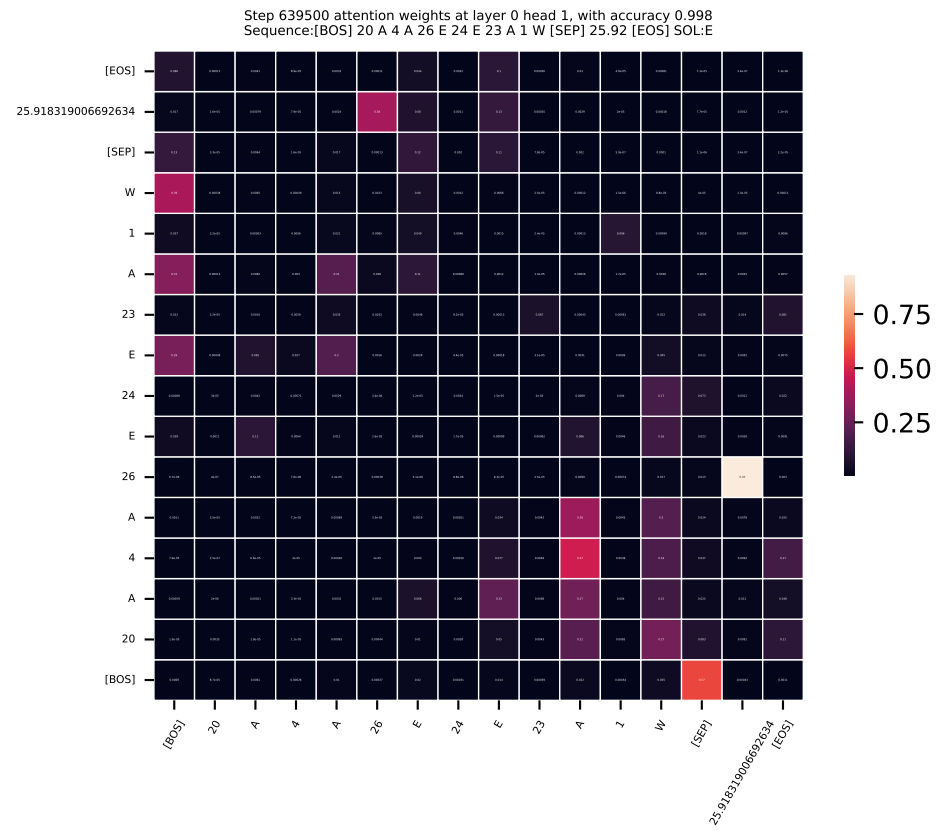

Step 630000 attention weights at layer 1 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

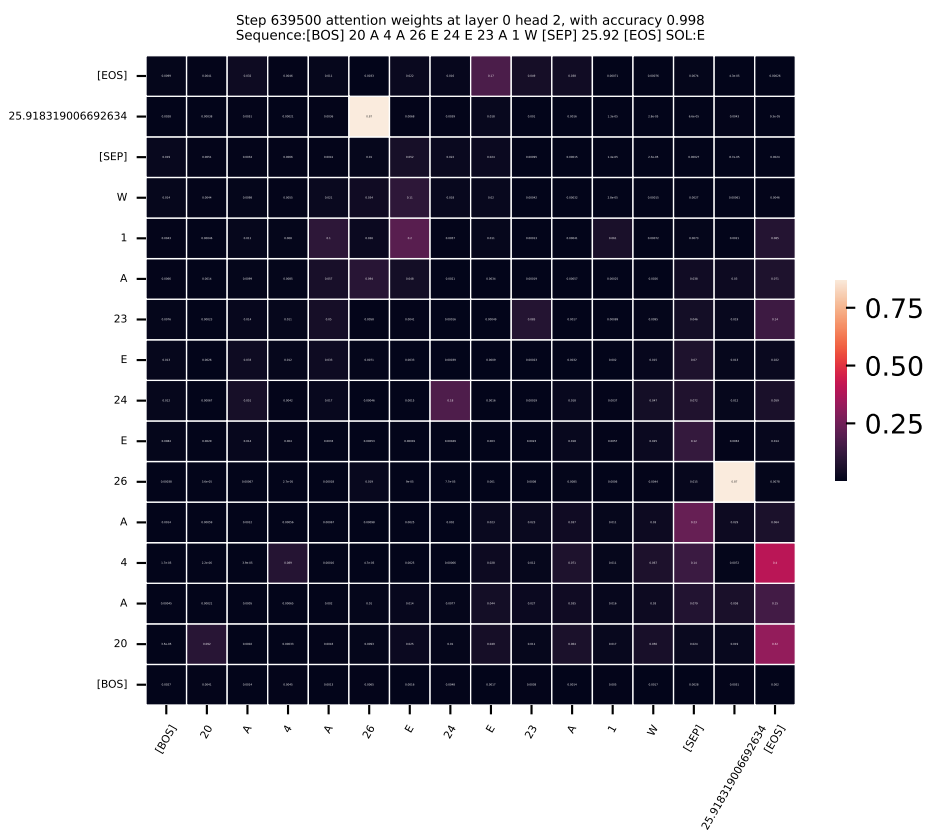

Step 630000 attention weights at layer 1 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

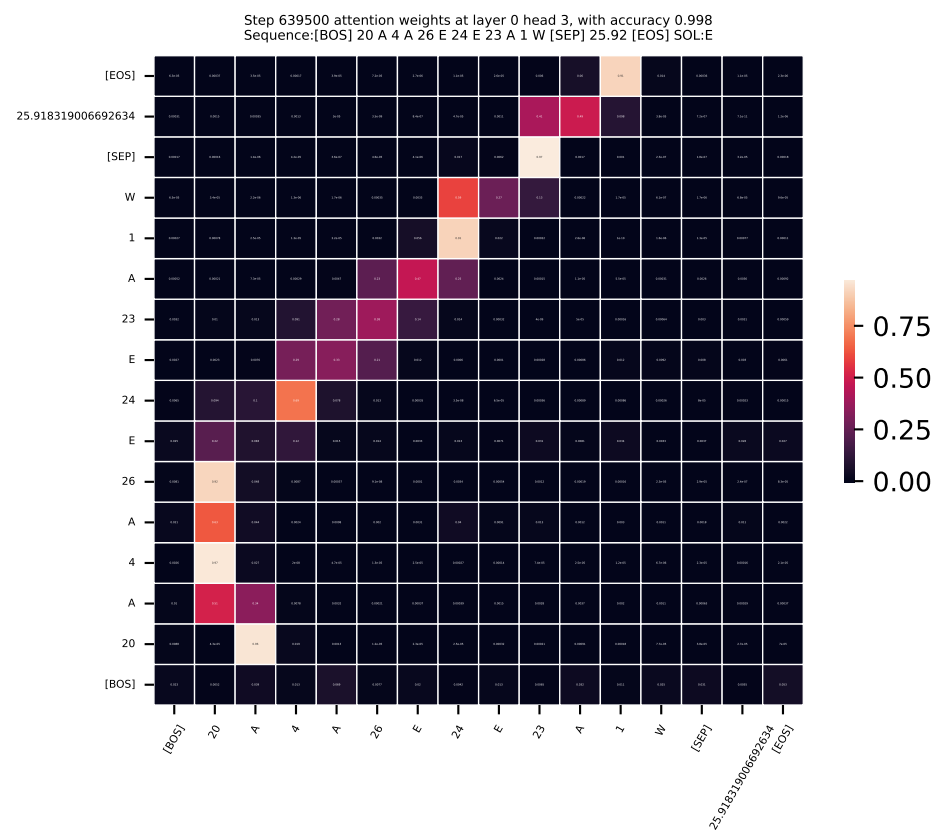

Step 630000 attention weights at layer 1 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

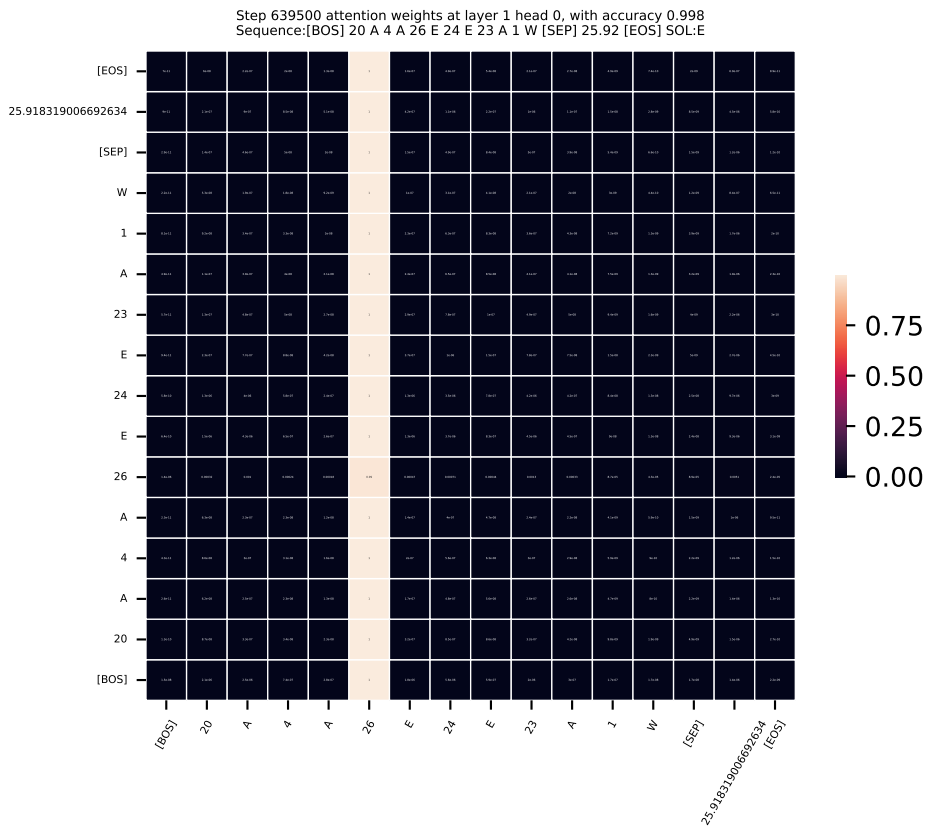

Step 630000 attention weights at layer 1 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

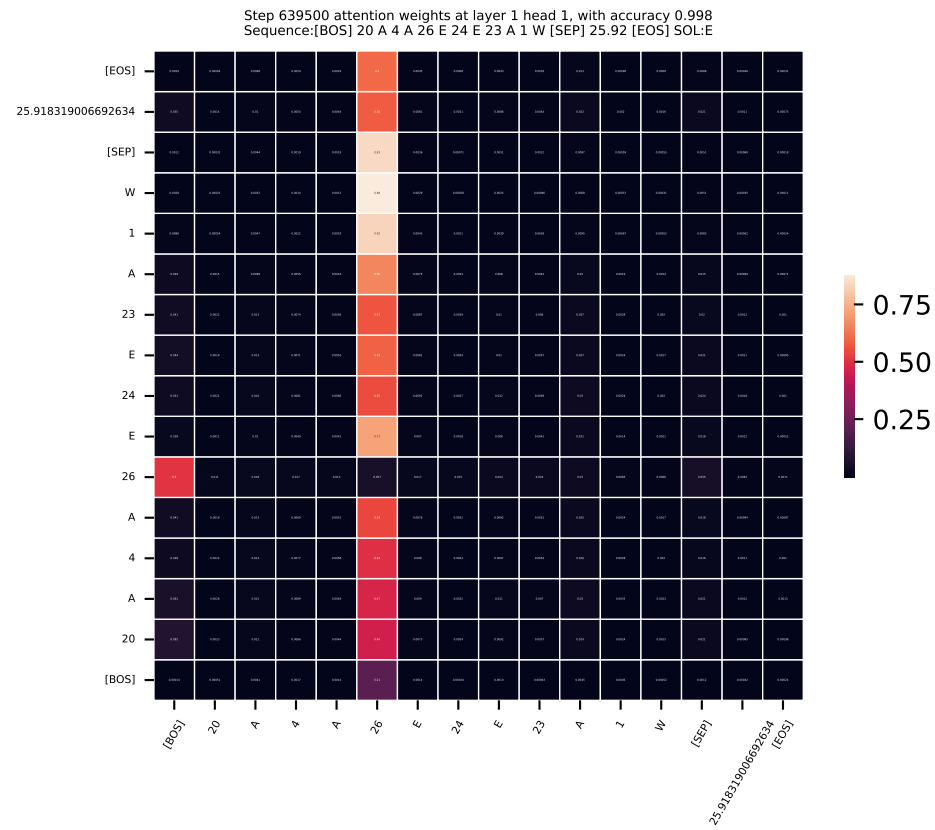

Step 639500 attention weights at layer 0 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

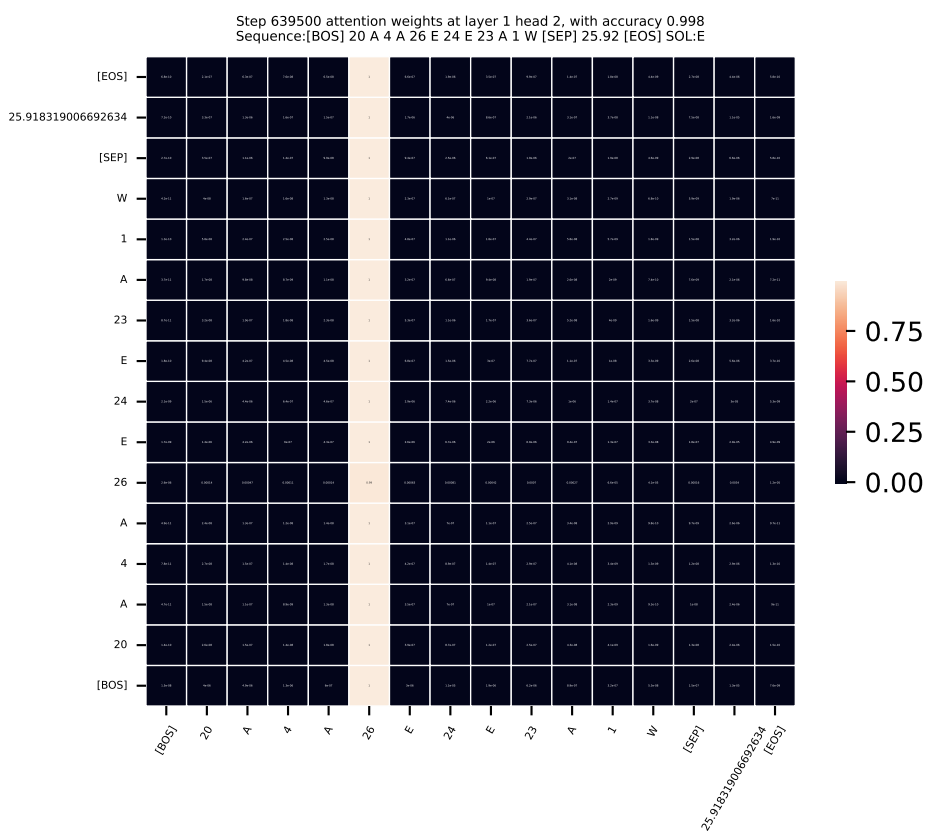

Step 639500 attention weights at layer 0 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

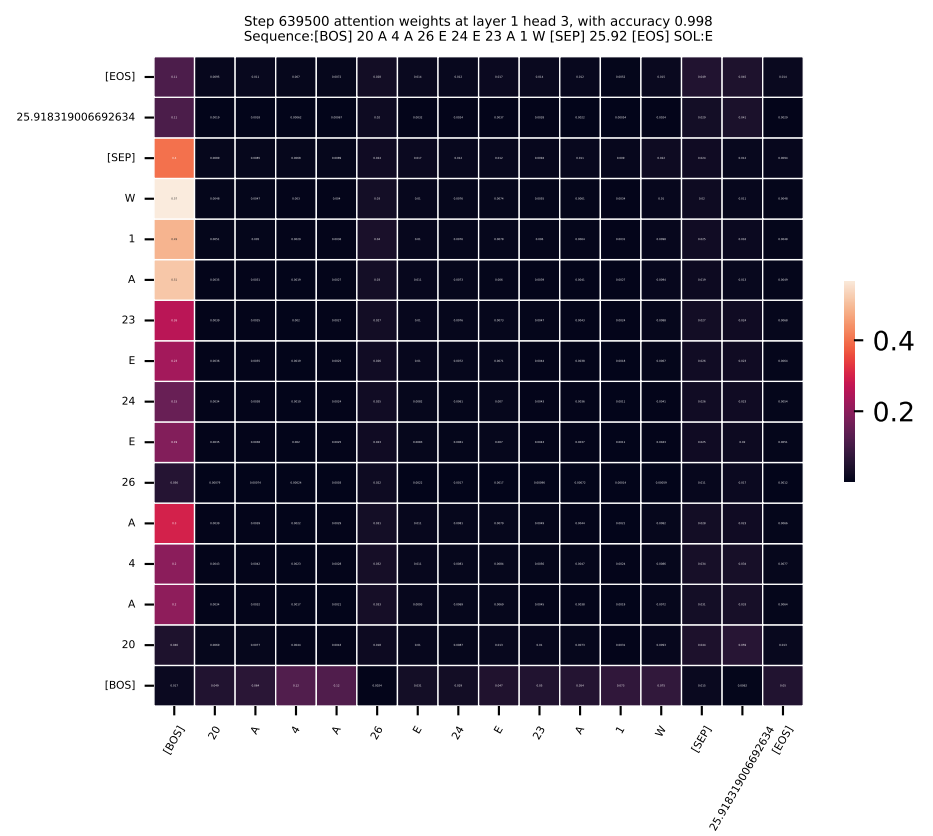

Step 639500 attention weights at layer 0 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

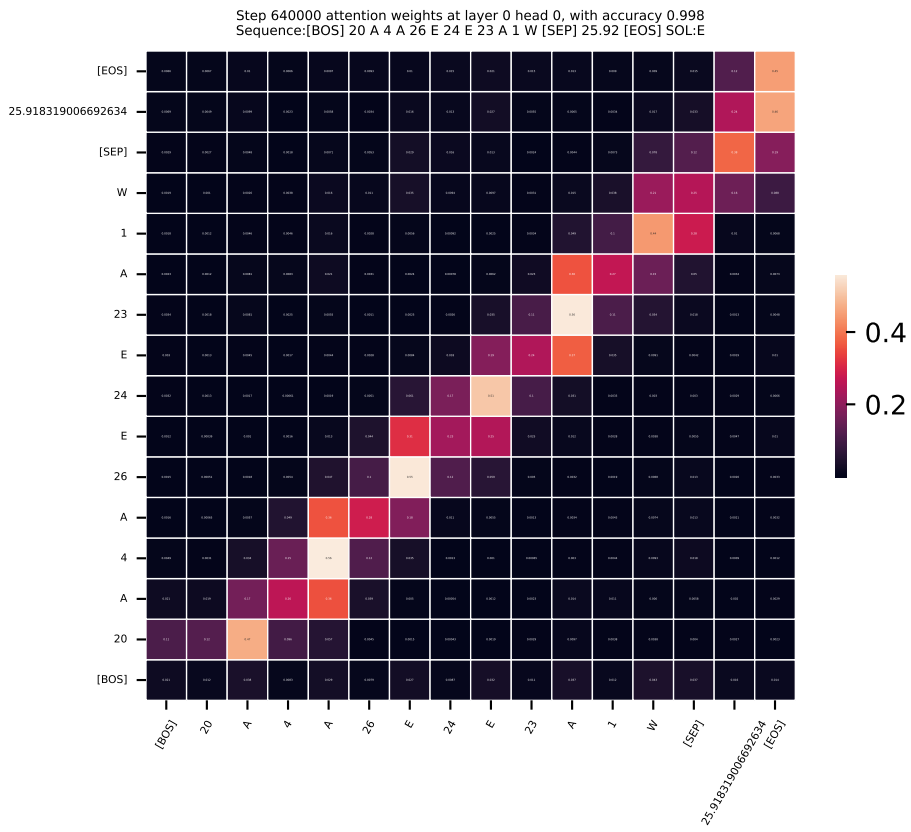

Step 639500 attention weights at layer 0 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

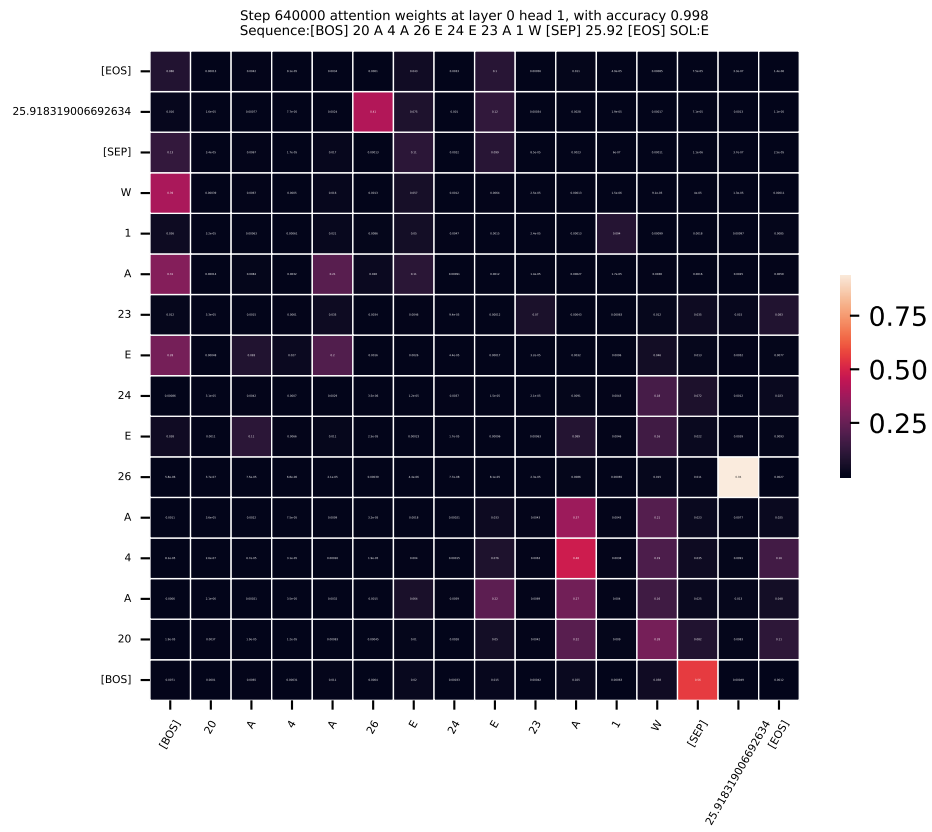

Step 639500 attention weights at layer 1 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

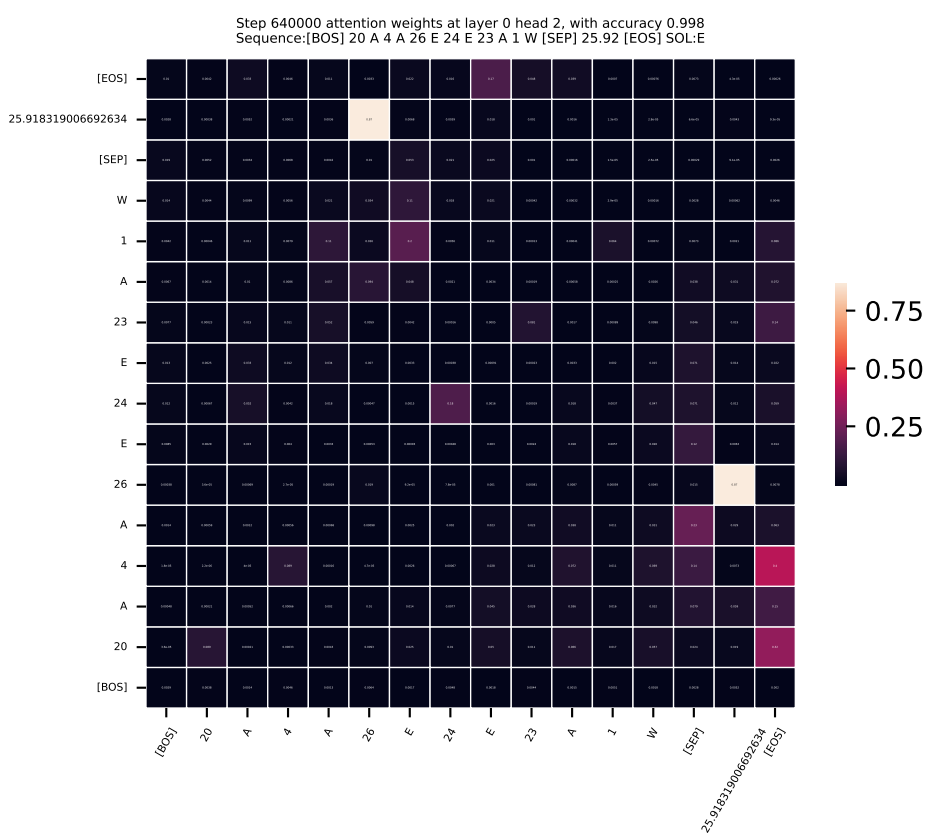

Step 639500 attention weights at layer 1 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

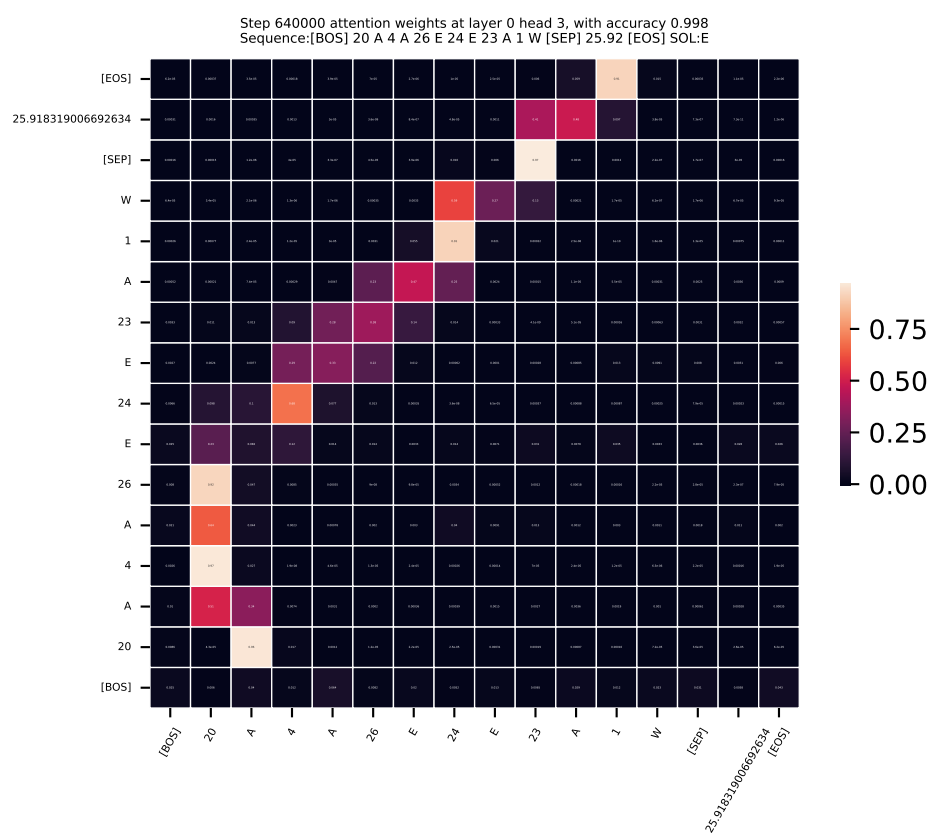

Step 639500 attention weights at layer 1 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

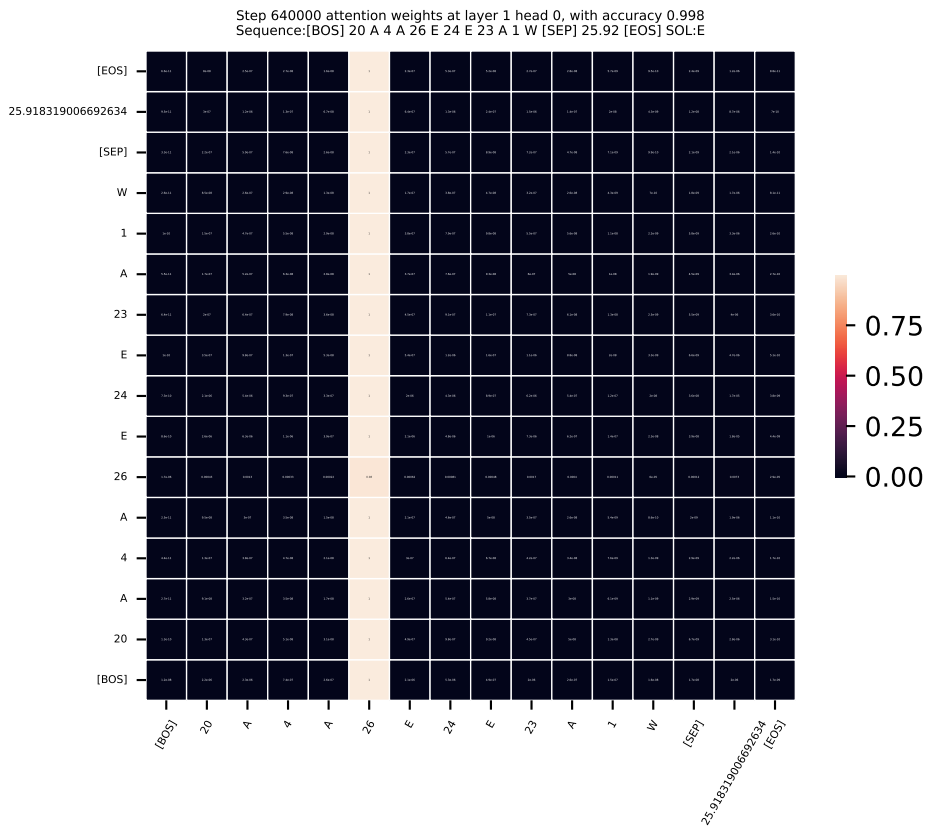

Step 639500 attention weights at layer 1 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

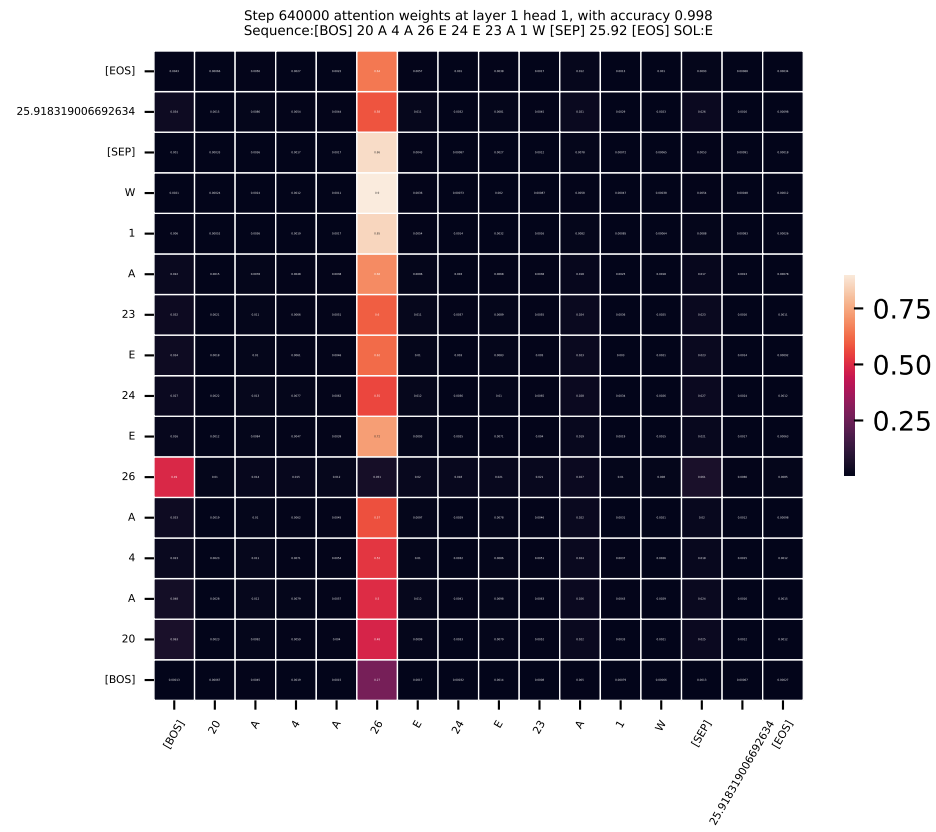

Step 640000 attention weights at layer 0 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

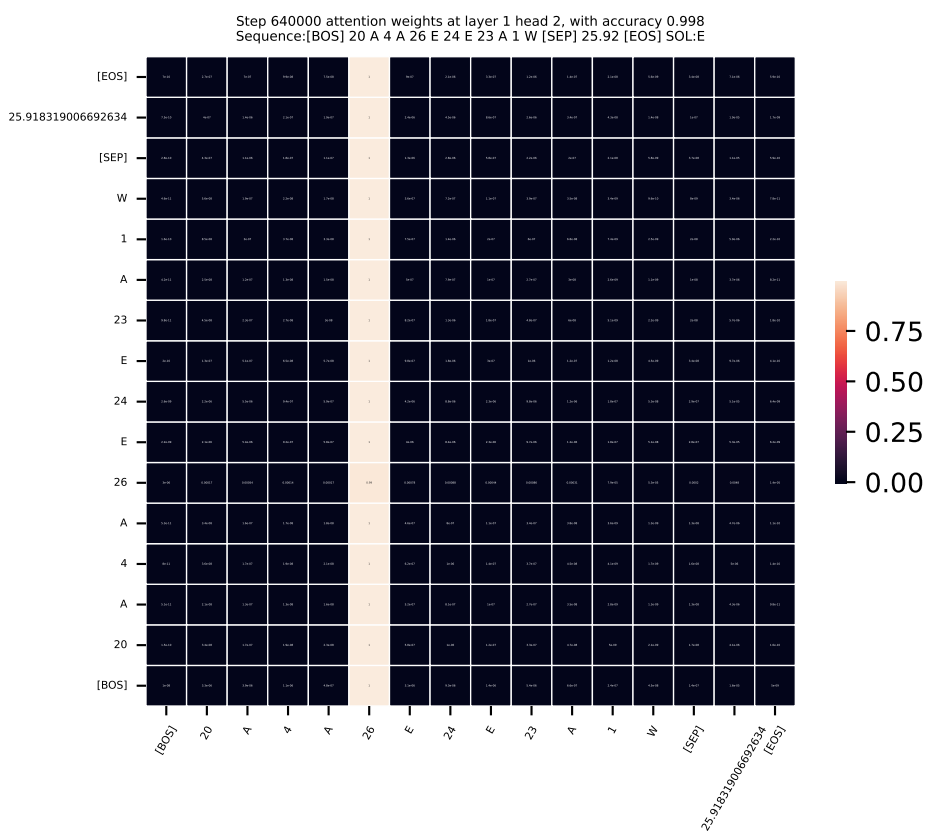

Step 640000 attention weights at layer 0 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

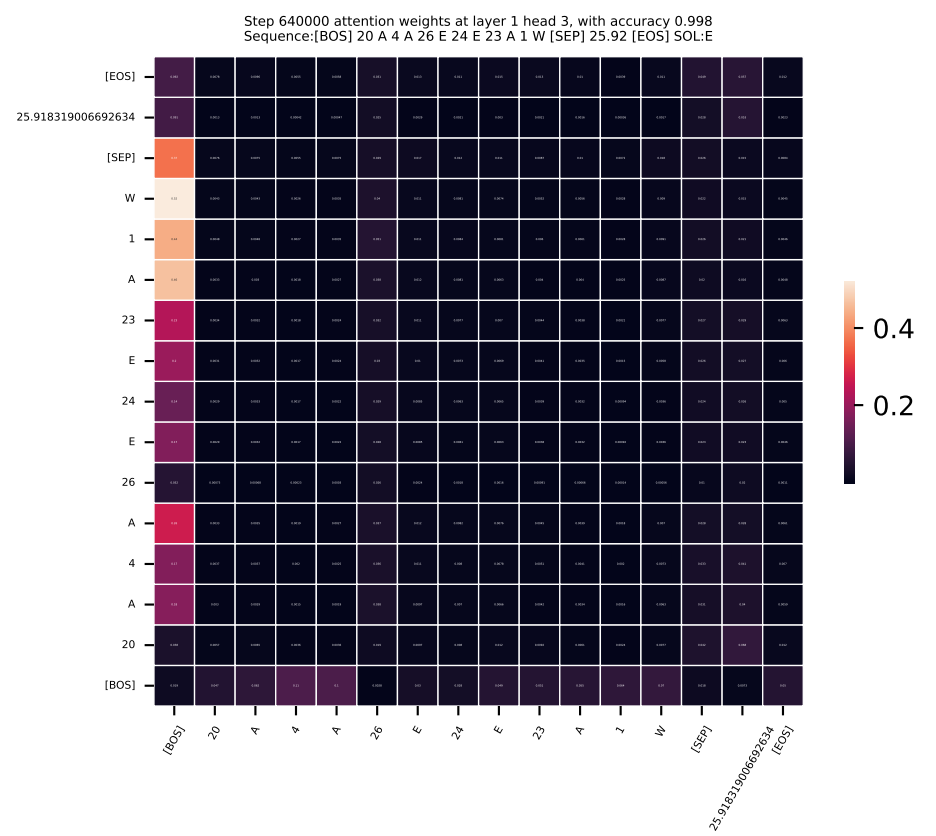

Step 640000 attention weights at layer 0 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

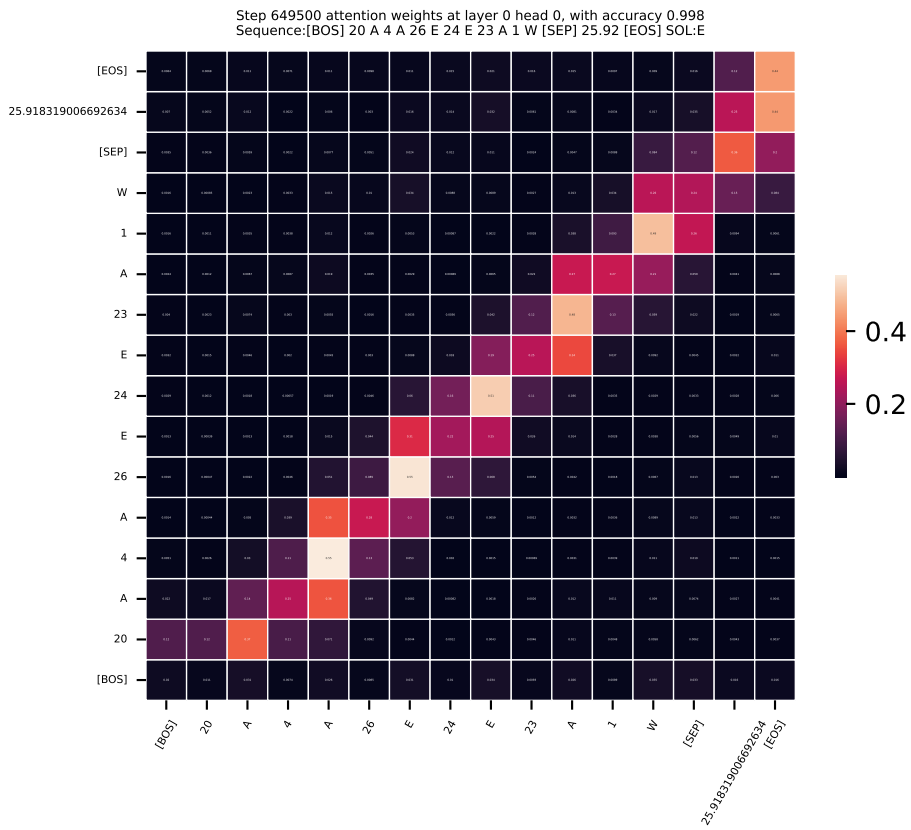

Step 640000 attention weights at layer 0 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

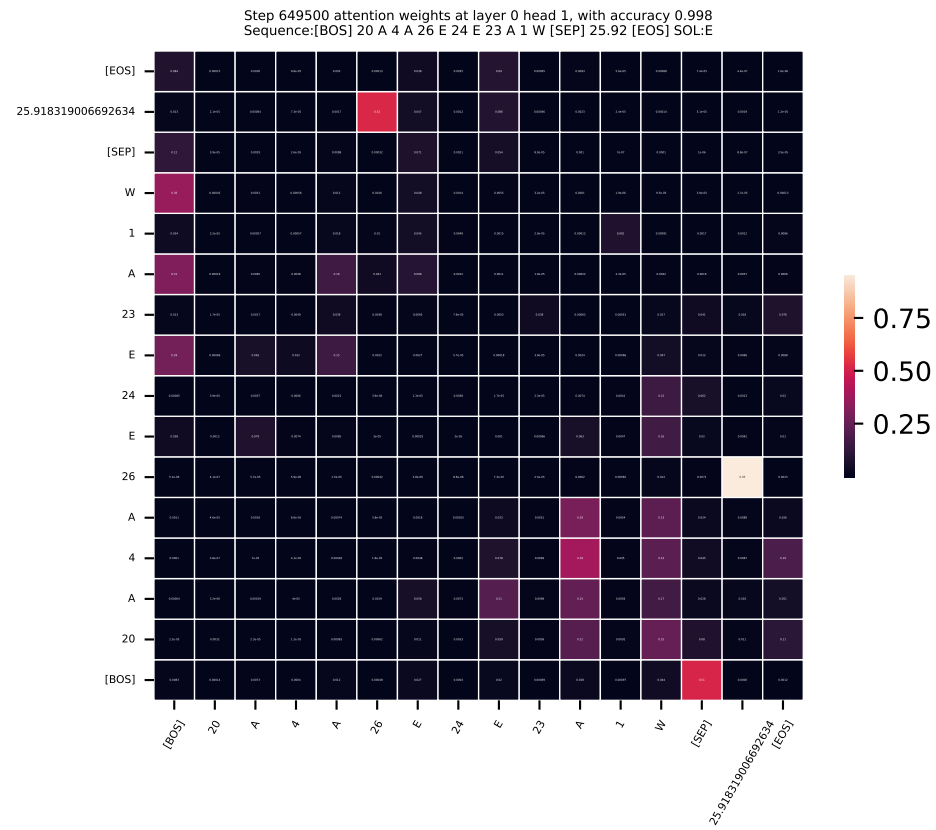

Step 640000 attention weights at layer 1 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

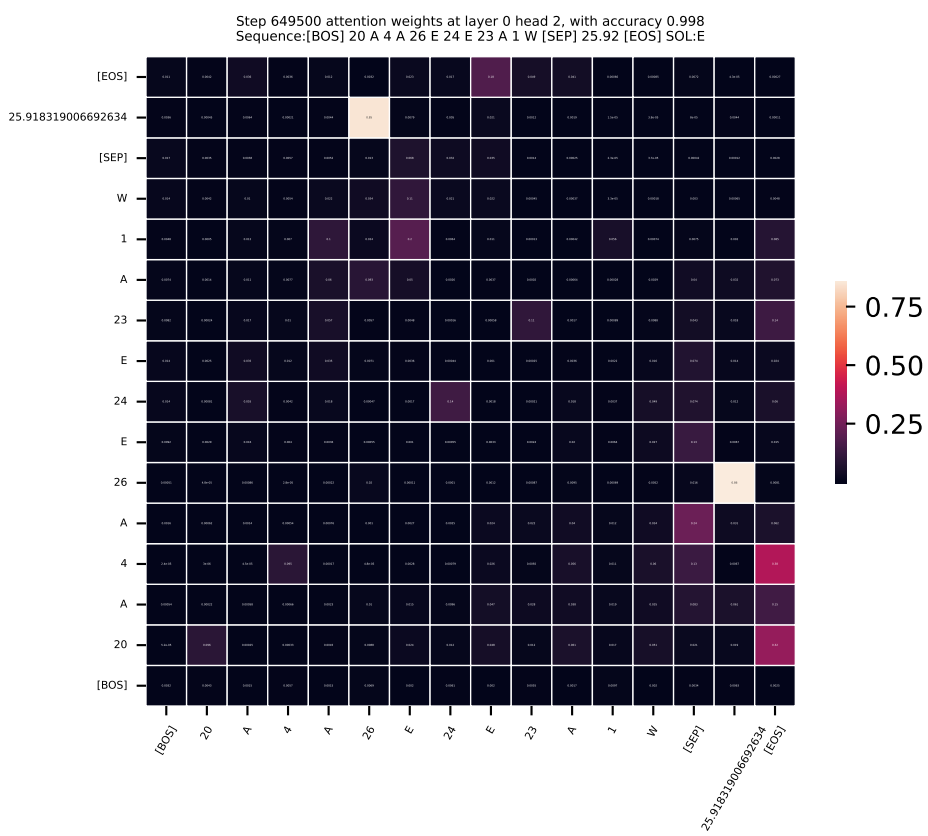

Step 640000 attention weights at layer 1 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

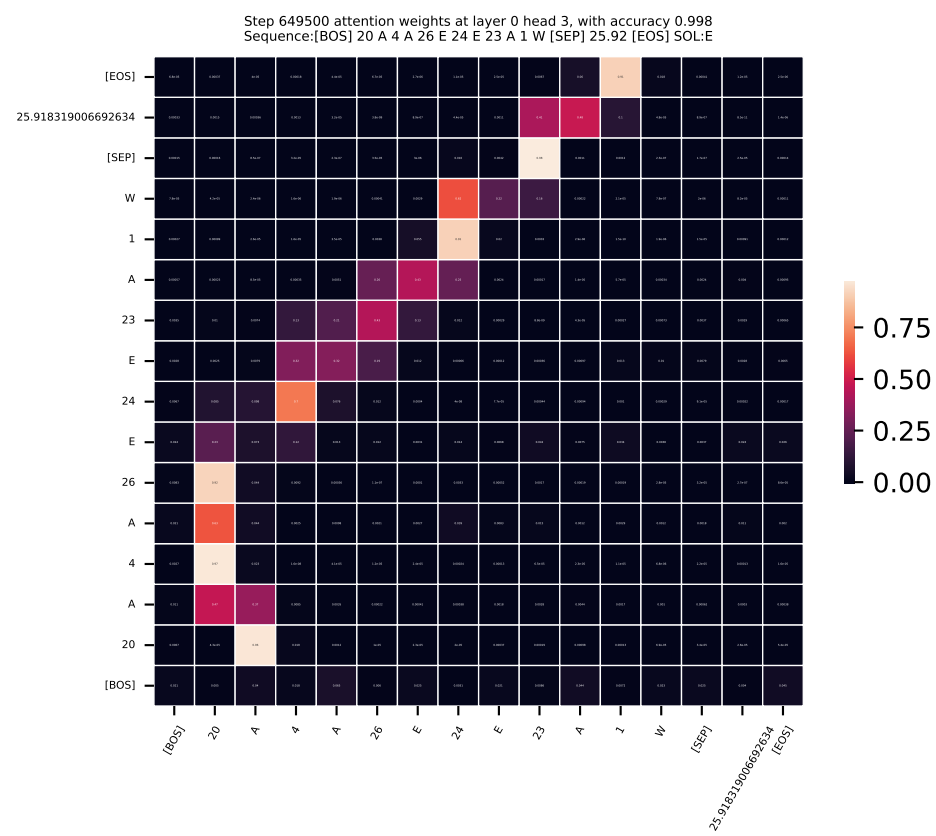

Step 640000 attention weights at layer 1 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

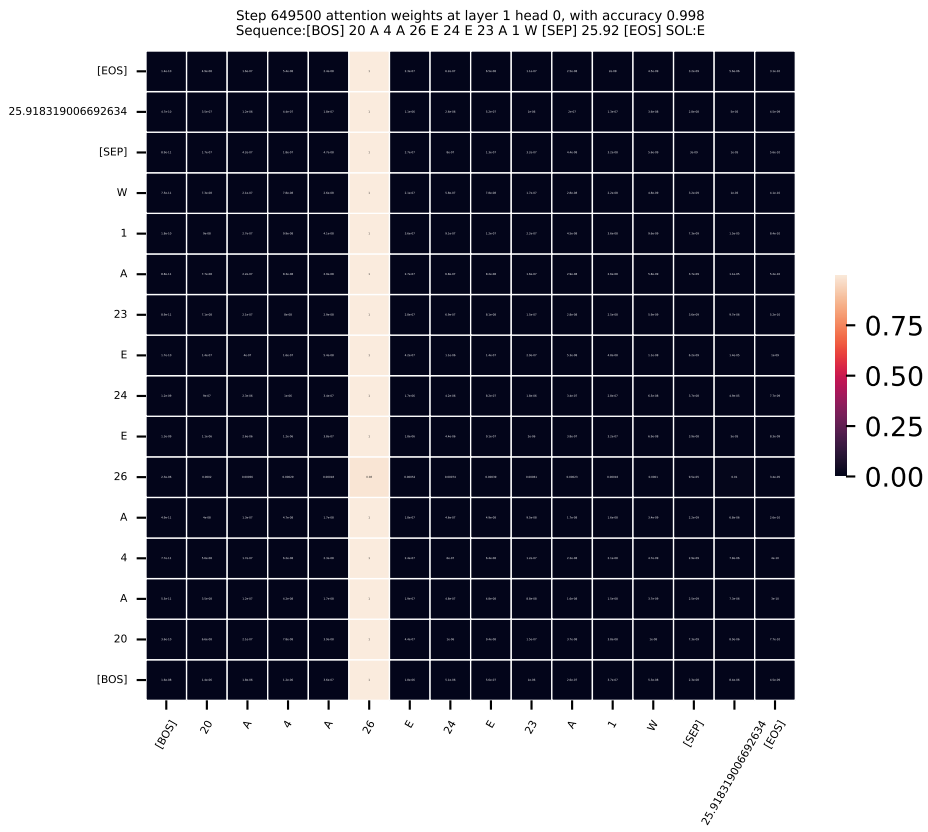

Step 640000 attention weights at layer 1 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

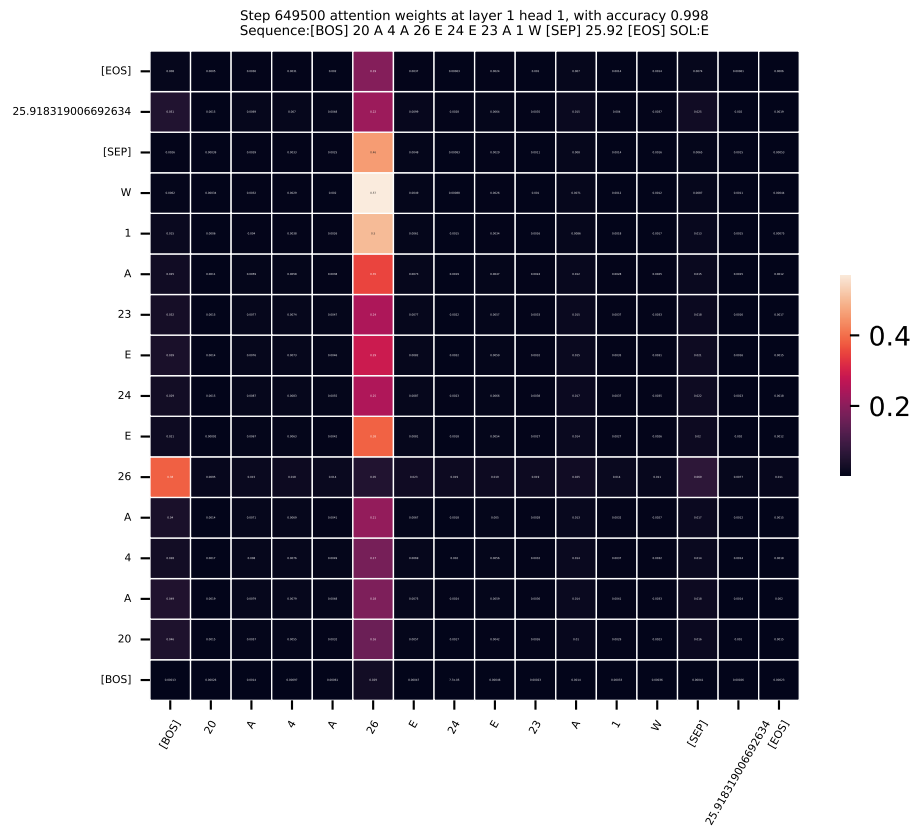

Step 649500 attention weights at layer 0 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

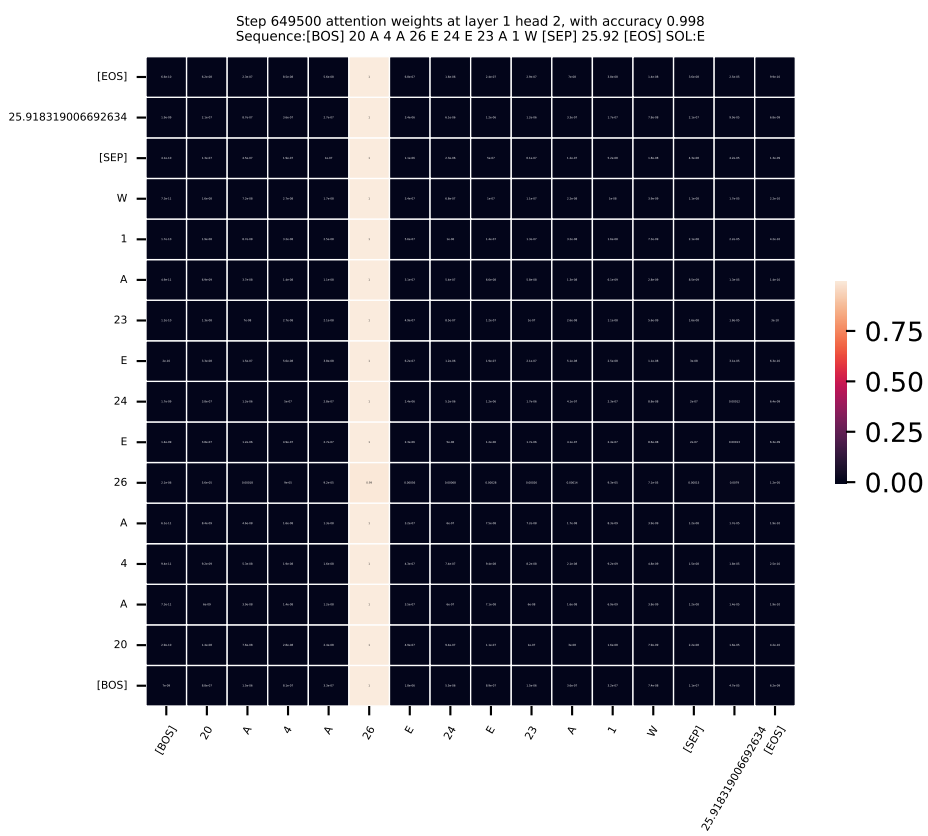

Step 649500 attention weights at layer 0 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

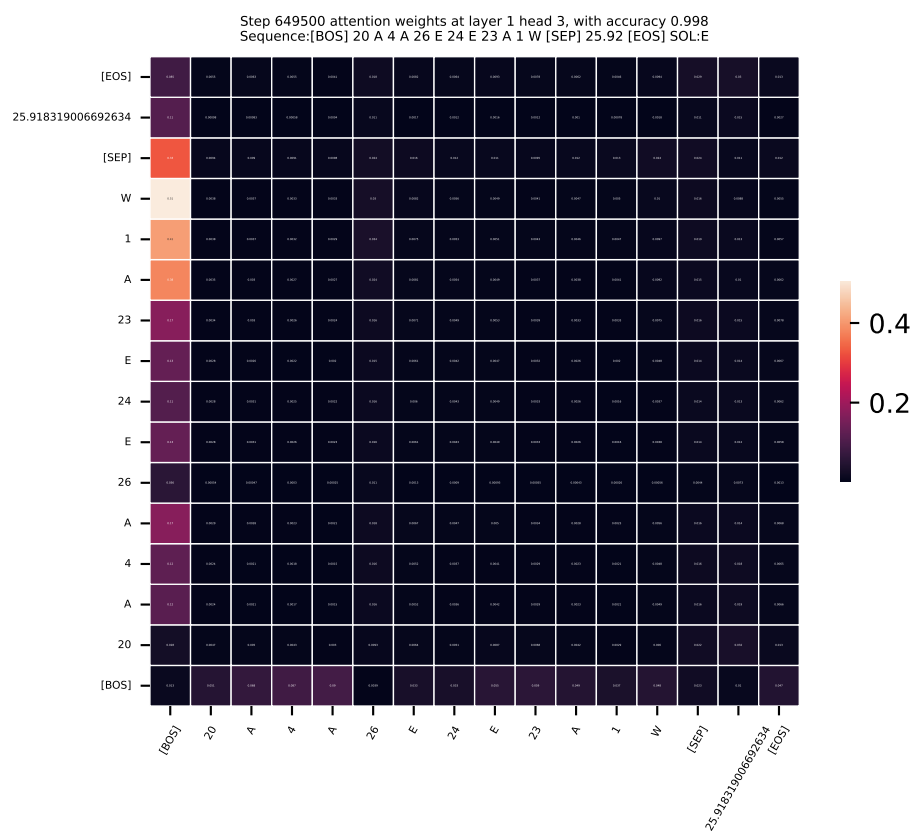

Step 649500 attention weights at layer 0 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

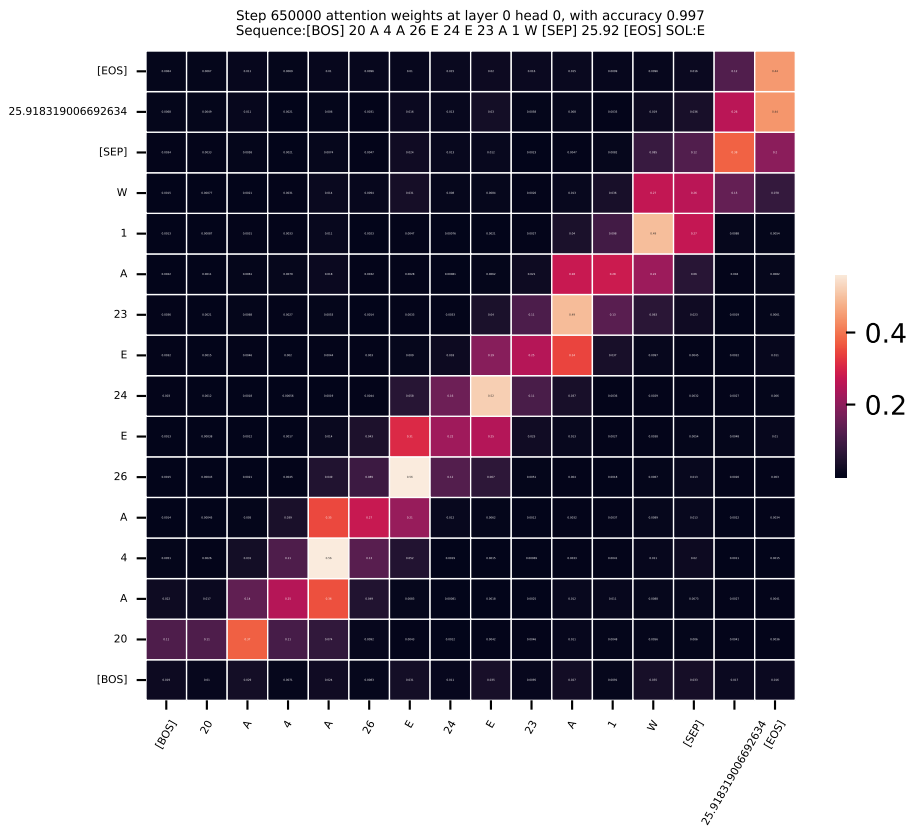

Step 649500 attention weights at layer 0 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25
0.00

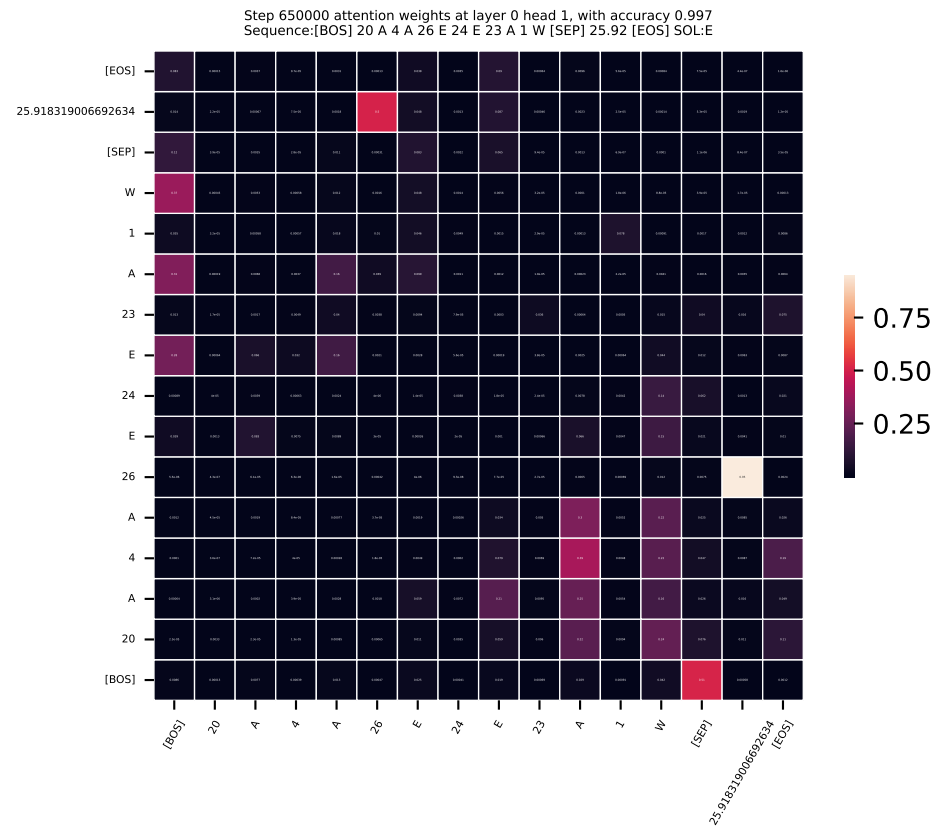

Step 649500 attention weights at layer 1 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

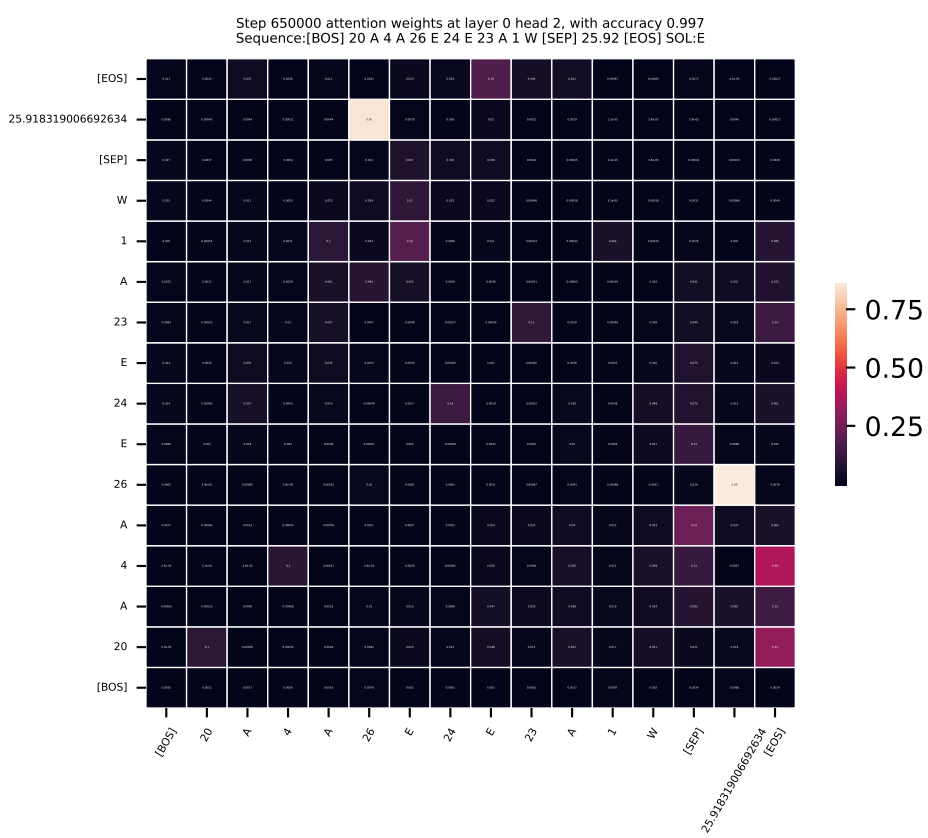

Step 649500 attention weights at layer 1 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

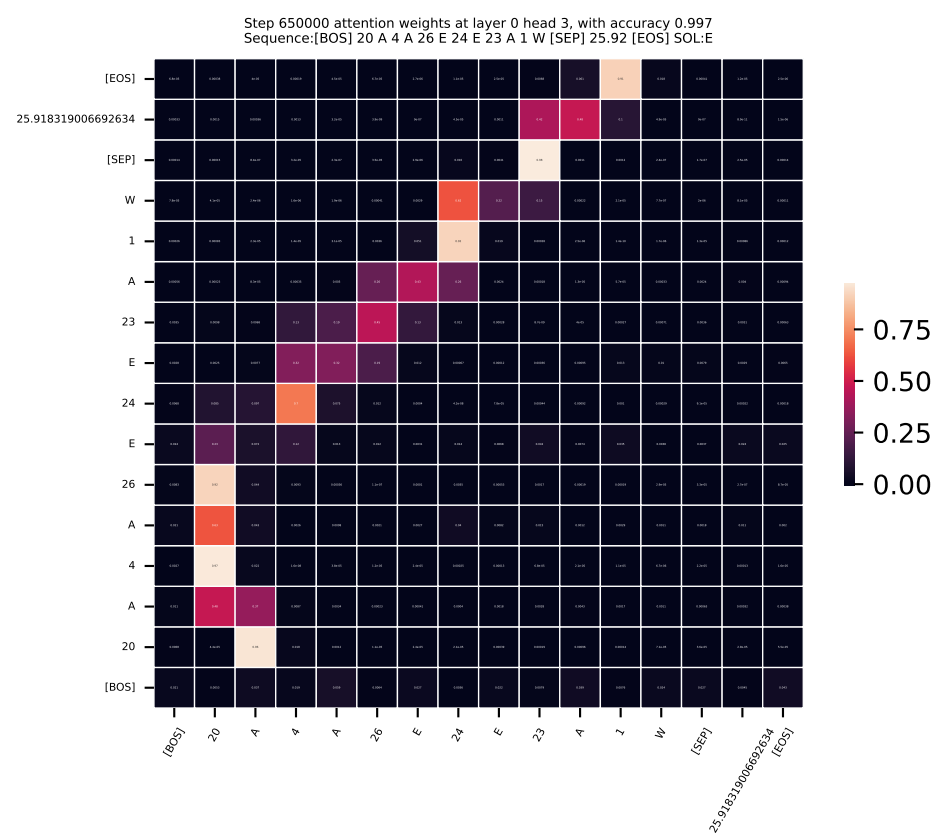

Step 649500 attention weights at layer 1 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

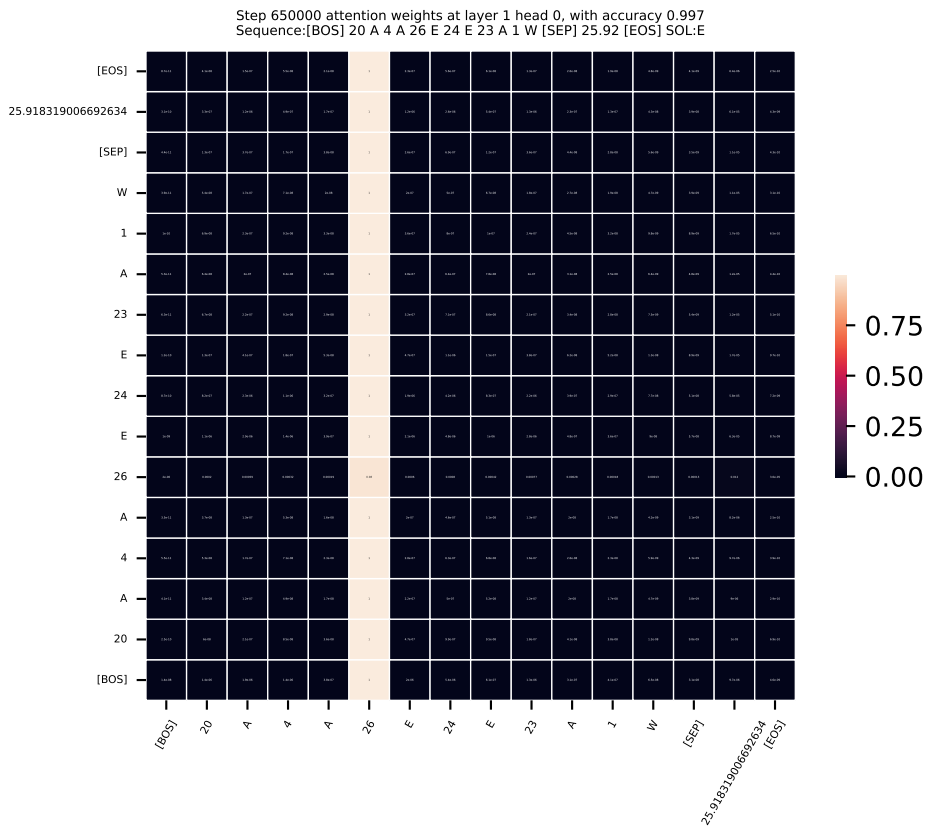

Step 649500 attention weights at layer 1 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

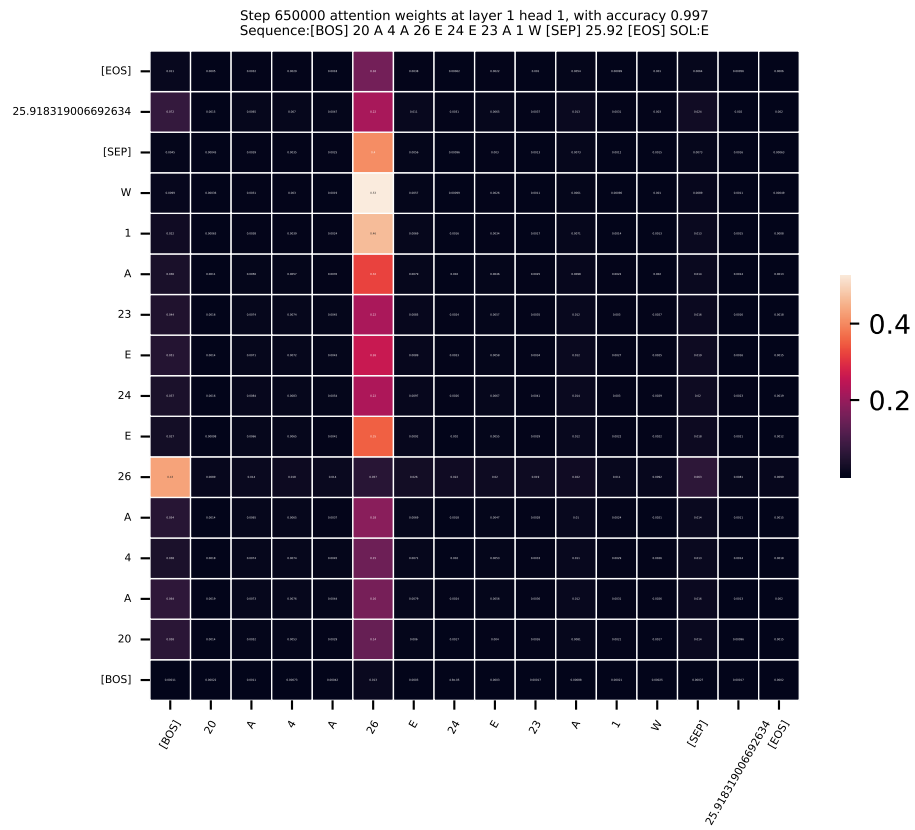

Step 650000 attention weights at layer 0 head 0, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

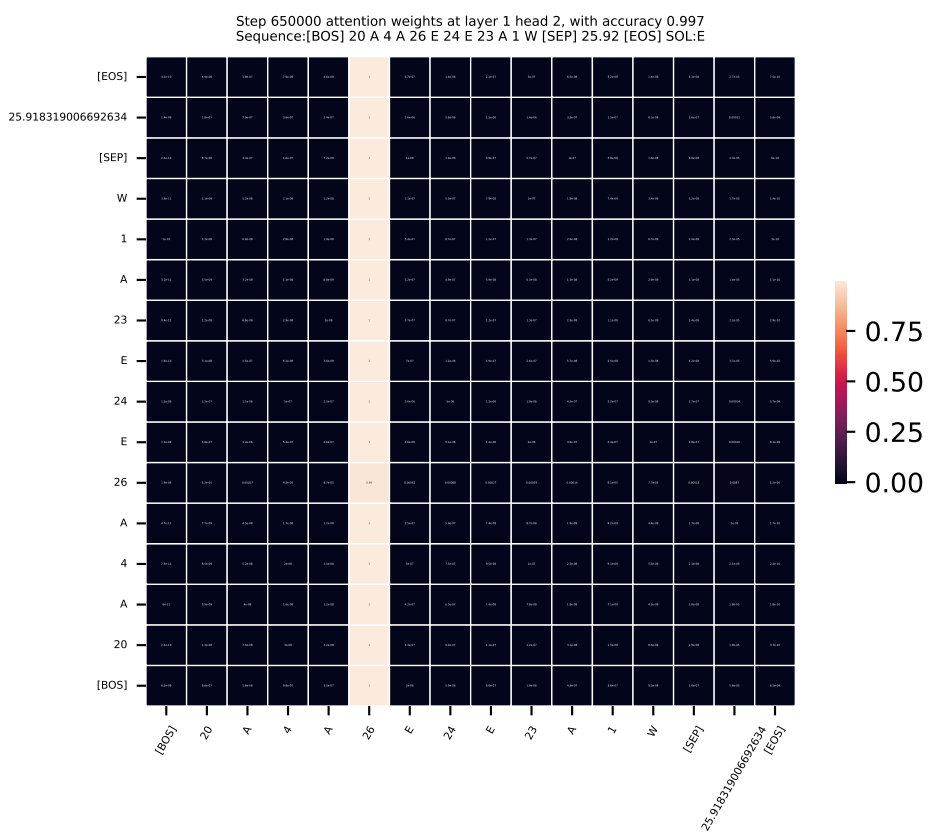

Step 650000 attention weights at layer 0 head 1, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

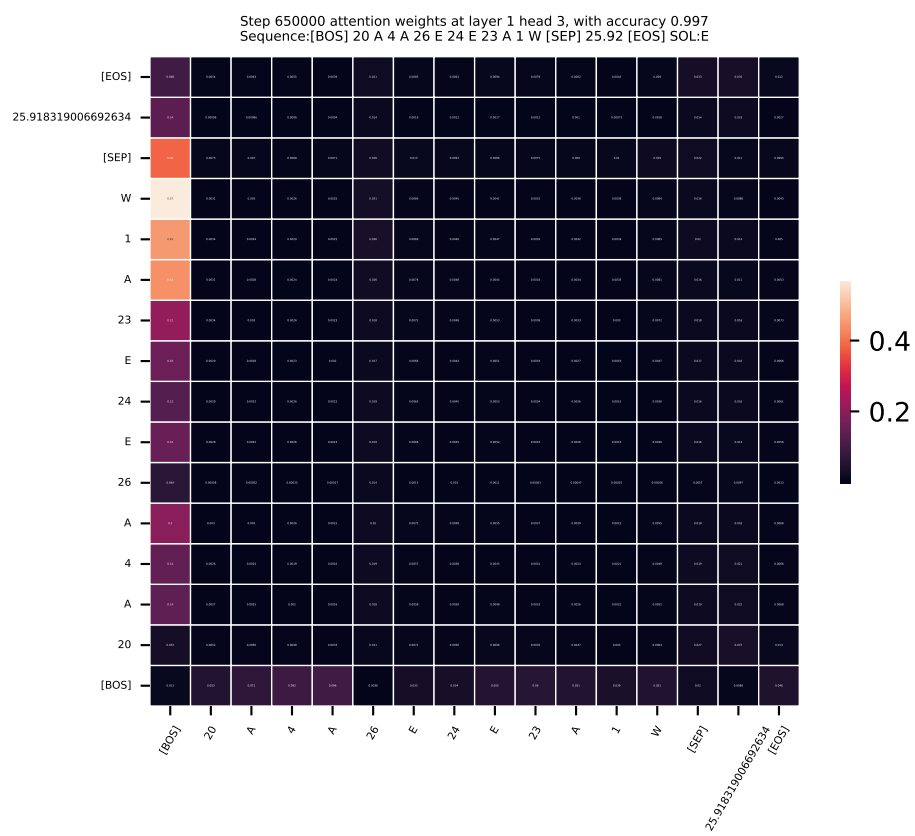

Step 650000 attention weights at layer 0 head 2, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

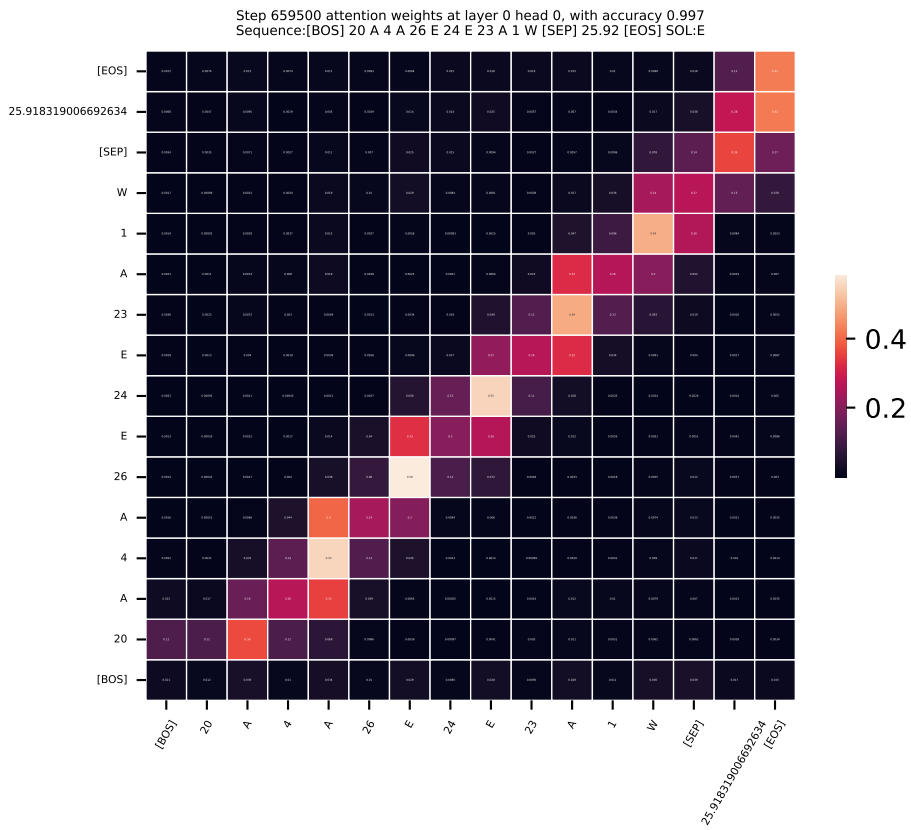

Step 650000 attention weights at layer 0 head 3, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

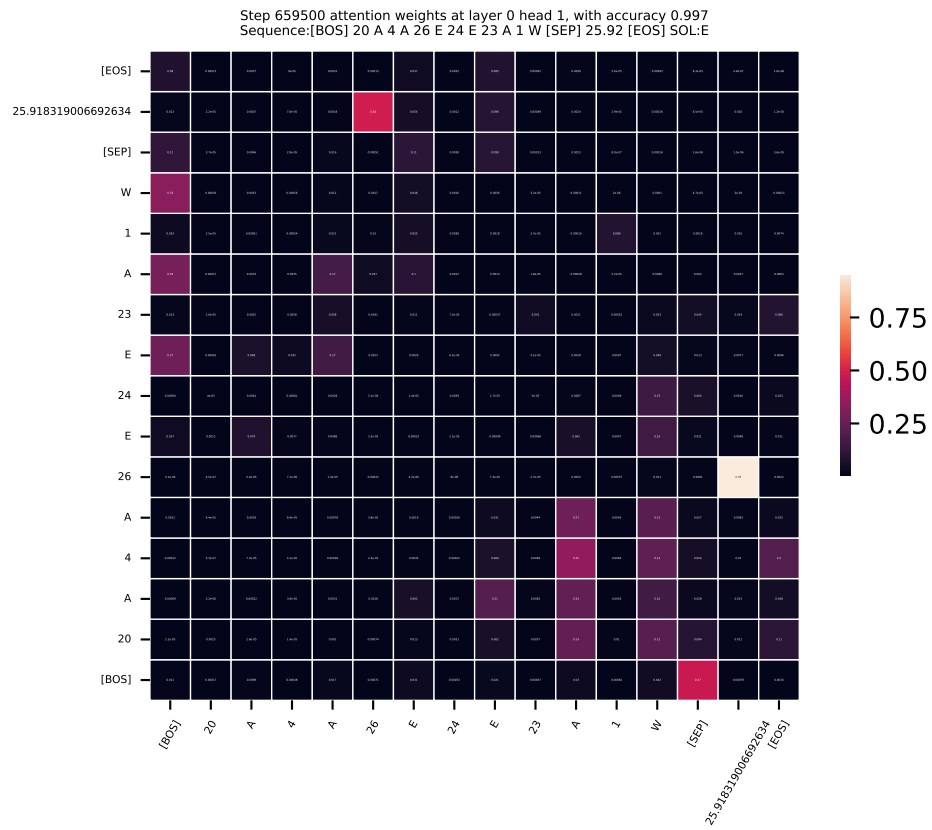

Step 650000 attention weights at layer 1 head 0, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

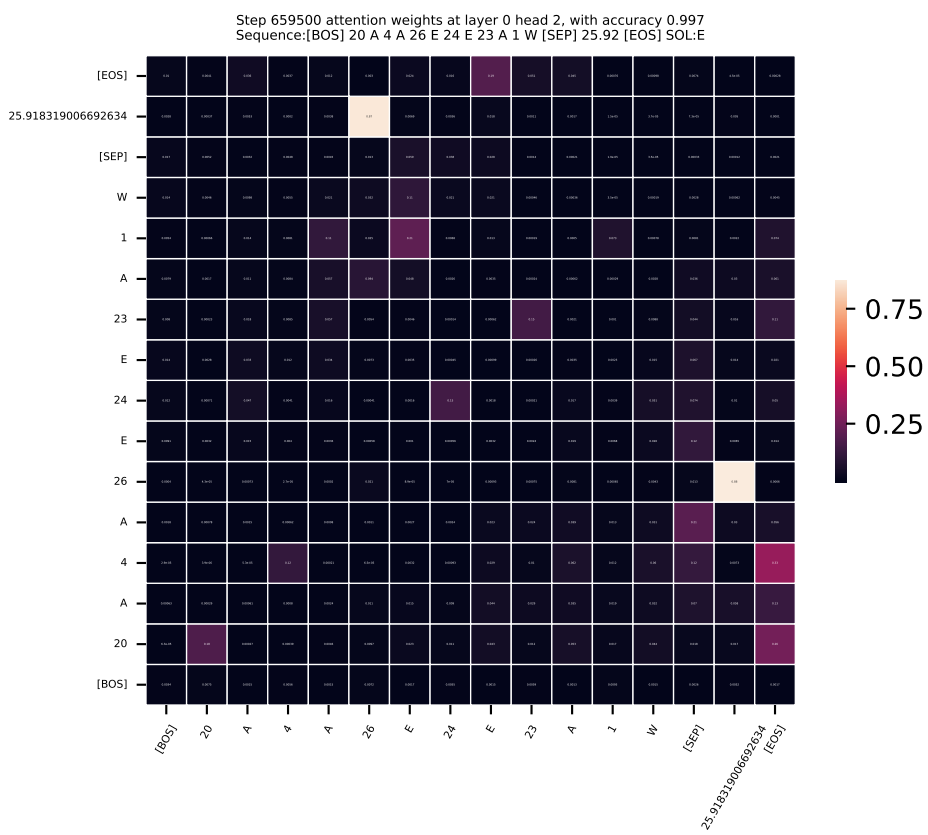

Step 650000 attention weights at layer 1 head 1, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

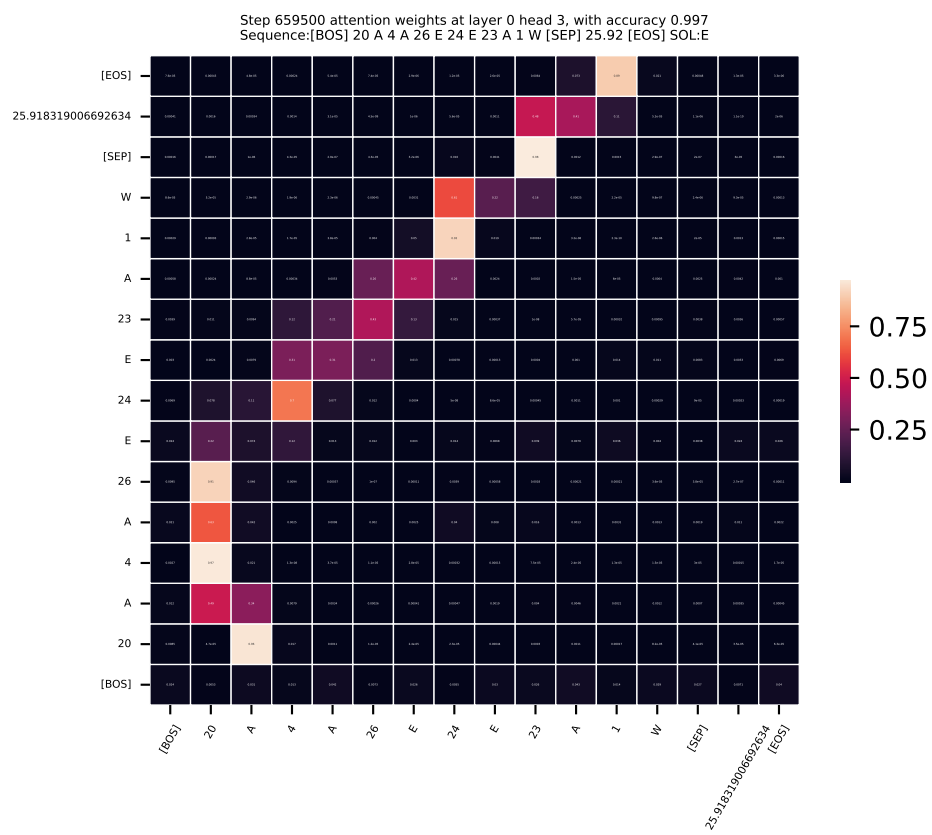

Step 650000 attention weights at layer 1 head 2, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

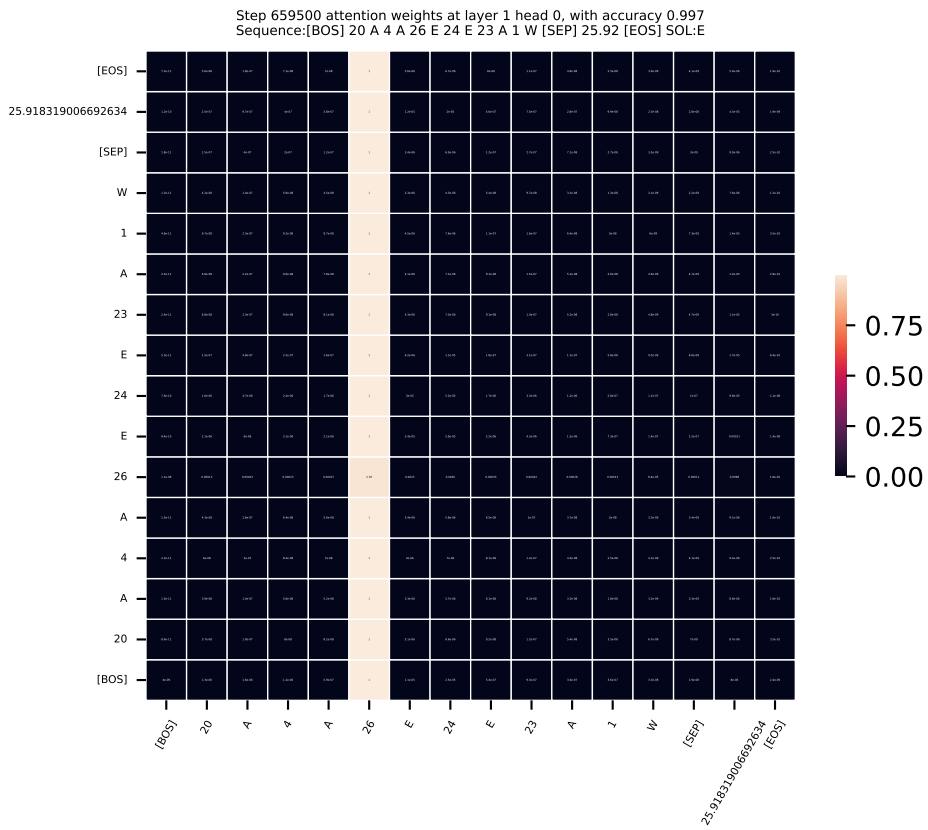

Step 650000 attention weights at layer 1 head 3, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

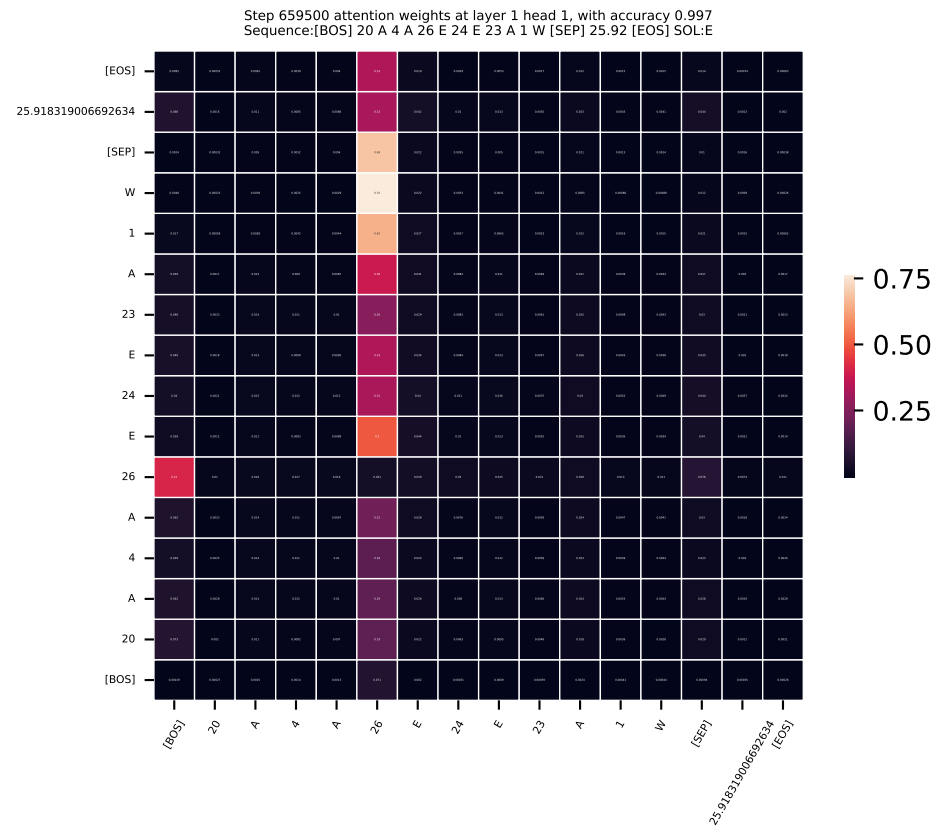

Step 659500 attention weights at layer 0 head 0, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

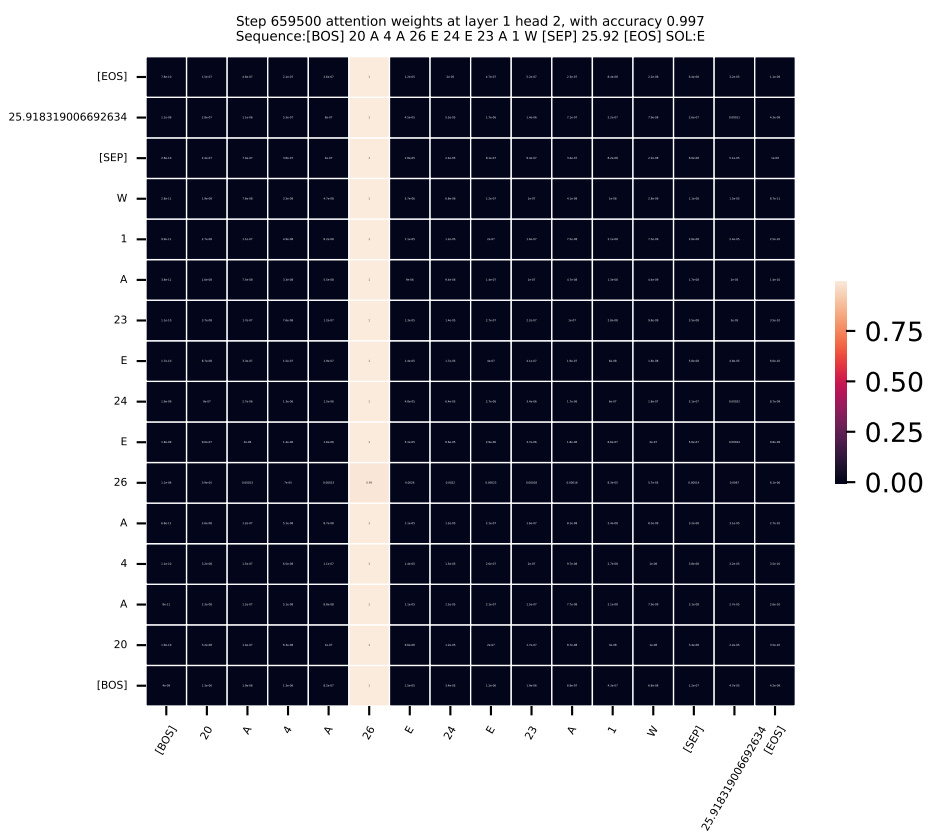

Step 659500 attention weights at layer 0 head 1, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

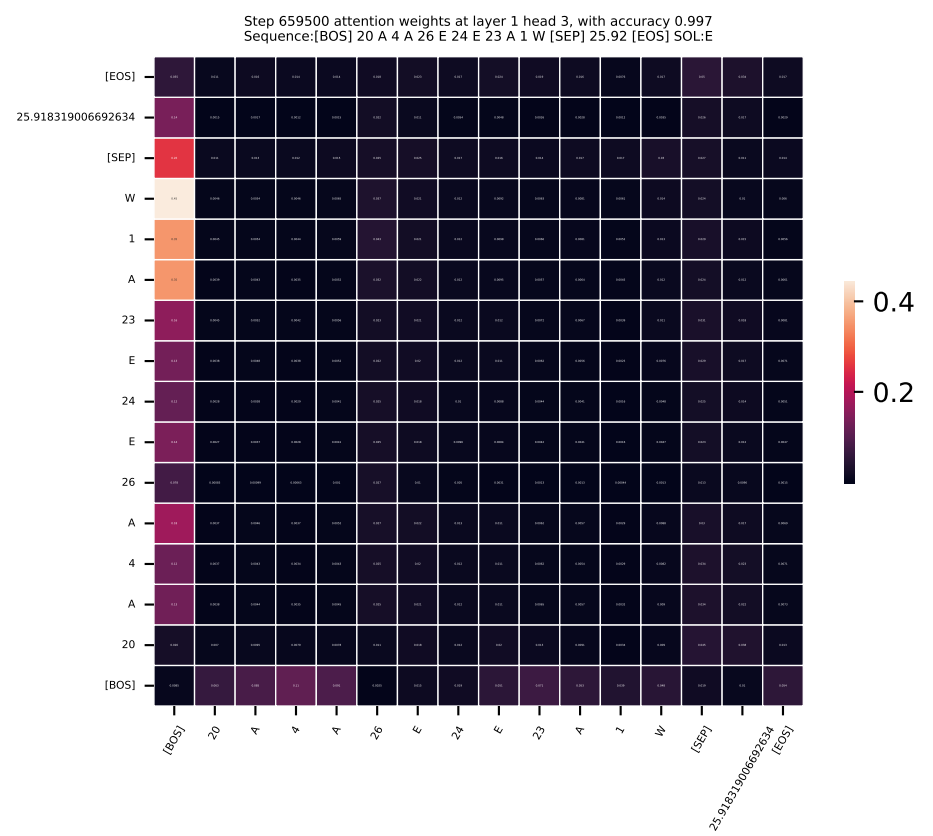

Step 659500 attention weights at layer 0 head 2, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

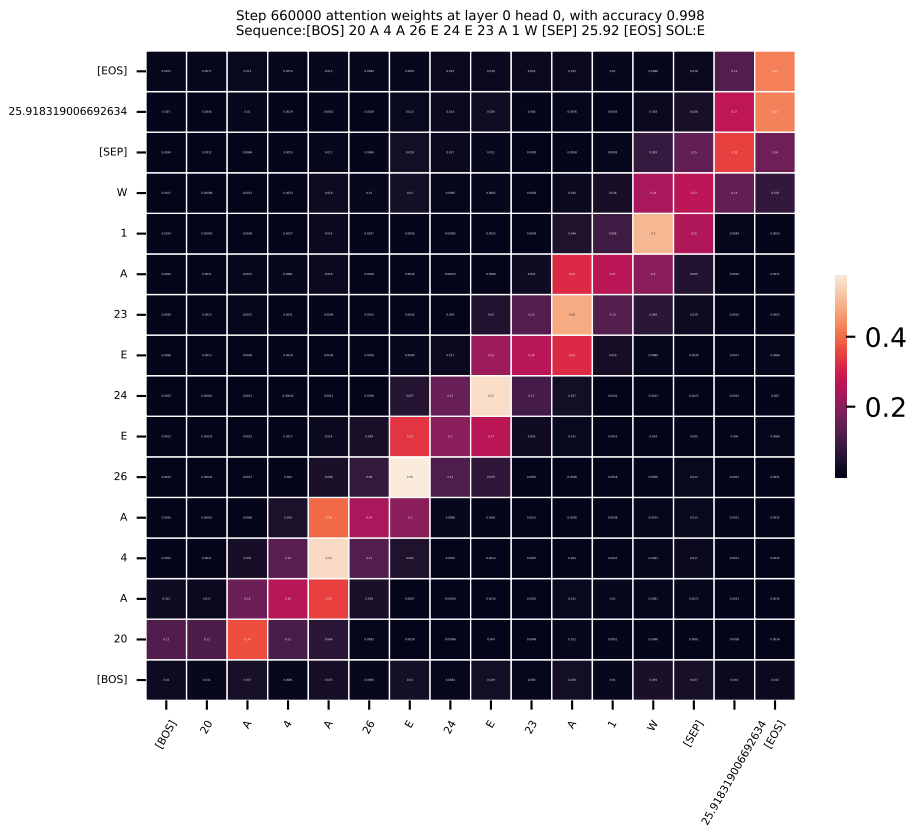

Step 659500 attention weights at layer 0 head 3, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

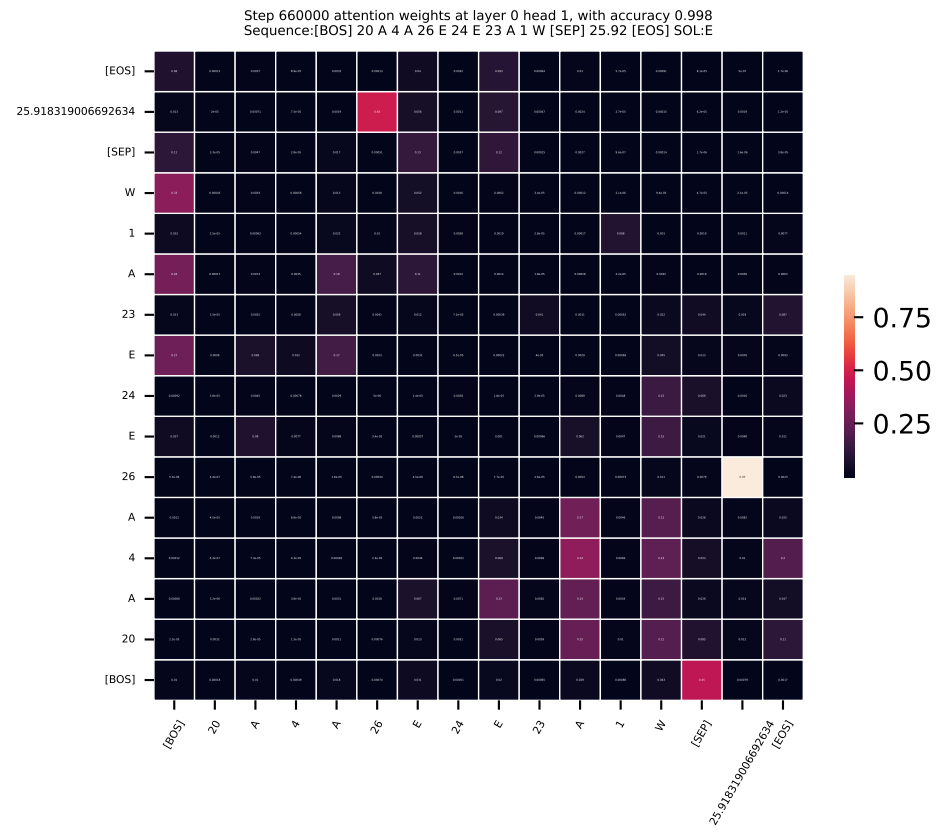

Step 659500 attention weights at layer 1 head 0, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

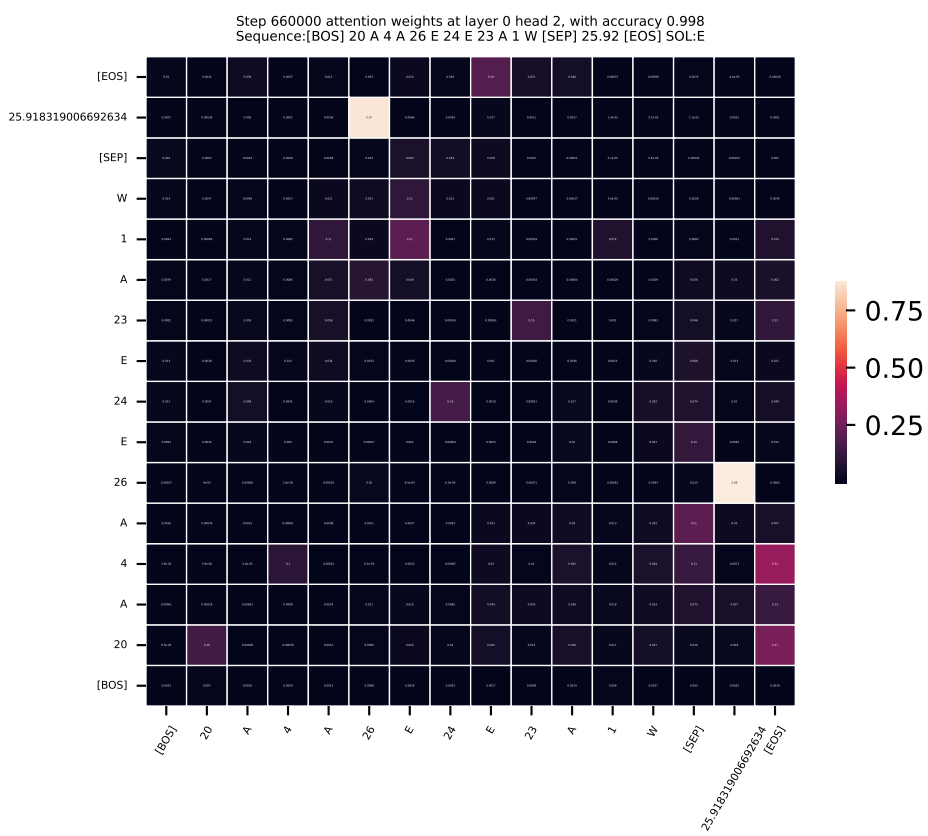

Step 659500 attention weights at layer 1 head 1, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

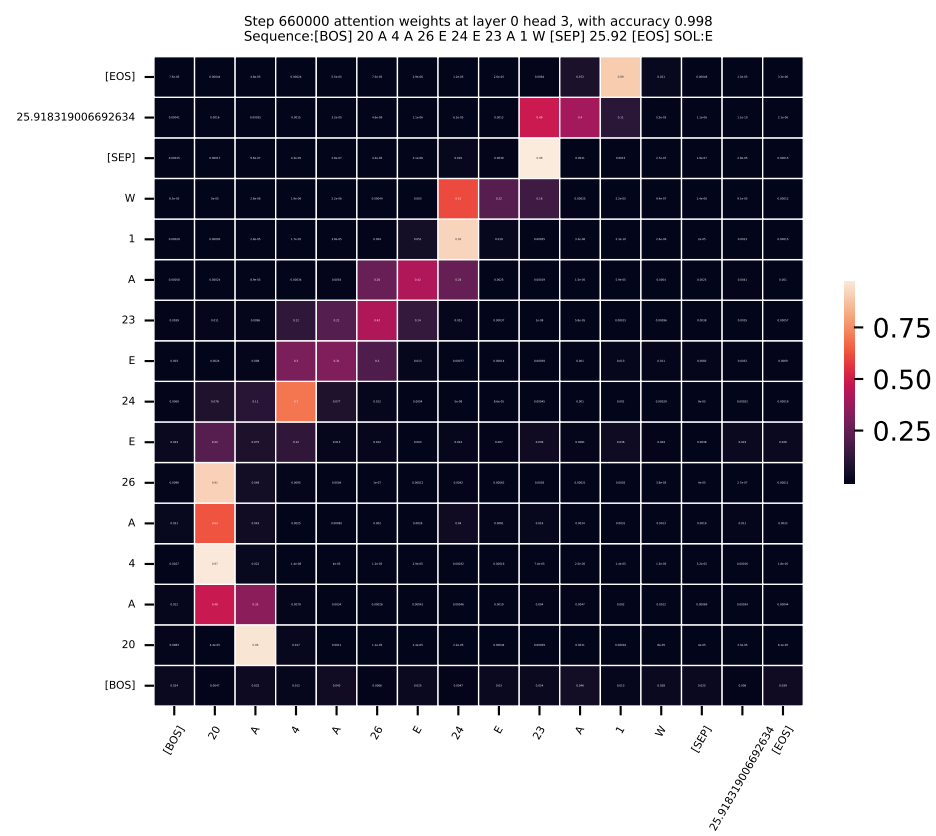

Step 659500 attention weights at layer 1 head 2, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

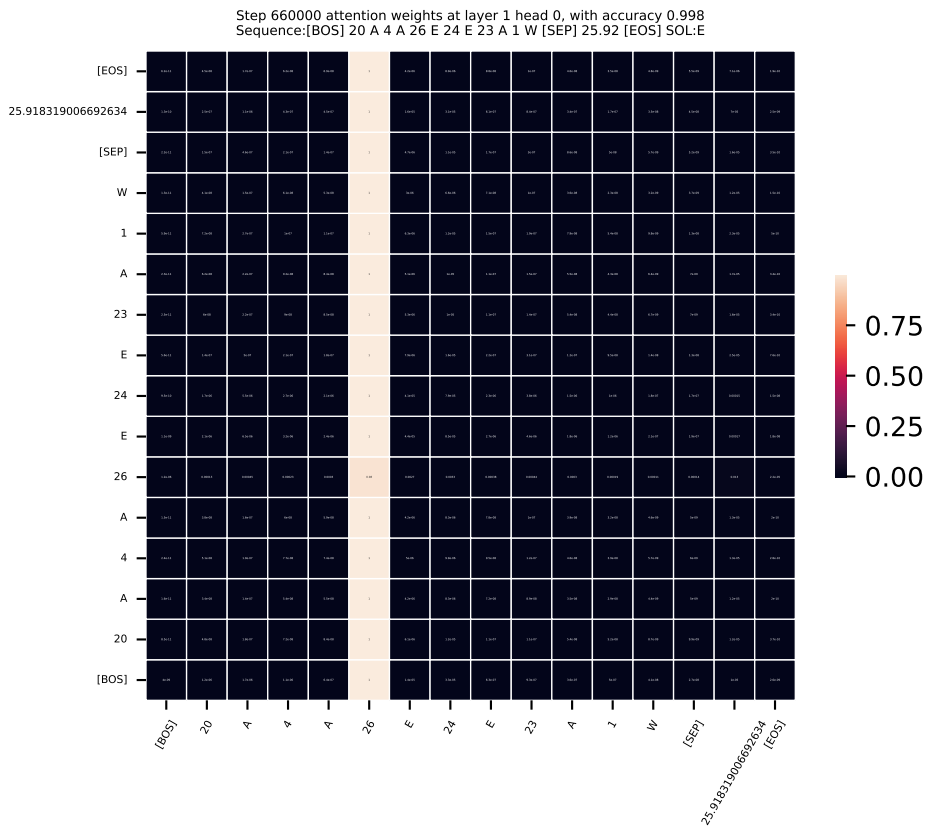

Step 659500 attention weights at layer 1 head 3, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

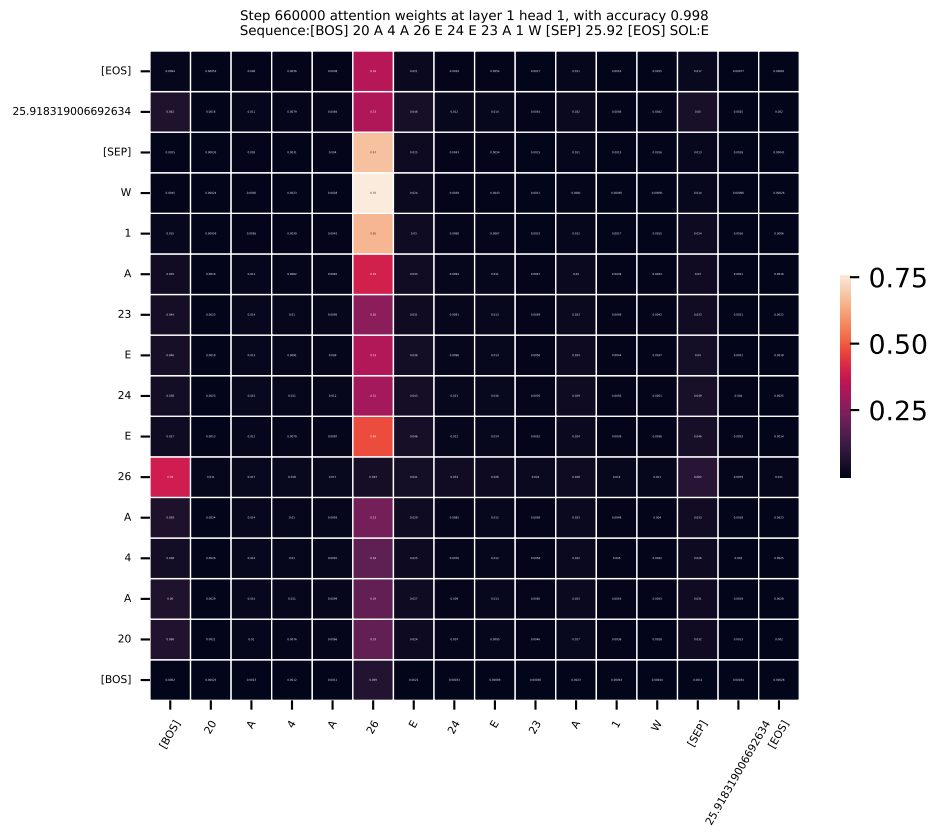

Step 660000 attention weights at layer 0 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

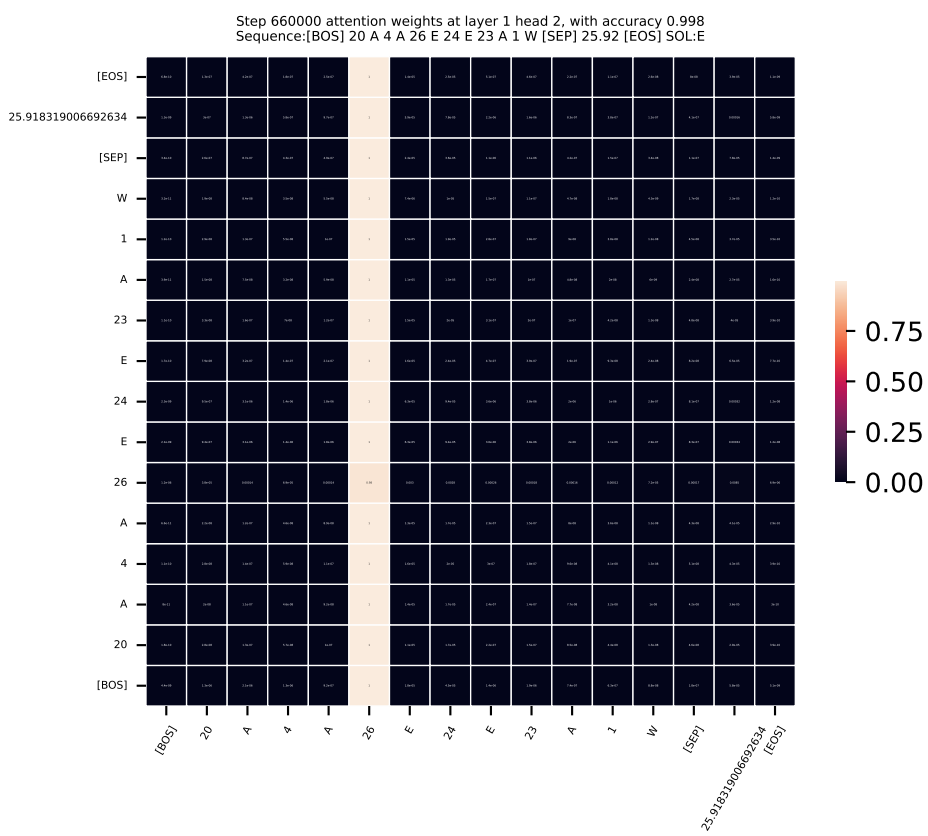

Step 660000 attention weights at layer 0 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

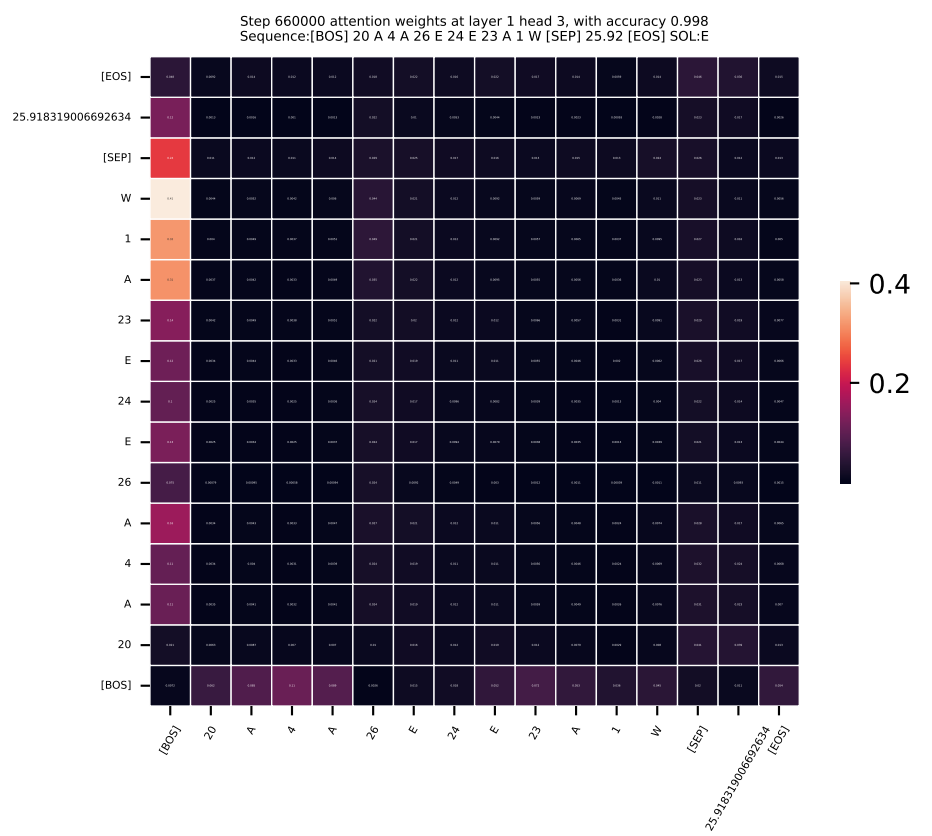

Step 660000 attention weights at layer 0 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

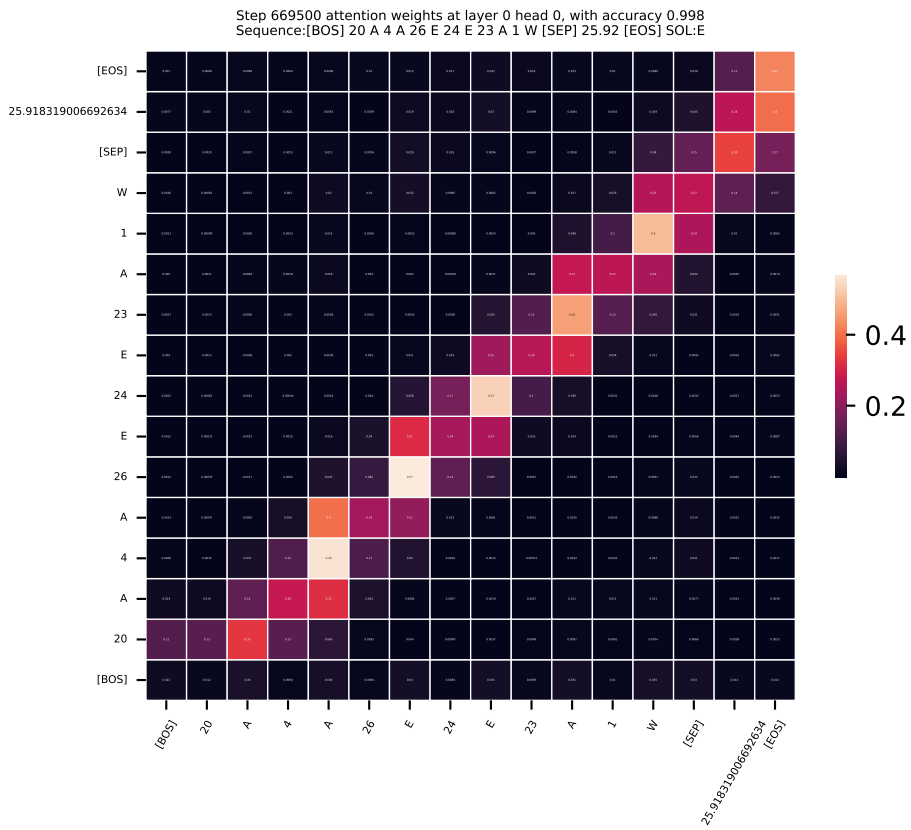

Step 660000 attention weights at layer 0 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

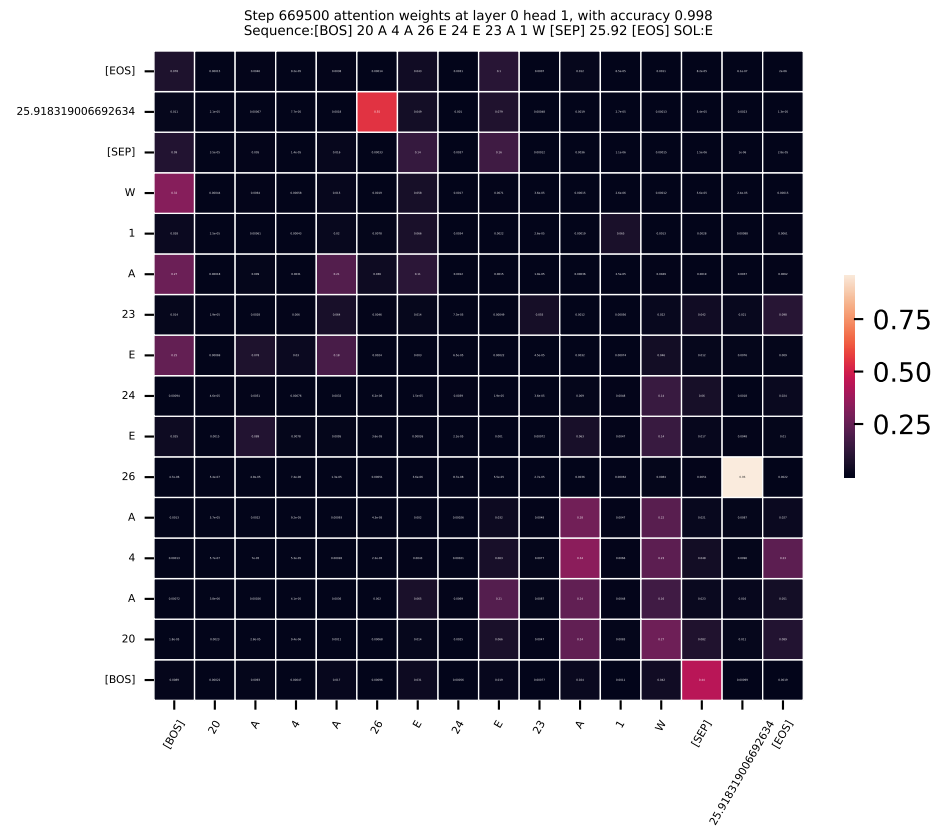

Step 660000 attention weights at layer 1 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

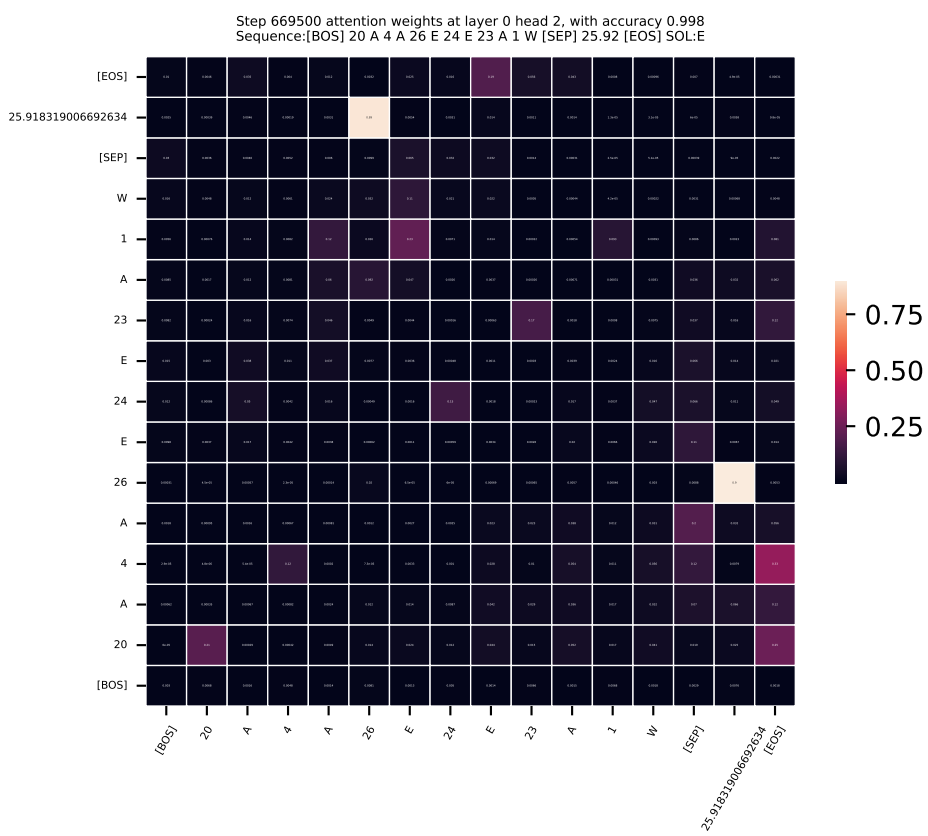

Step 660000 attention weights at layer 1 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

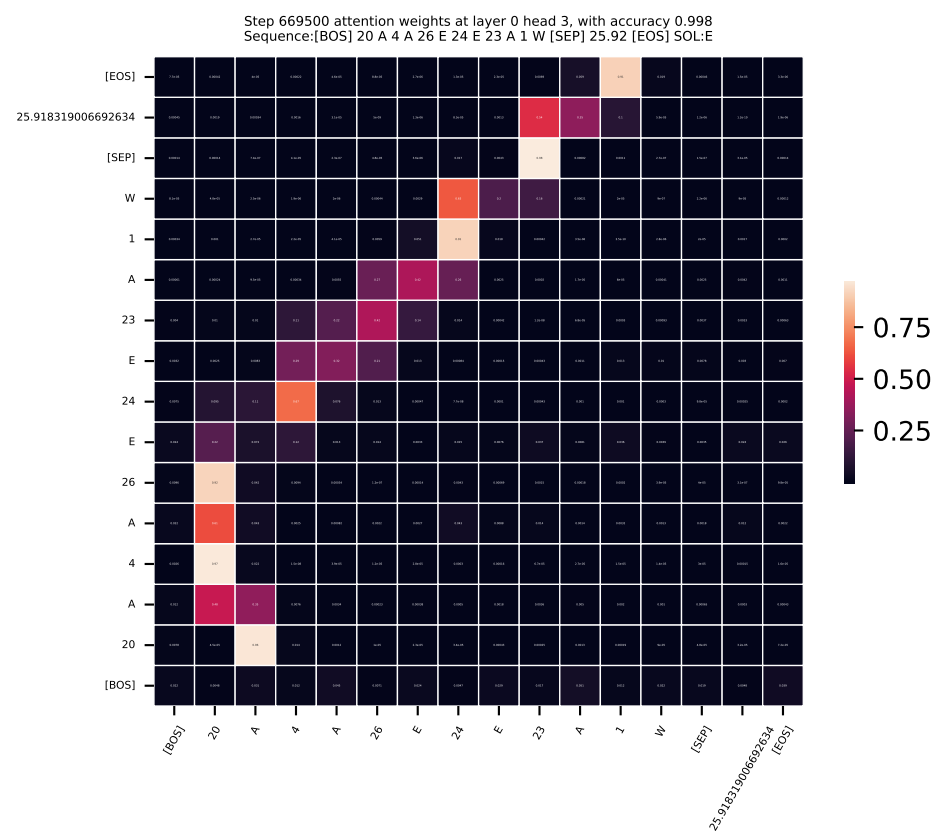

Step 660000 attention weights at layer 1 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

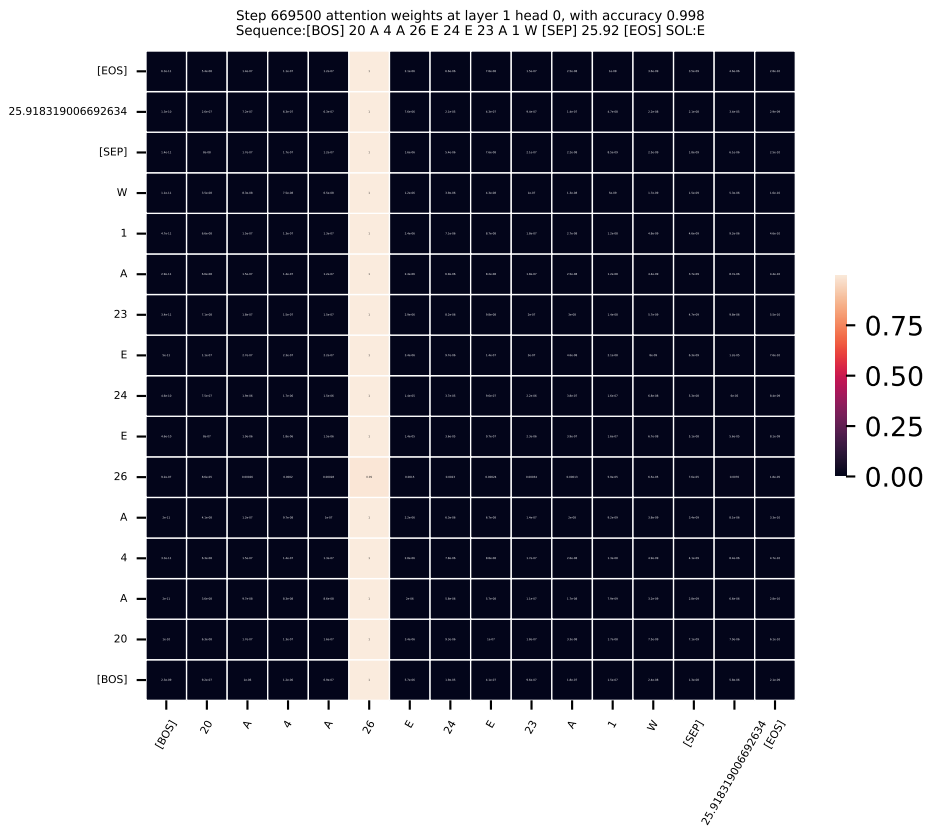

Step 660000 attention weights at layer 1 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

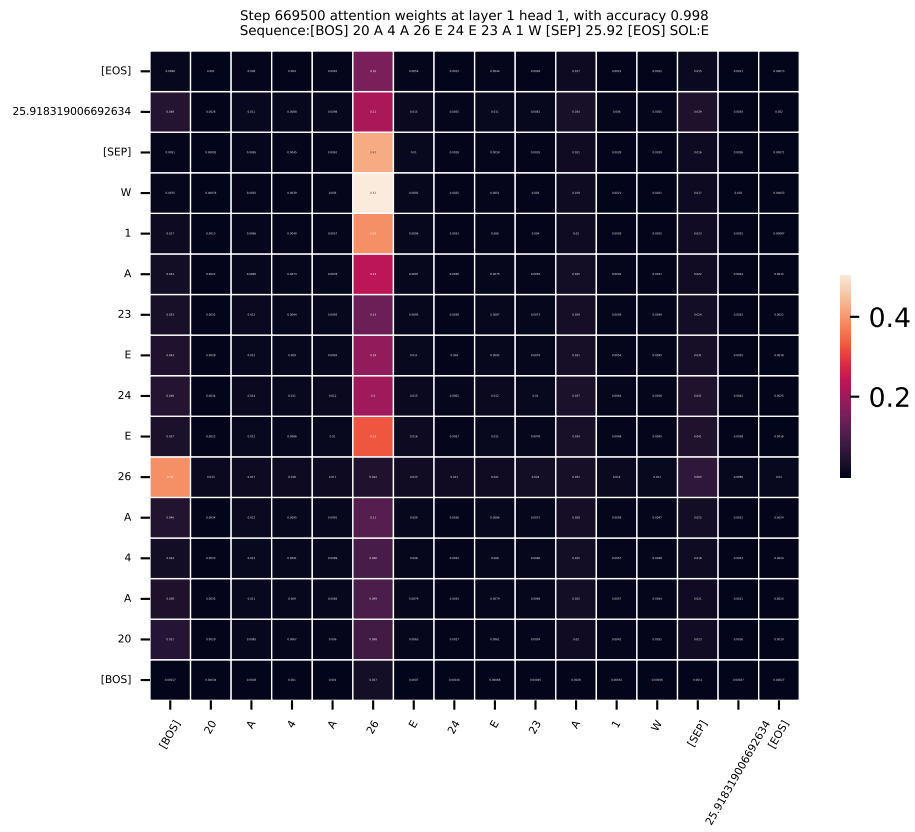

Step 669500 attention weights at layer 0 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

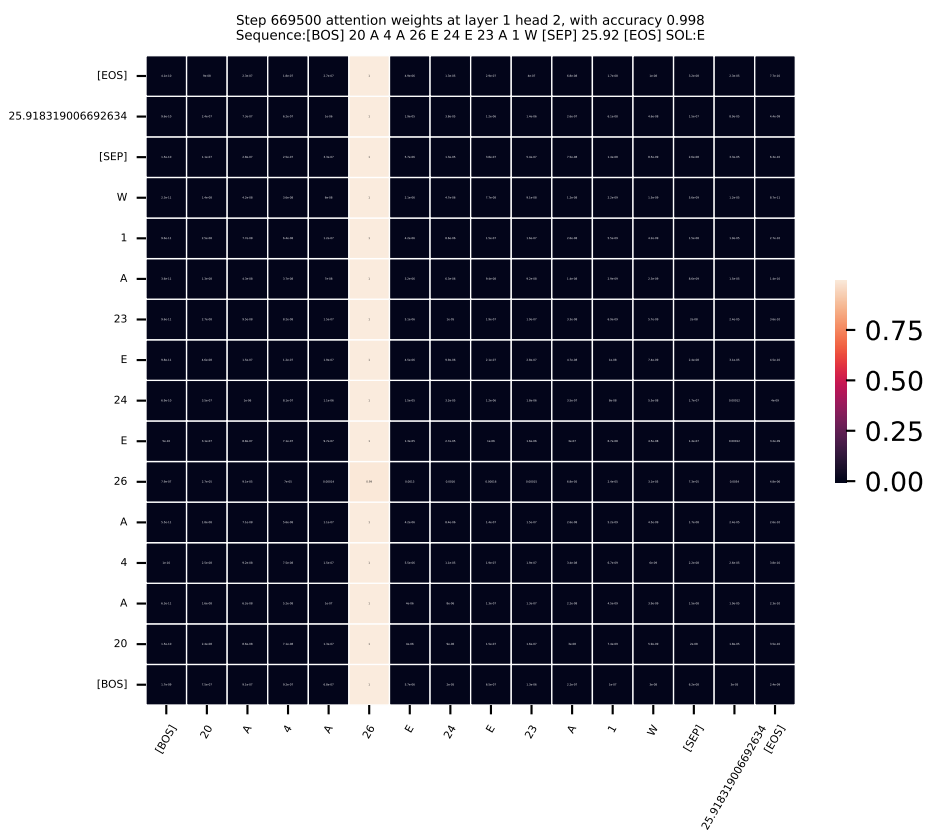

Step 669500 attention weights at layer 0 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

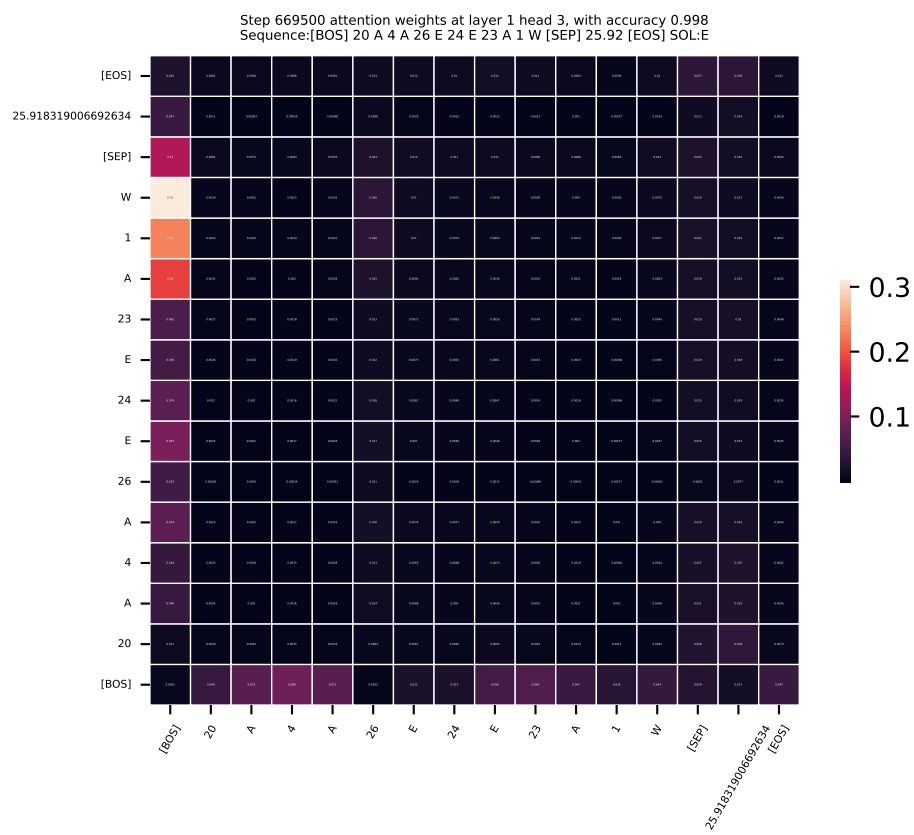

Step 669500 attention weights at layer 0 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

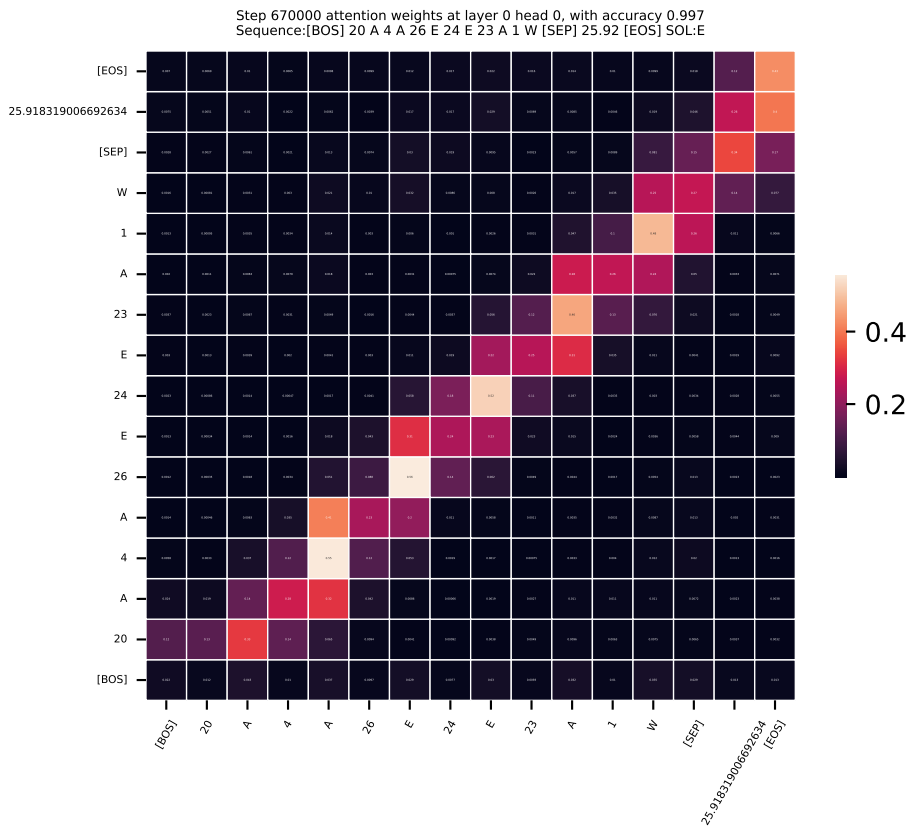

Step 669500 attention weights at layer 0 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

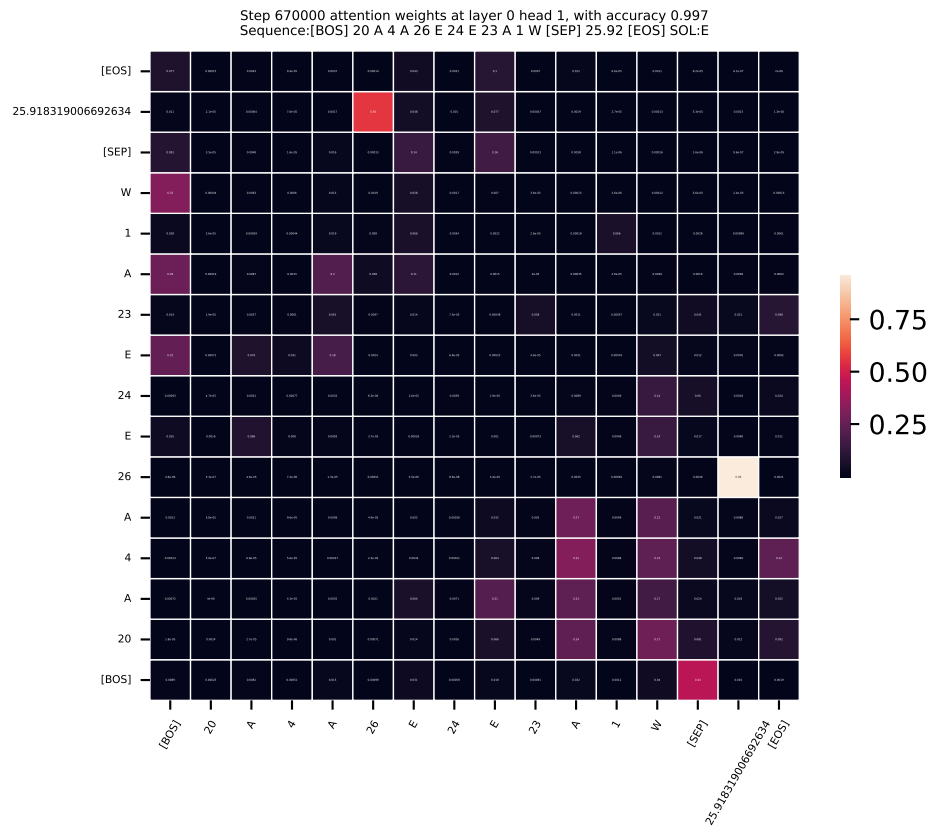

Step 669500 attention weights at layer 1 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

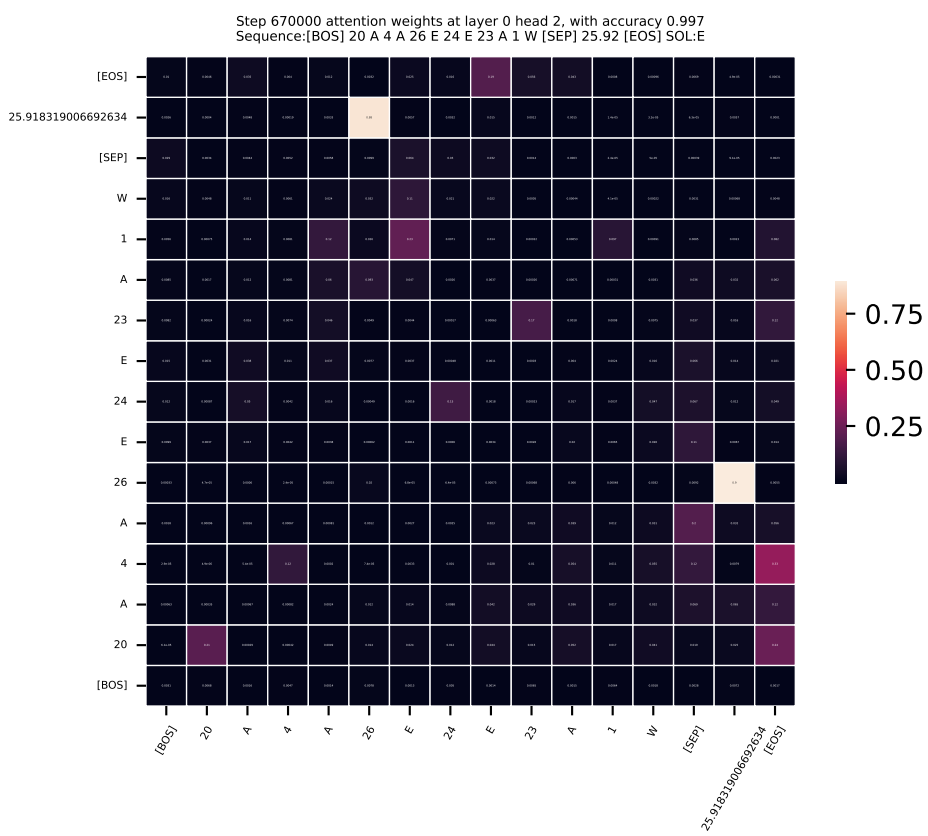

Step 669500 attention weights at layer 1 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

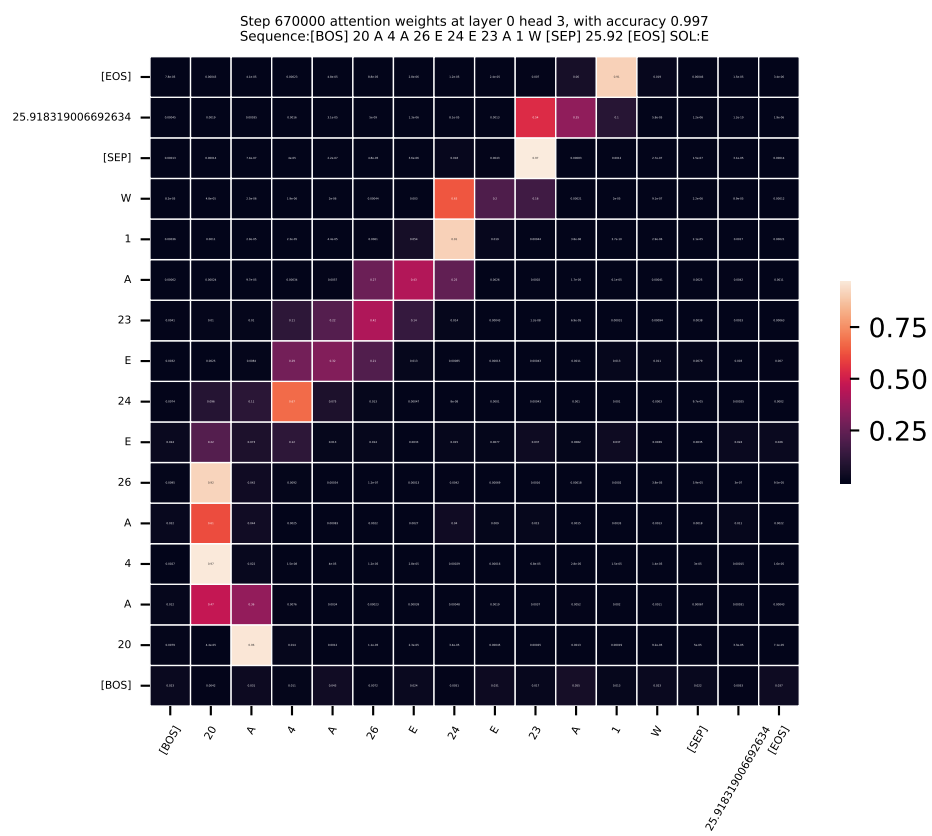

Step 669500 attention weights at layer 1 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

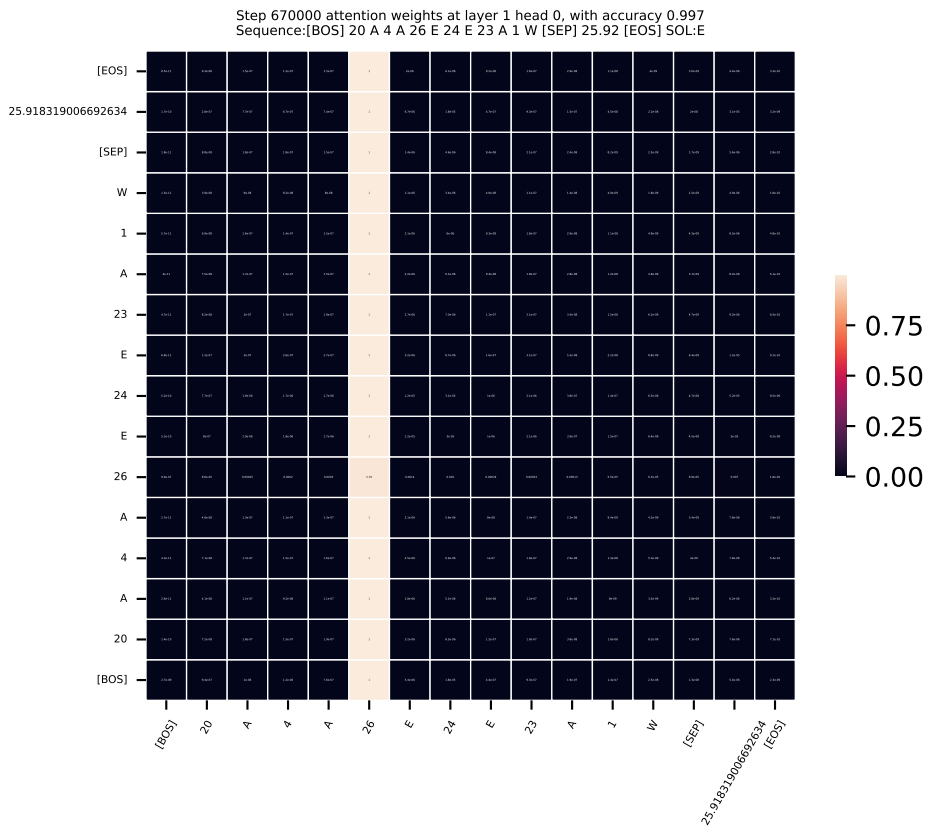

Step 669500 attention weights at layer 1 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

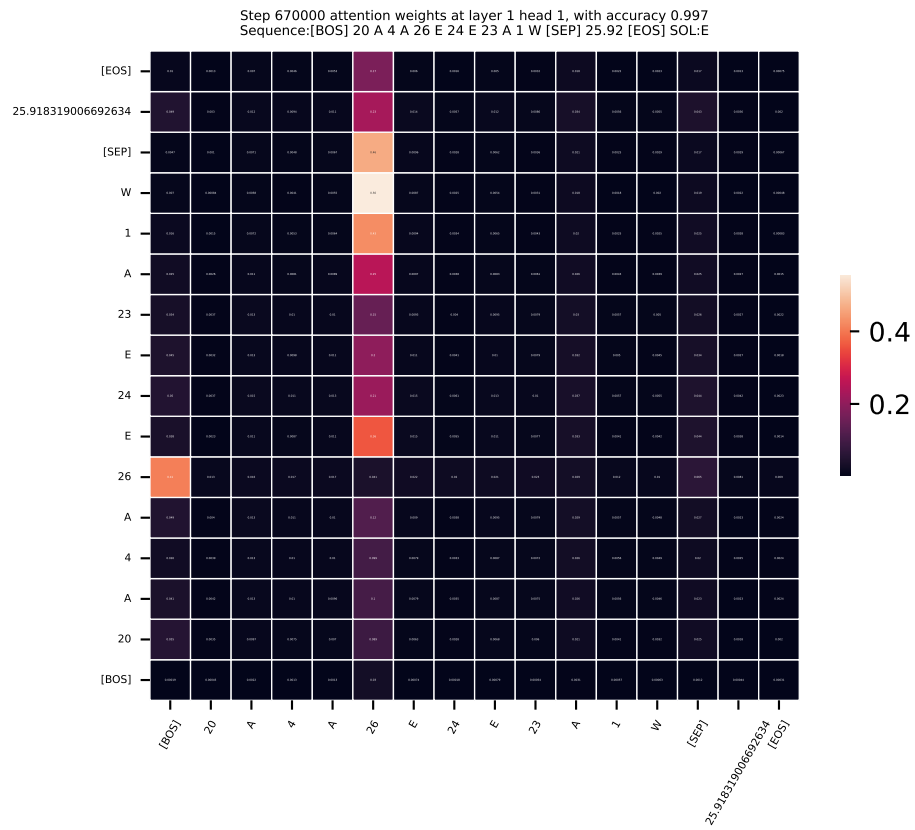

Step 670000 attention weights at layer 0 head 0, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.4
0.2

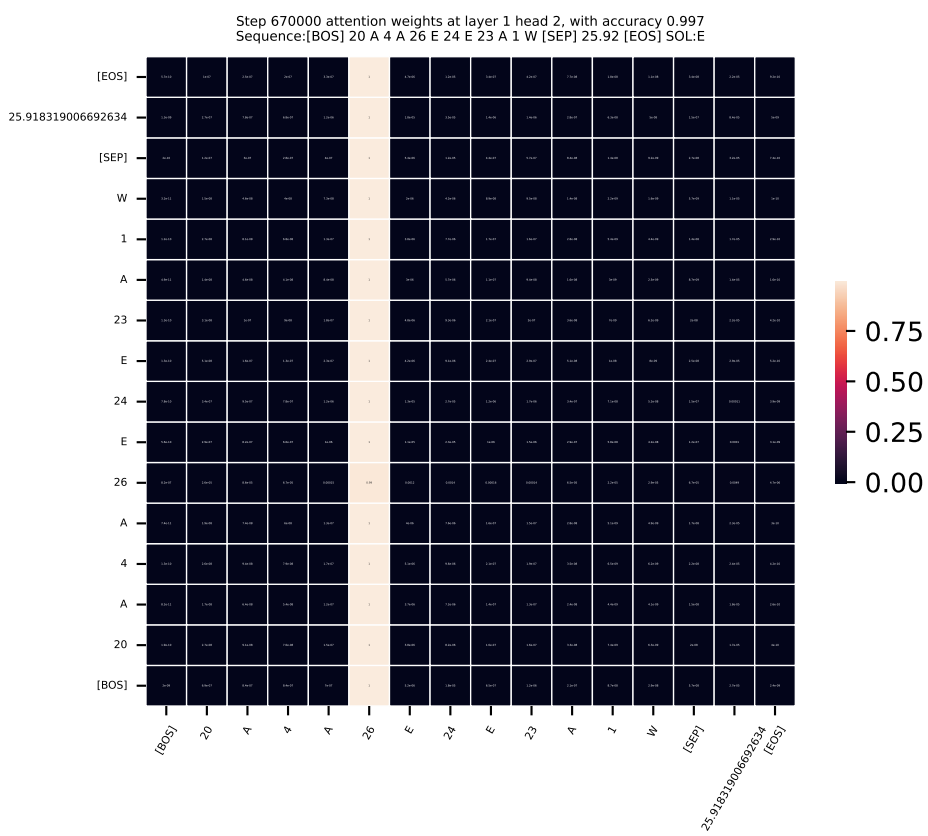

Step 670000 attention weights at layer 0 head 1, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

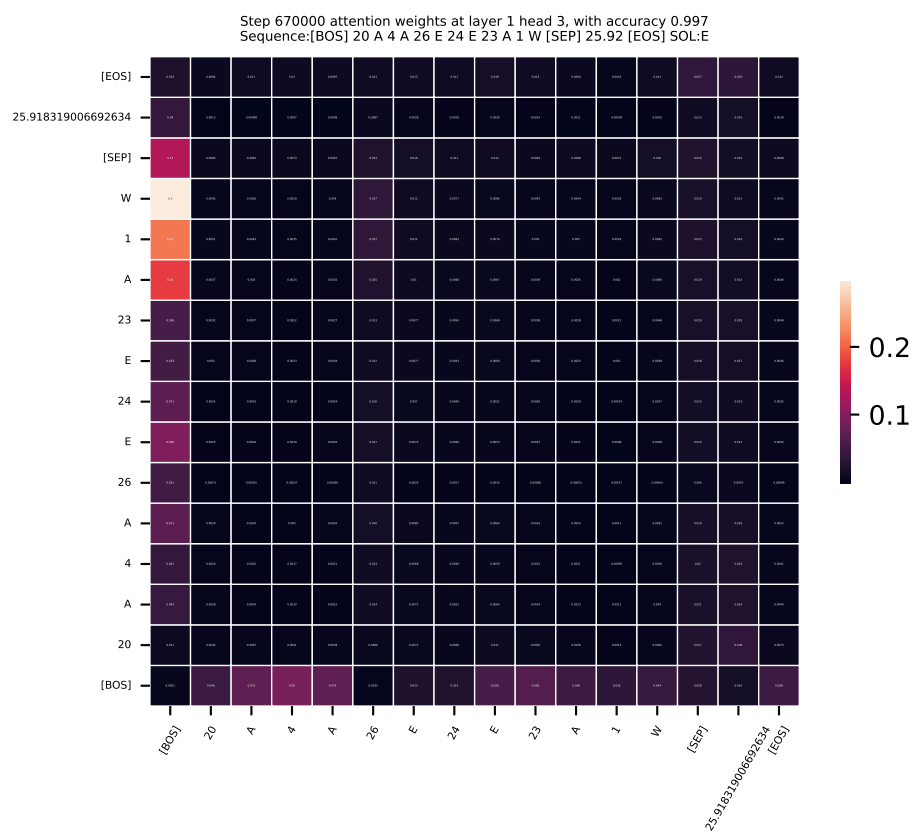

Step 670000 attention weights at layer 0 head 2, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

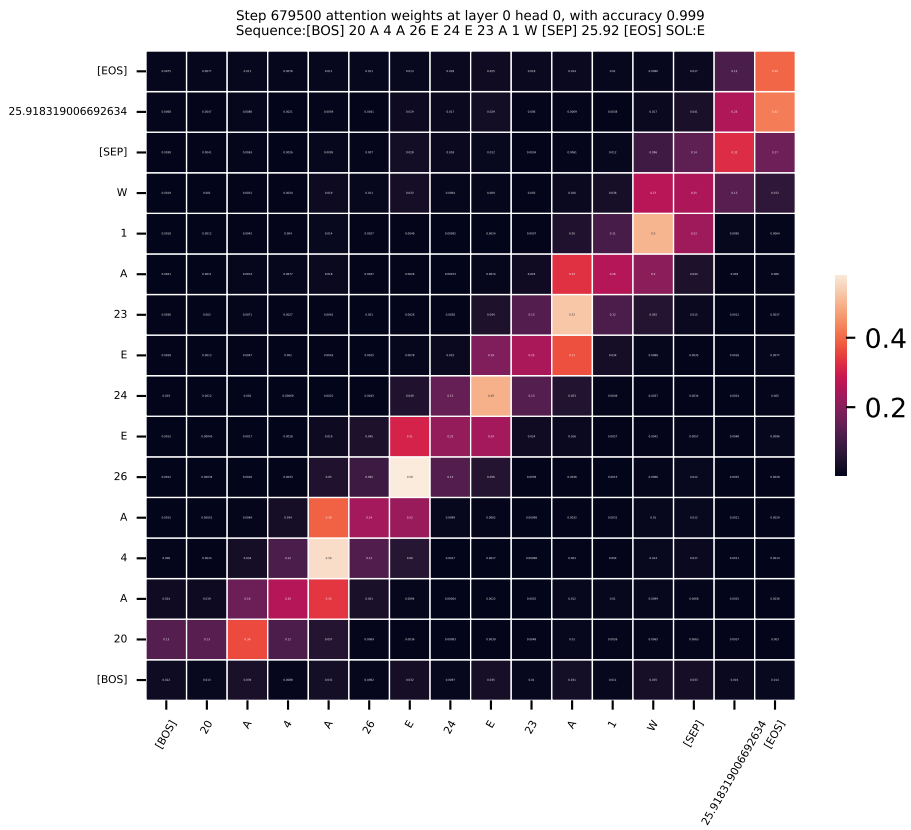

Step 670000 attention weights at layer 0 head 3, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

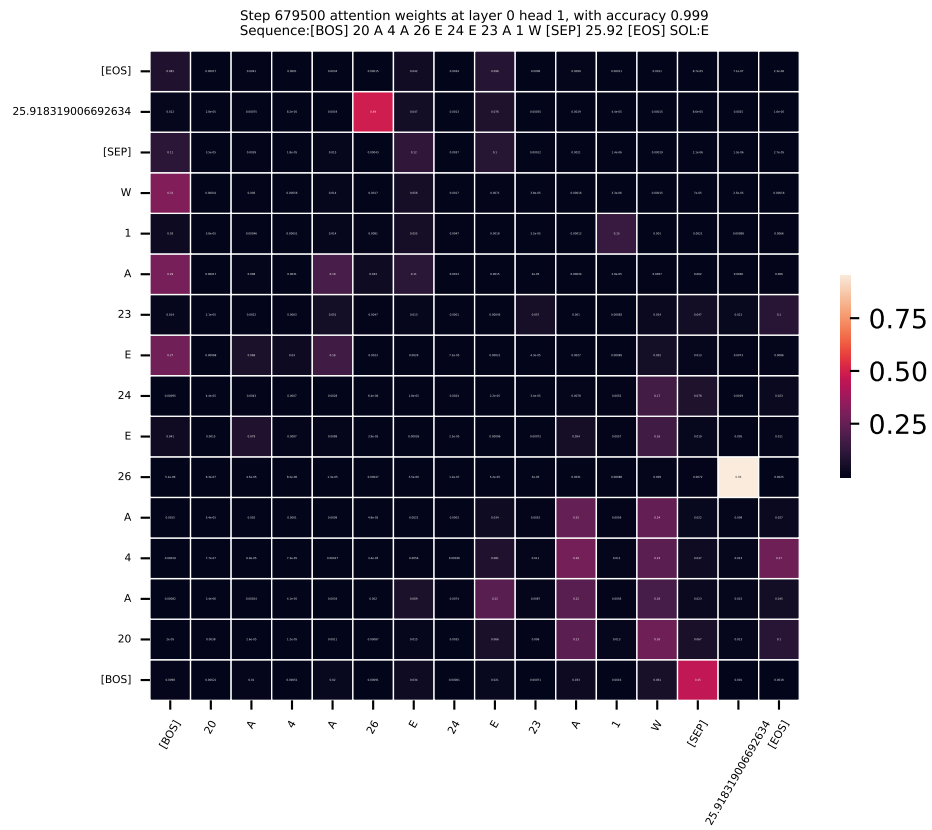

Step 670000 attention weights at layer 1 head 0, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

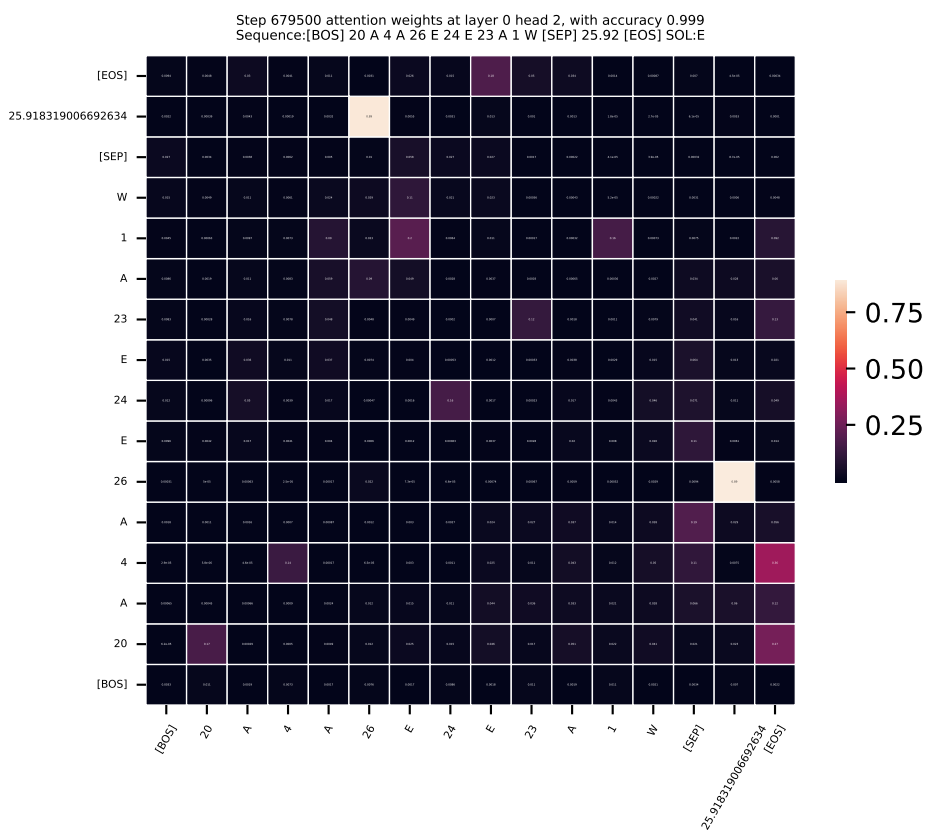

Step 670000 attention weights at layer 1 head 1, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

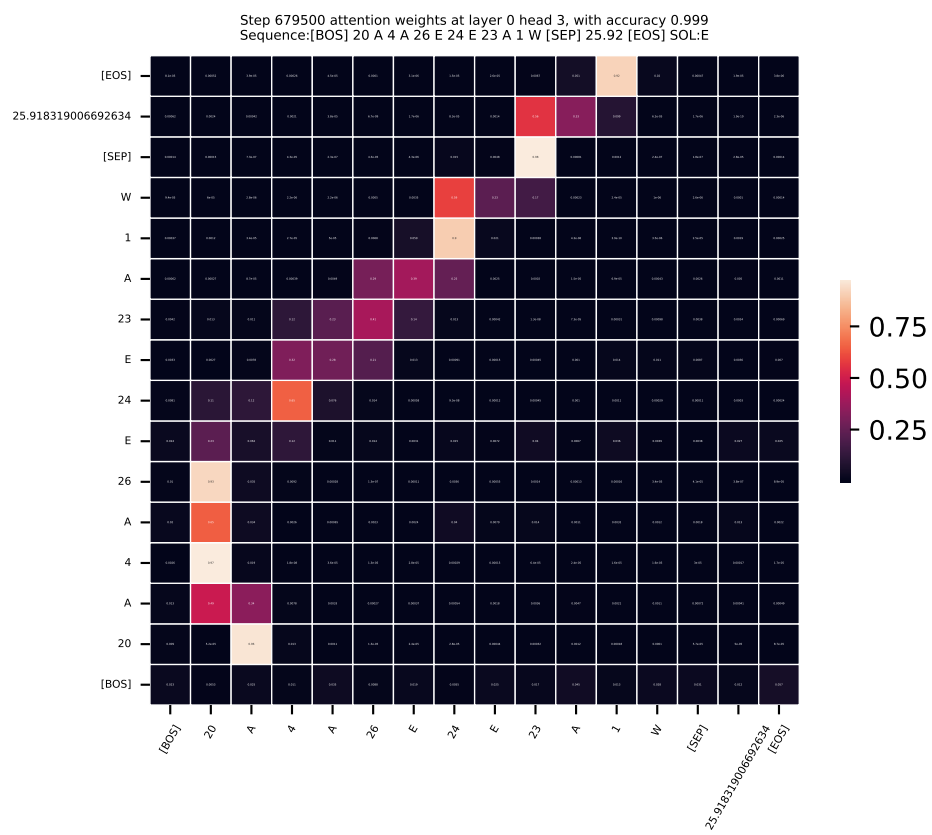

Step 670000 attention weights at layer 1 head 2, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

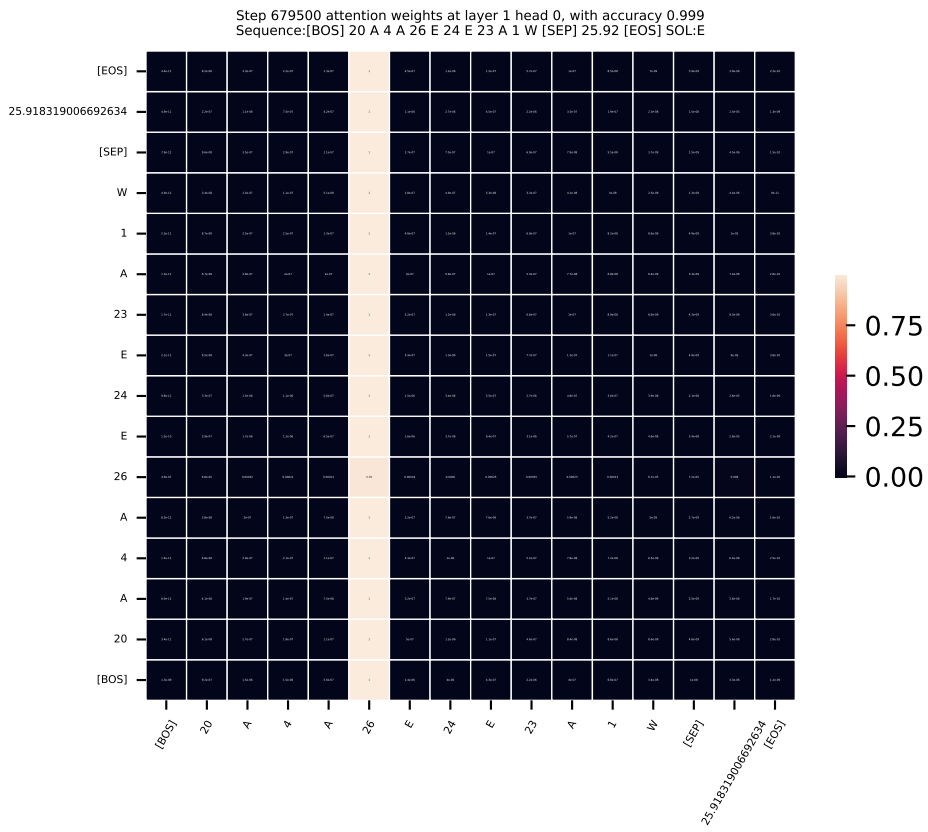

Step 670000 attention weights at layer 1 head 3, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.2
0.1

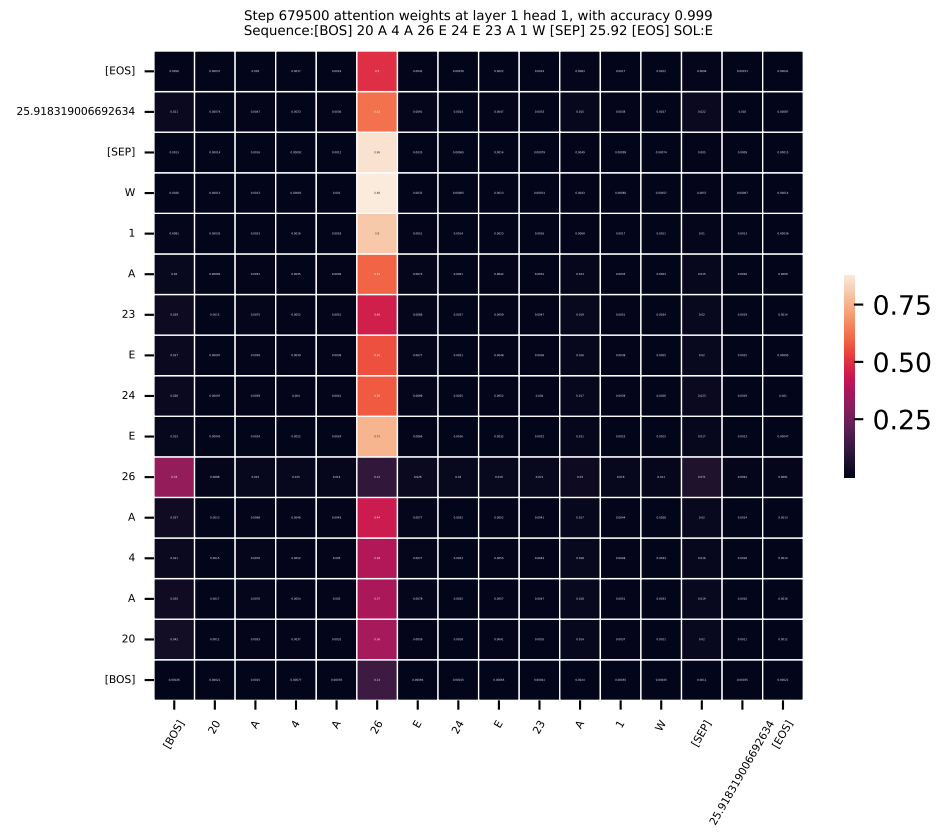

Step 679500 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

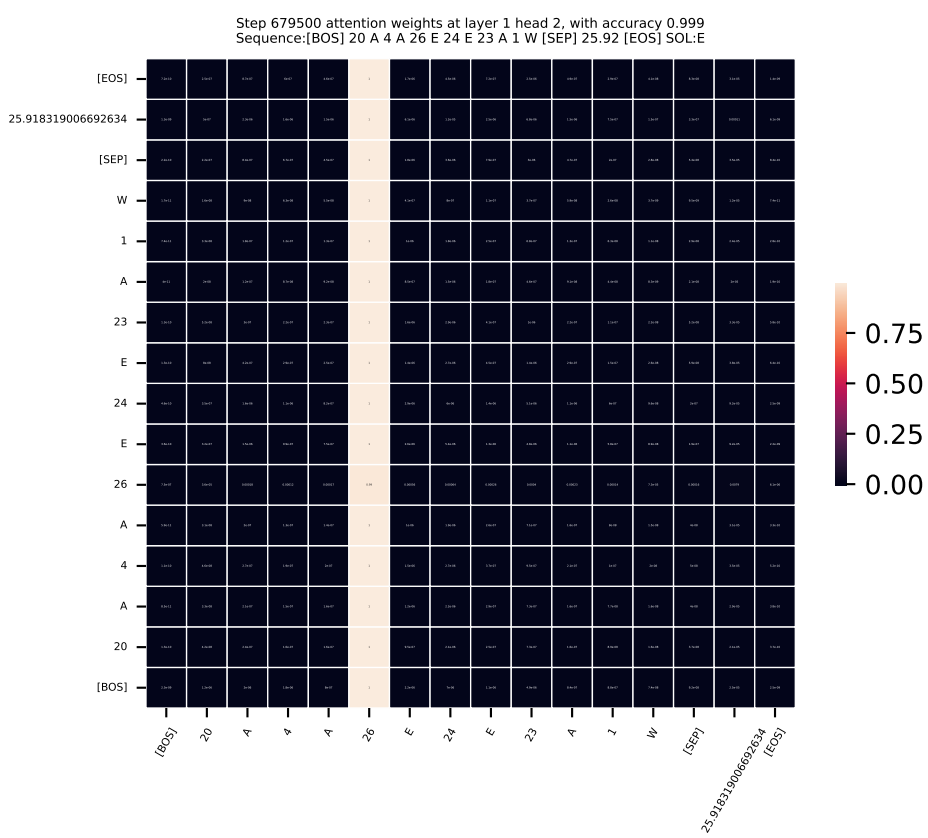

Step 679500 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

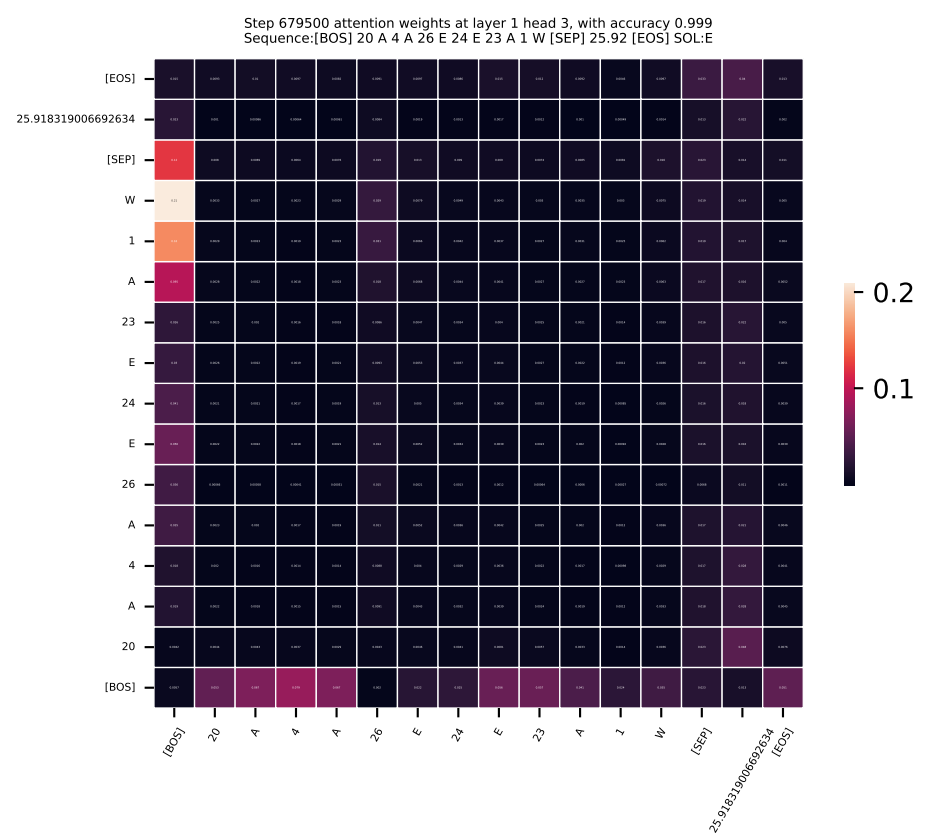

Step 679500 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

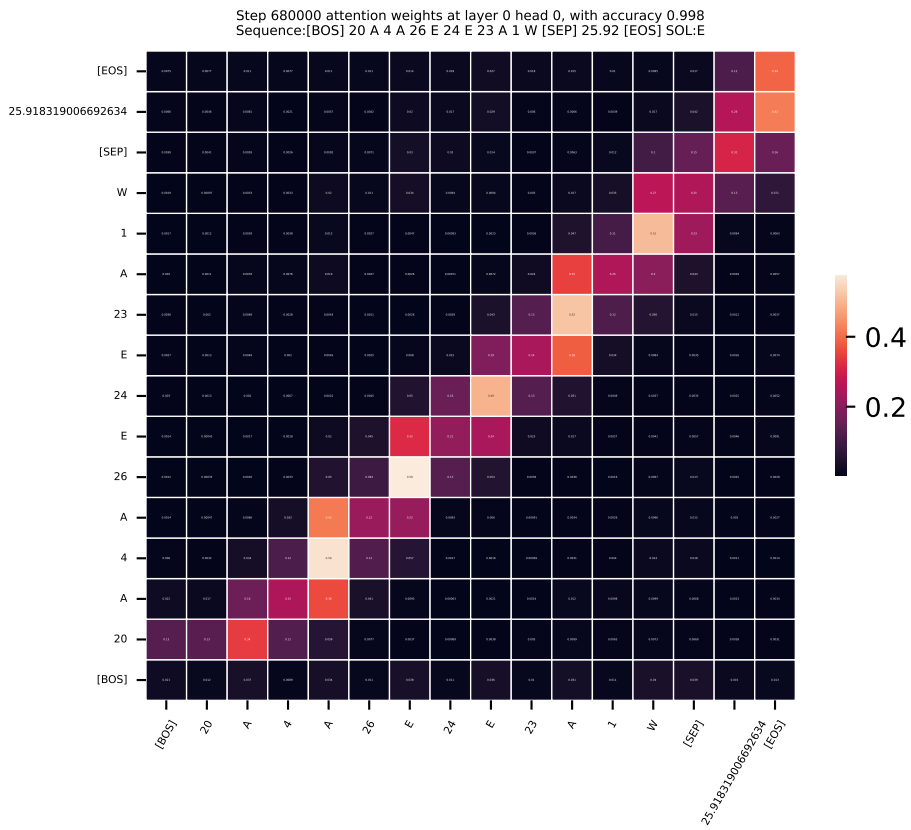

Step 679500 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

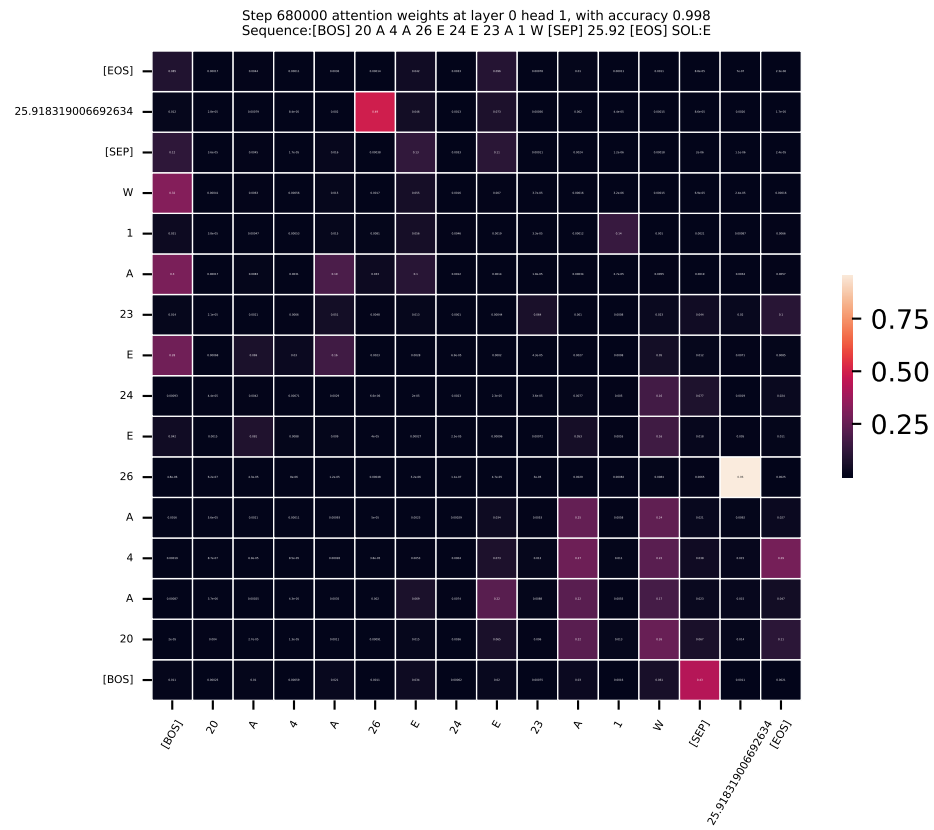

Step 679500 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

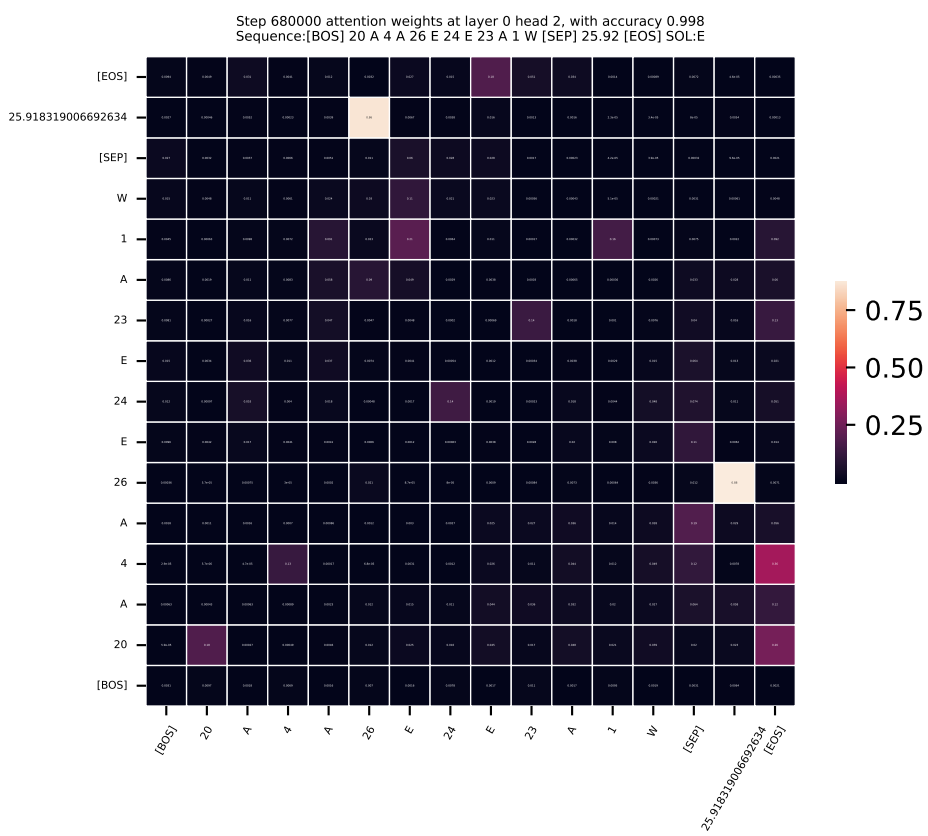

Step 679500 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

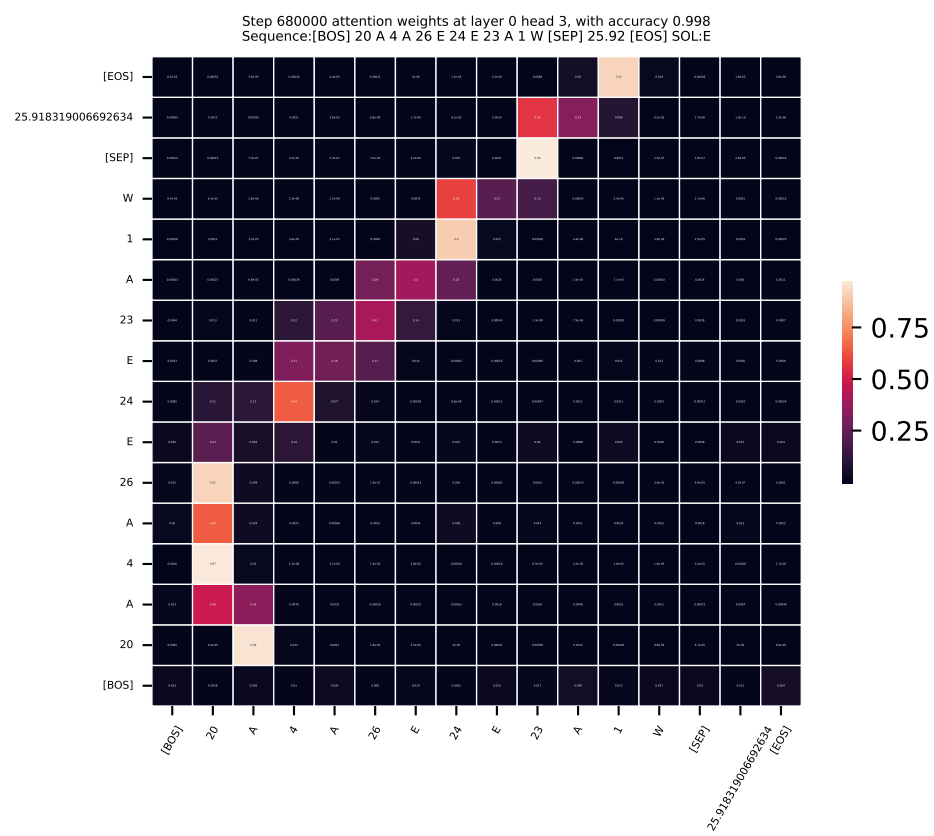

Step 679500 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

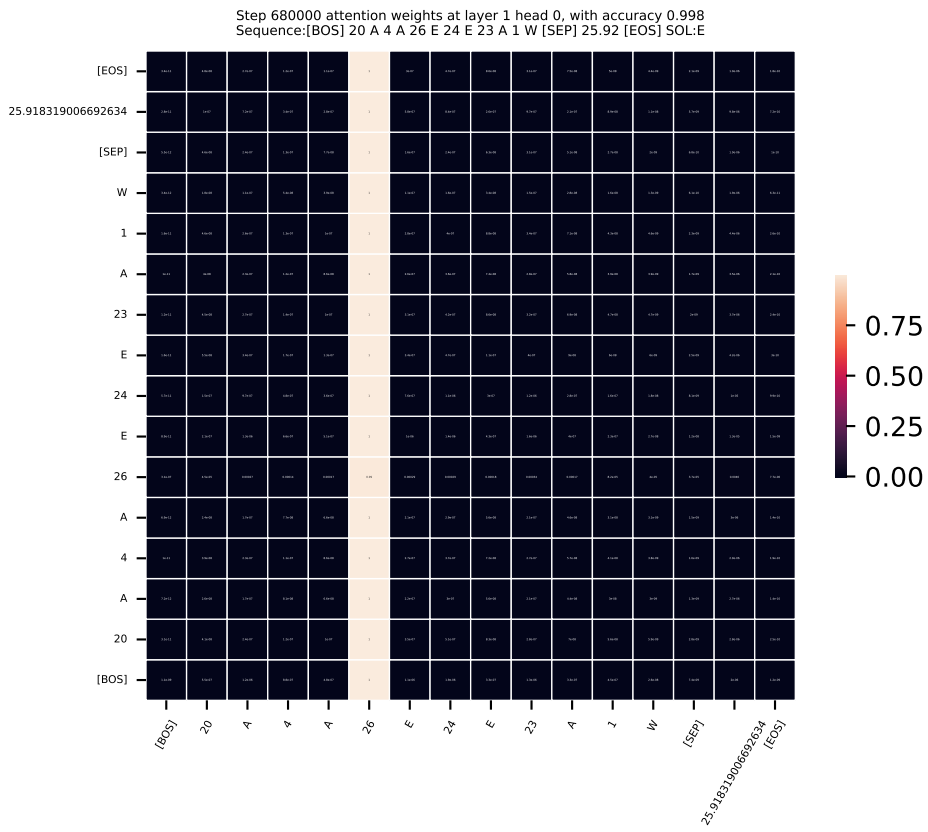

Step 679500 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.2
0.1

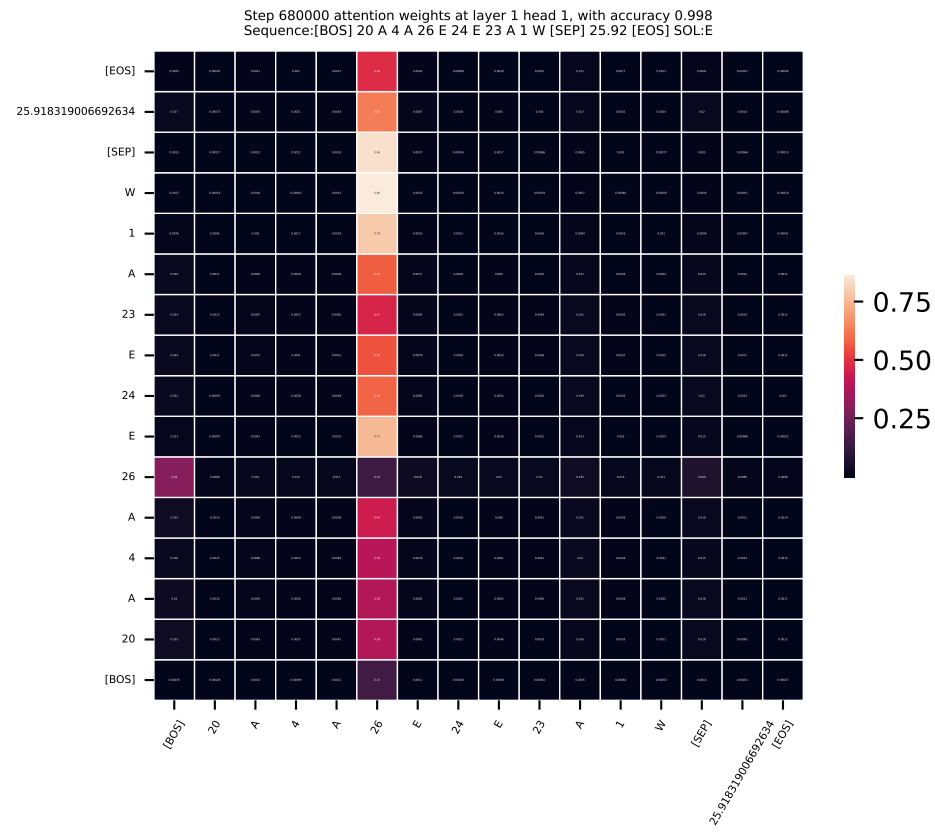

Step 680000 attention weights at layer 0 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

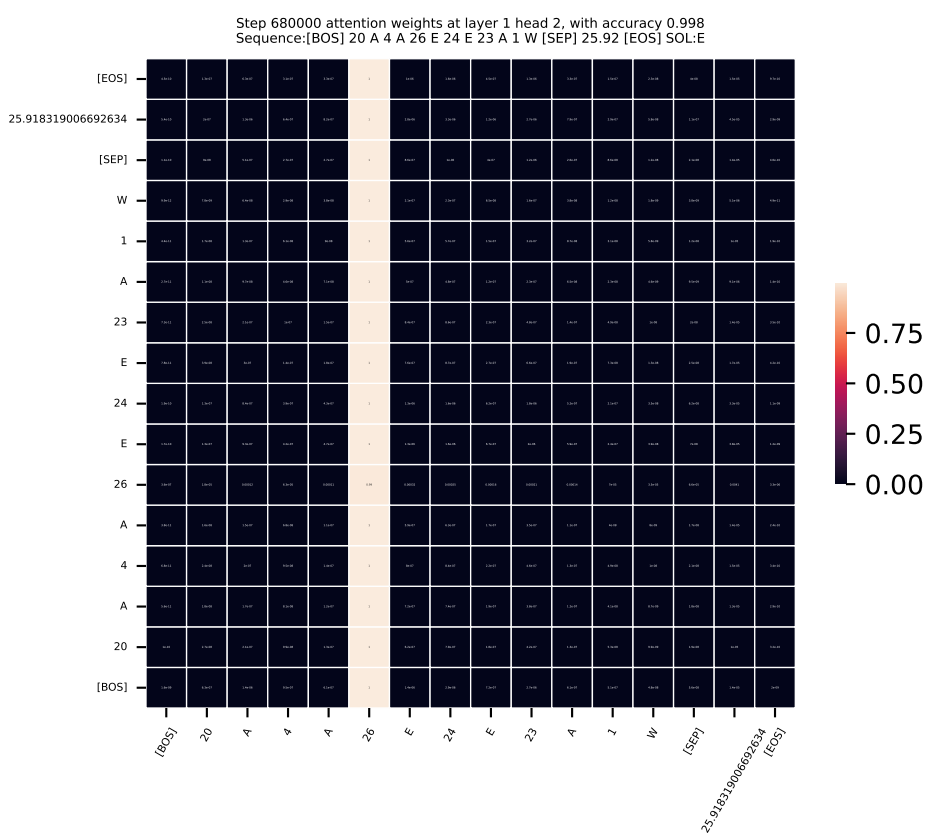

Step 680000 attention weights at layer 0 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

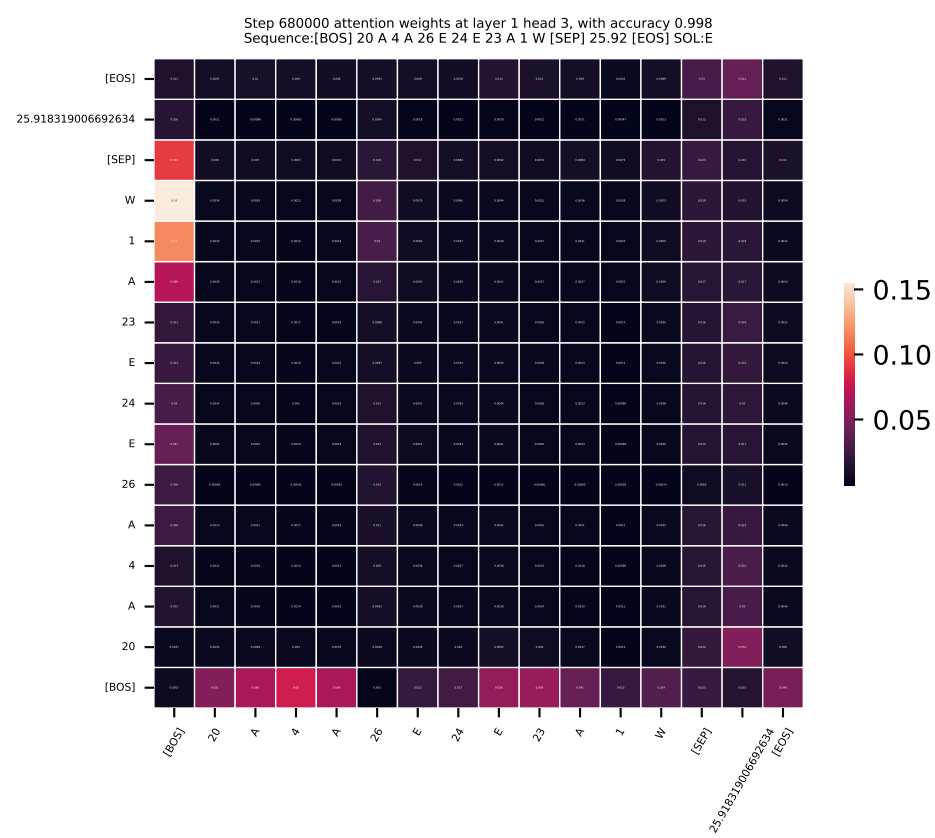

Step 680000 attention weights at layer 0 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

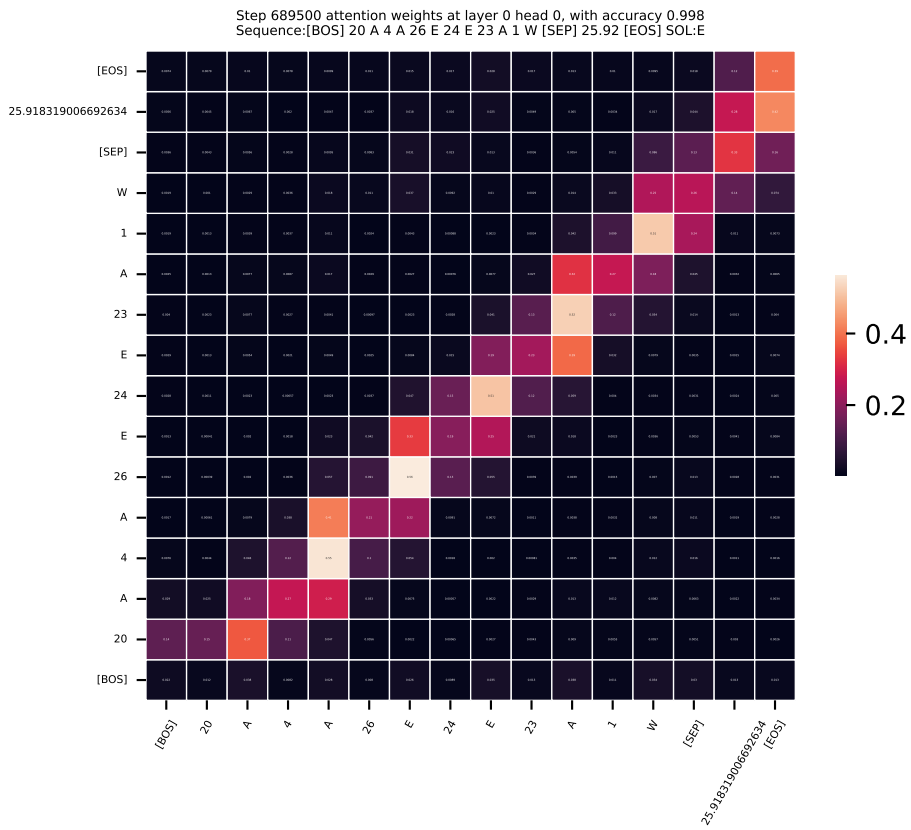

Step 680000 attention weights at layer 0 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

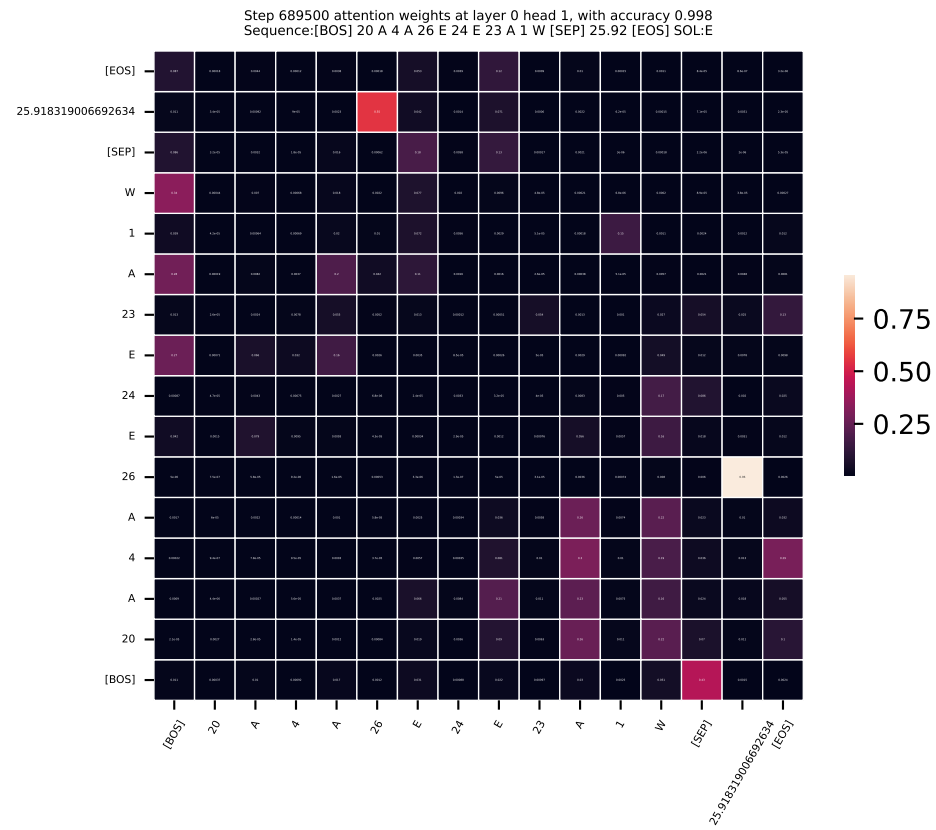

Step 680000 attention weights at layer 1 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

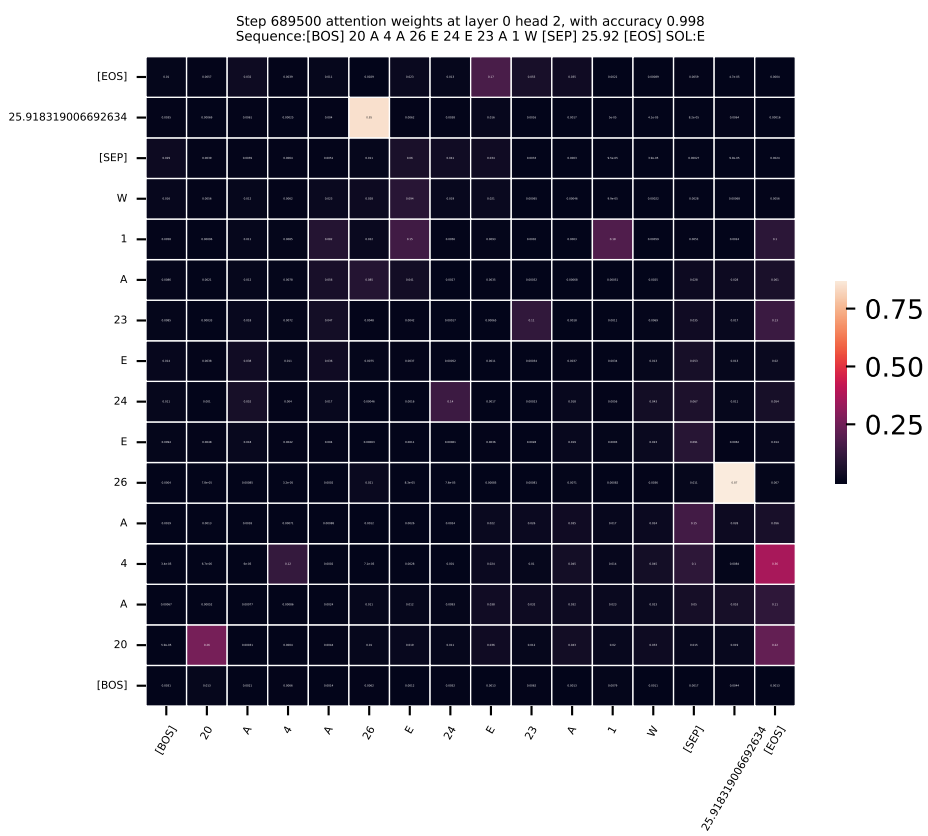

Step 680000 attention weights at layer 1 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

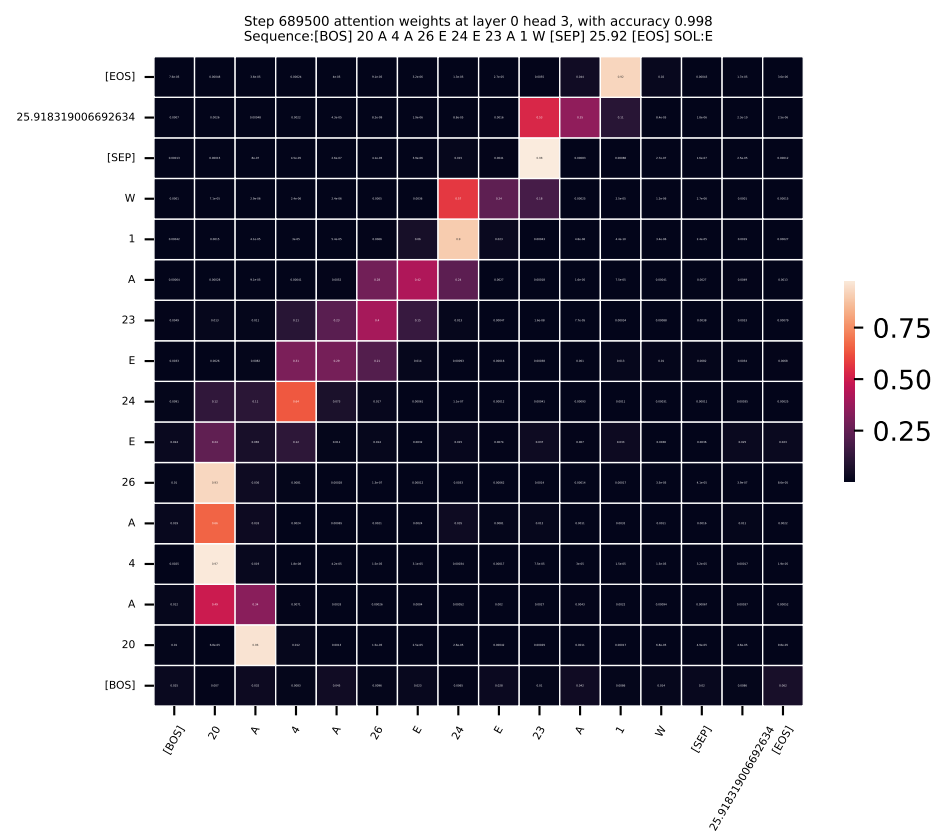

Step 680000 attention weights at layer 1 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

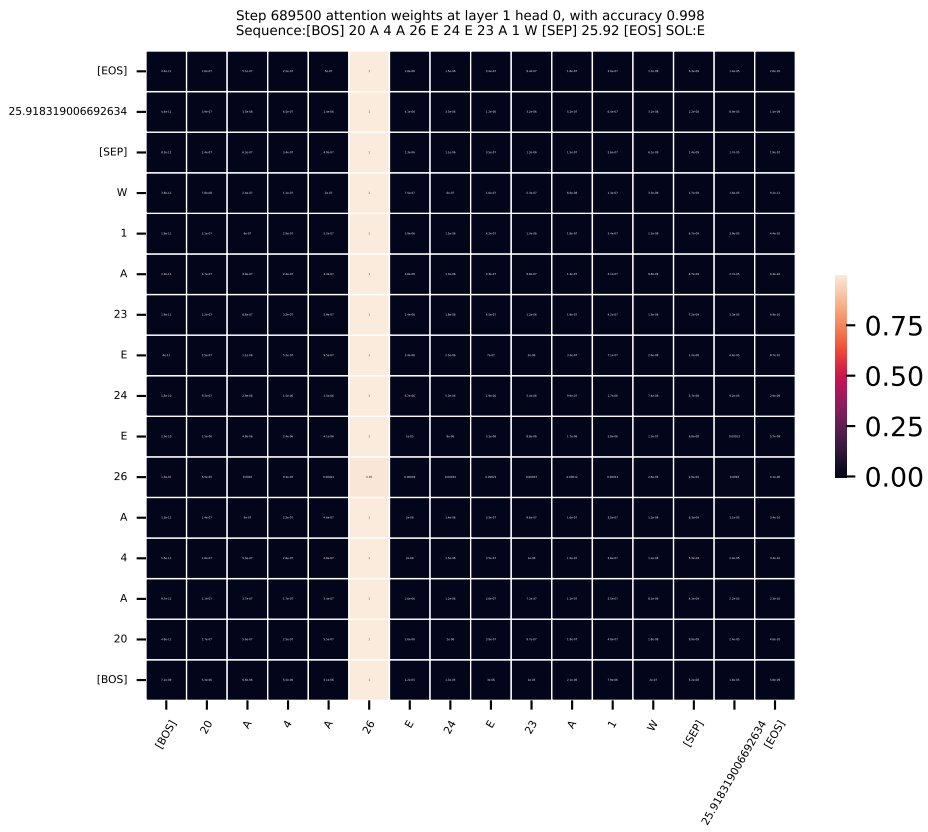

Step 680000 attention weights at layer 1 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.15
0.10
0.05

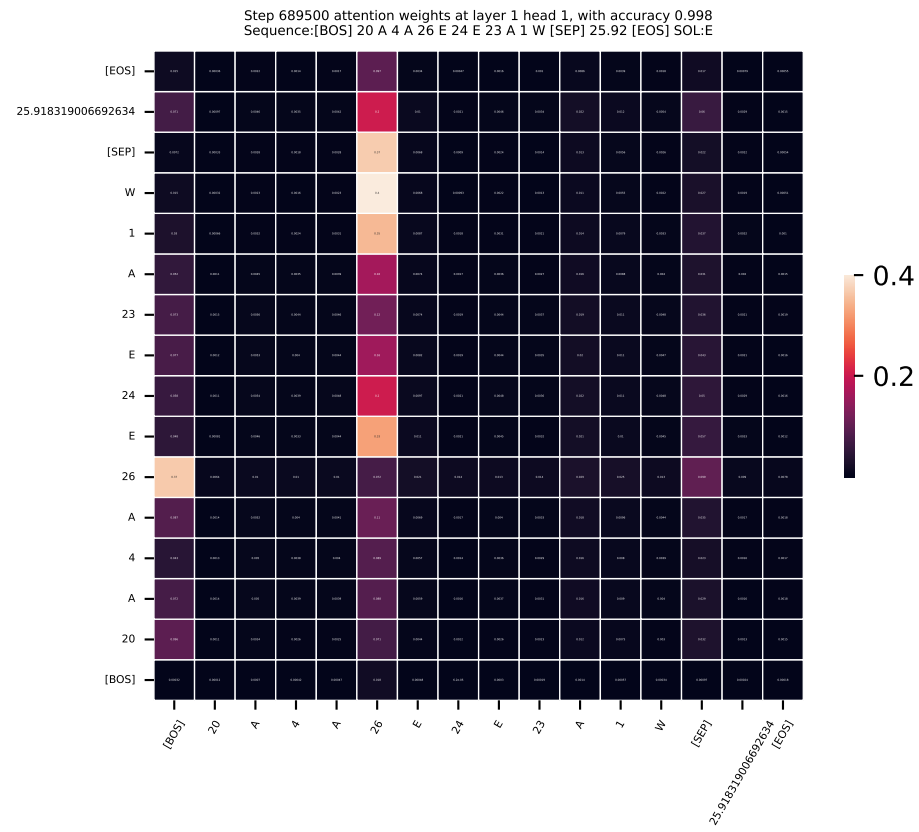

Step 689500 attention weights at layer 0 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

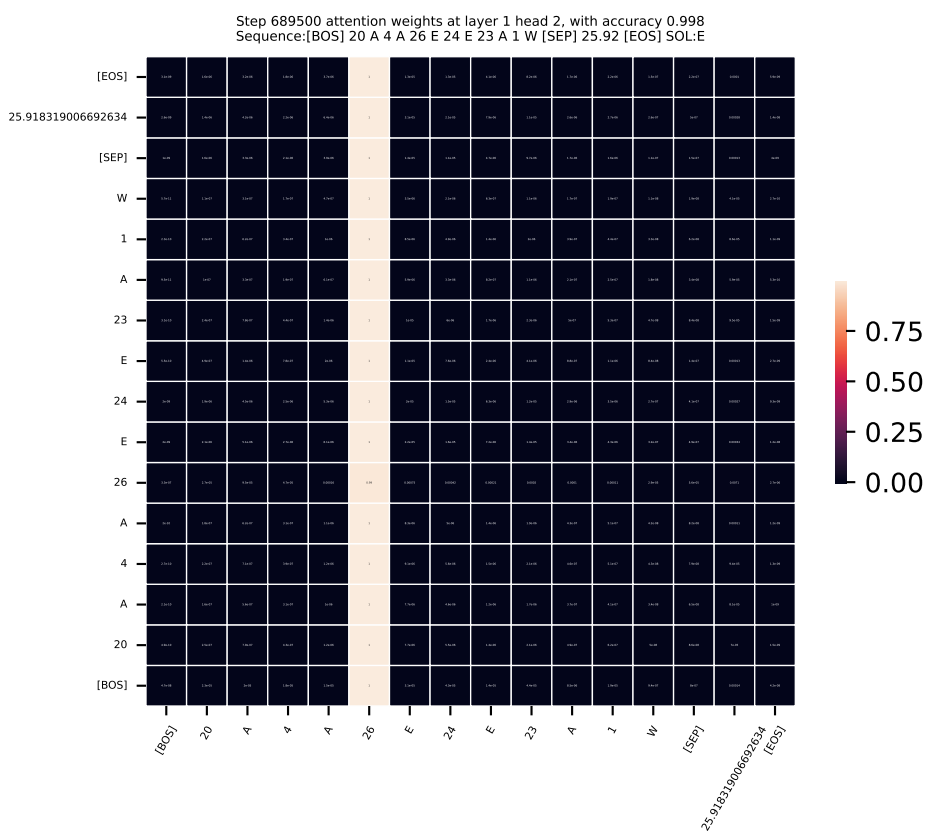

Step 689500 attention weights at layer 0 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

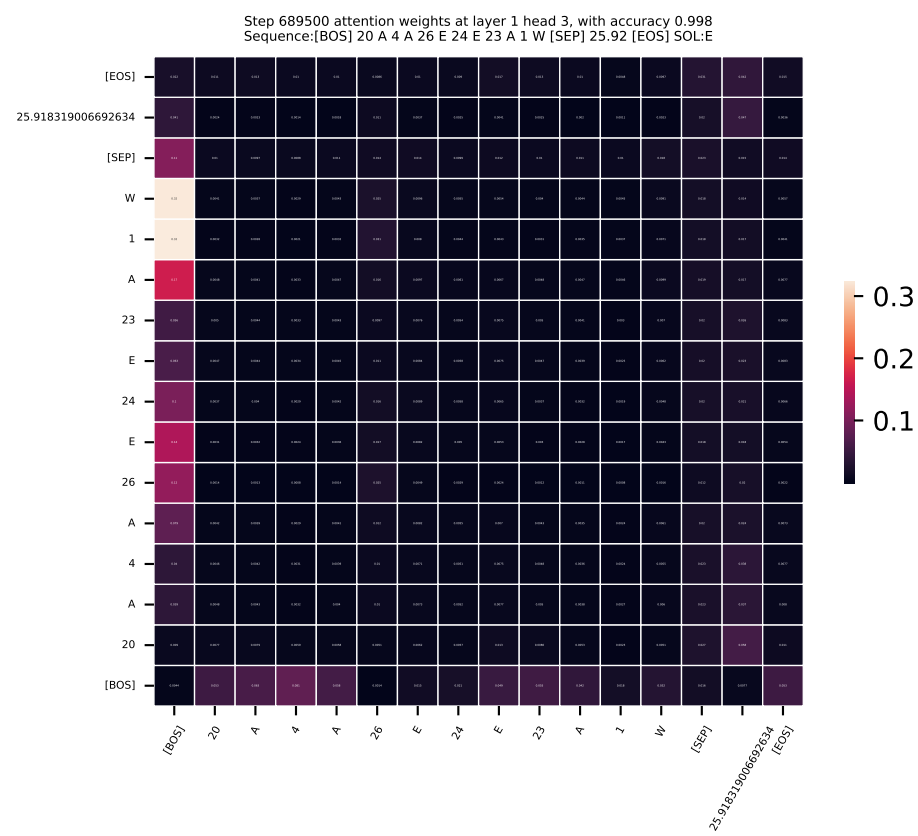

Step 689500 attention weights at layer 0 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

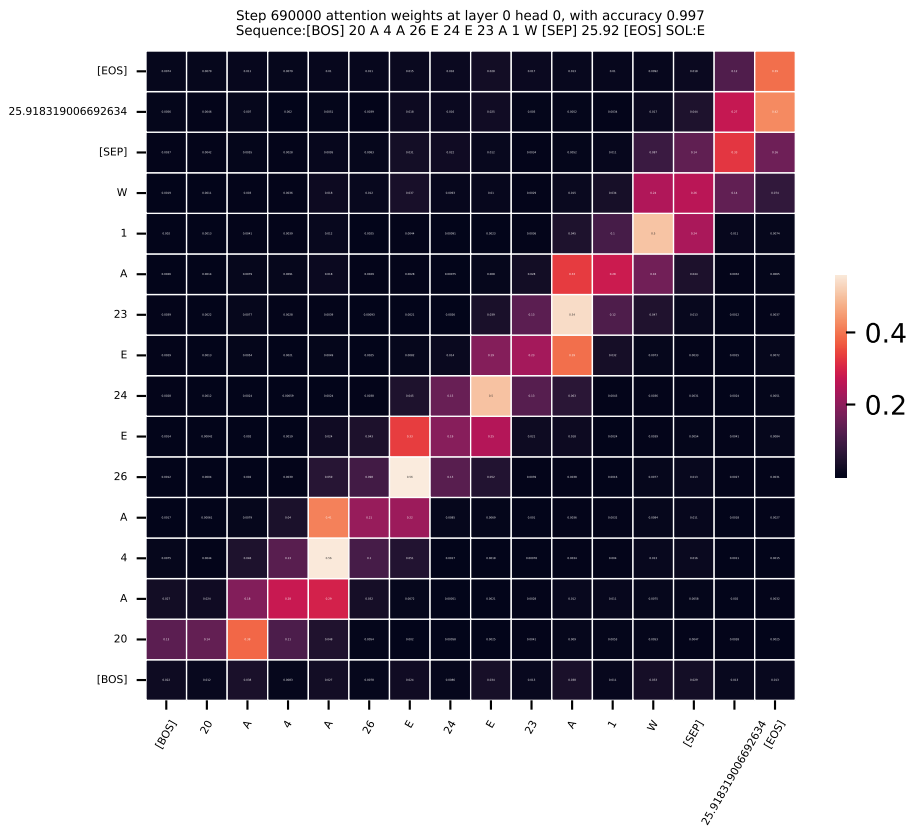

Step 689500 attention weights at layer 0 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

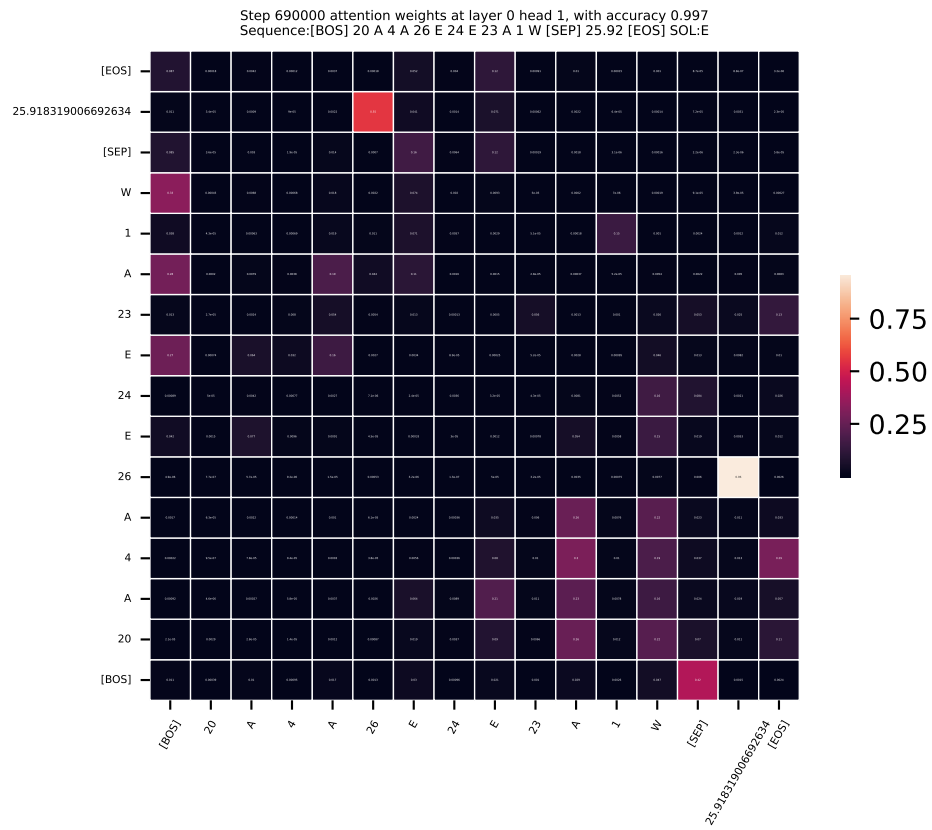

Step 689500 attention weights at layer 1 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

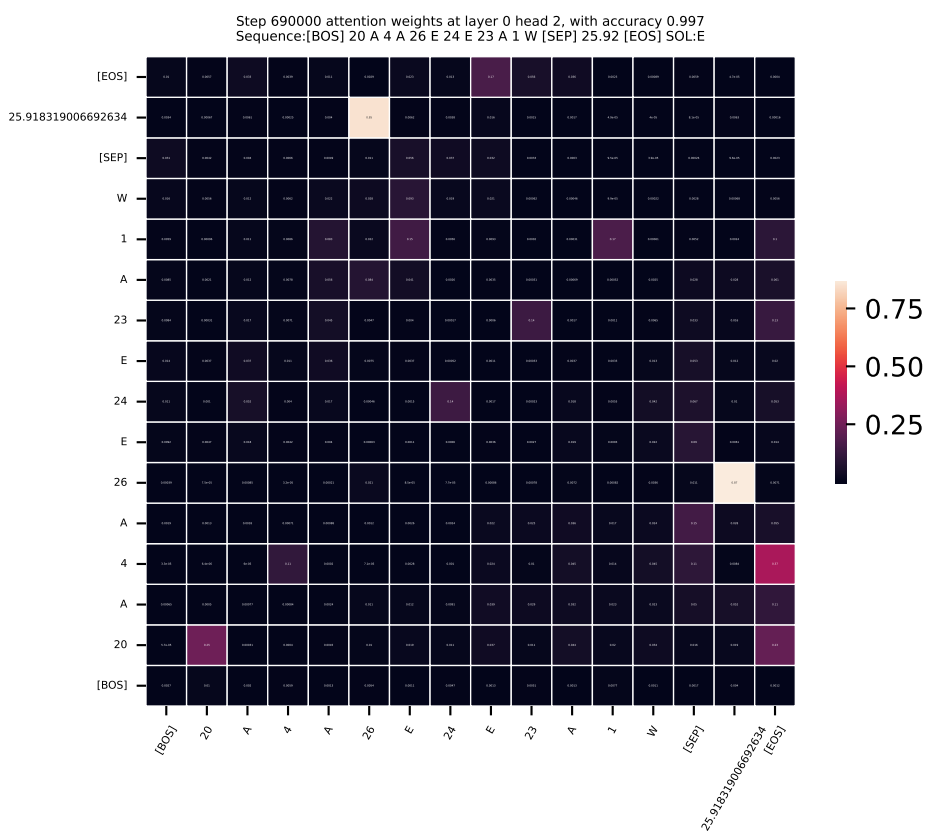

Step 689500 attention weights at layer 1 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

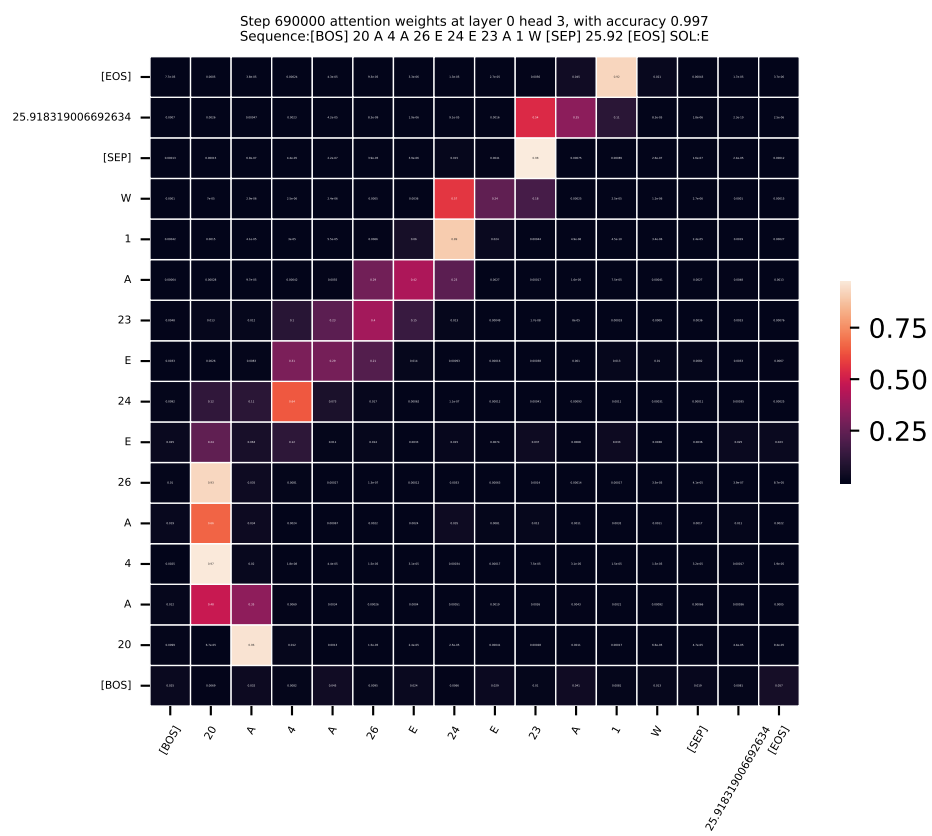

Step 689500 attention weights at layer 1 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

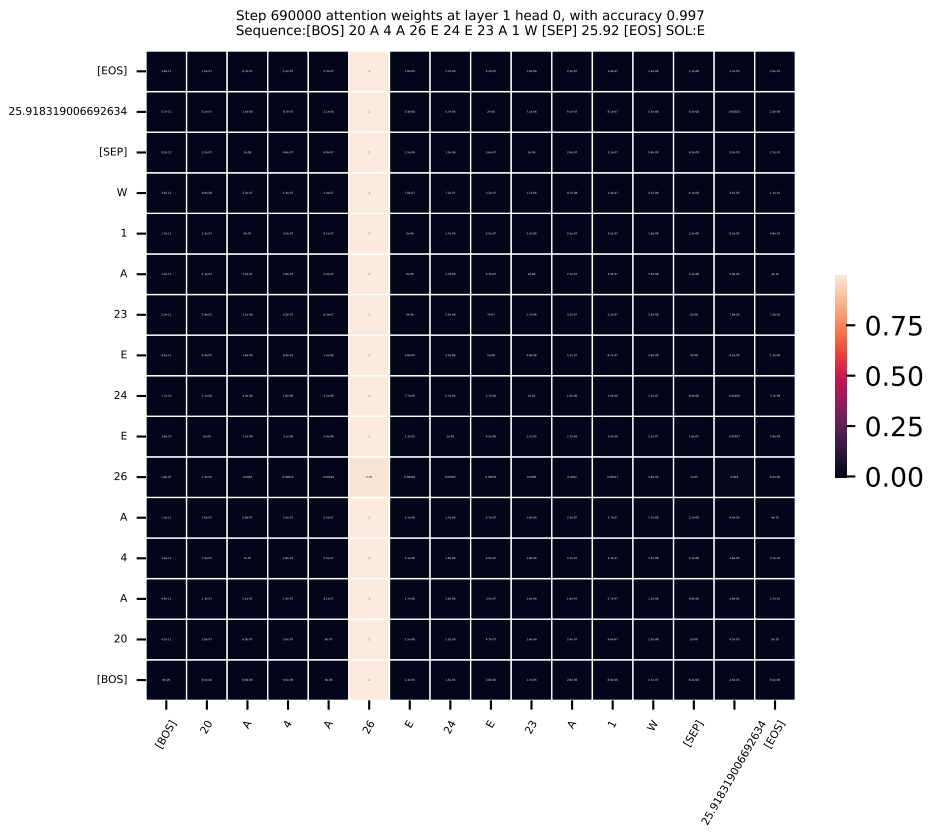

Step 689500 attention weights at layer 1 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

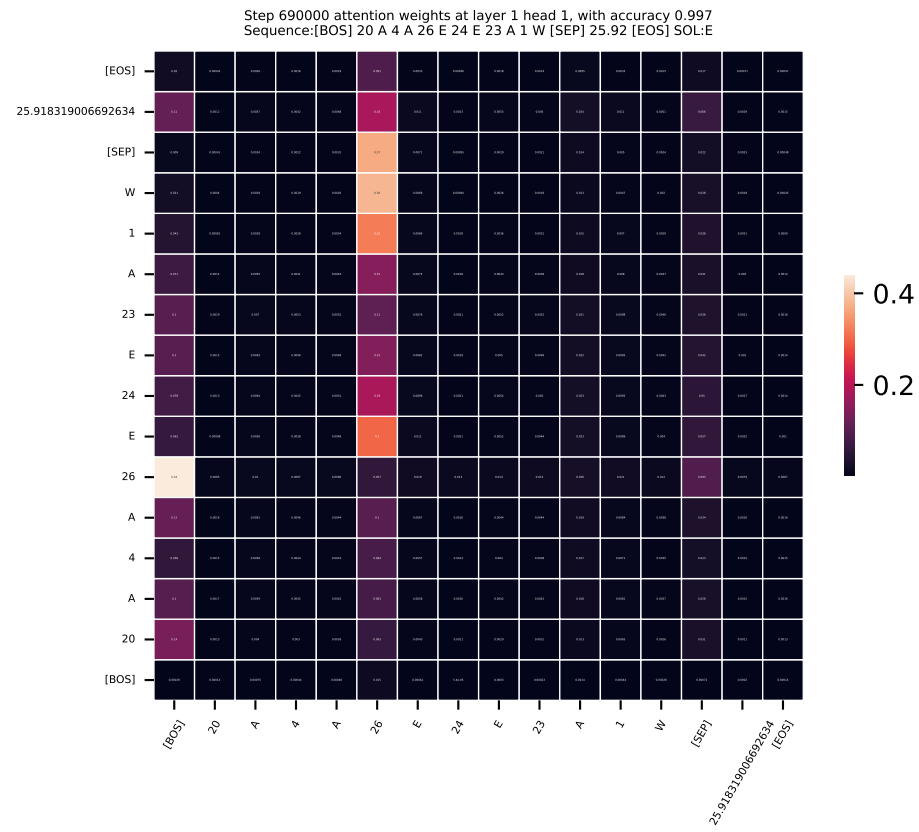

Step 690000 attention weights at layer 0 head 0, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

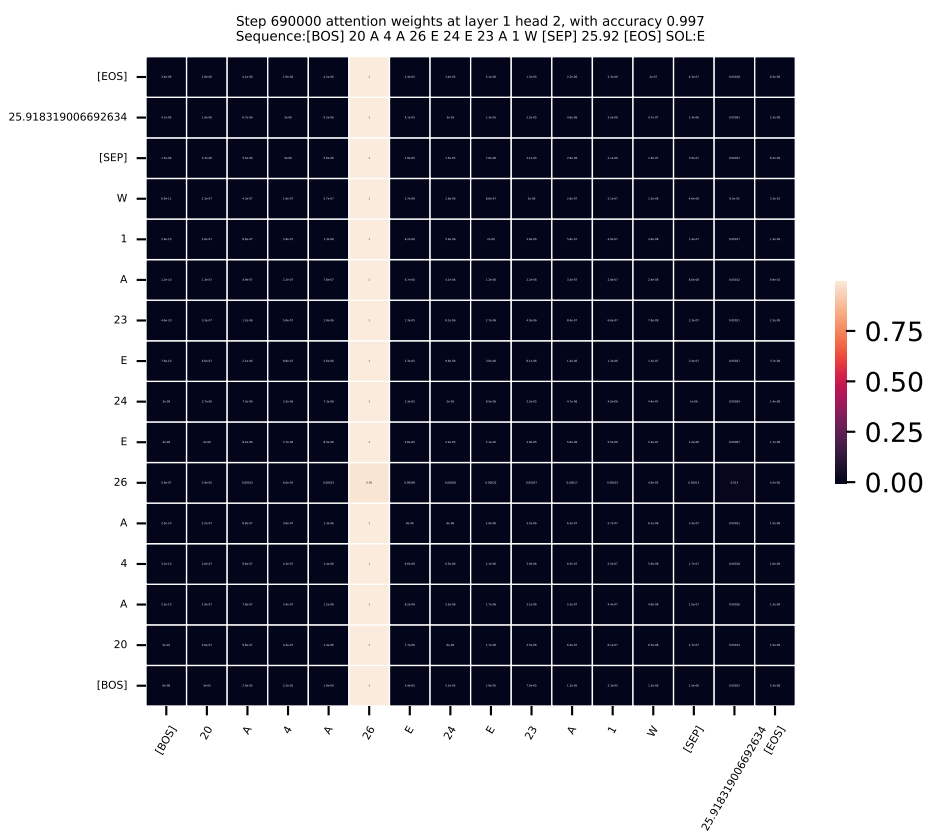

Step 690000 attention weights at layer 0 head 1, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

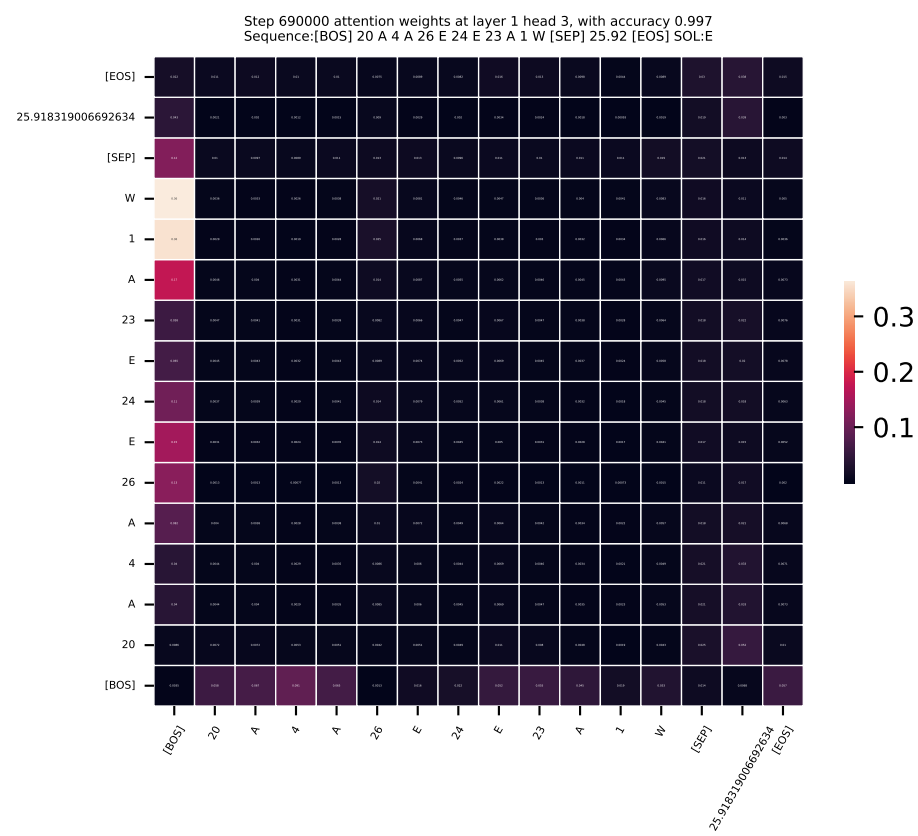

Step 690000 attention weights at layer 0 head 2, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

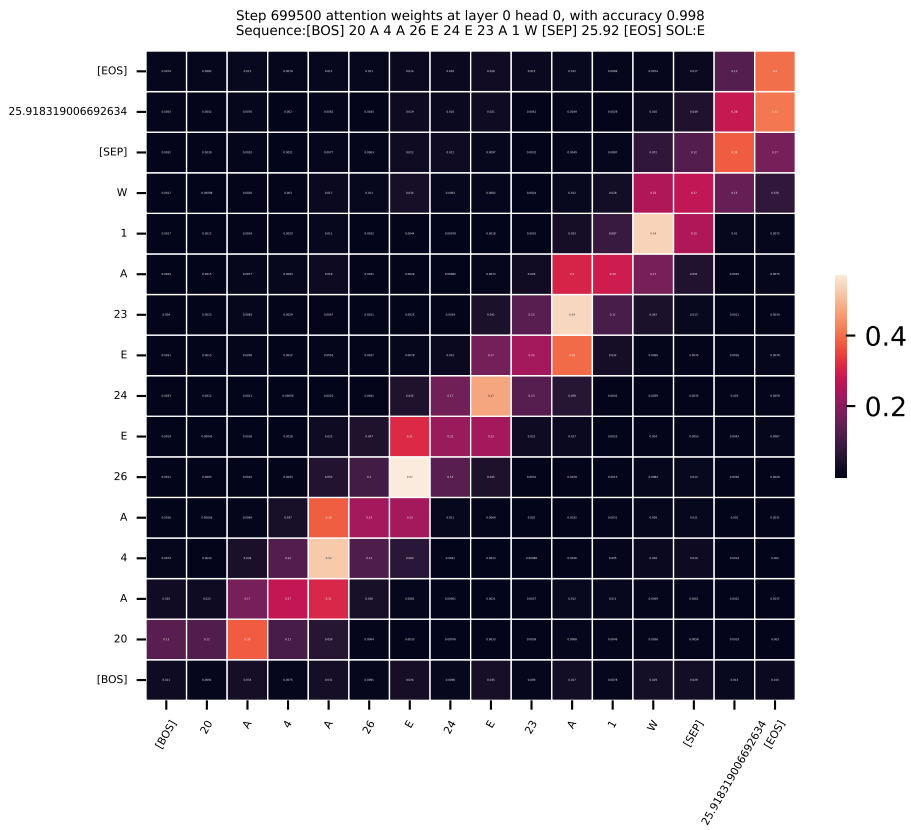

Step 690000 attention weights at layer 0 head 3, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

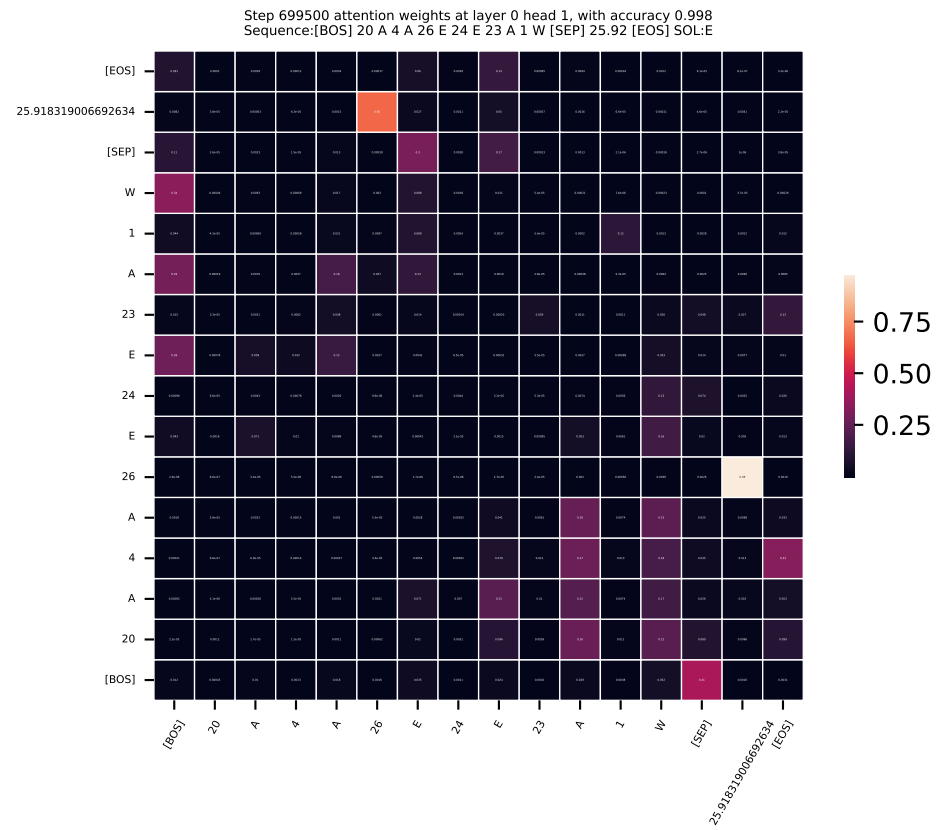

Step 690000 attention weights at layer 1 head 0, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

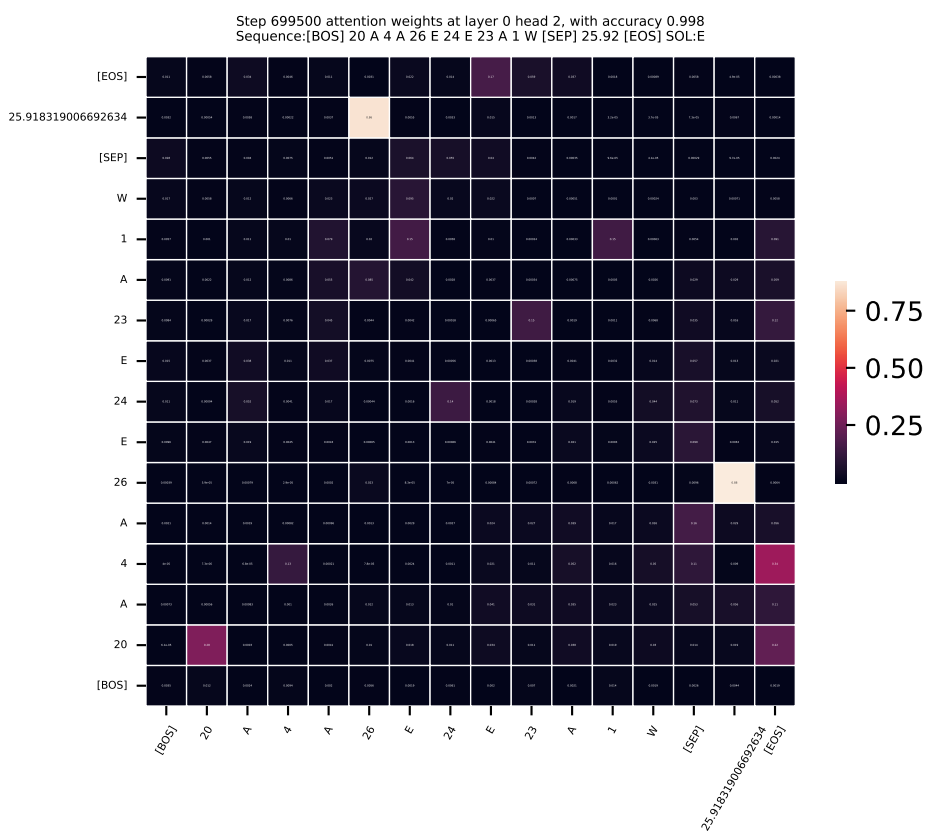

Step 690000 attention weights at layer 1 head 1, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

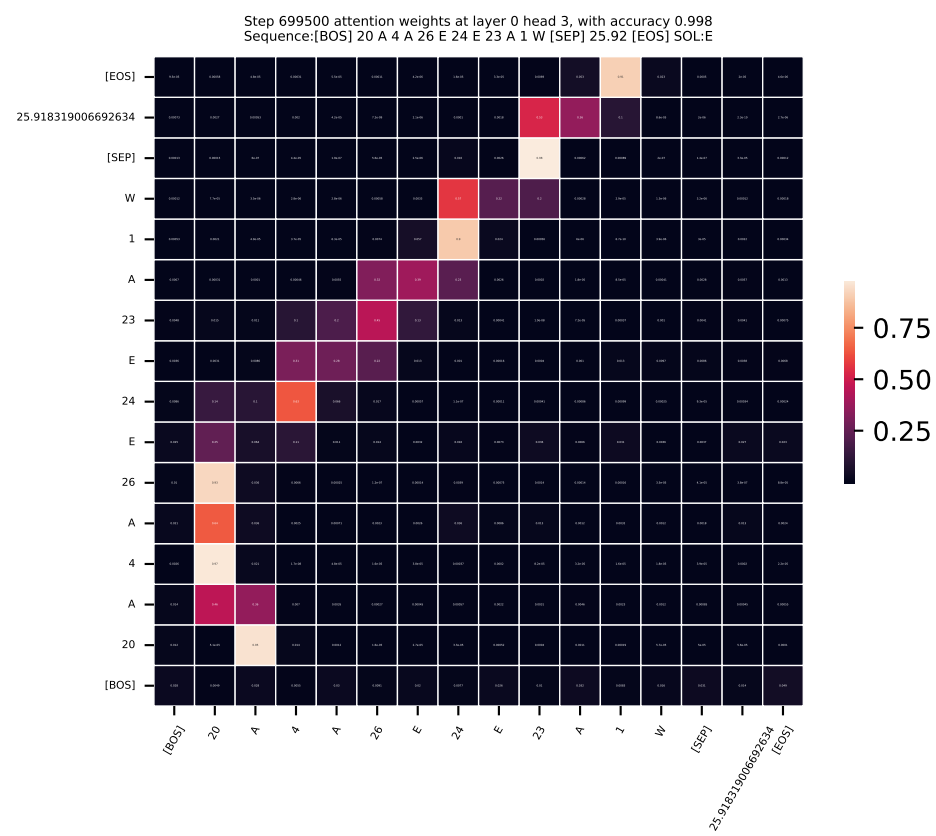

Step 690000 attention weights at layer 1 head 2, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

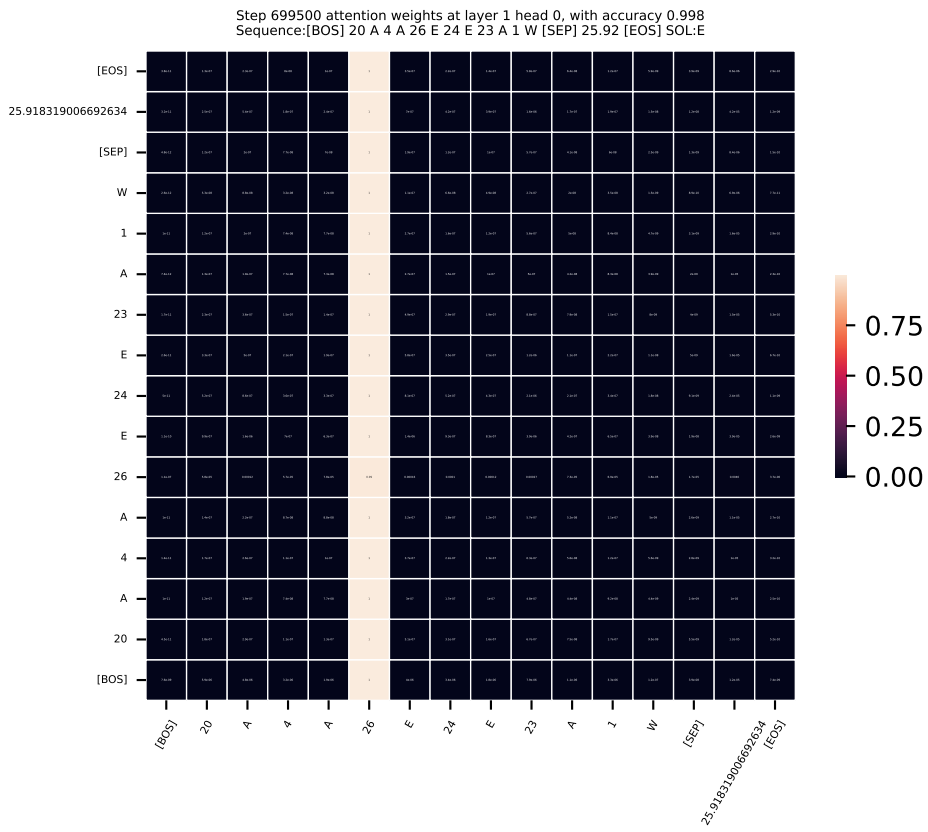

Step 690000 attention weights at layer 1 head 3, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

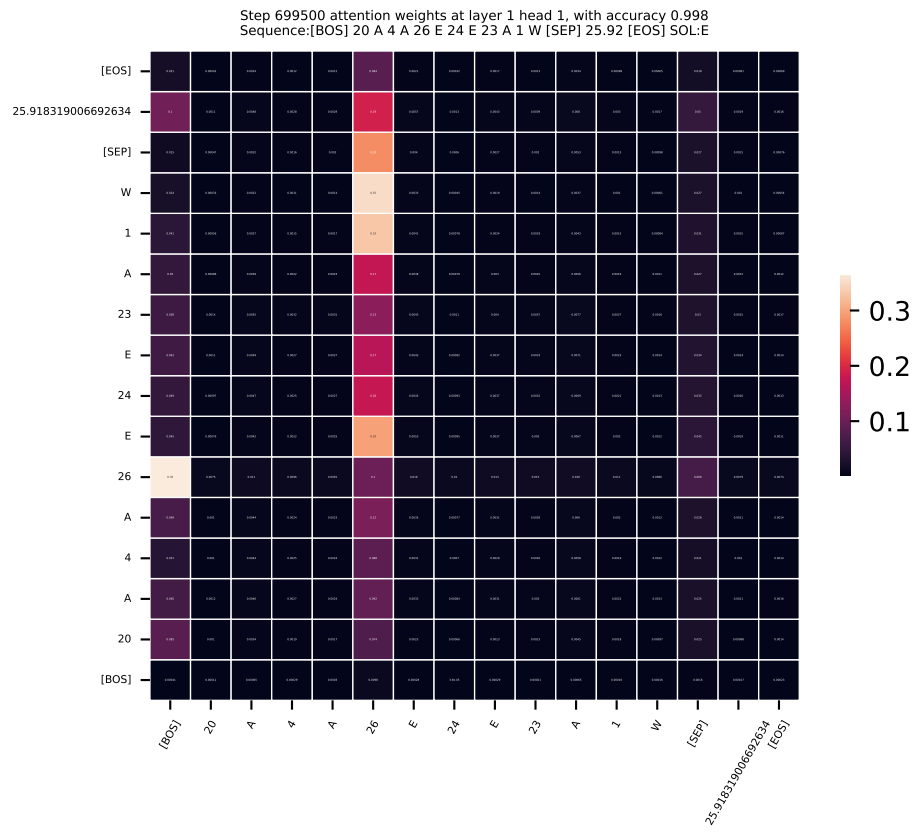

Step 699500 attention weights at layer 0 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

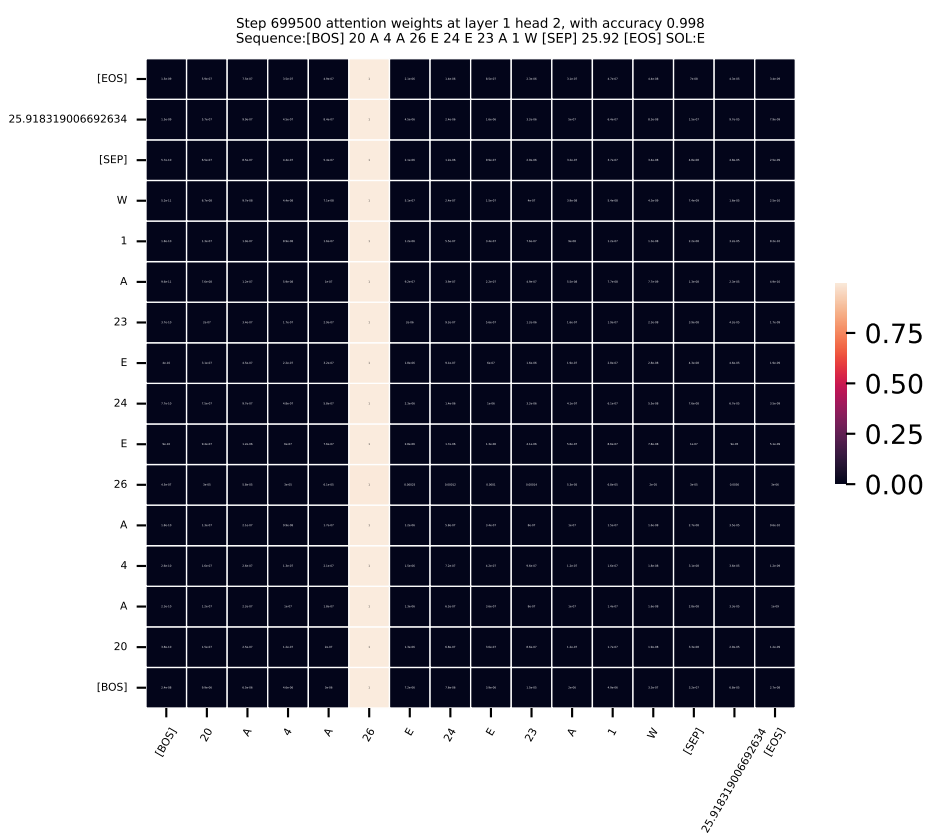

Step 699500 attention weights at layer 0 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

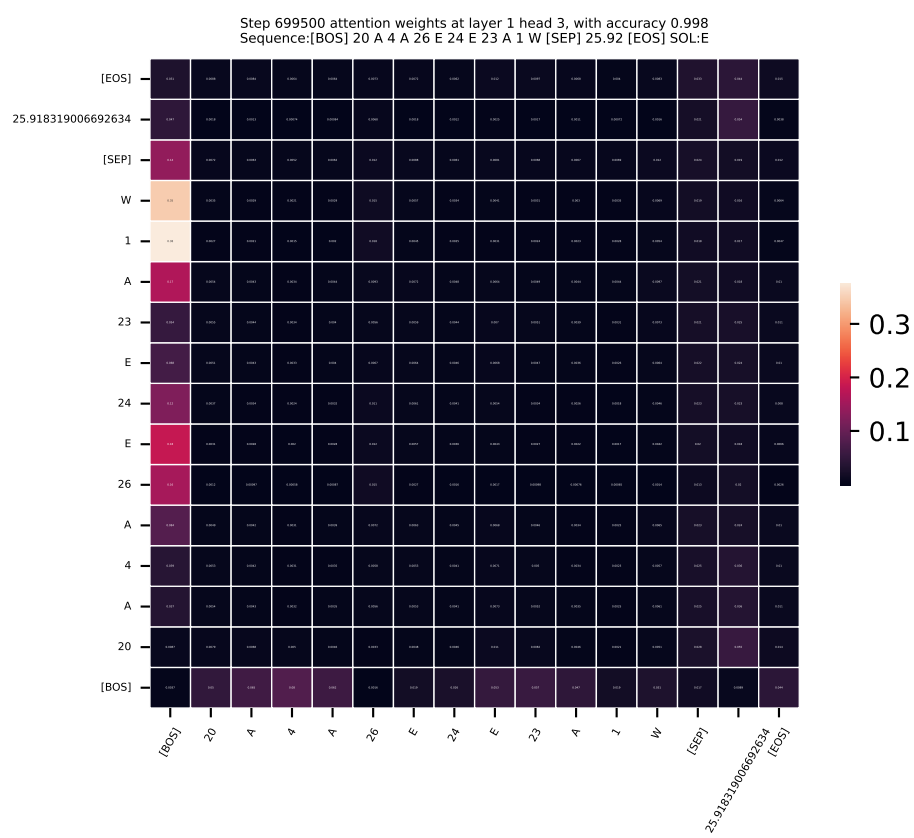

Step 699500 attention weights at layer 0 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

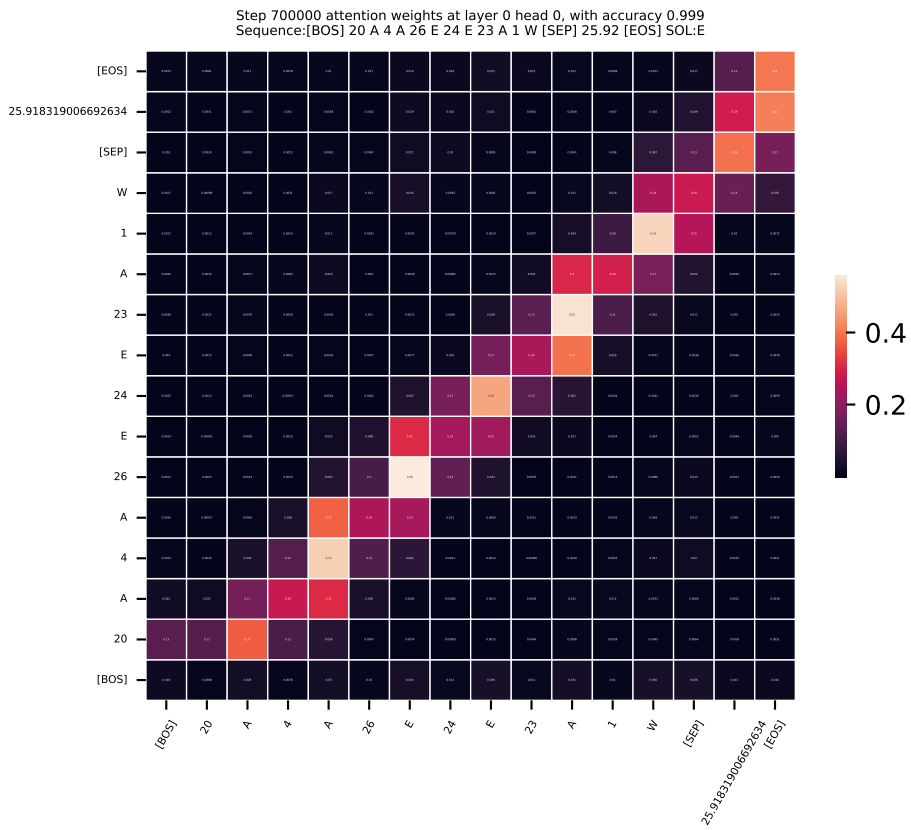

Step 699500 attention weights at layer 0 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

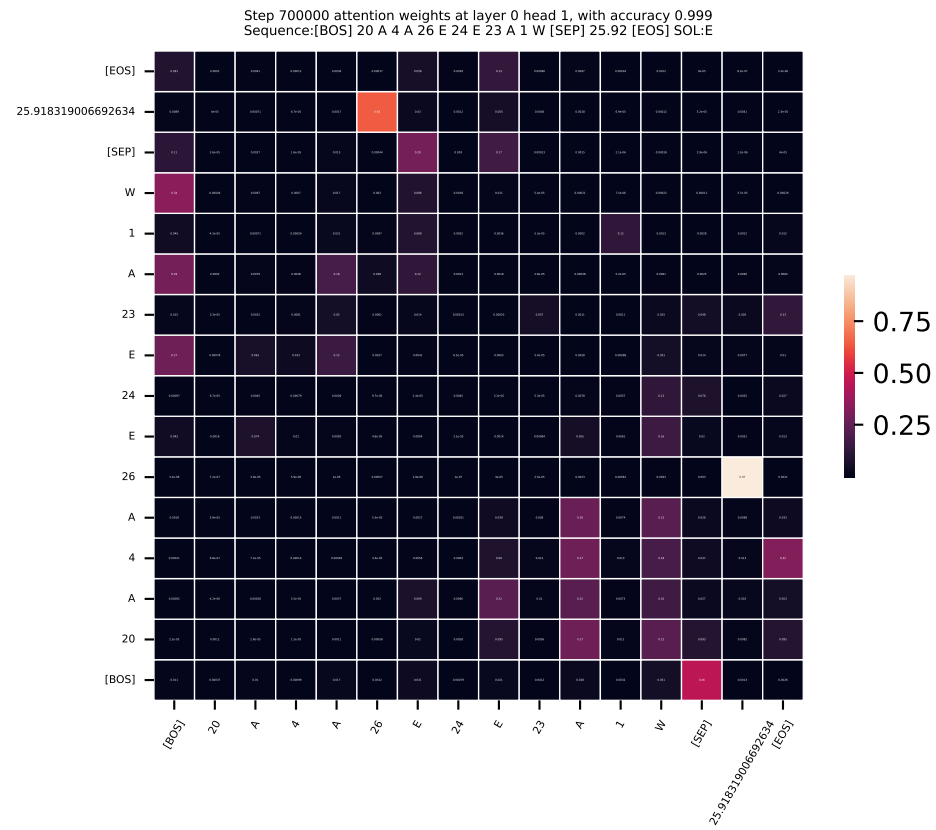

Step 699500 attention weights at layer 1 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25
0.00

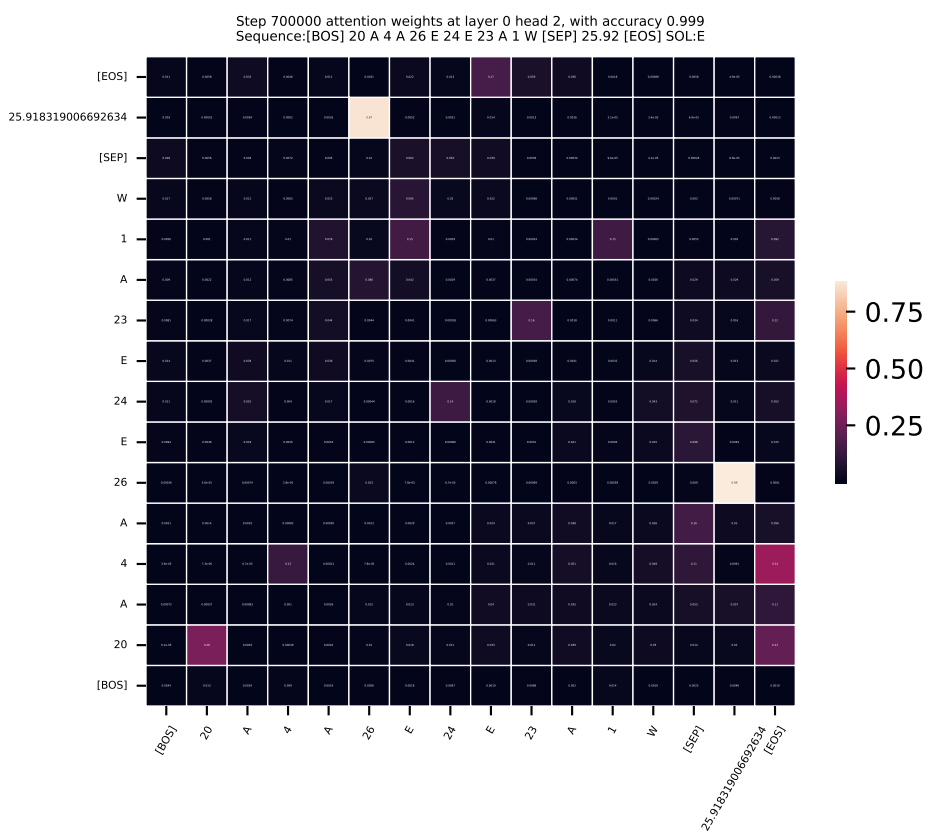

Step 699500 attention weights at layer 1 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.3
0.2
0.1

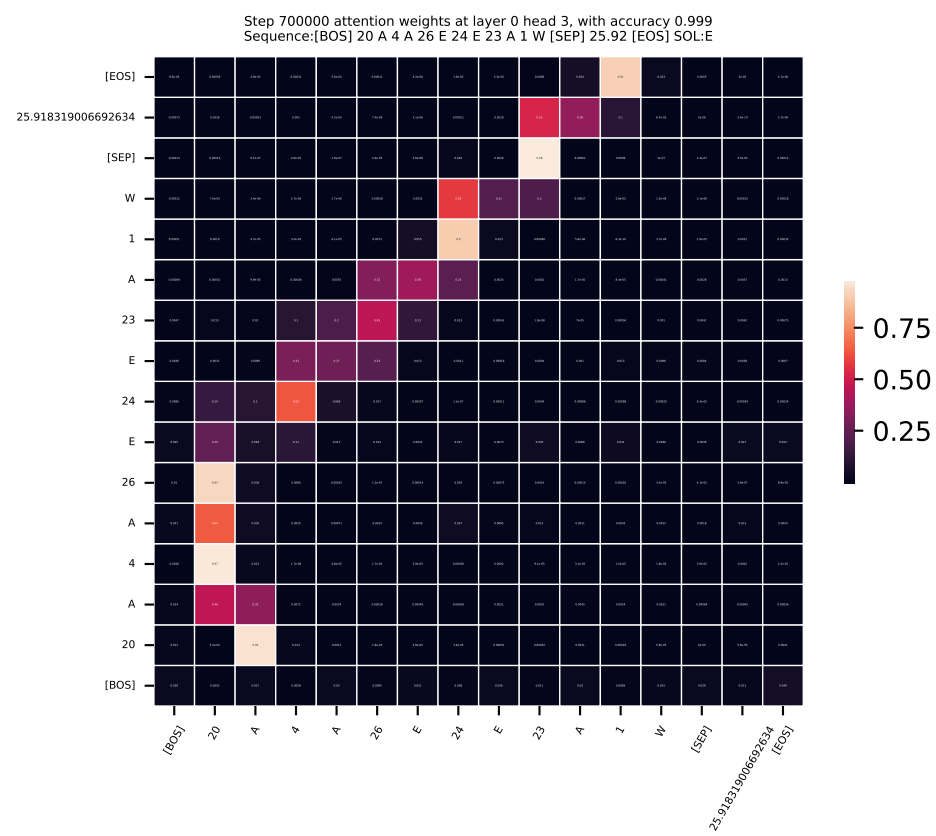

Step 699500 attention weights at layer 1 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25
0.00

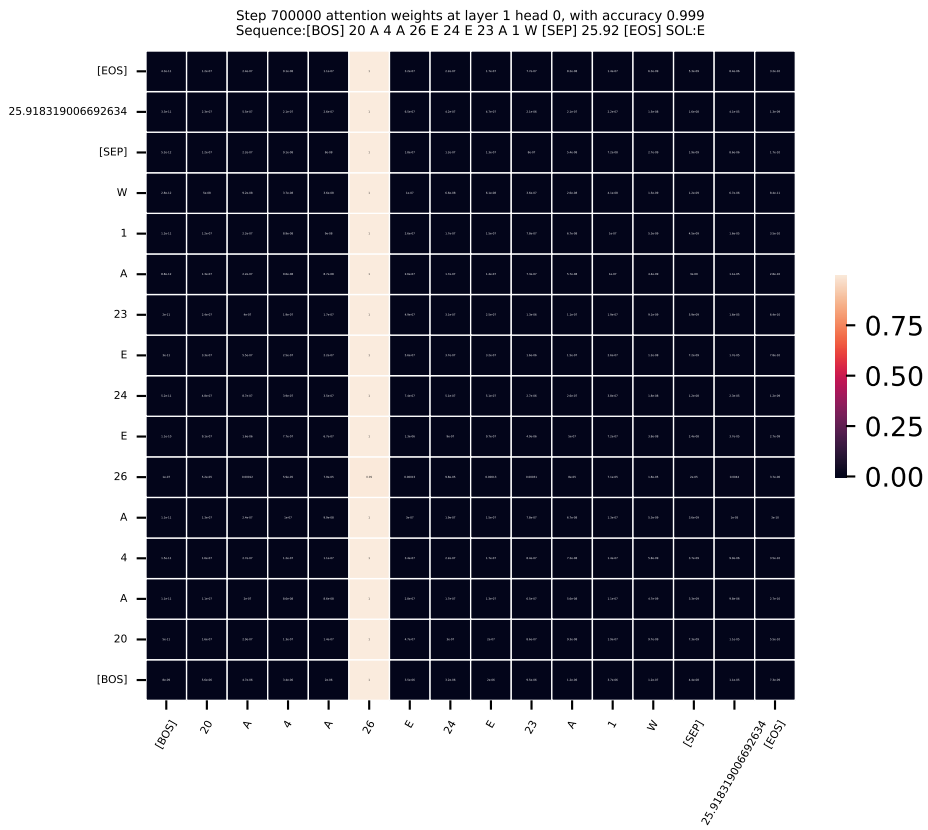

Step 699500 attention weights at layer 1 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.3
0.2
0.1

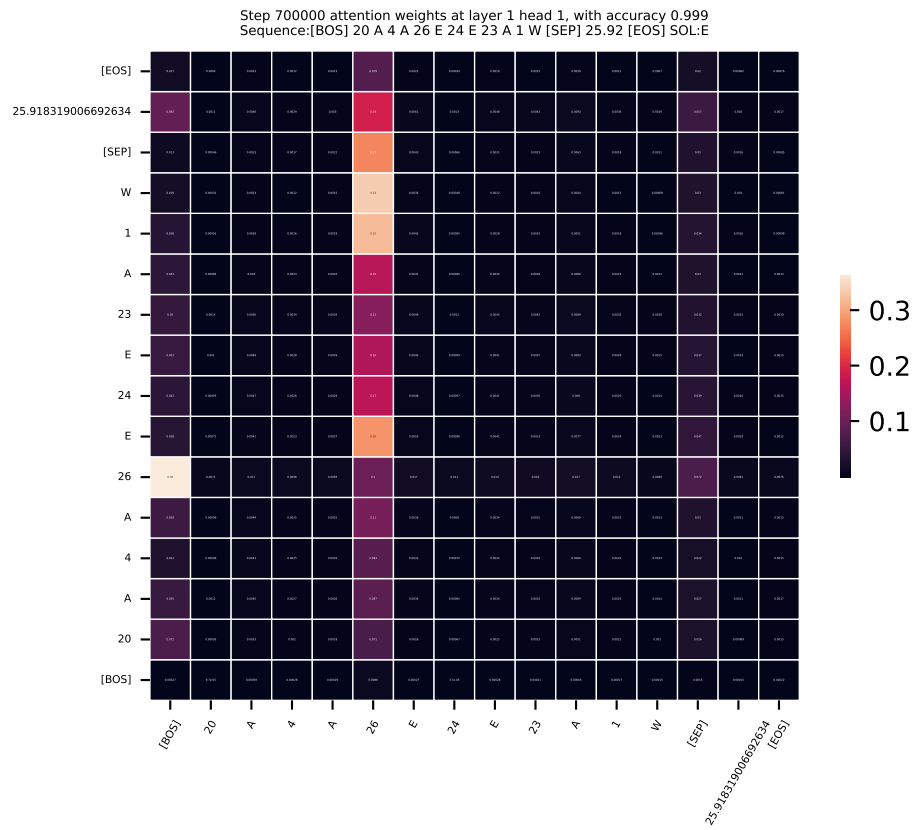

Step 700000 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

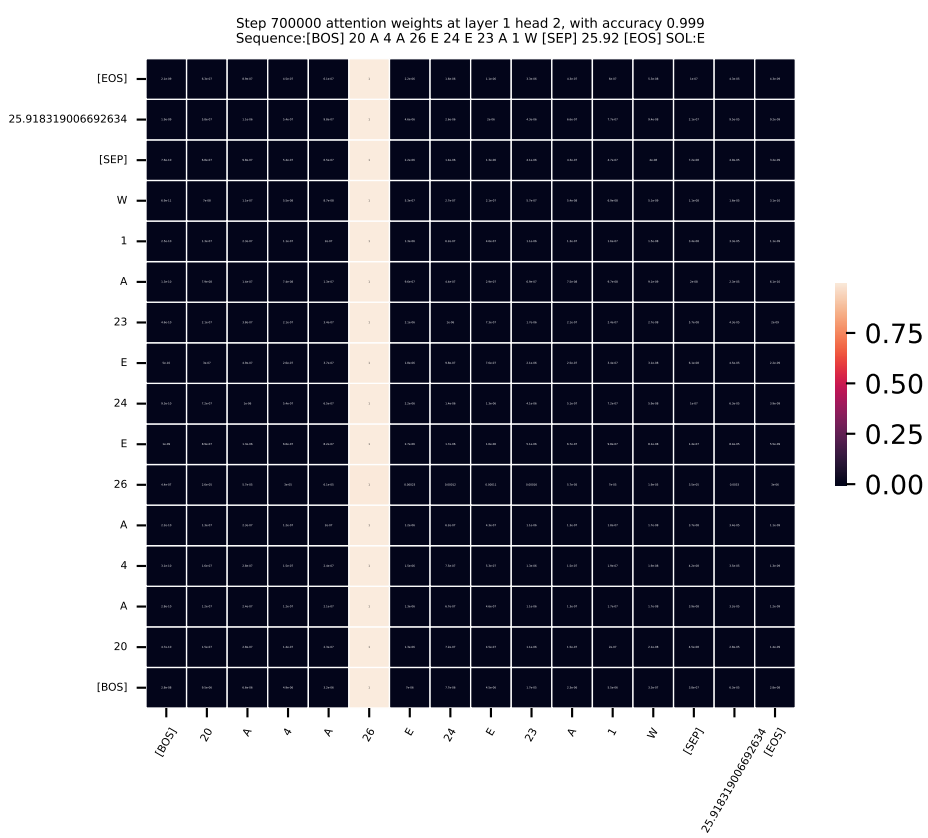

Step 700000 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

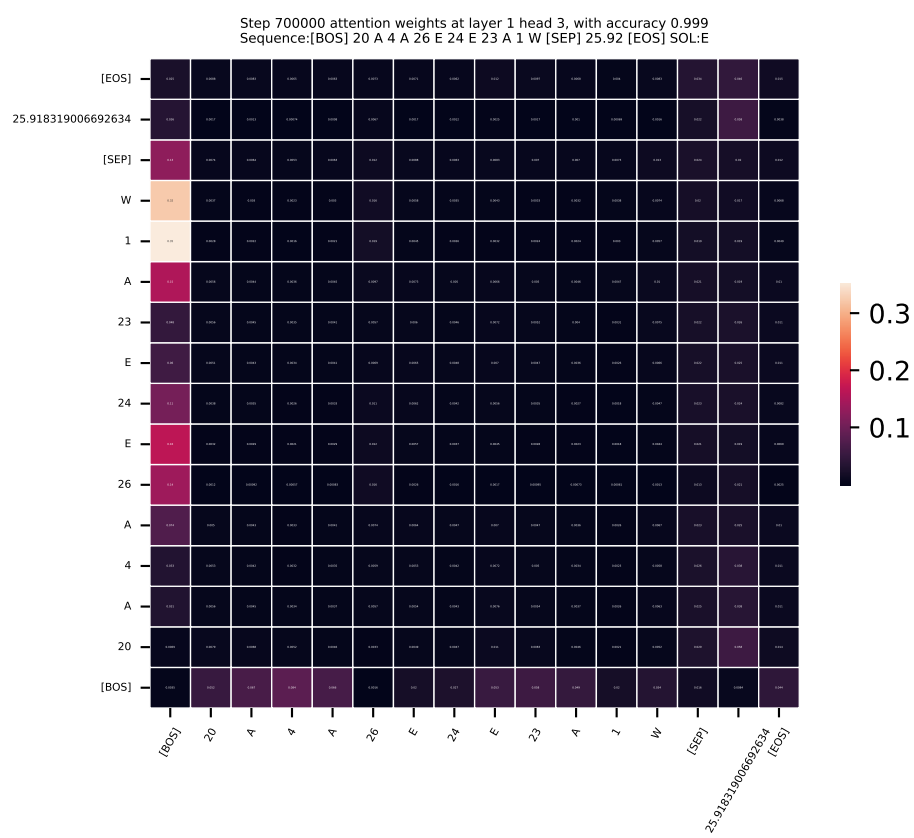

Step 700000 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

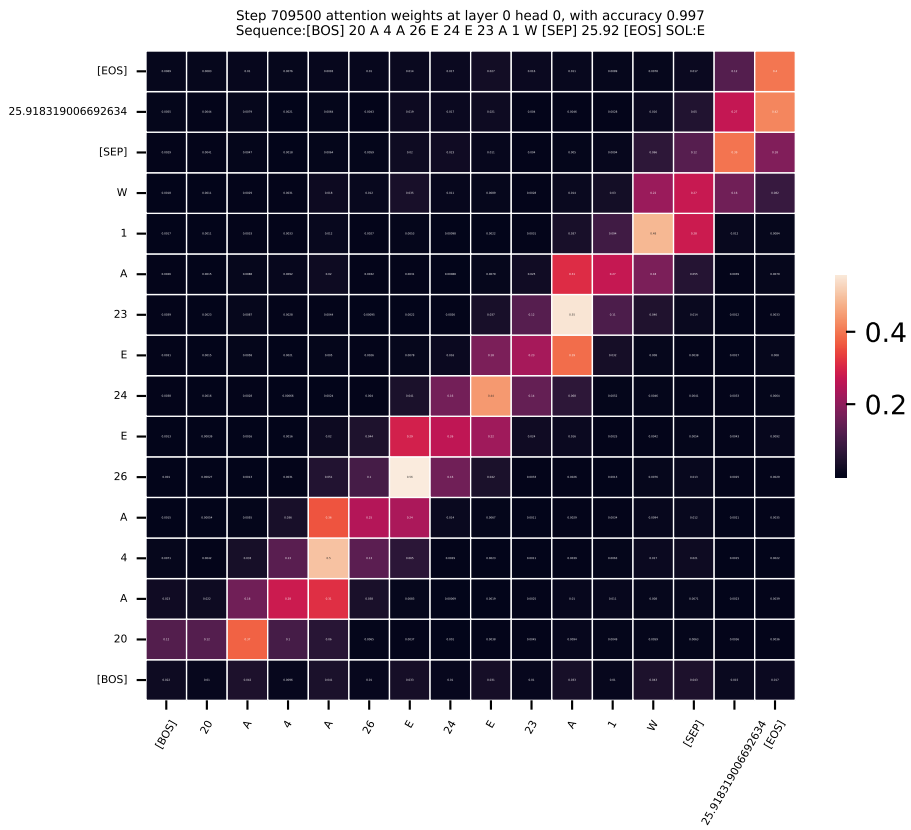

Step 700000 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

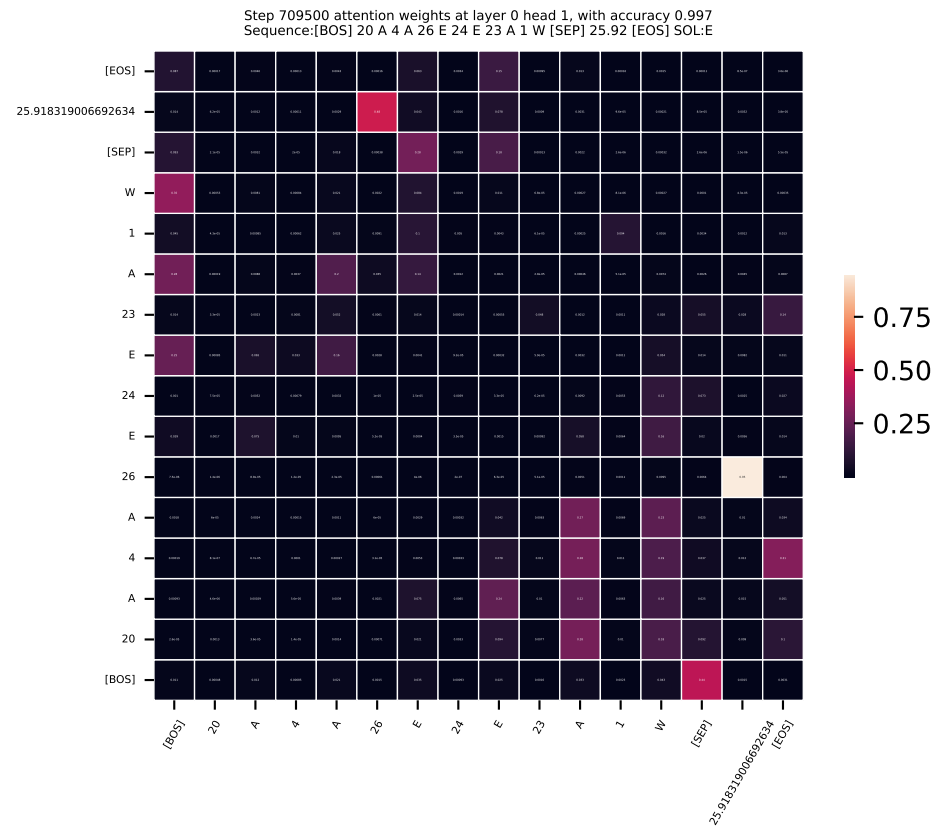

Step 700000 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

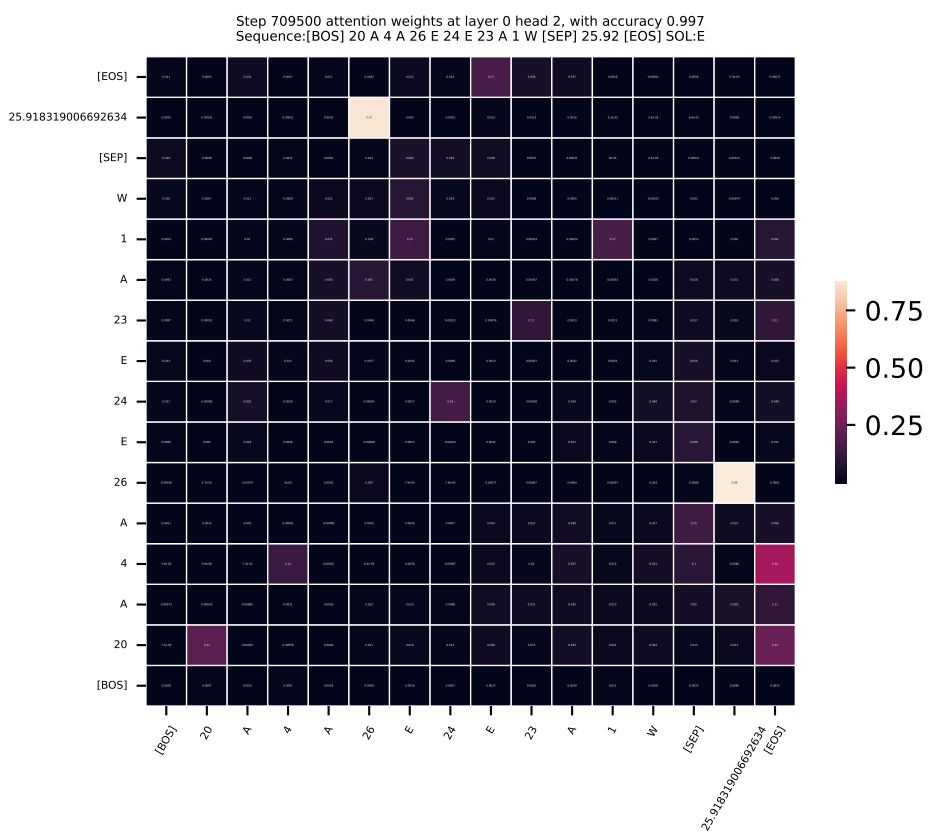

Step 700000 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

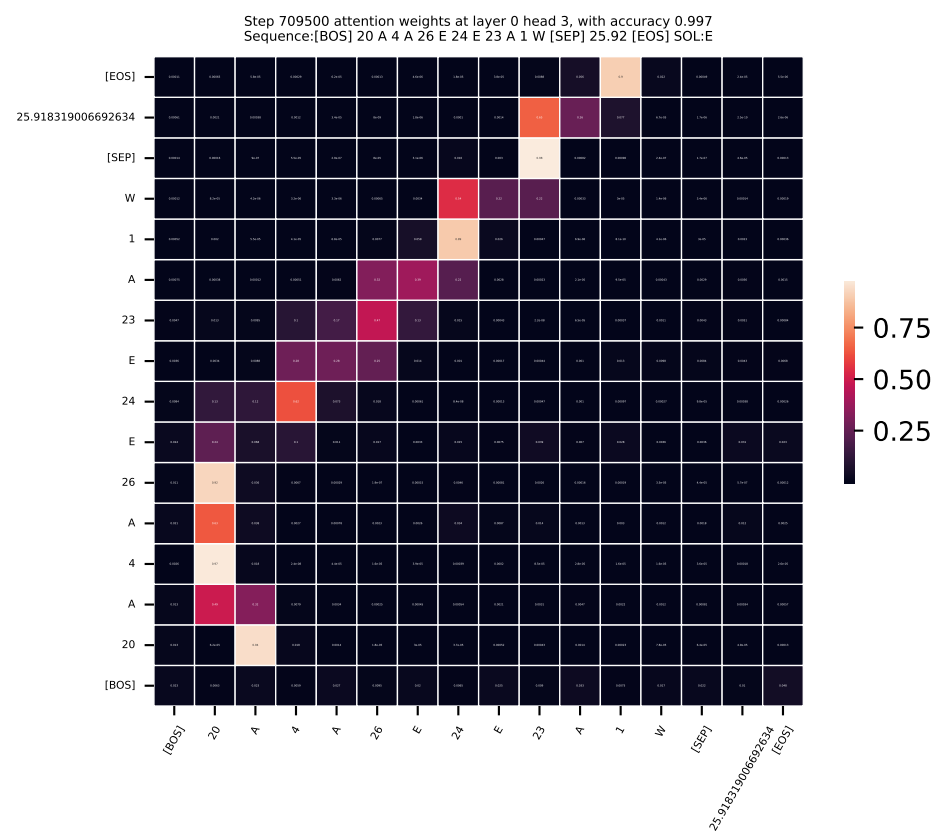

Step 700000 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

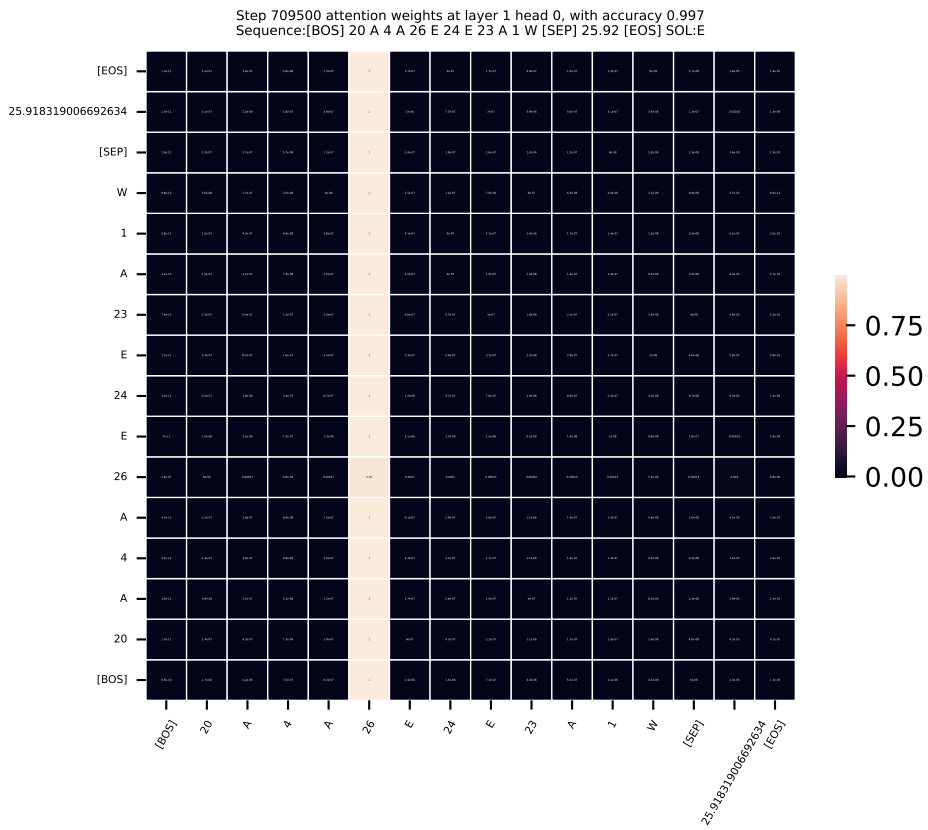

Step 700000 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

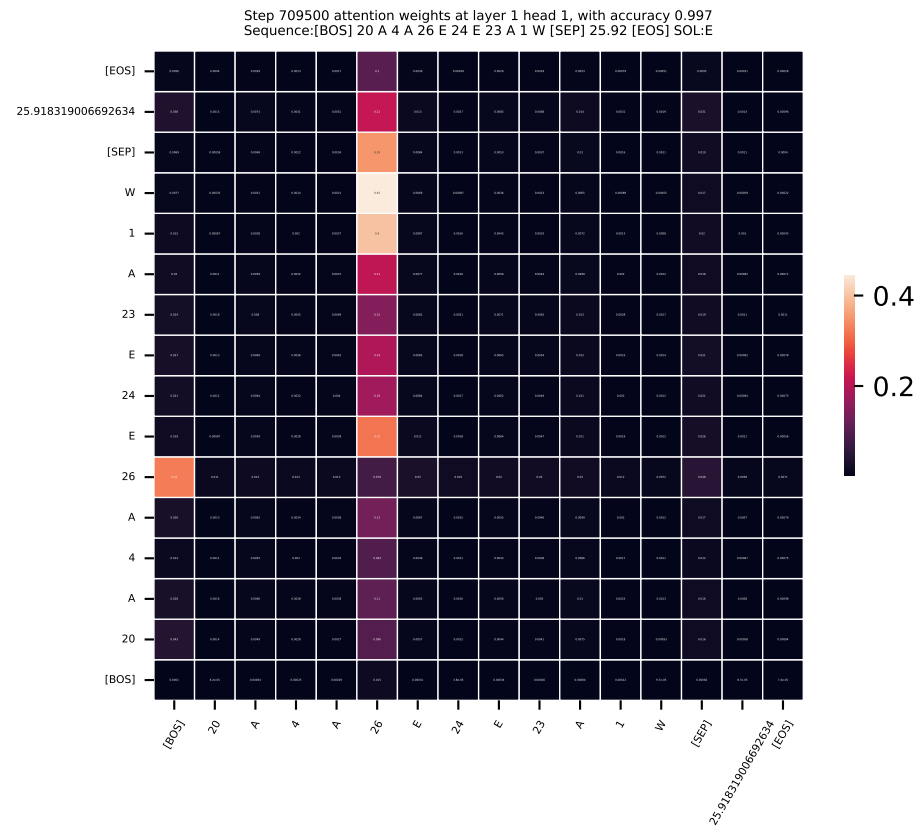

Step 709500 attention weights at layer 0 head 0, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

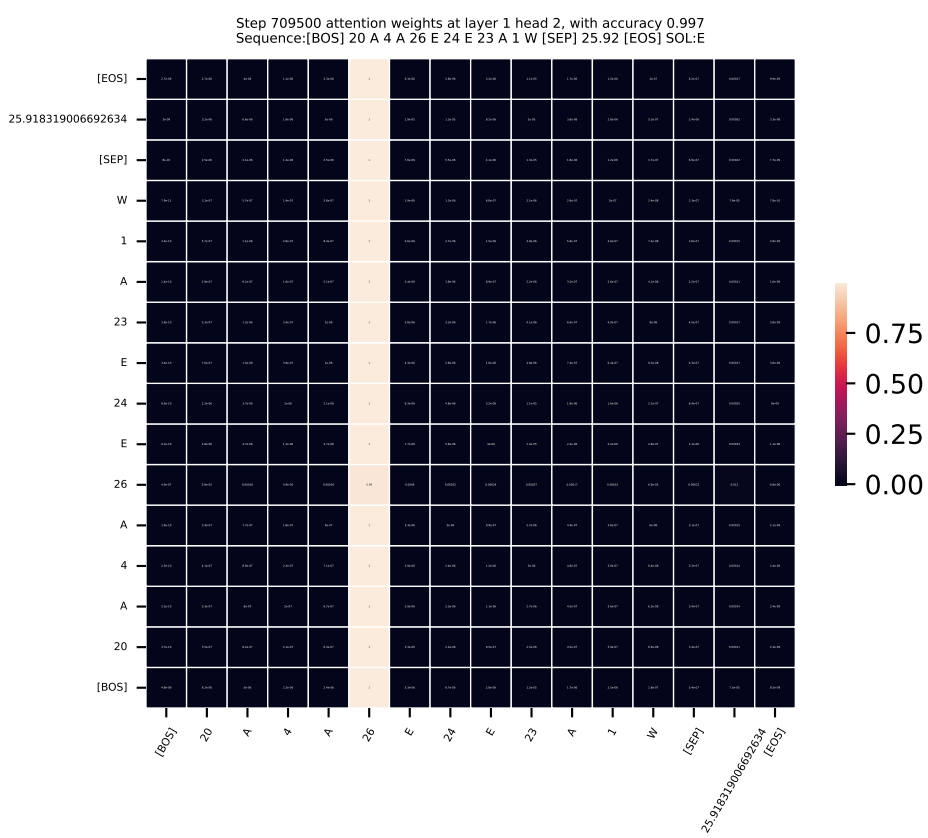

Step 709500 attention weights at layer 0 head 1, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

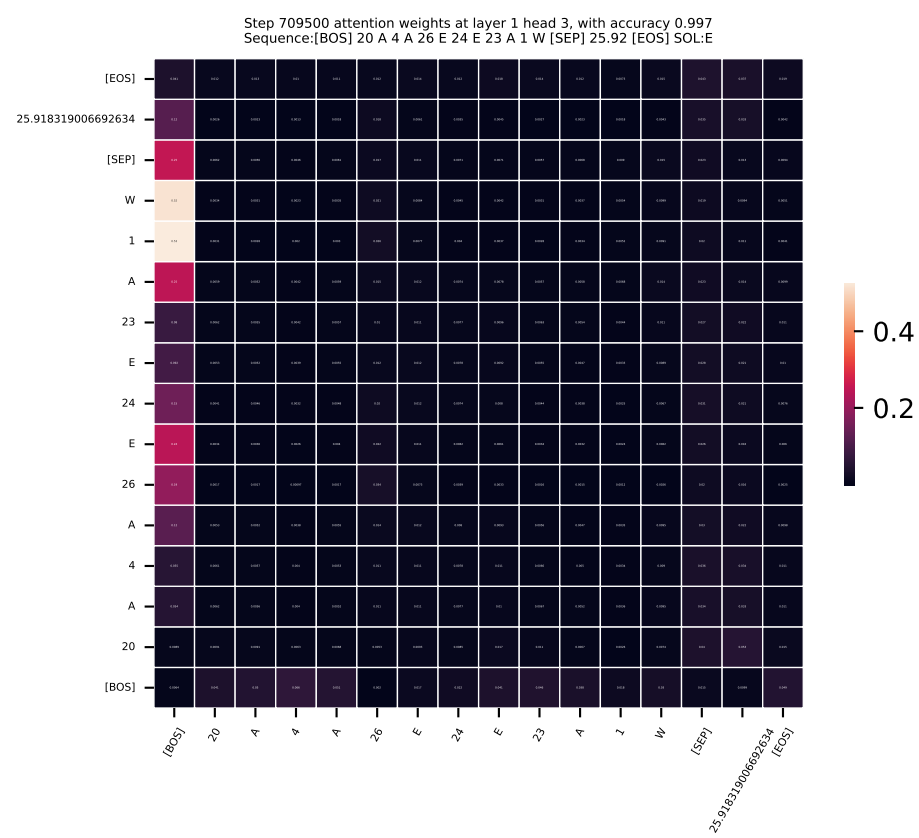

Step 709500 attention weights at layer 0 head 2, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

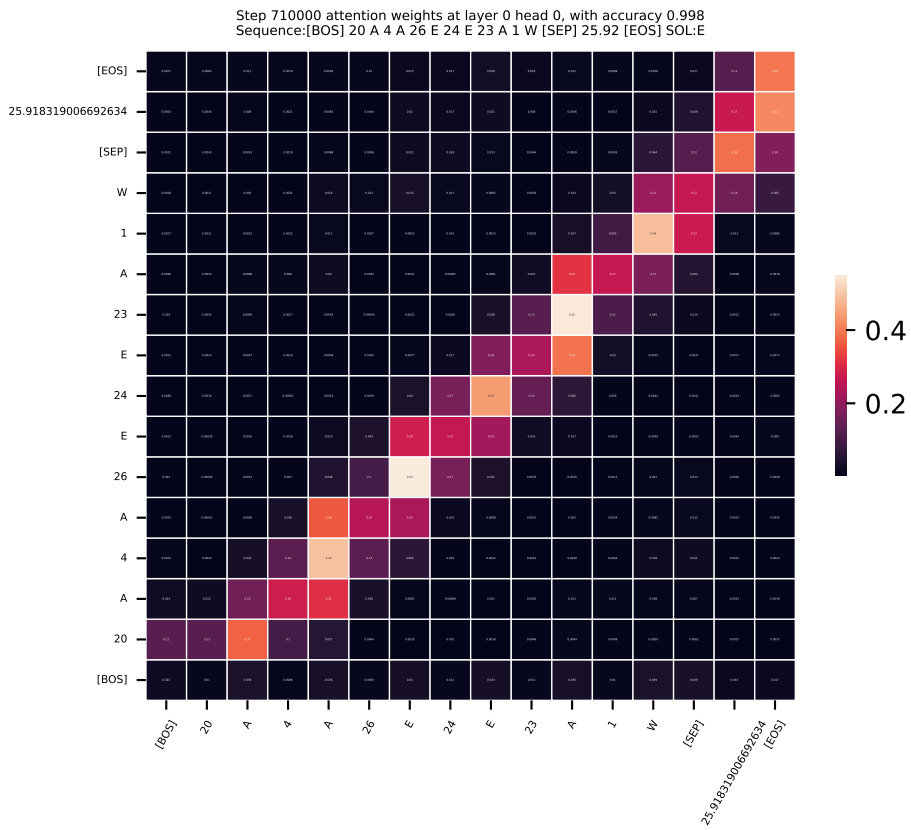

Step 709500 attention weights at layer 0 head 3, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

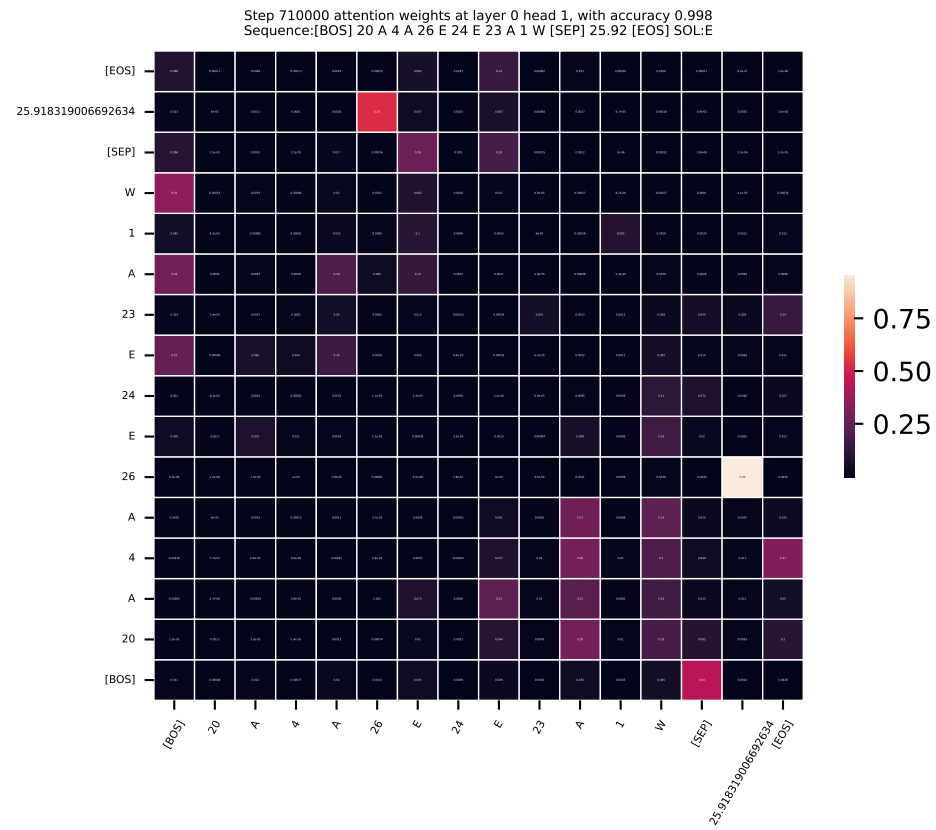

Step 709500 attention weights at layer 1 head 0, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

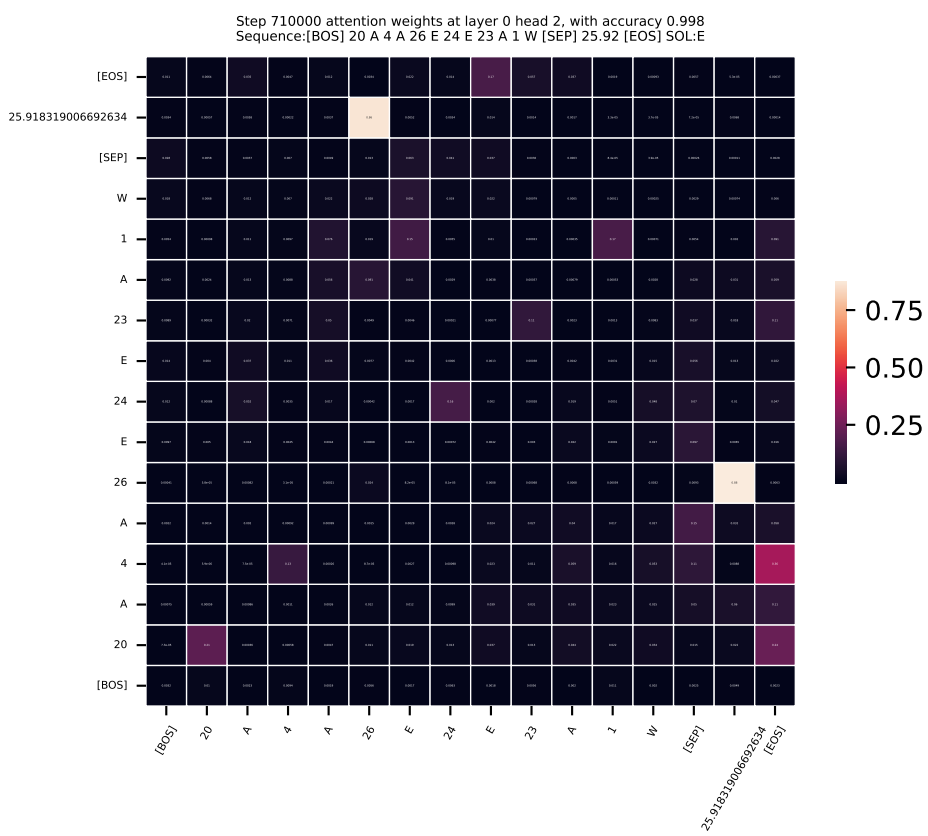

Step 709500 attention weights at layer 1 head 1, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

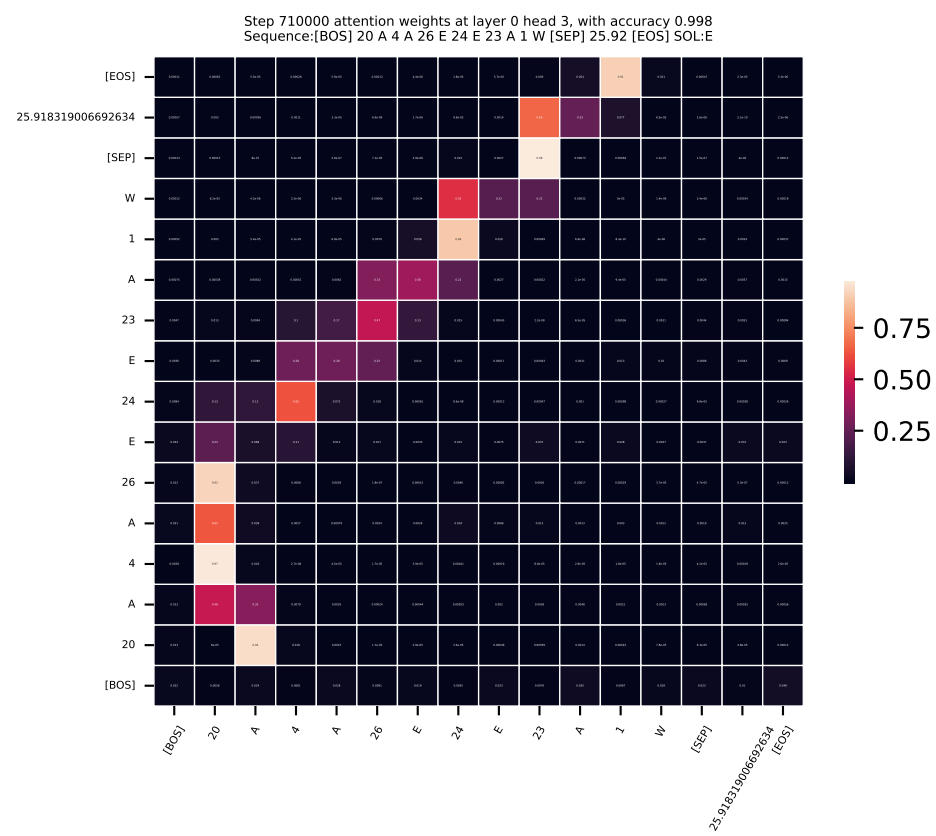

Step 709500 attention weights at layer 1 head 2, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

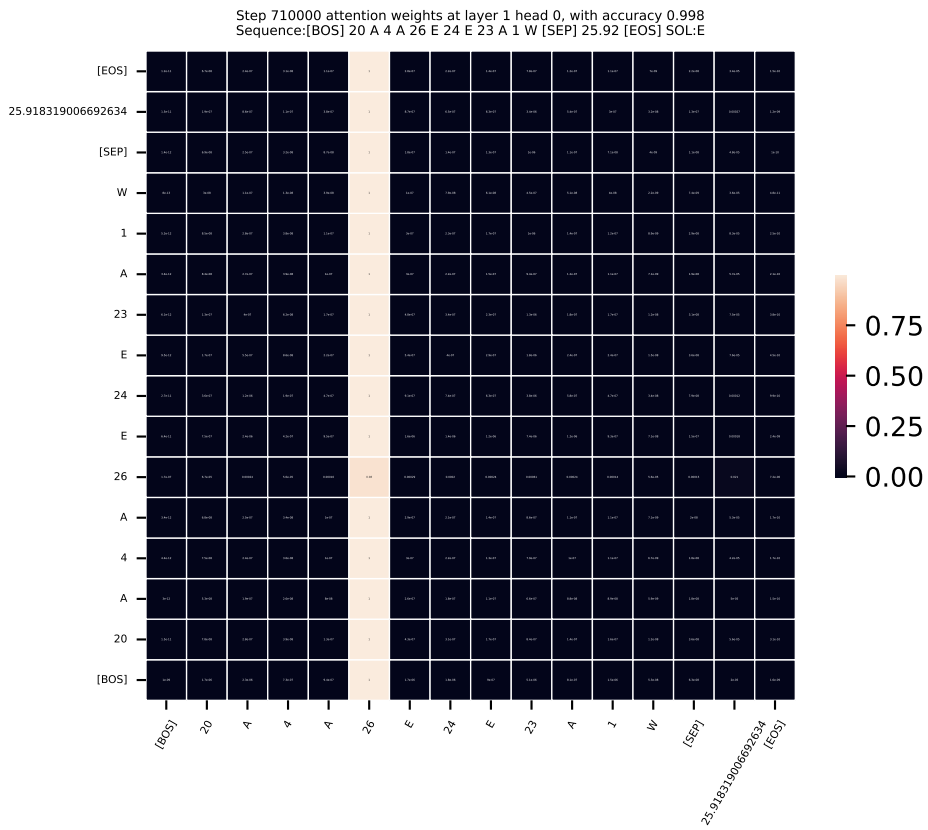

Step 709500 attention weights at layer 1 head 3, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

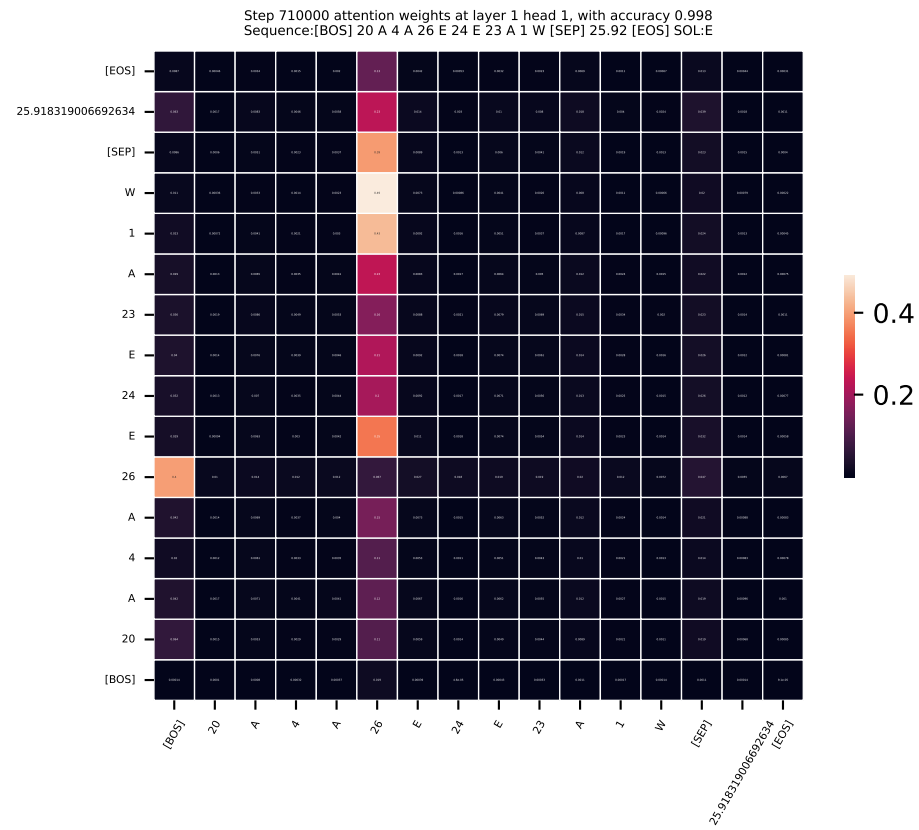

Step 710000 attention weights at layer 0 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

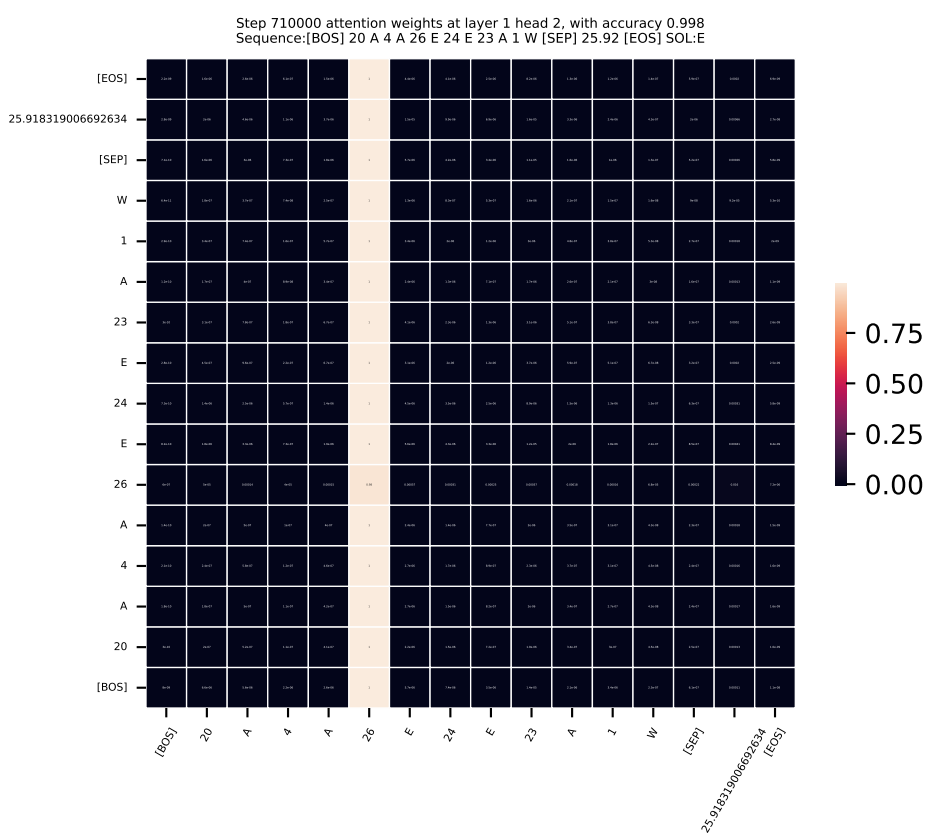

Step 710000 attention weights at layer 0 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

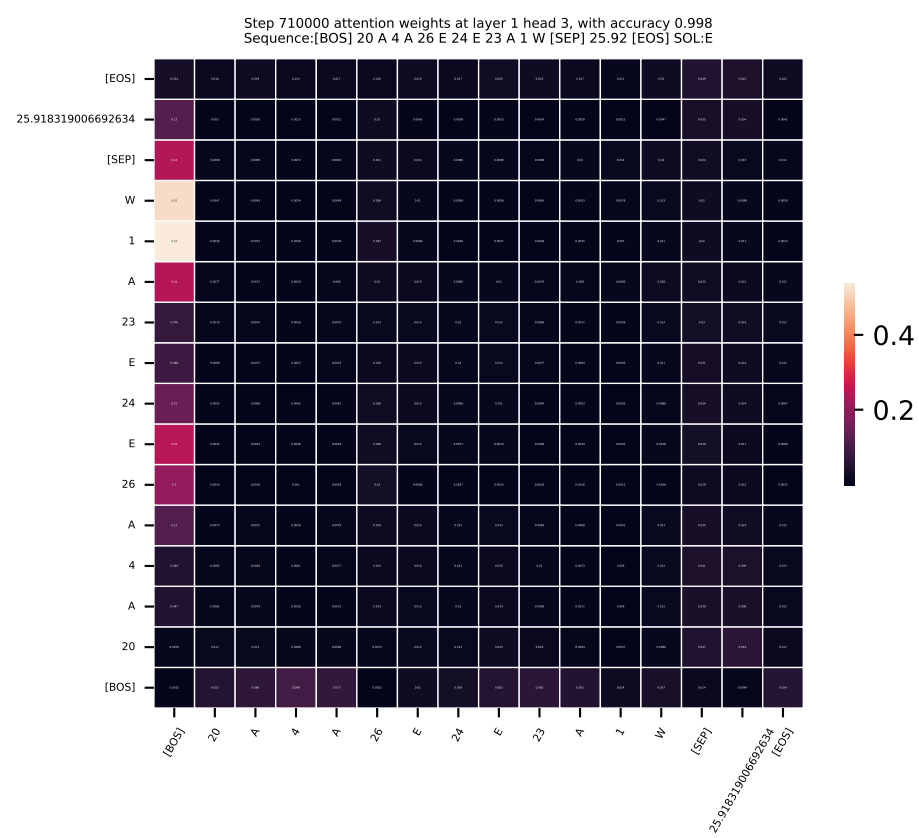

Step 710000 attention weights at layer 0 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

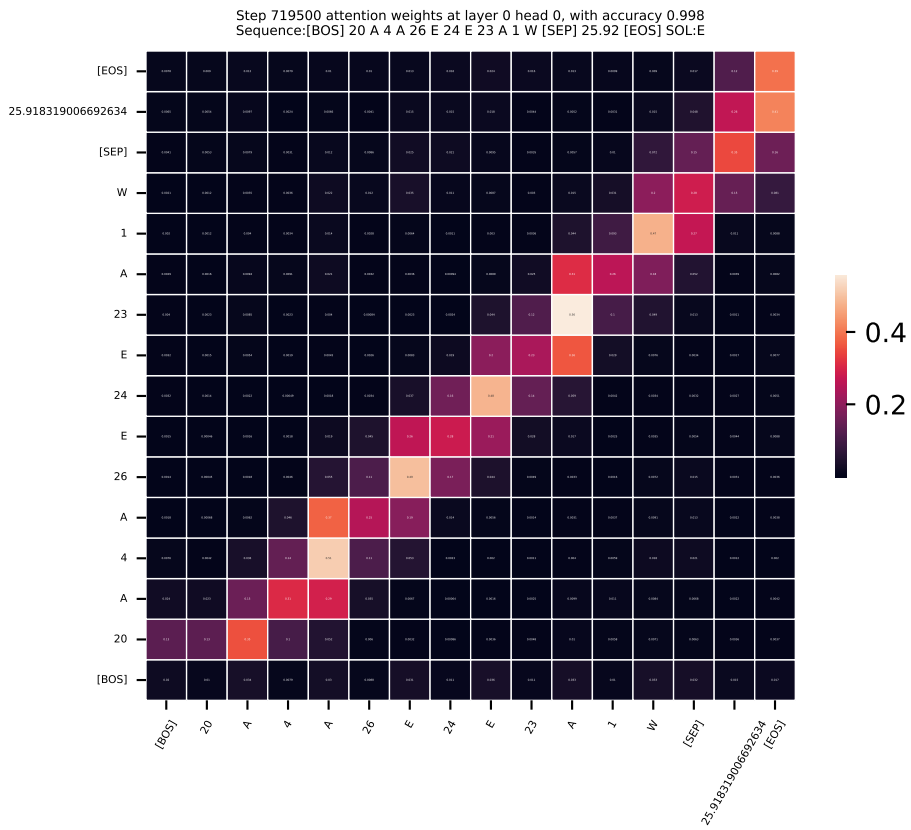

Step 710000 attention weights at layer 0 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

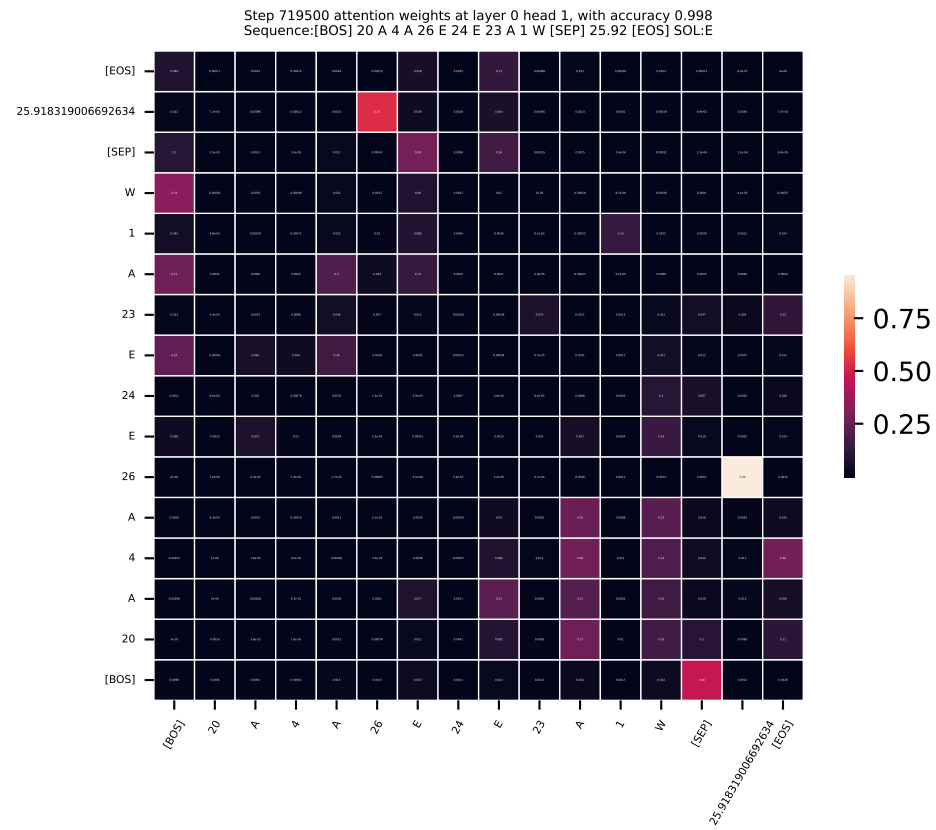

Step 710000 attention weights at layer 1 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

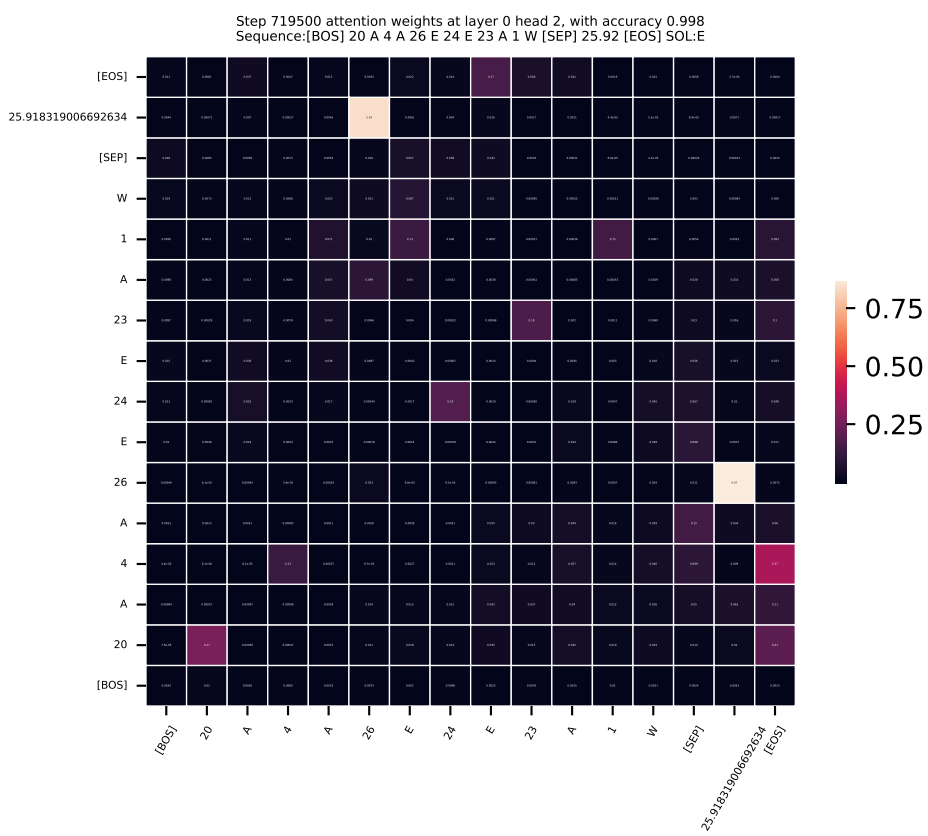

Step 710000 attention weights at layer 1 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

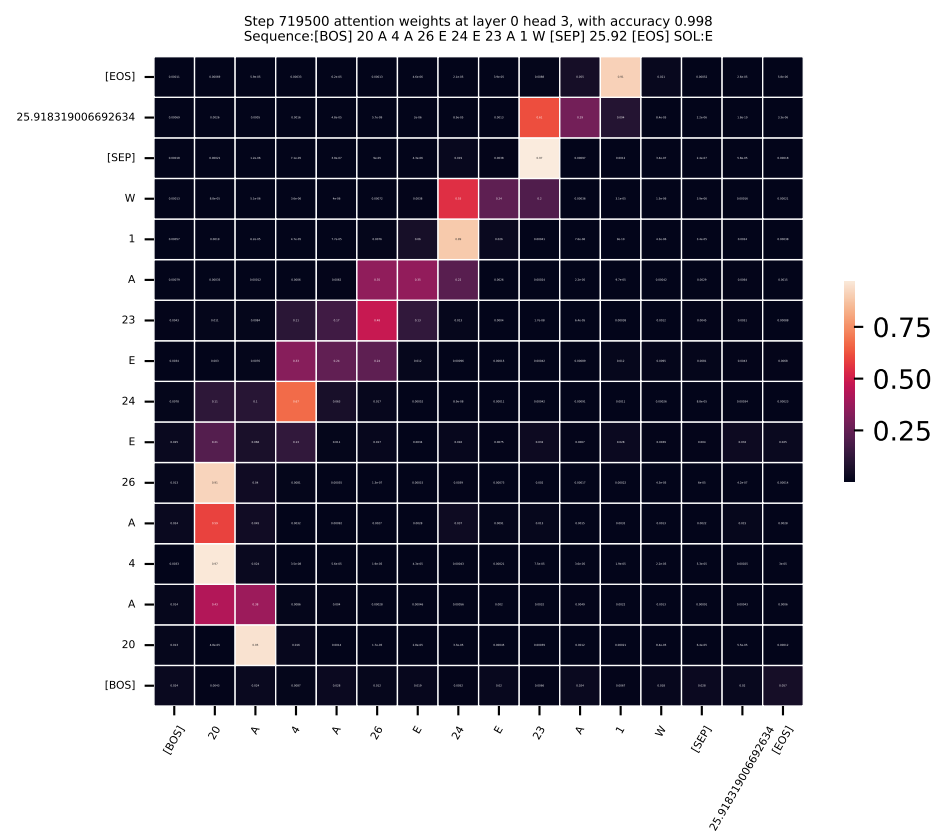

Step 710000 attention weights at layer 1 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

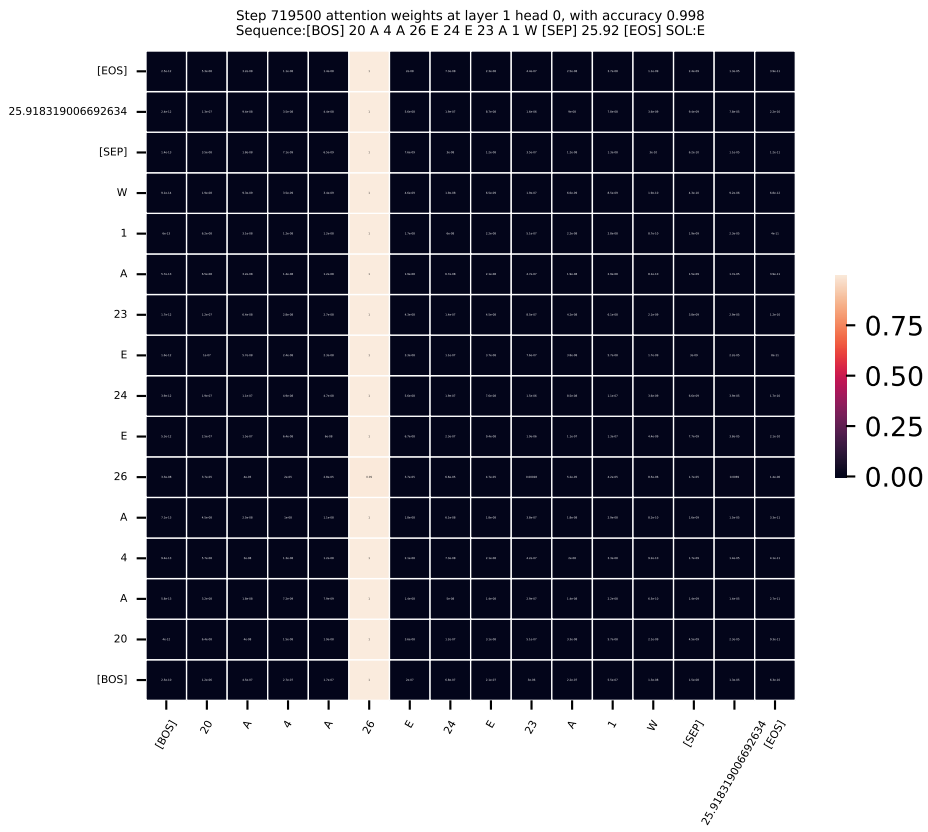

Step 710000 attention weights at layer 1 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

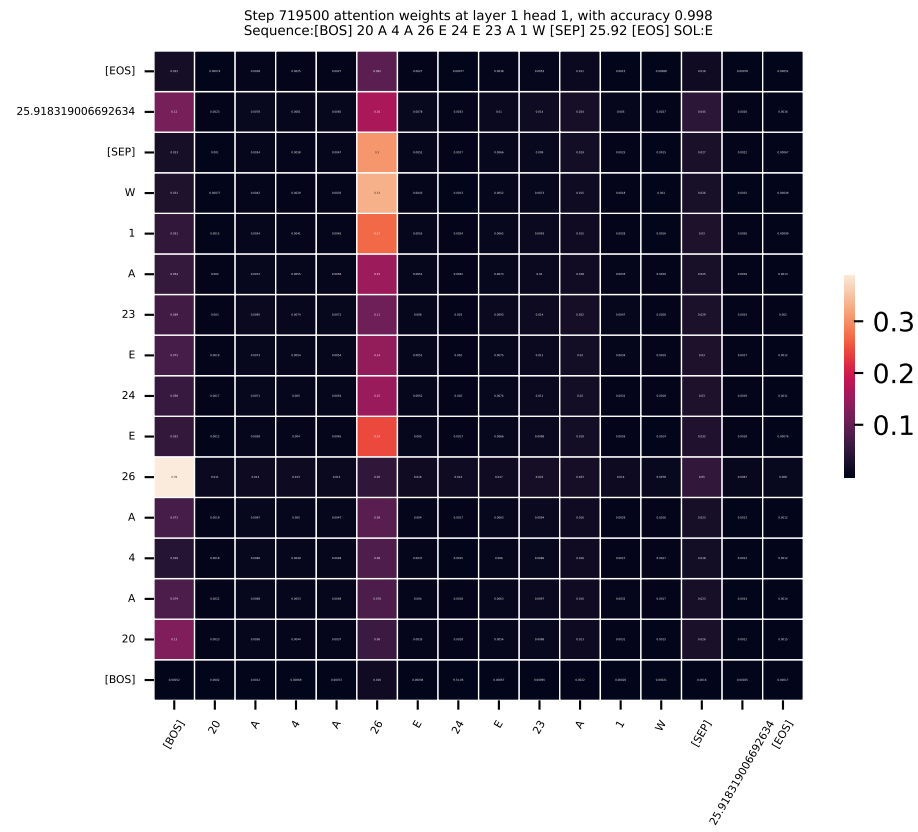

Step 719500 attention weights at layer 0 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

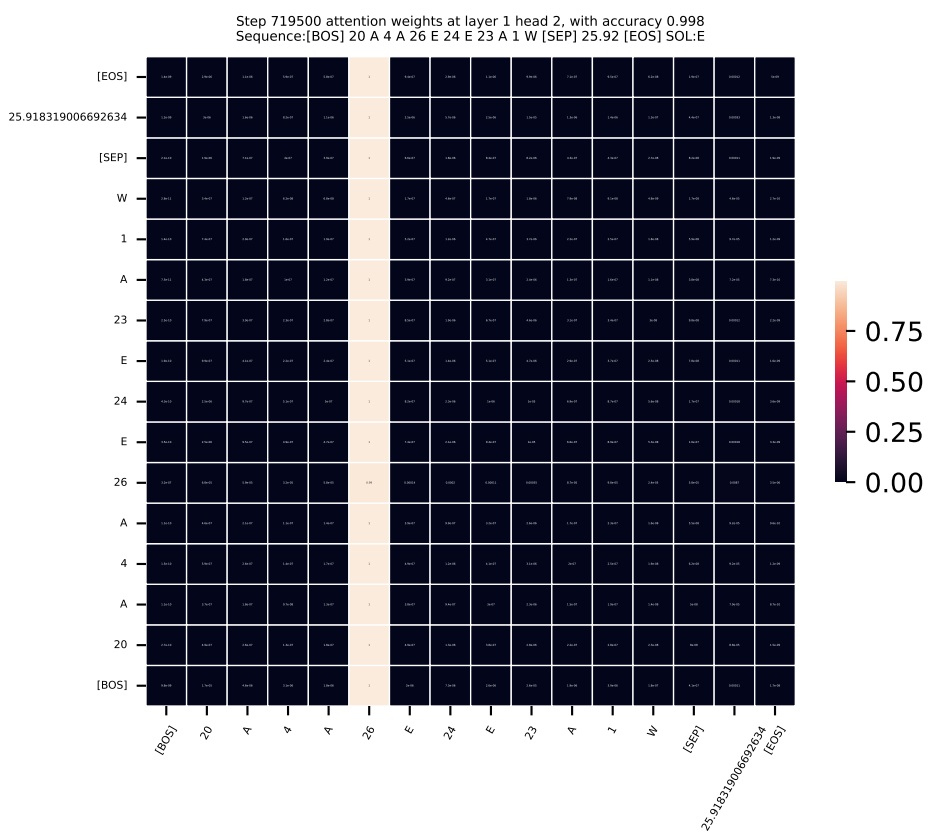

Step 719500 attention weights at layer 0 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

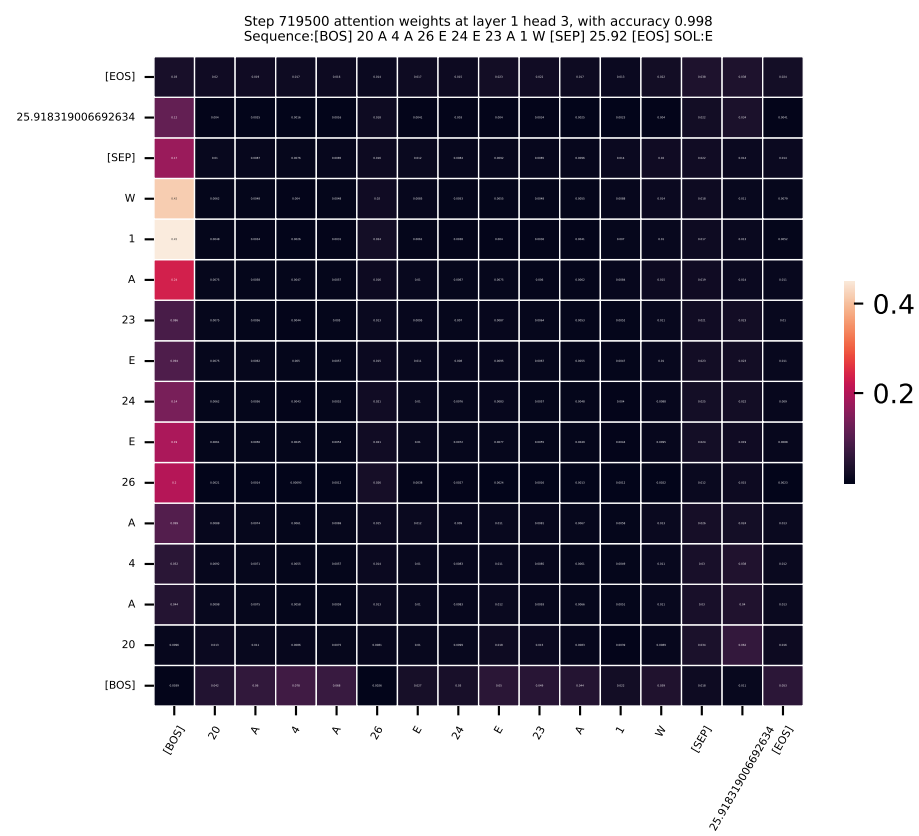

Step 719500 attention weights at layer 0 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

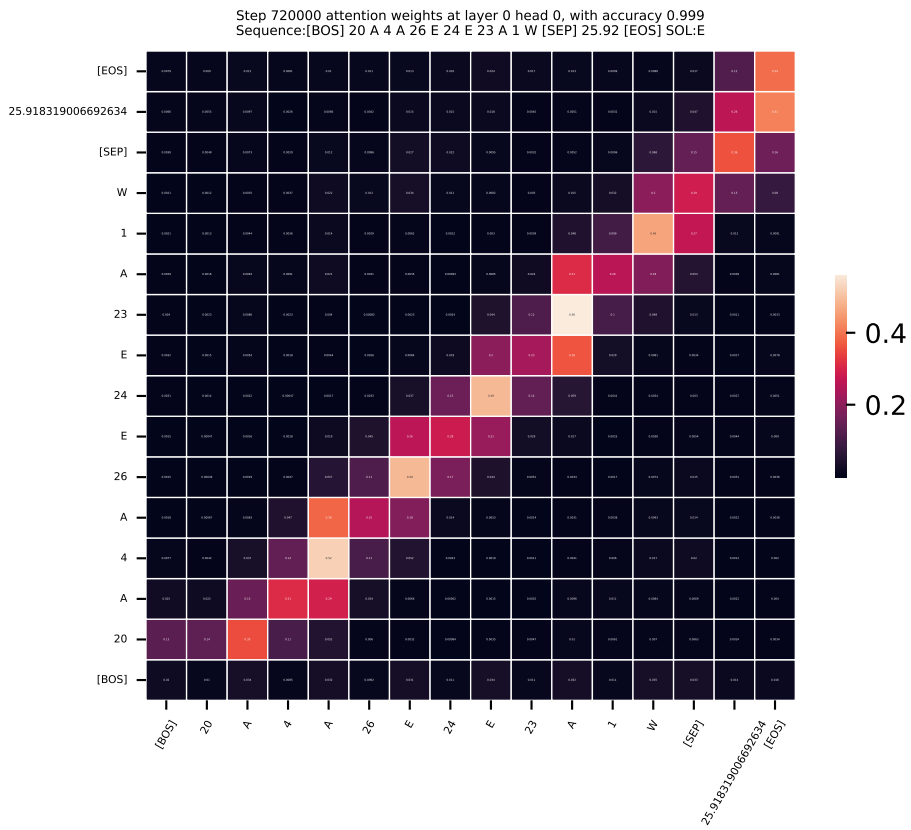

Step 719500 attention weights at layer 0 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

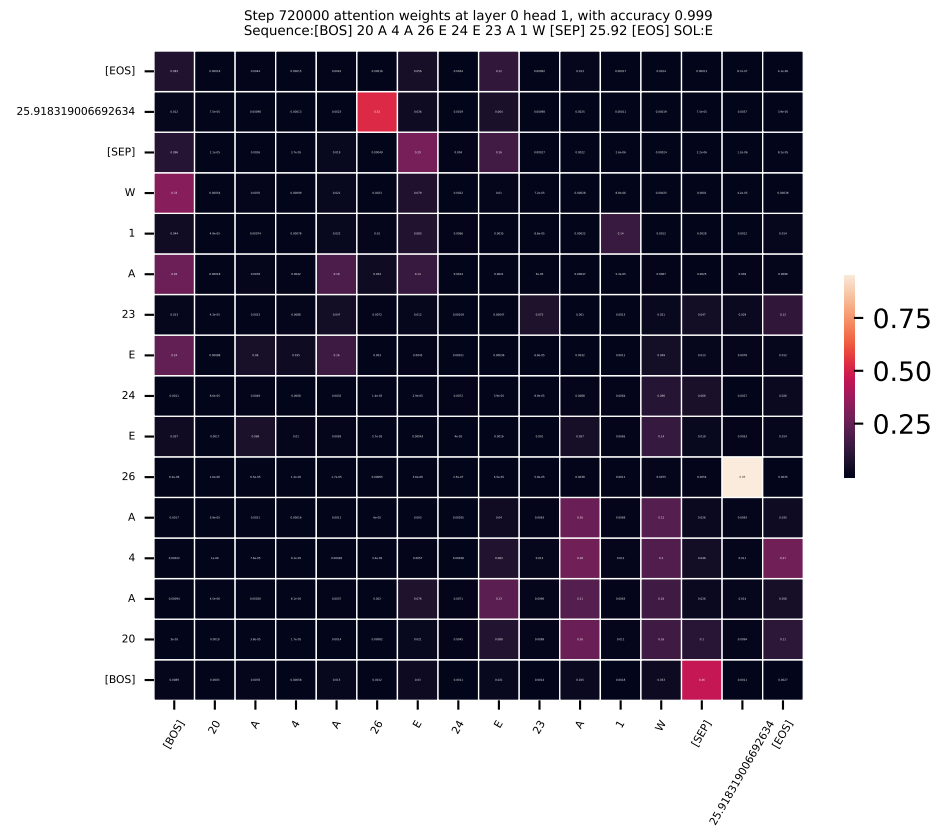

Step 719500 attention weights at layer 1 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

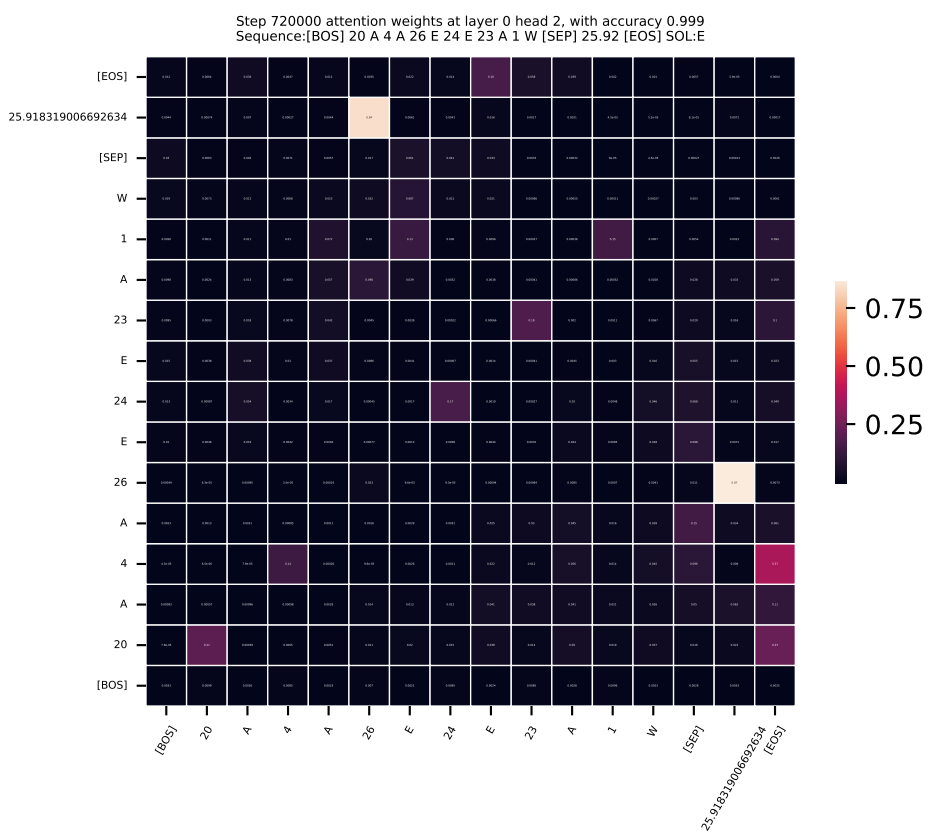

Step 719500 attention weights at layer 1 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

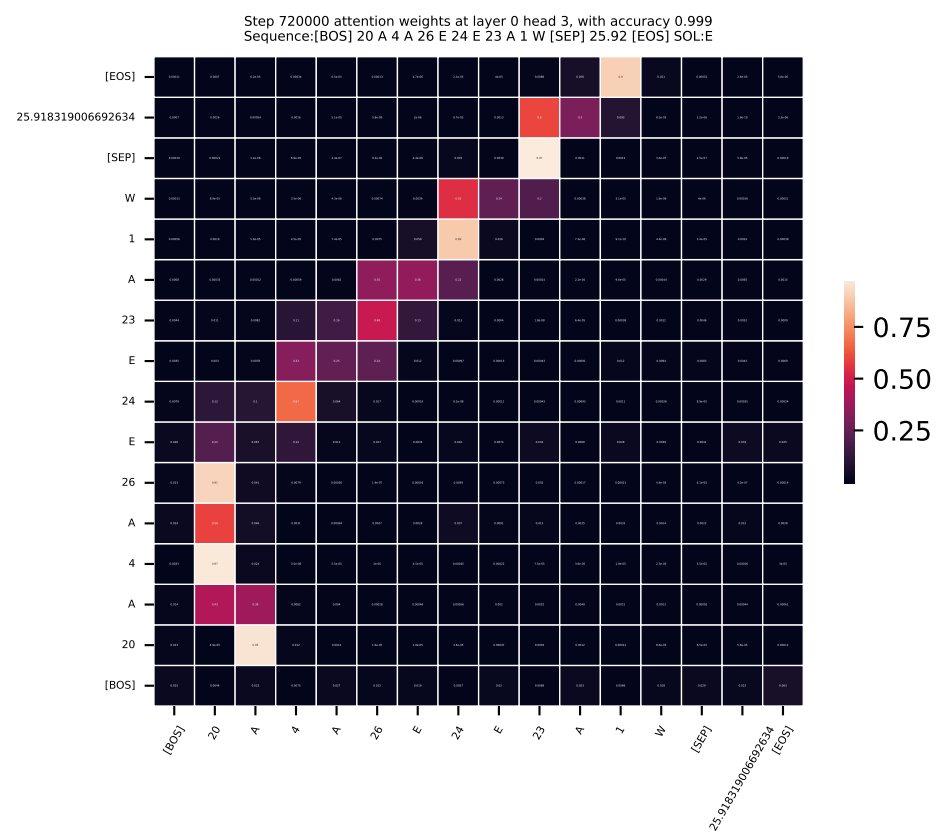

Step 719500 attention weights at layer 1 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

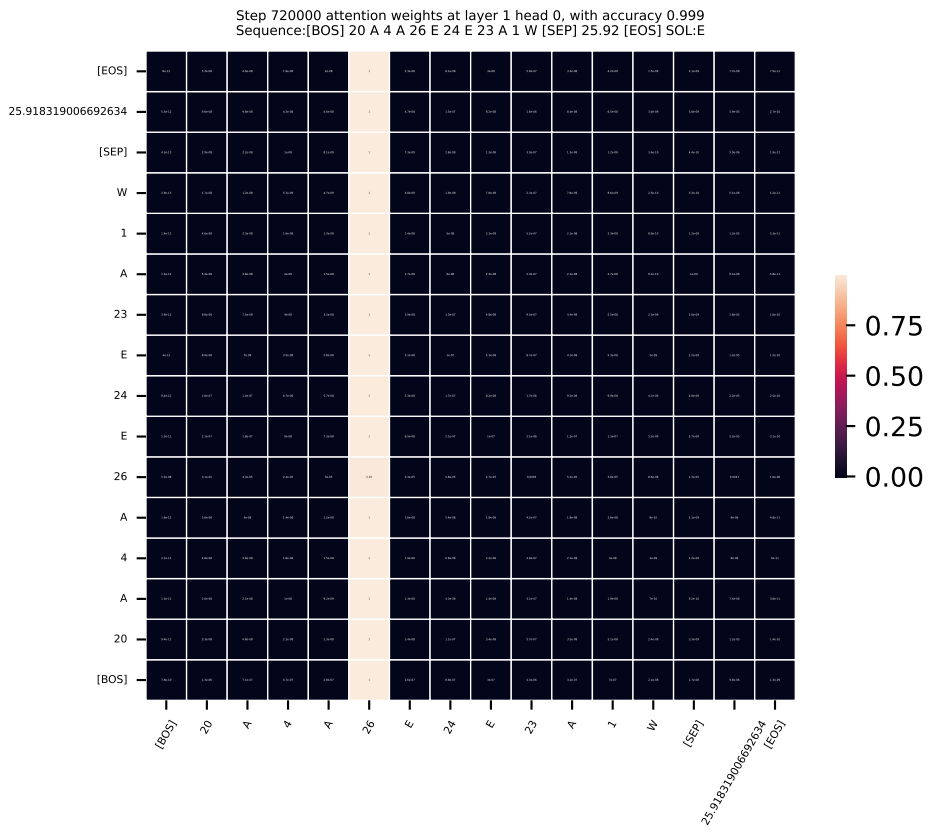

Step 719500 attention weights at layer 1 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

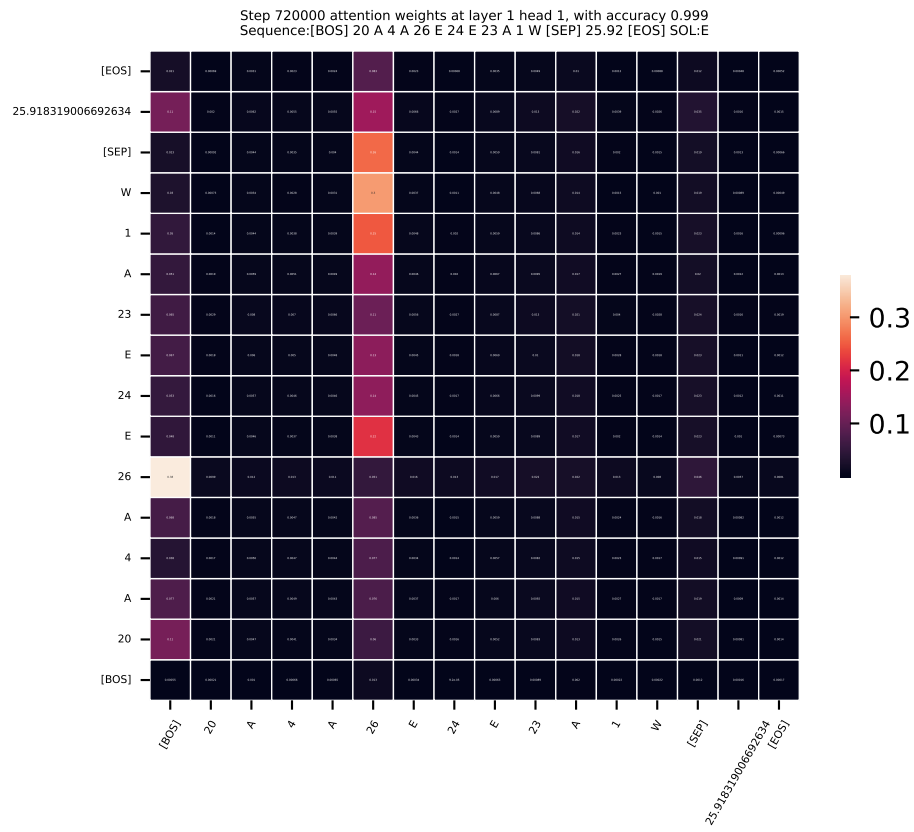

Step 720000 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

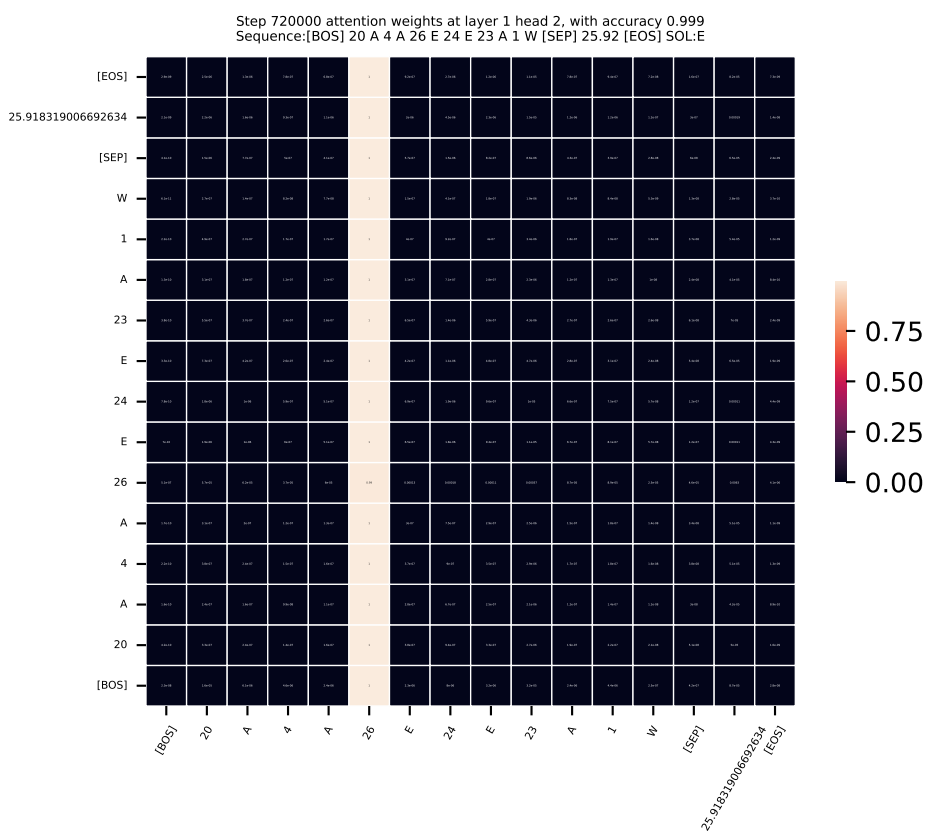

Step 720000 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

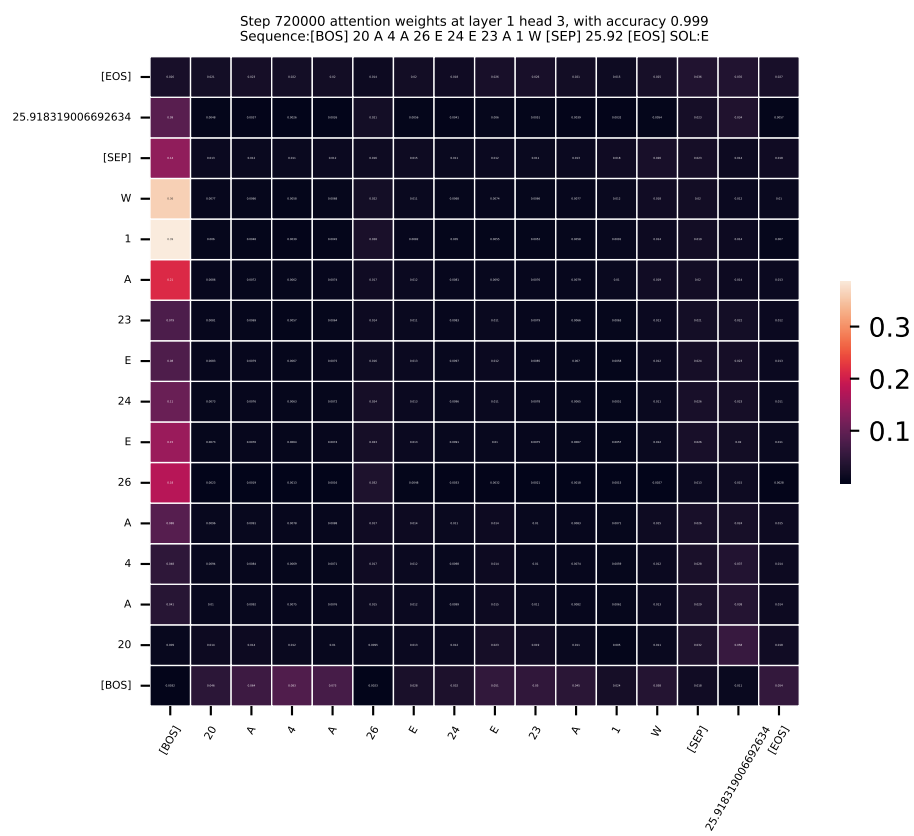

Step 720000 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

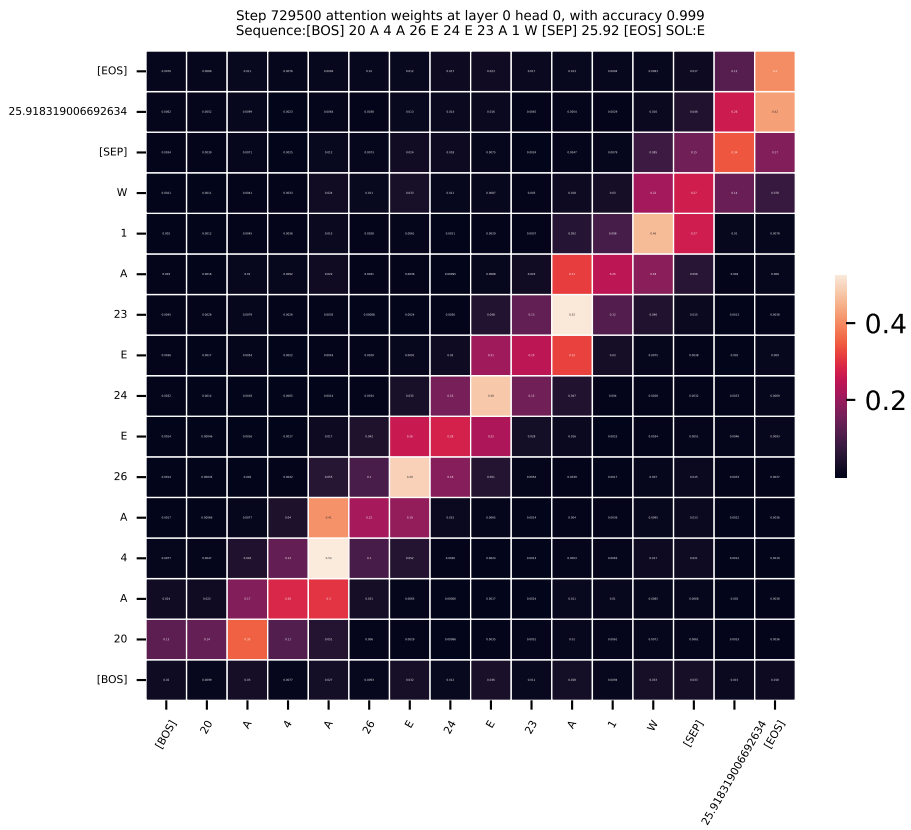

Step 720000 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

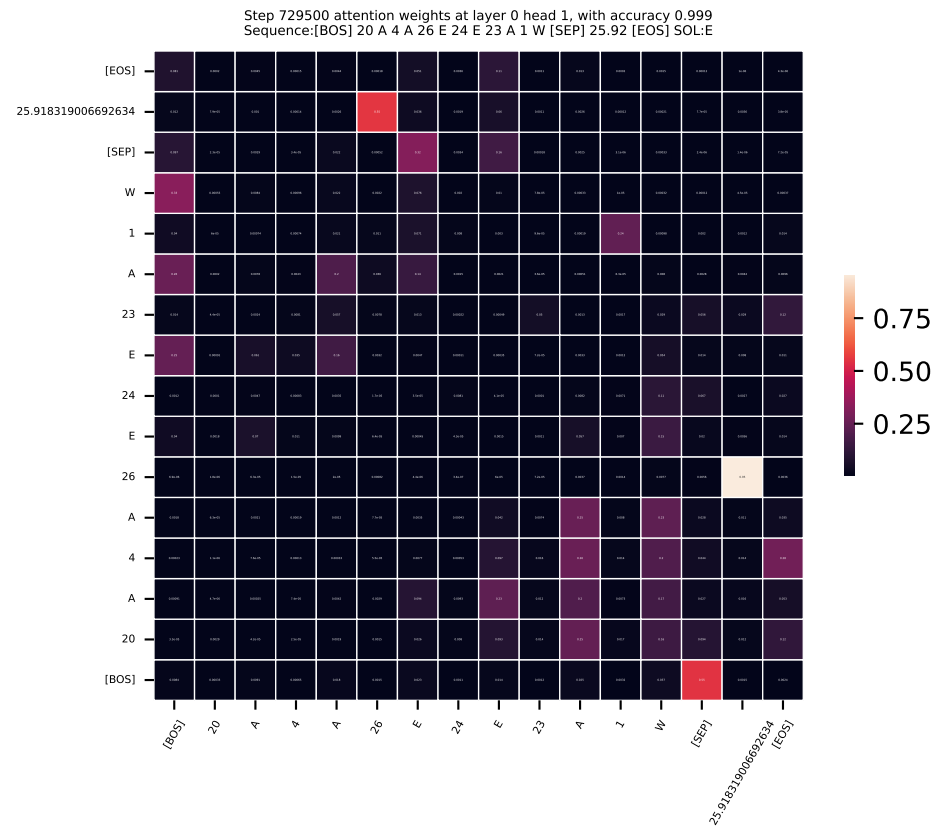

Step 720000 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

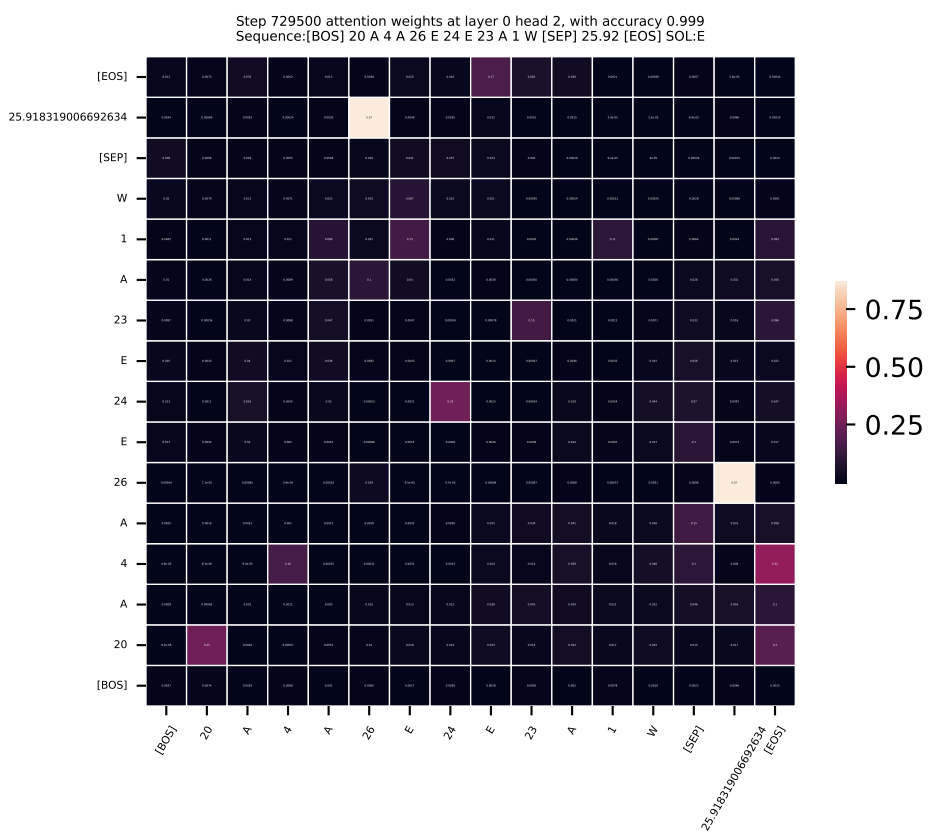

Step 720000 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

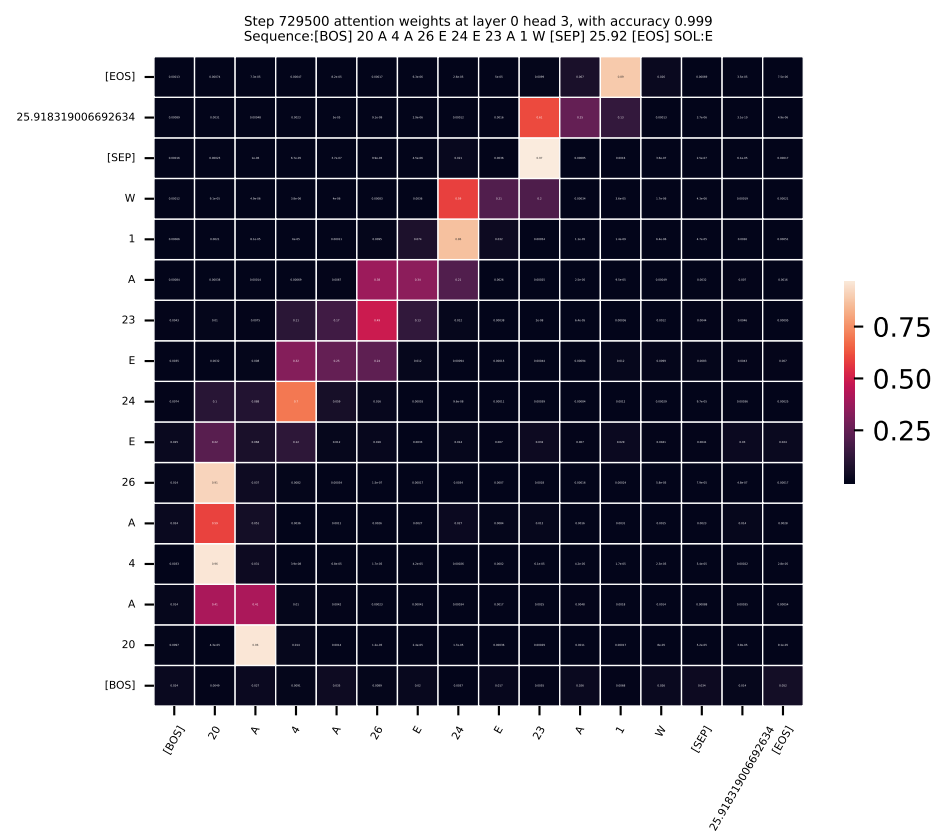

Step 720000 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

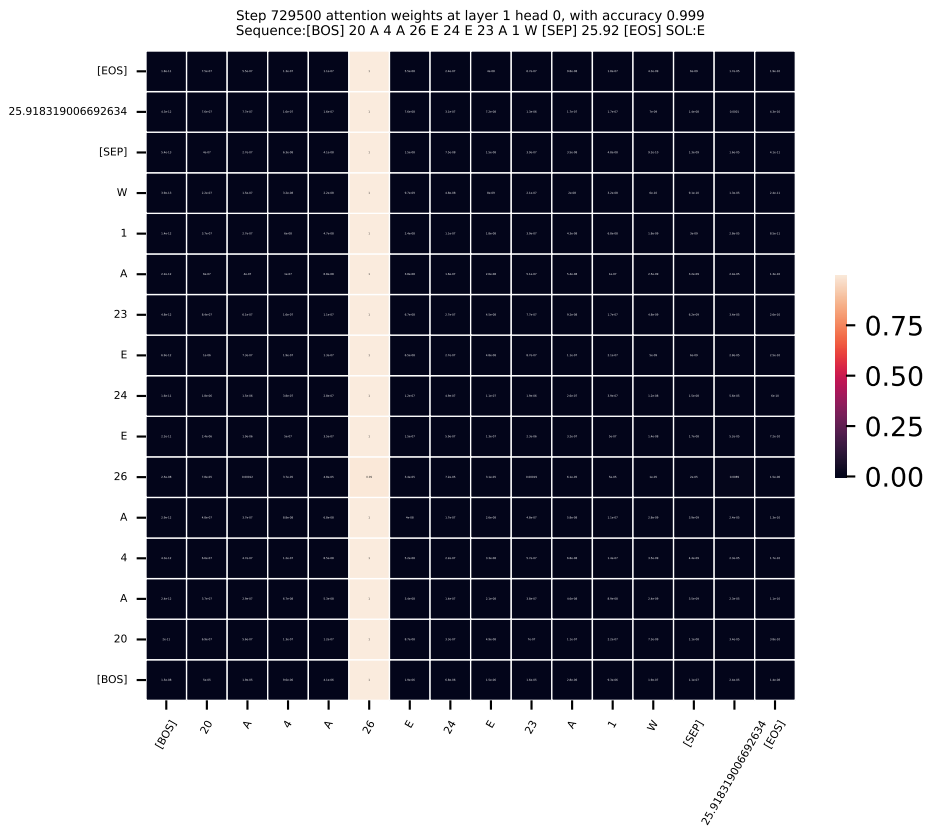

Step 720000 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

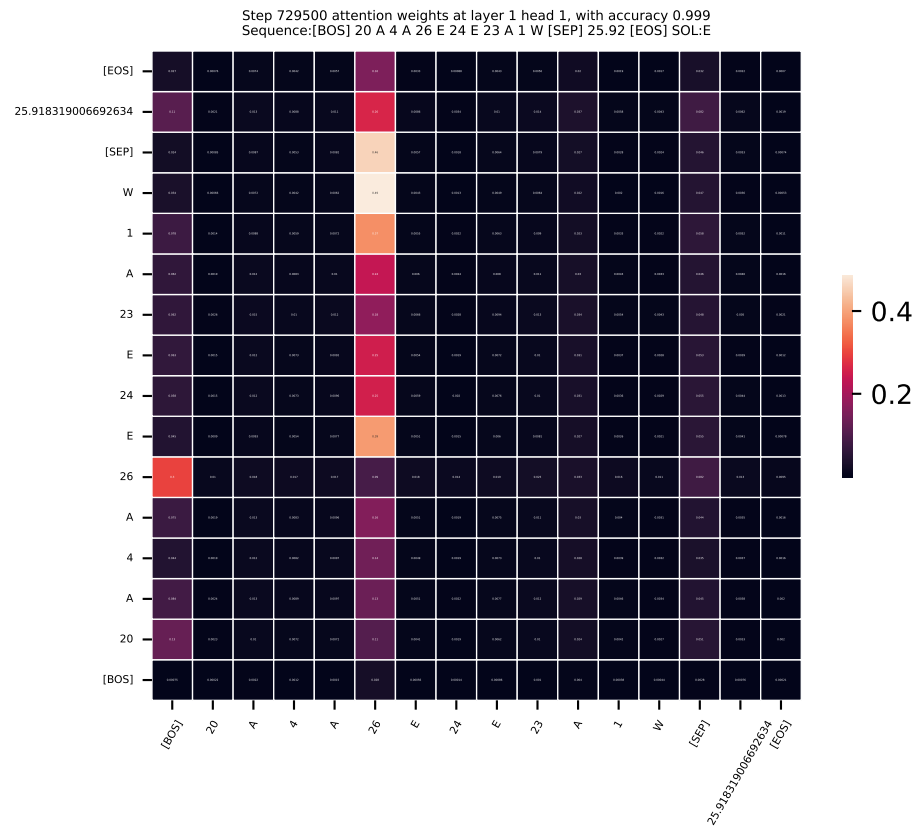

Step 729500 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

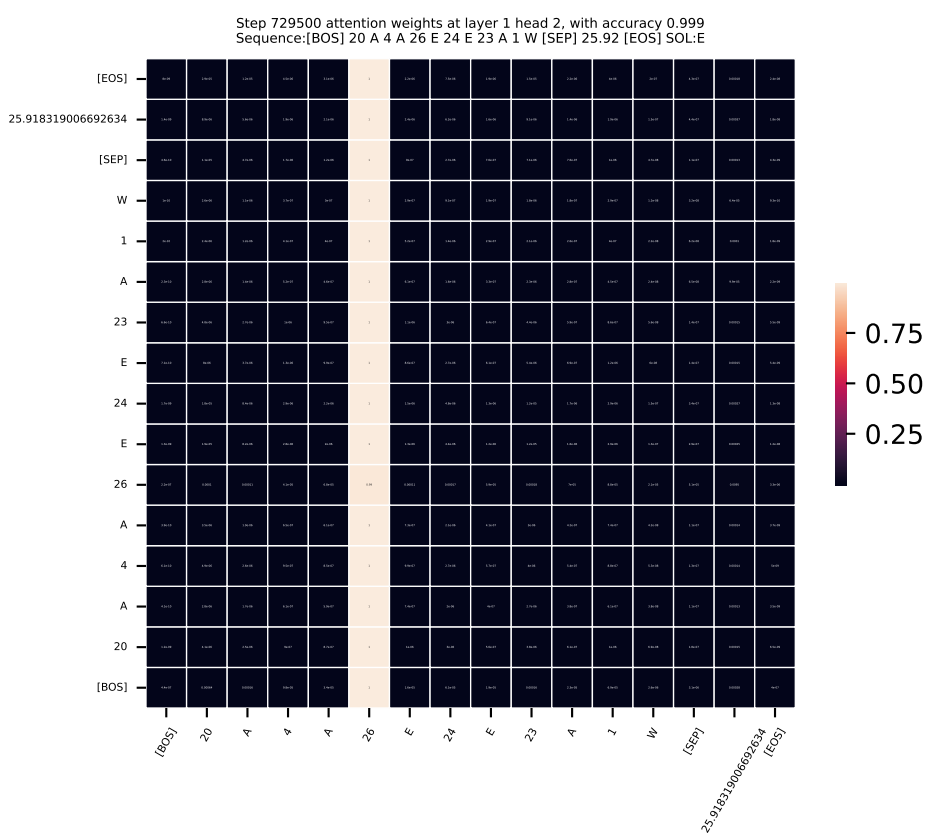

Step 729500 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

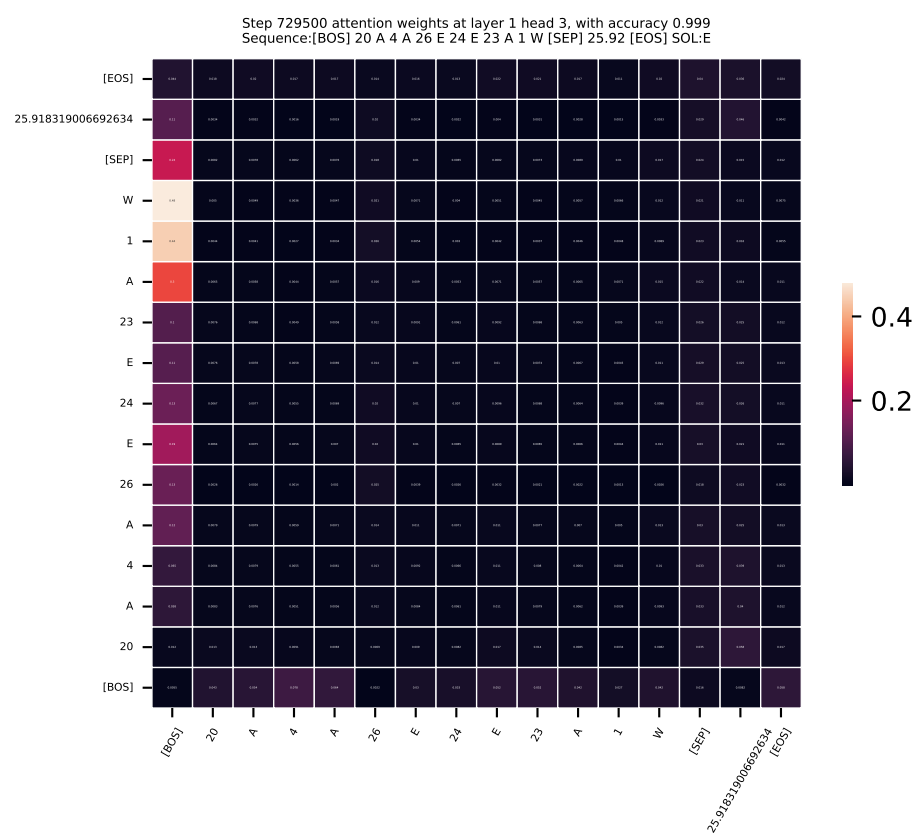

Step 729500 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

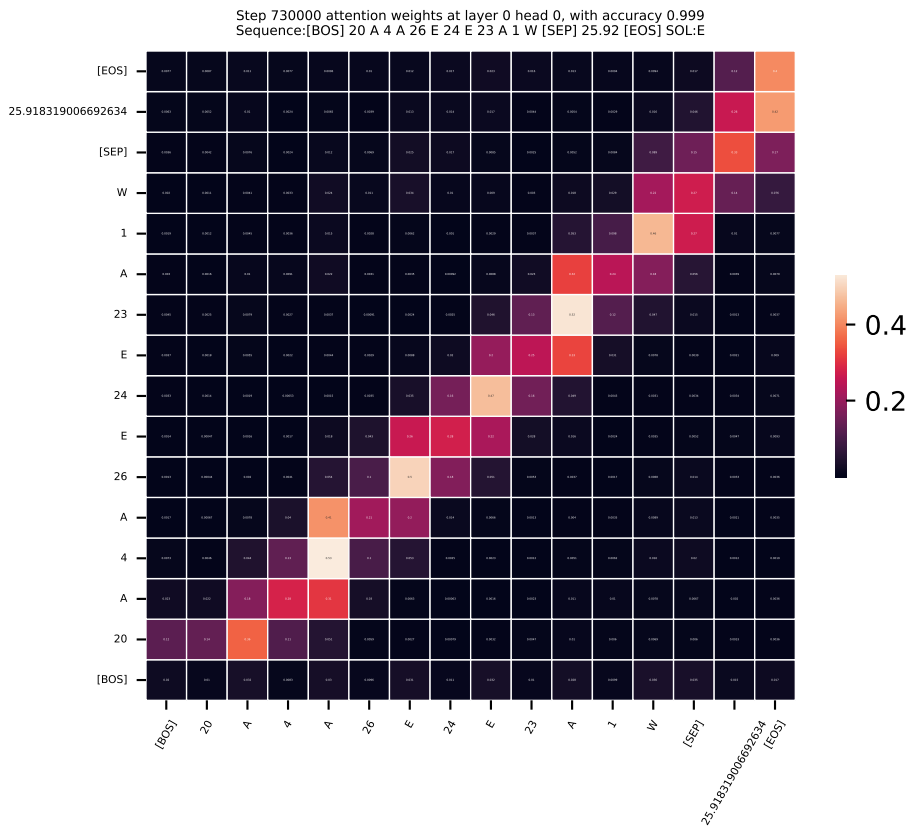

Step 729500 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

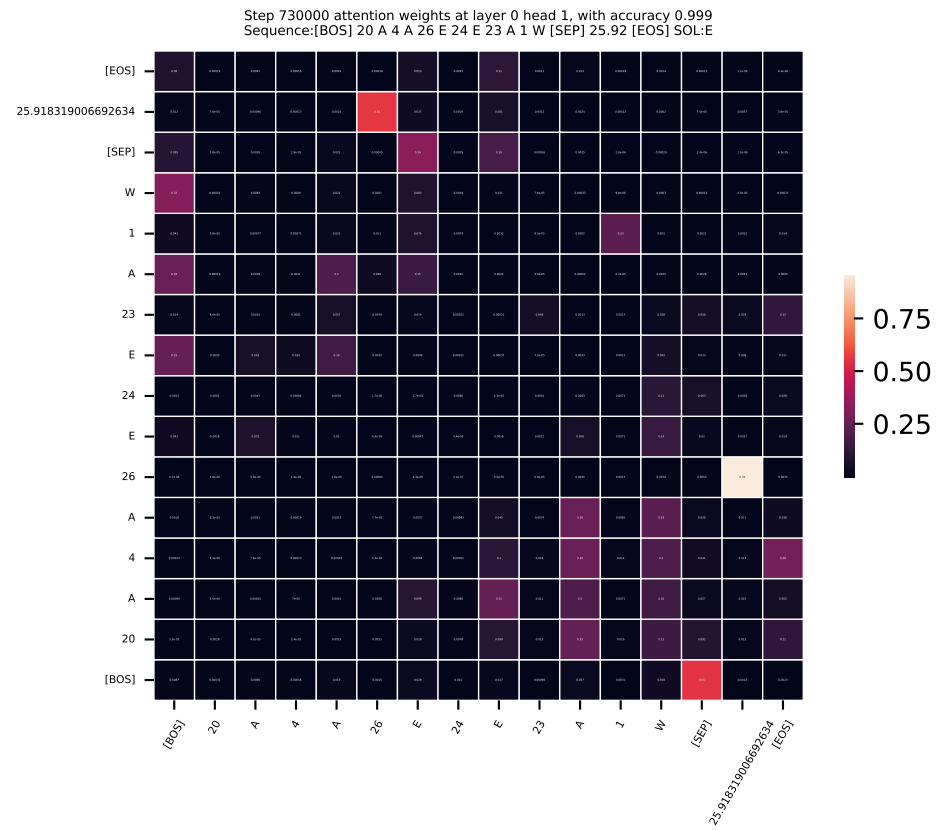

Step 729500 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

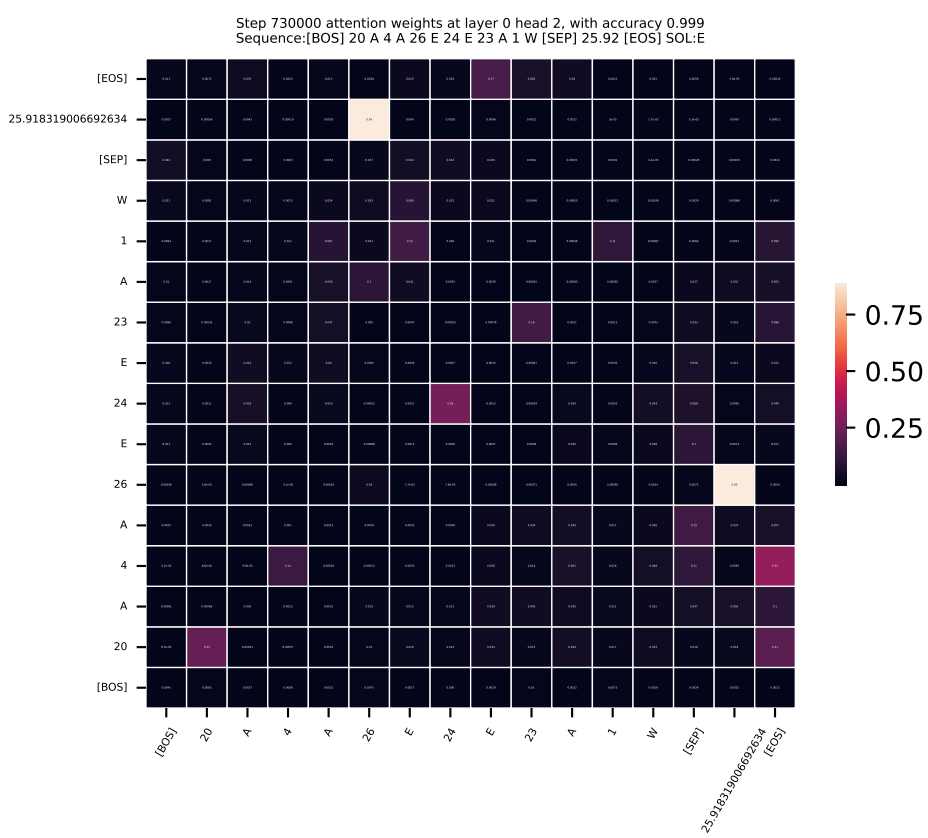

Step 729500 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

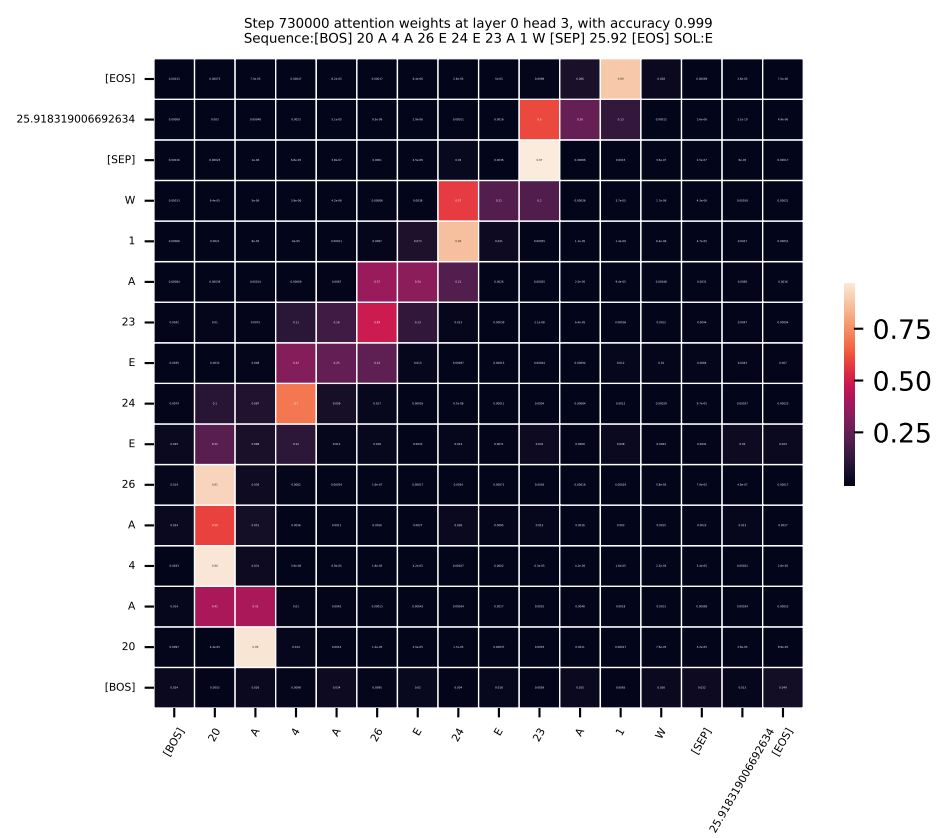

Step 729500 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

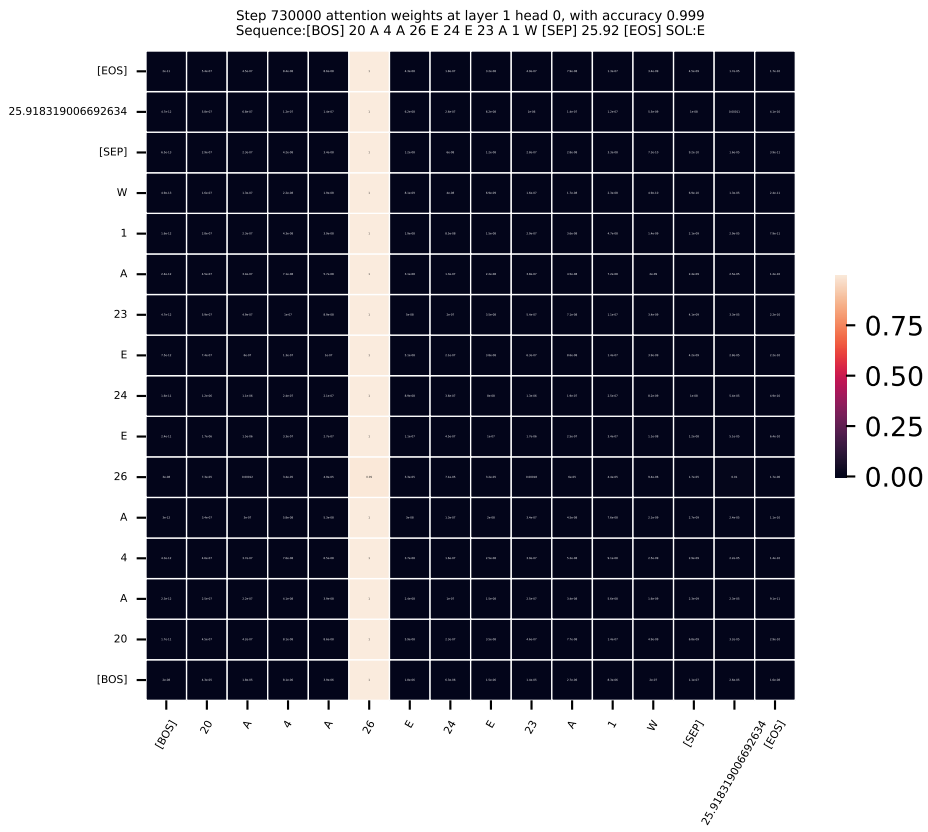

Step 729500 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

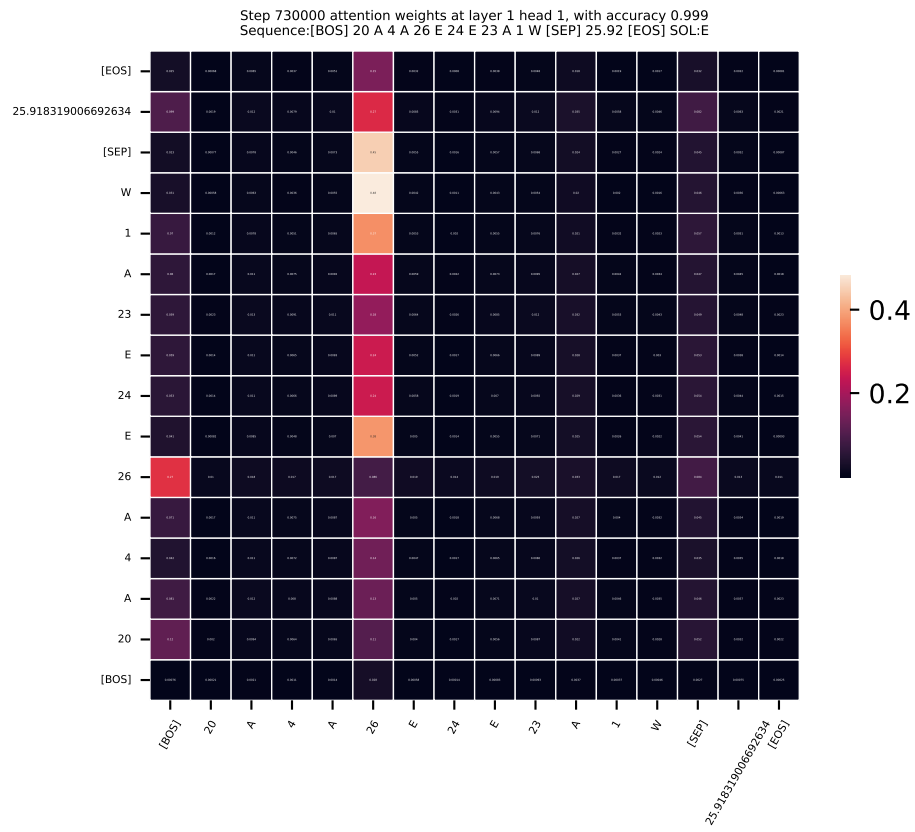

Step 730000 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

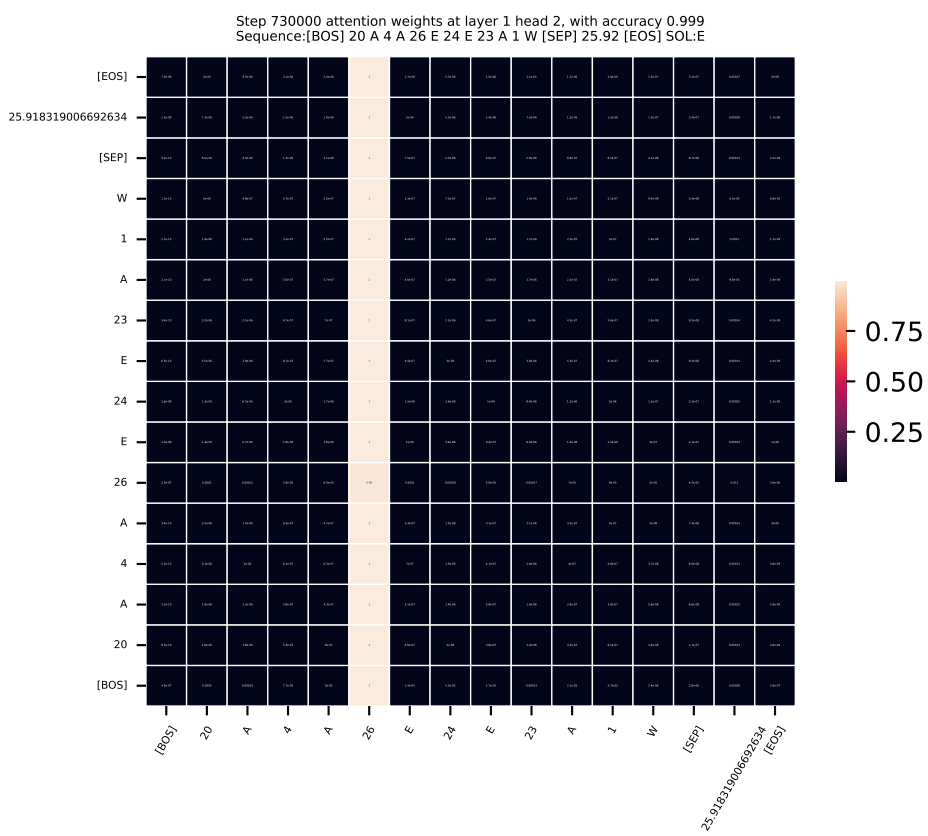

Step 730000 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

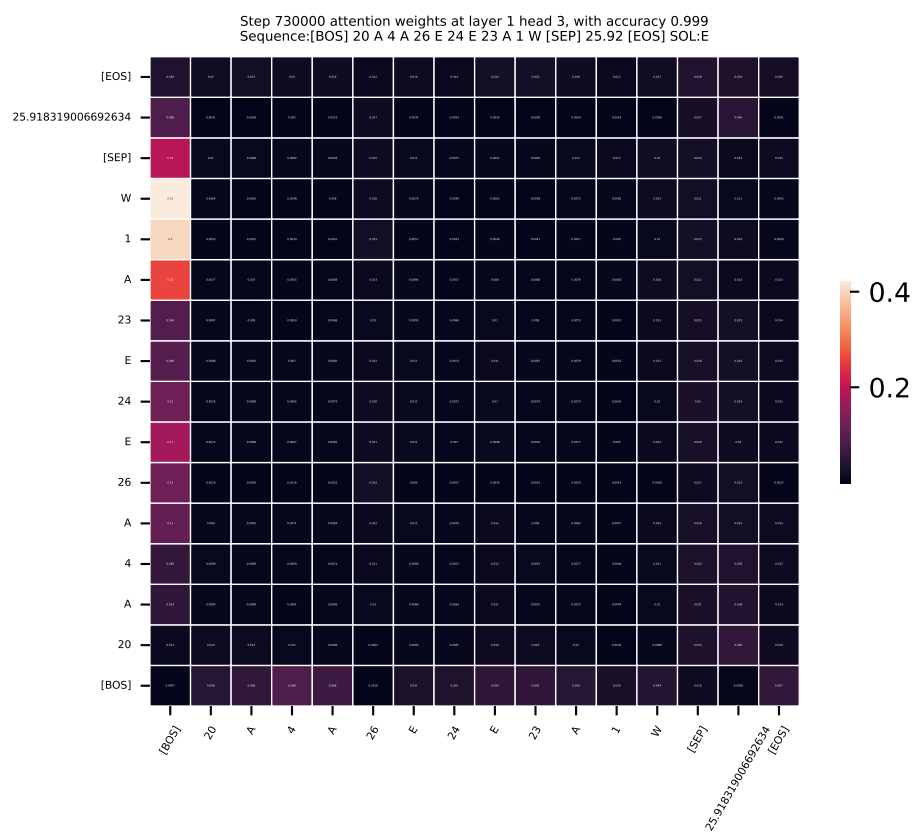

Step 730000 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

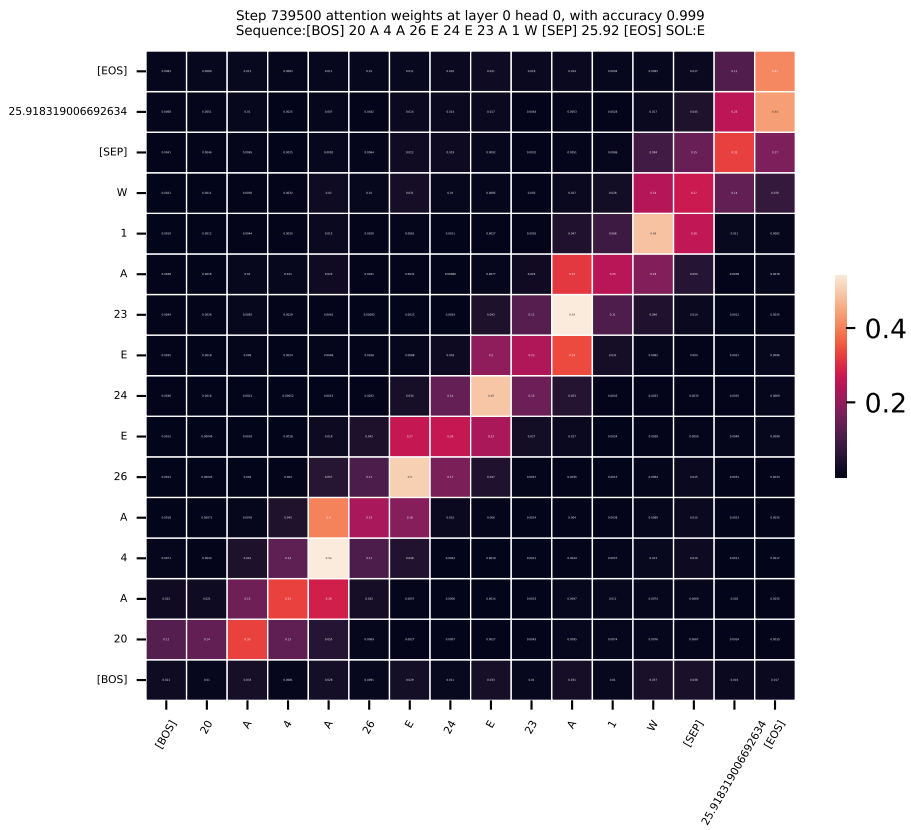

Step 730000 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

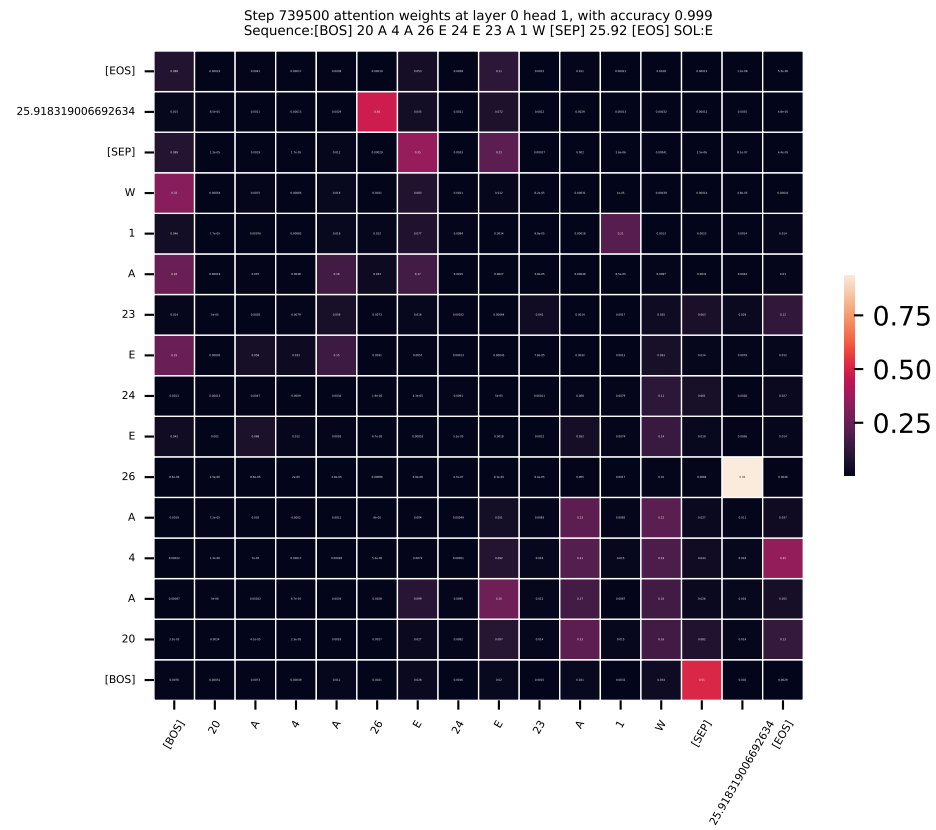

Step 730000 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

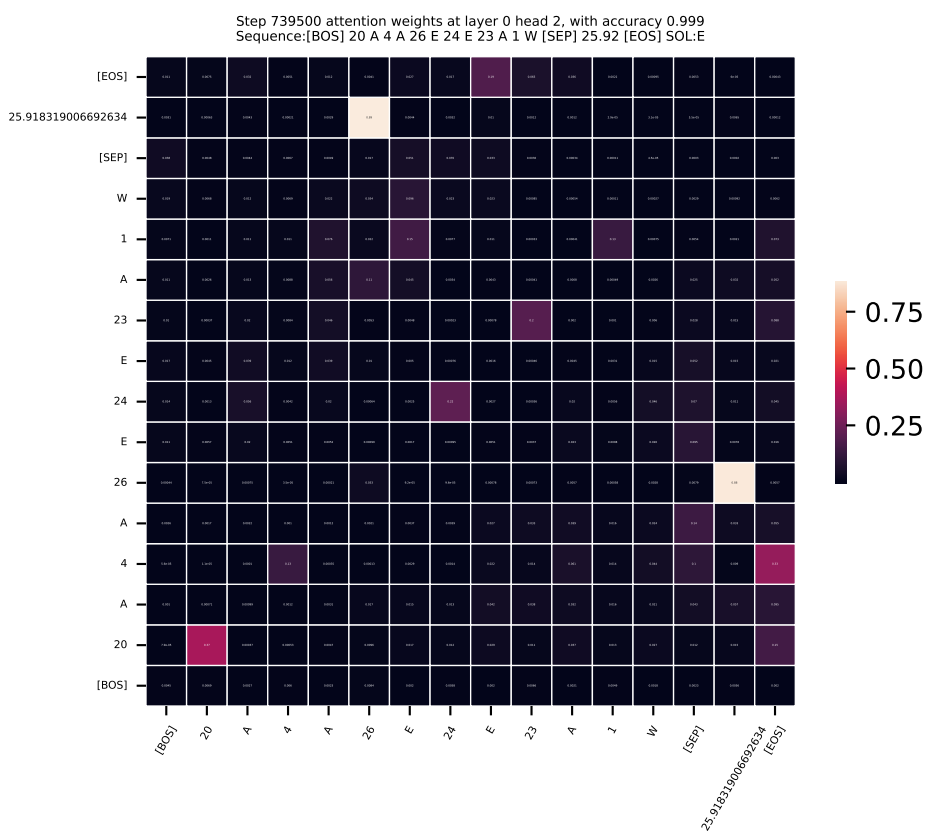

Step 730000 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

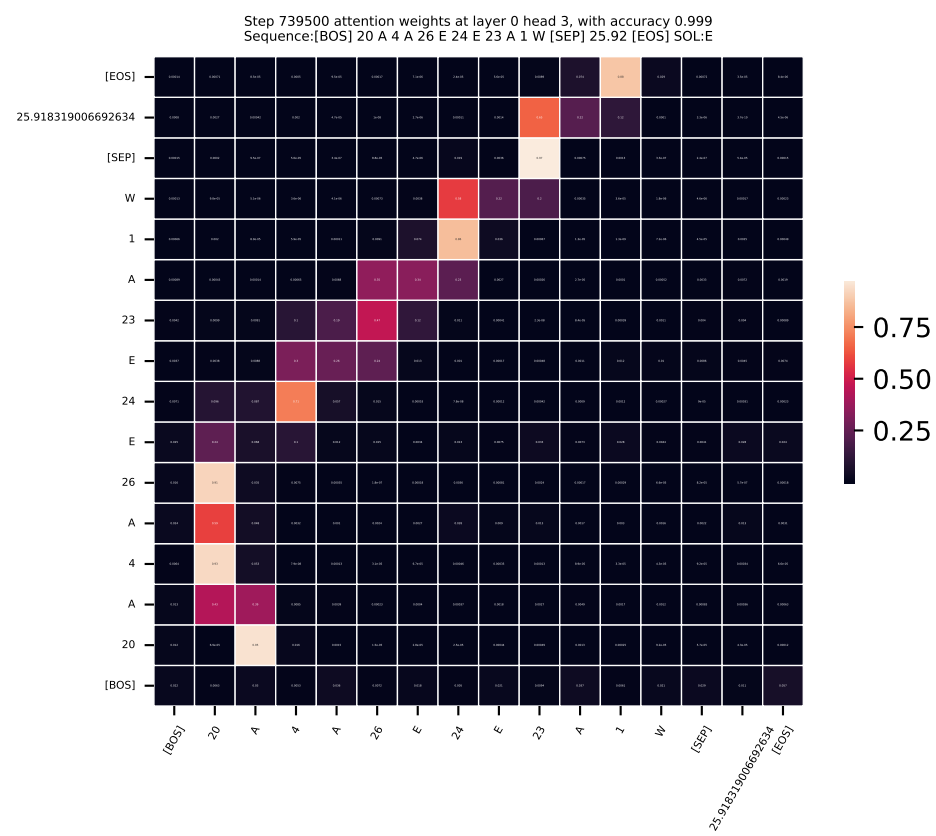

Step 730000 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

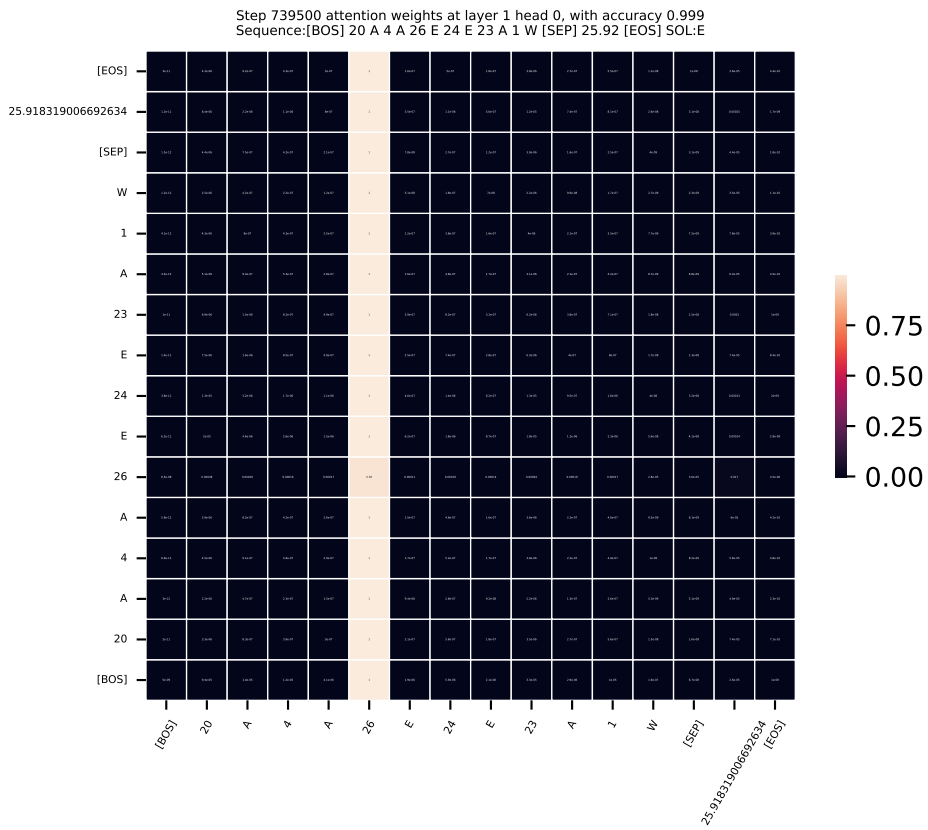

Step 730000 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

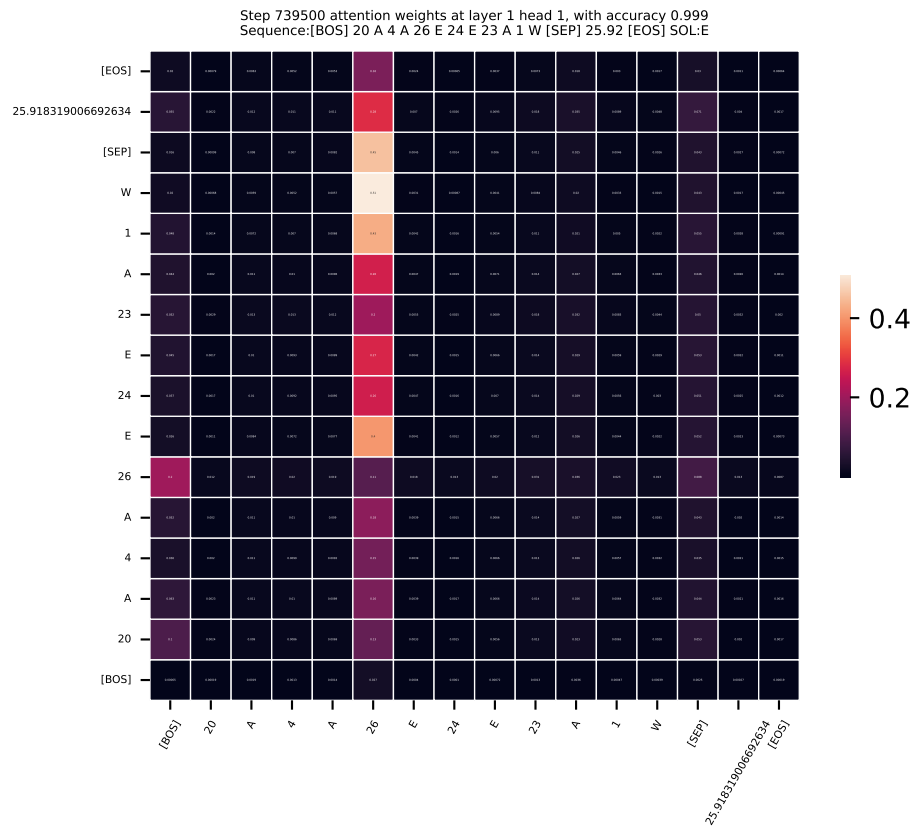

Step 739500 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

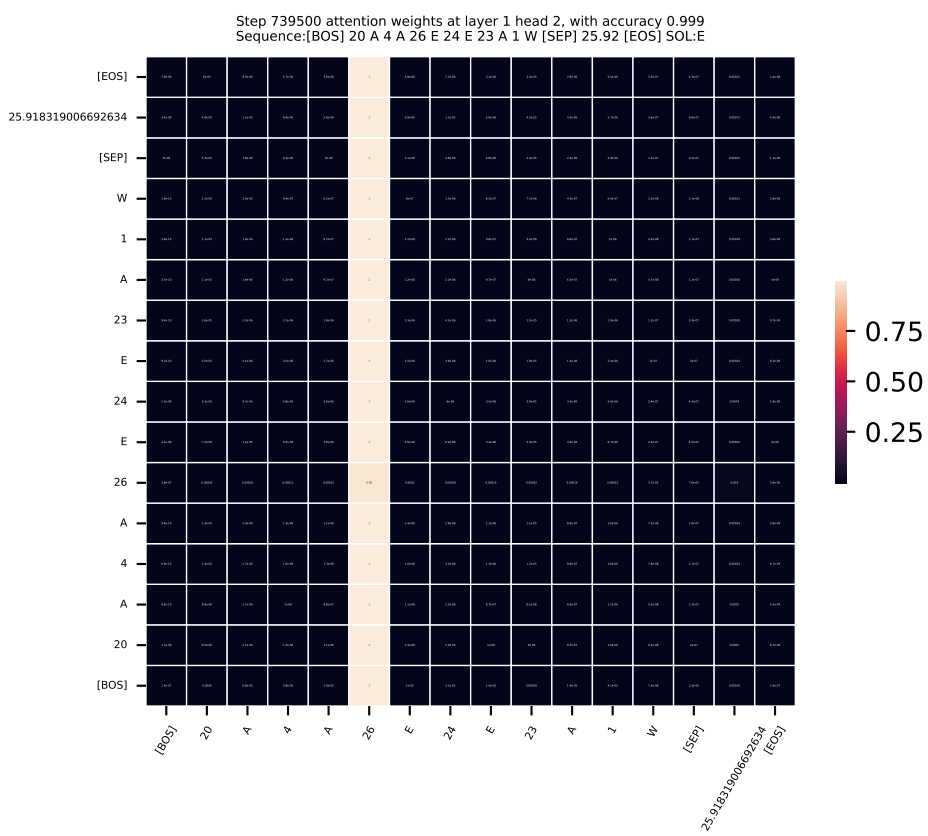

Step 739500 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

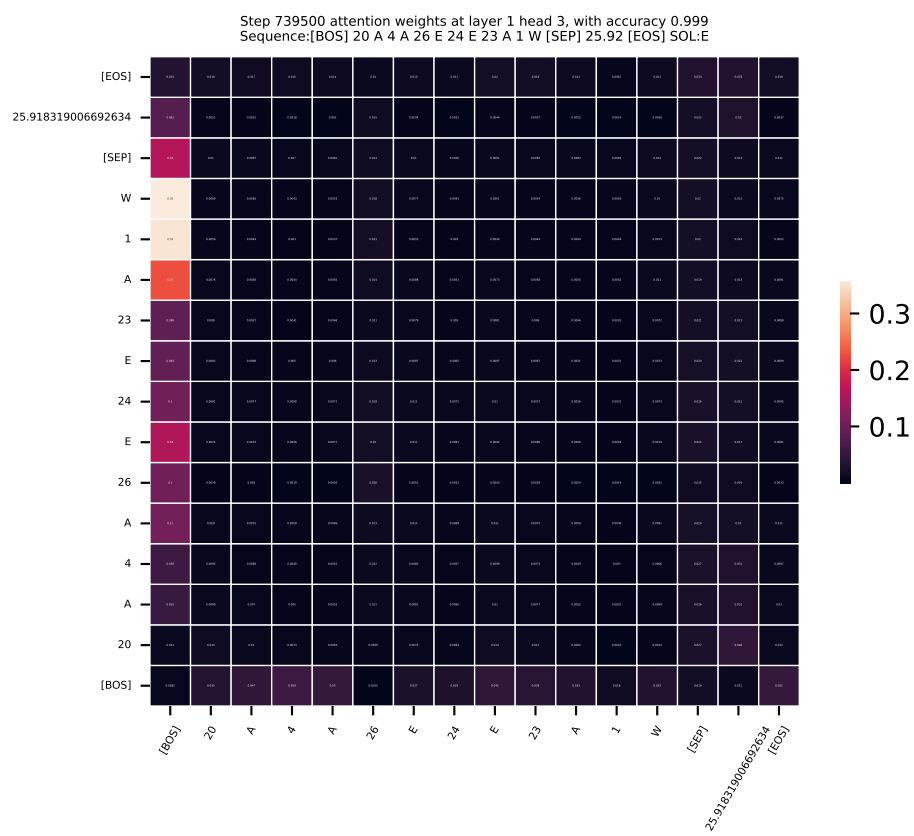

Step 739500 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

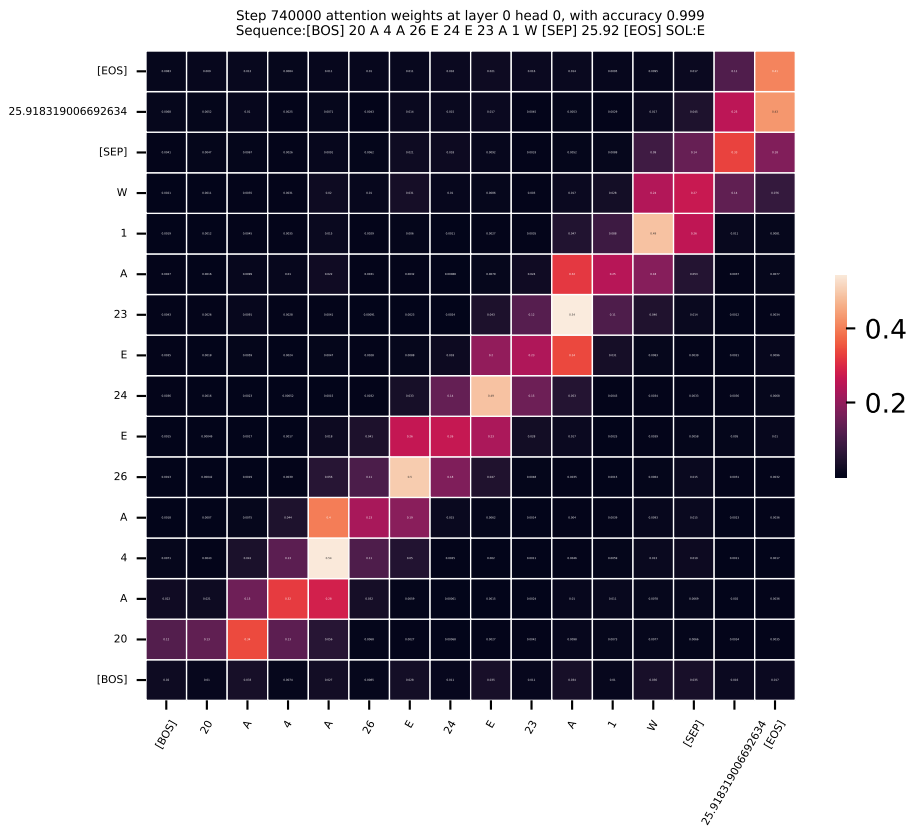

Step 739500 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

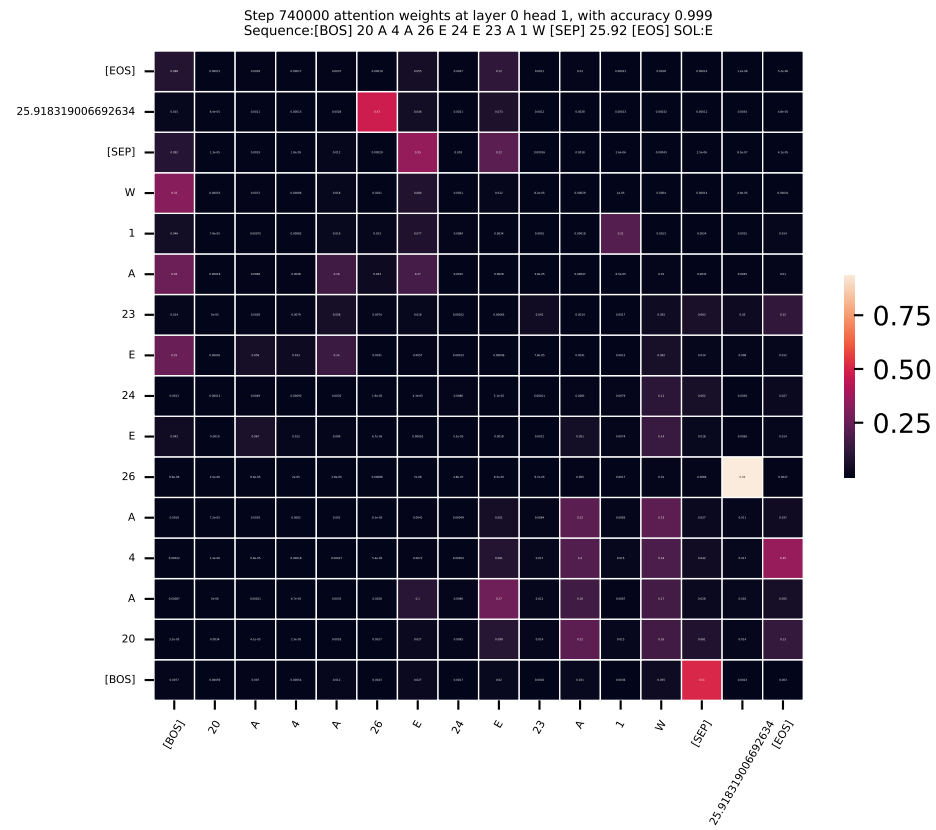

Step 739500 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

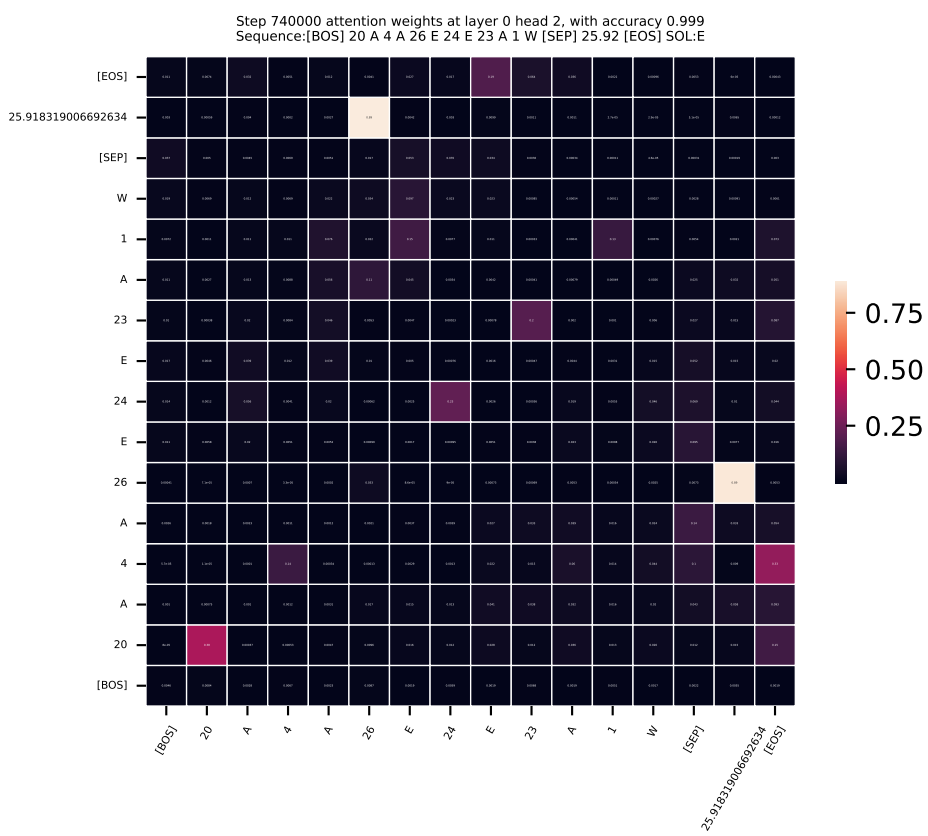

Step 739500 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

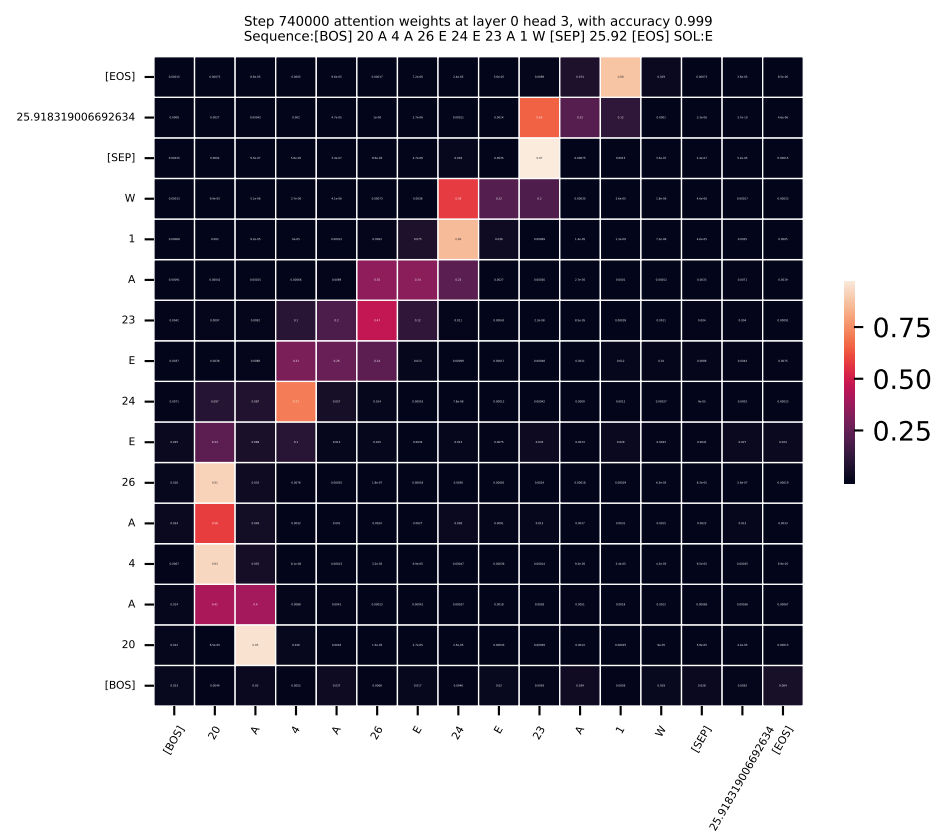

Step 739500 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

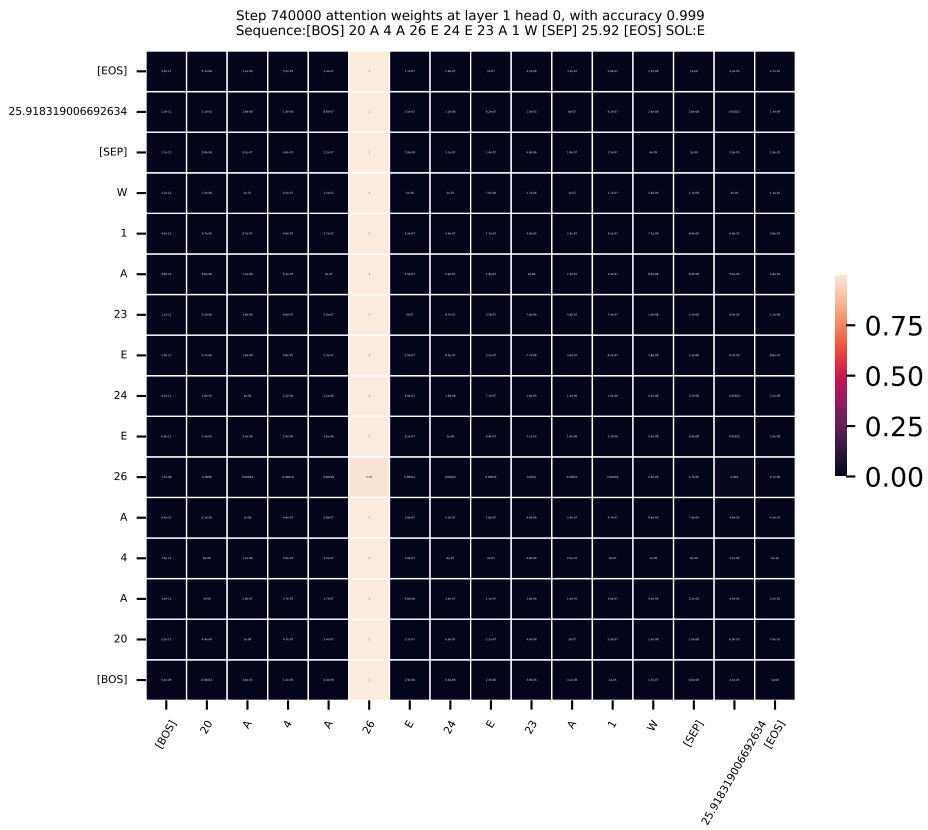

Step 739500 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

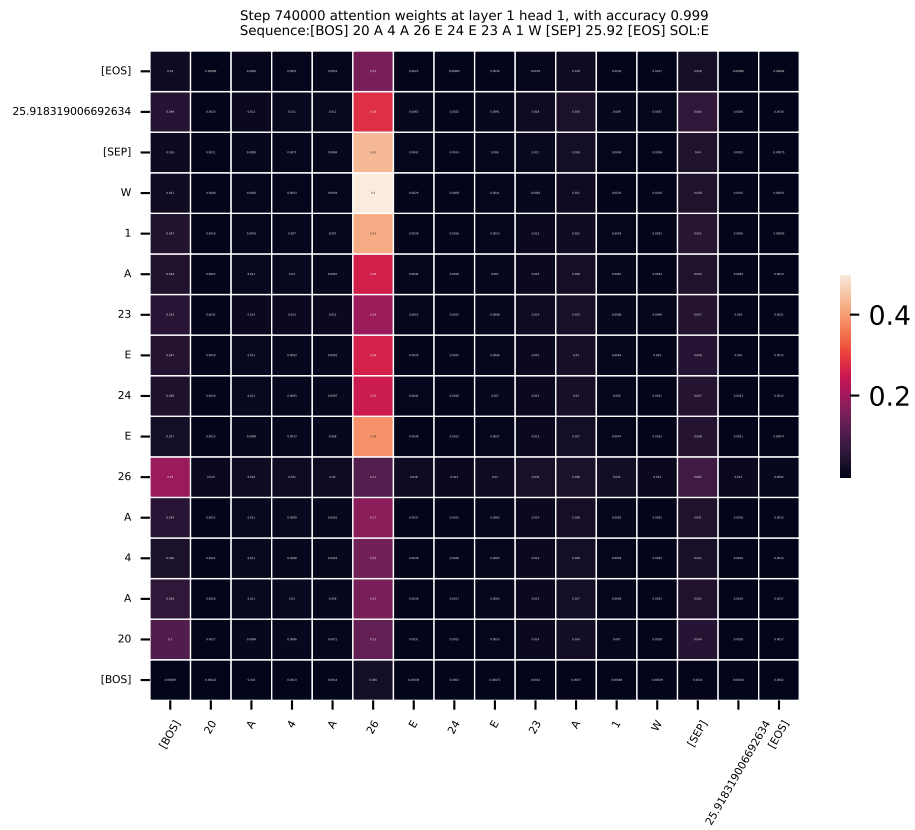

Step 740000 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

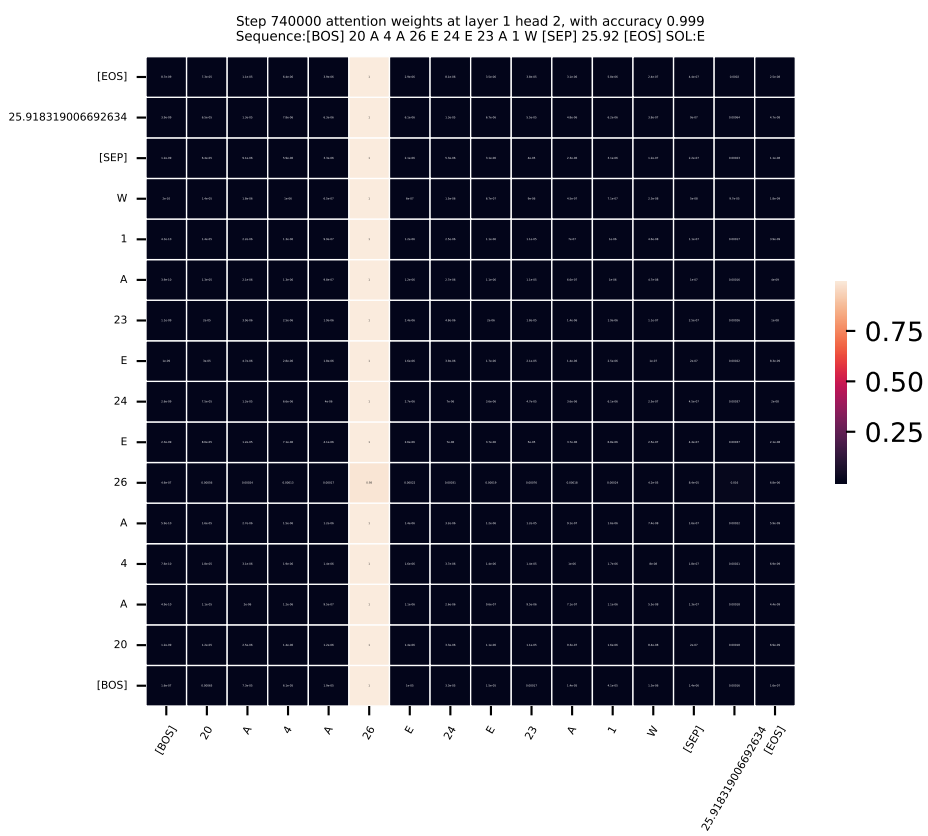

Step 740000 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

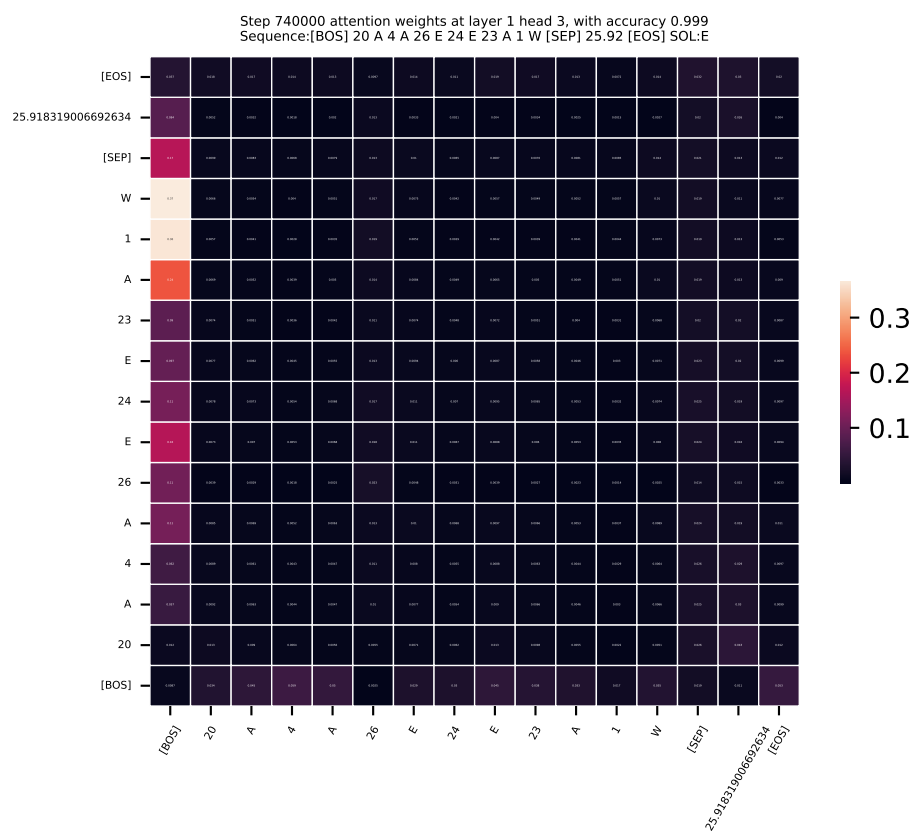

Step 740000 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

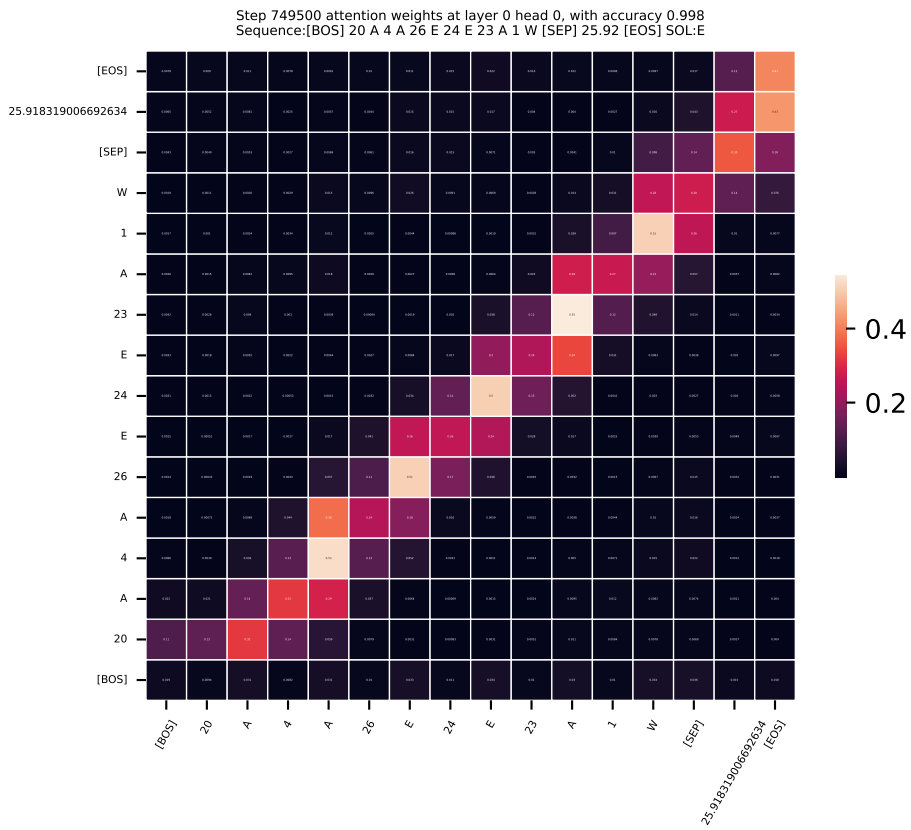

Step 740000 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

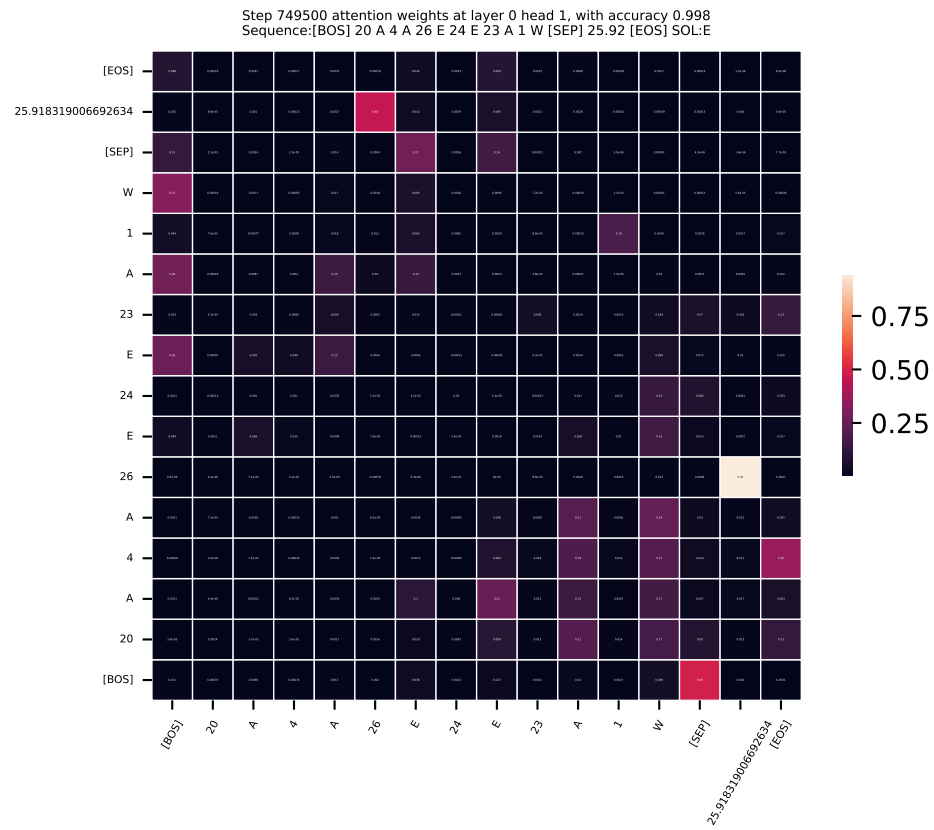

Step 740000 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

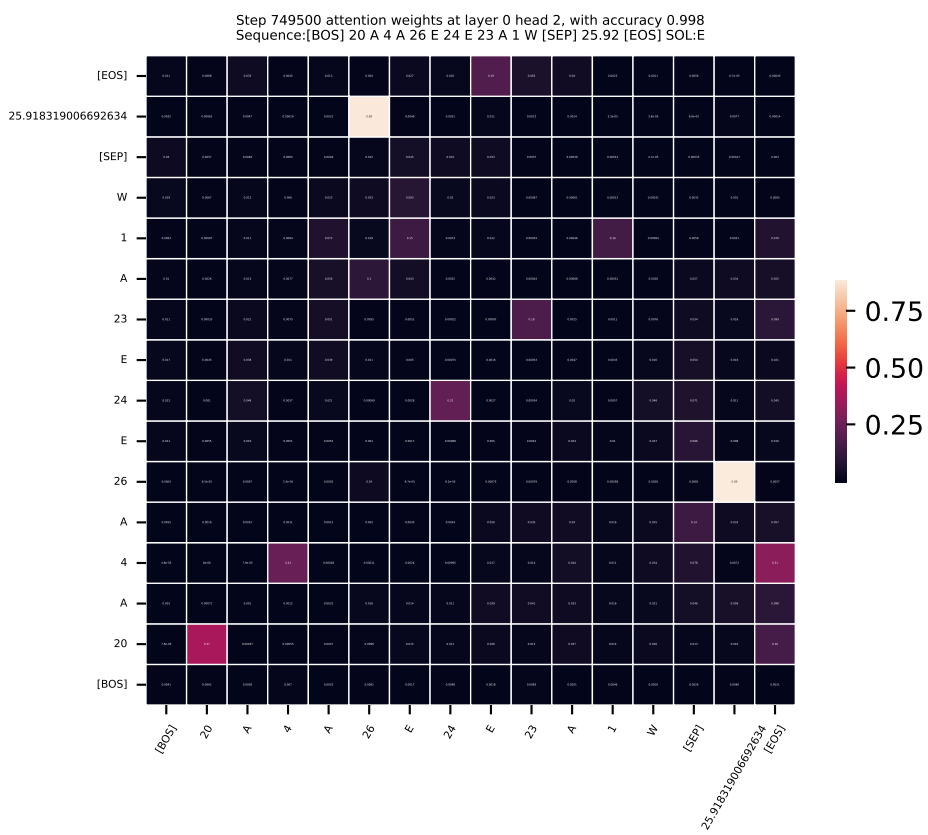

Step 740000 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

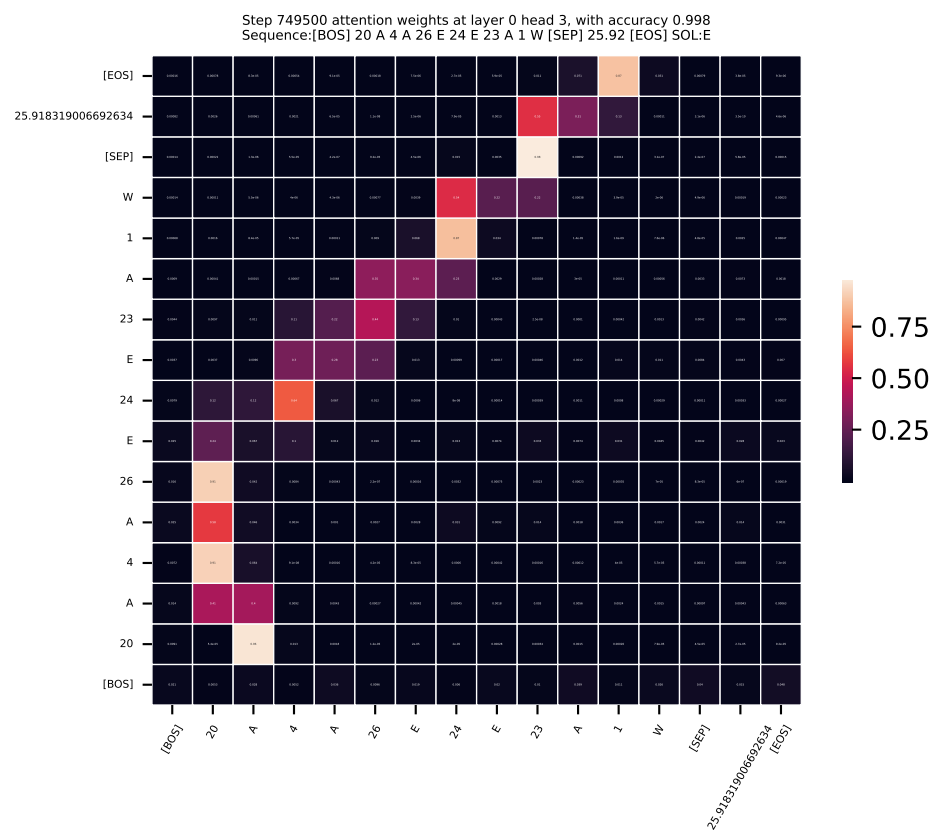

Step 740000 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

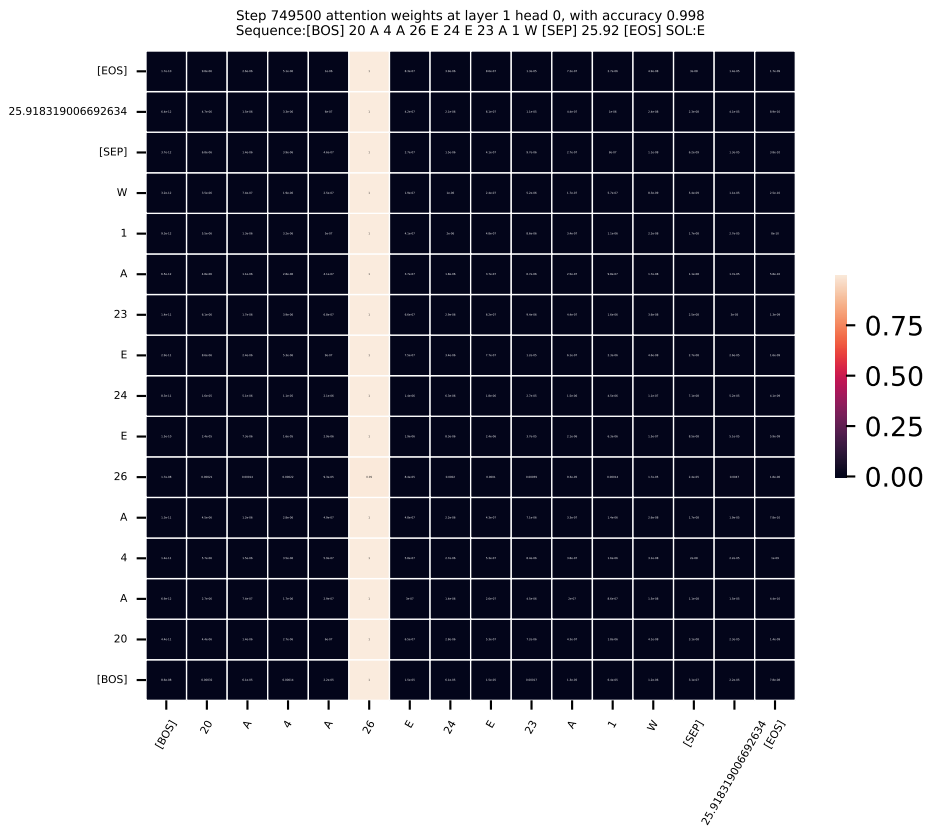

Step 740000 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

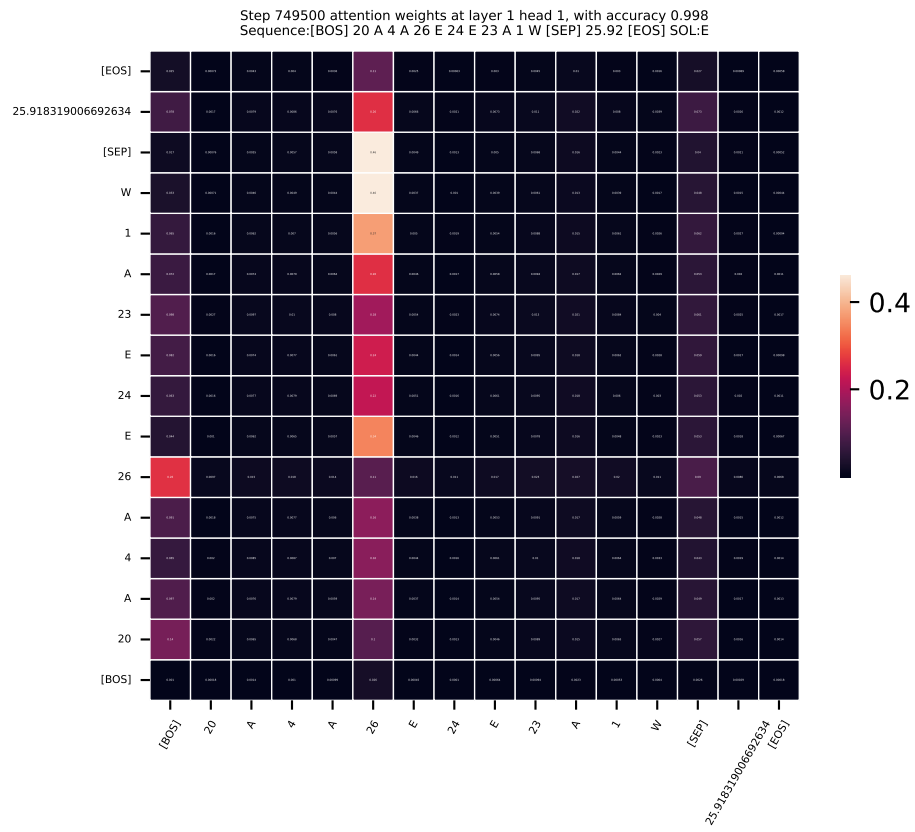

Step 749500 attention weights at layer 0 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

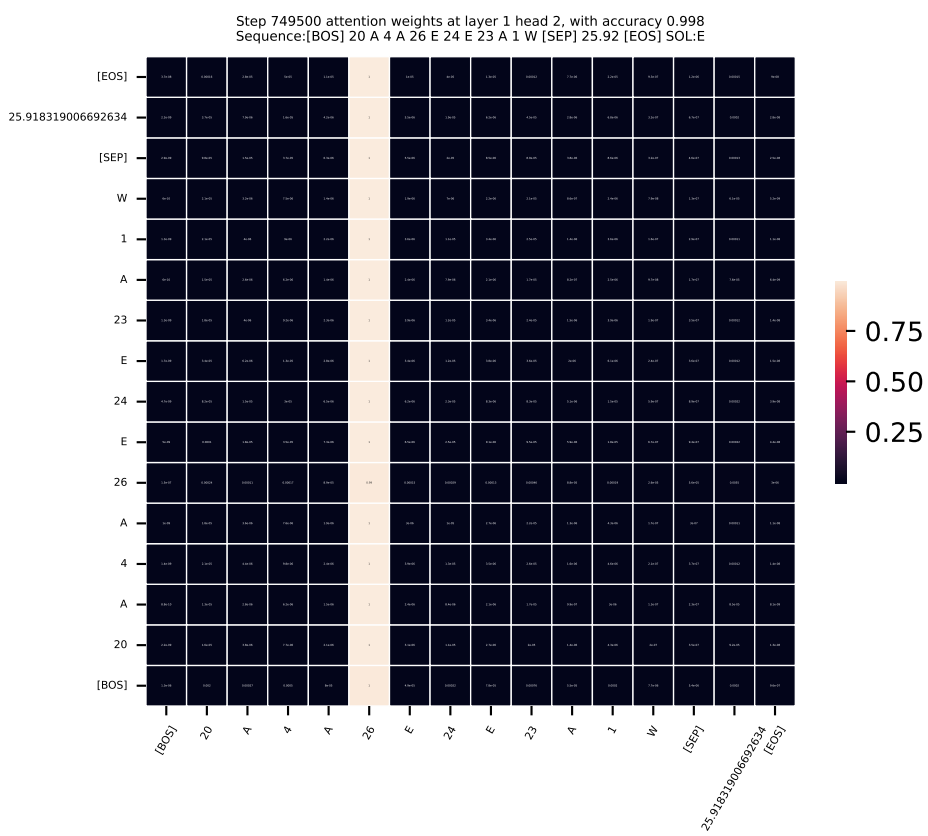

Step 749500 attention weights at layer 0 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

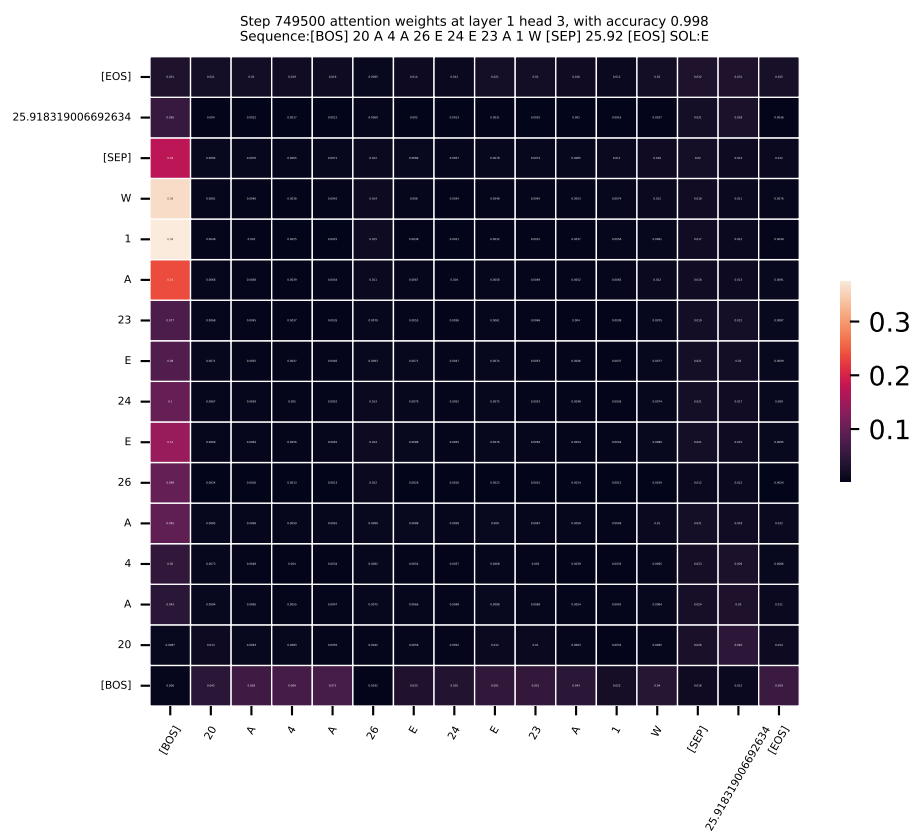

Step 749500 attention weights at layer 0 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

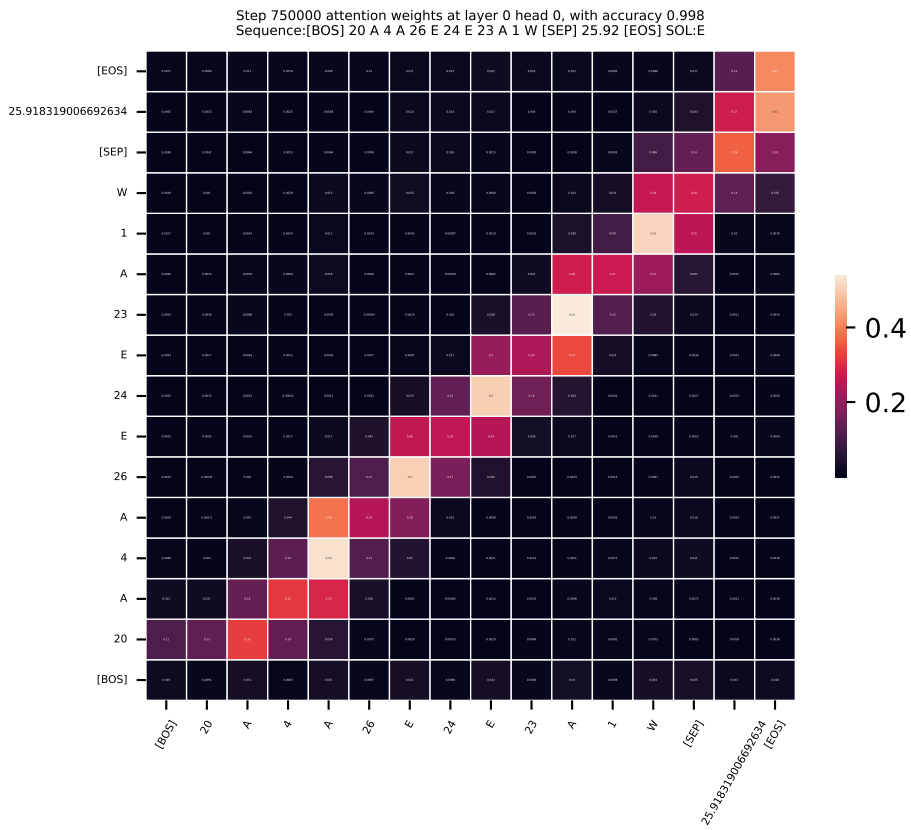

Step 749500 attention weights at layer 0 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

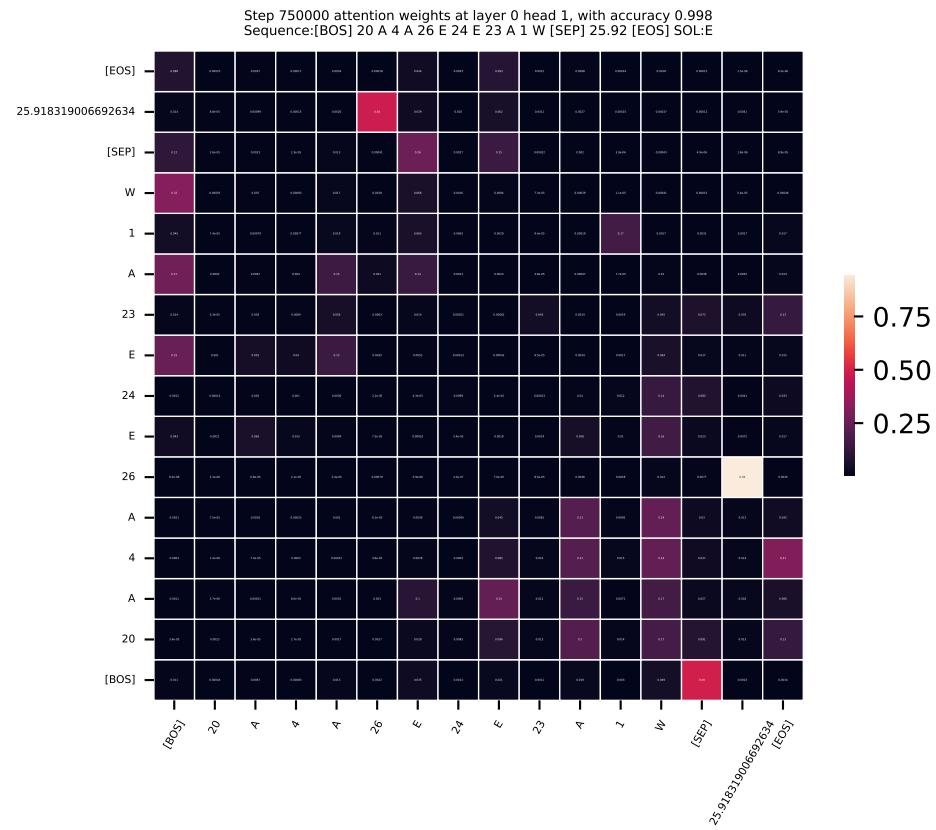

Step 749500 attention weights at layer 1 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

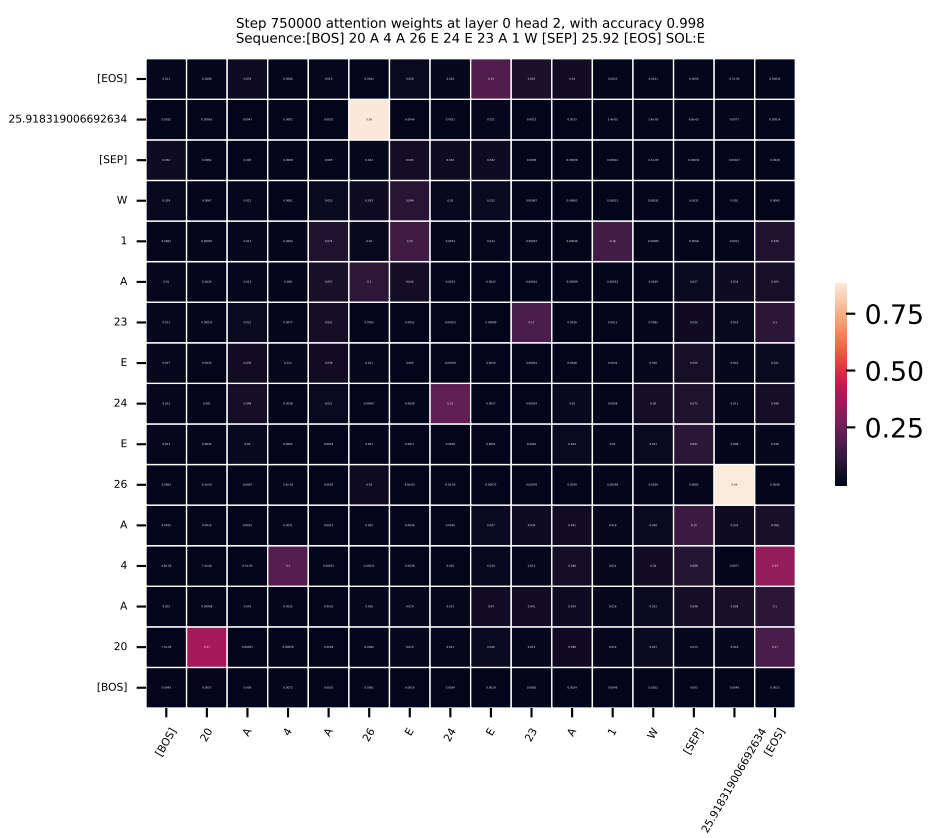

Step 749500 attention weights at layer 1 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

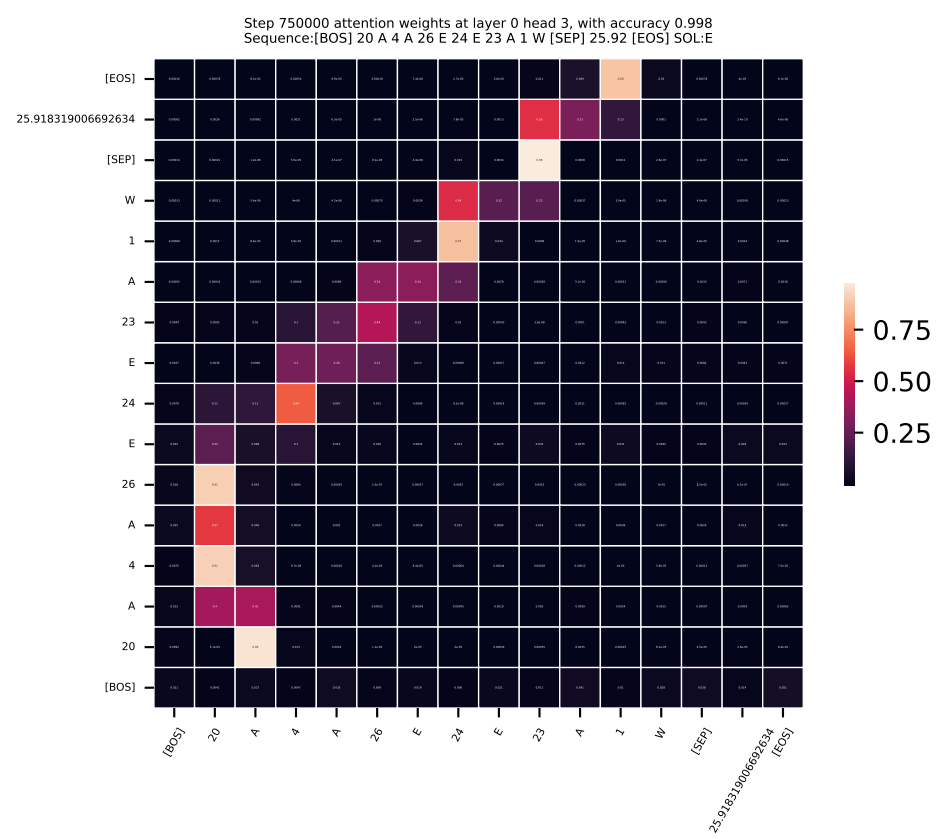

Step 749500 attention weights at layer 1 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

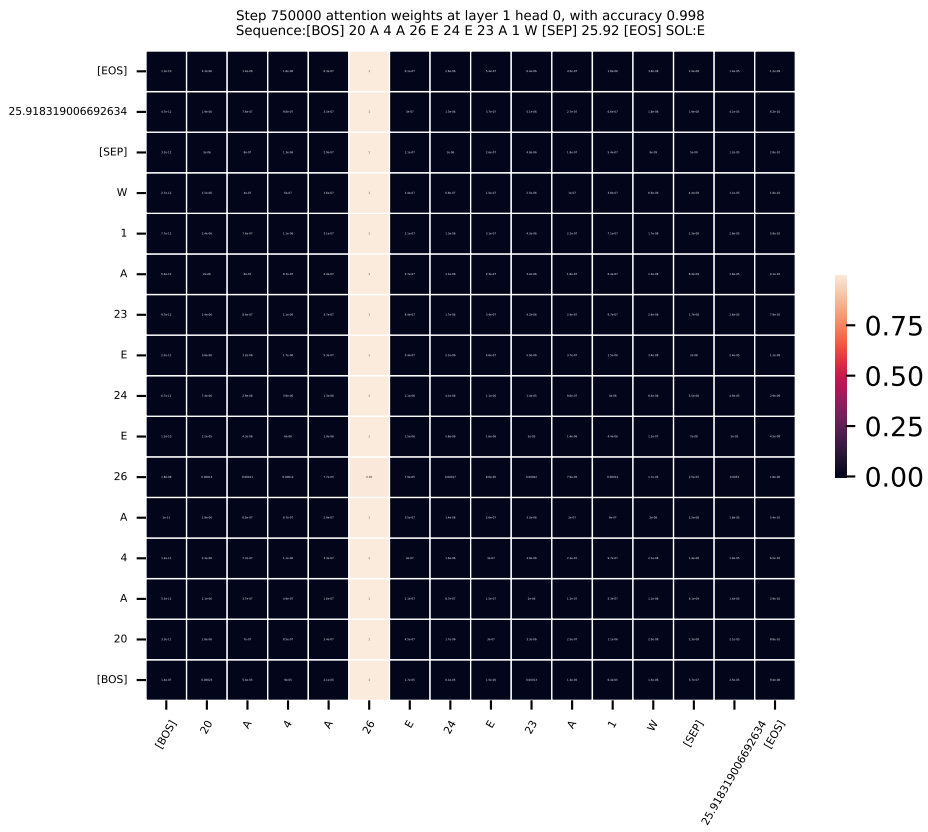

Step 749500 attention weights at layer 1 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

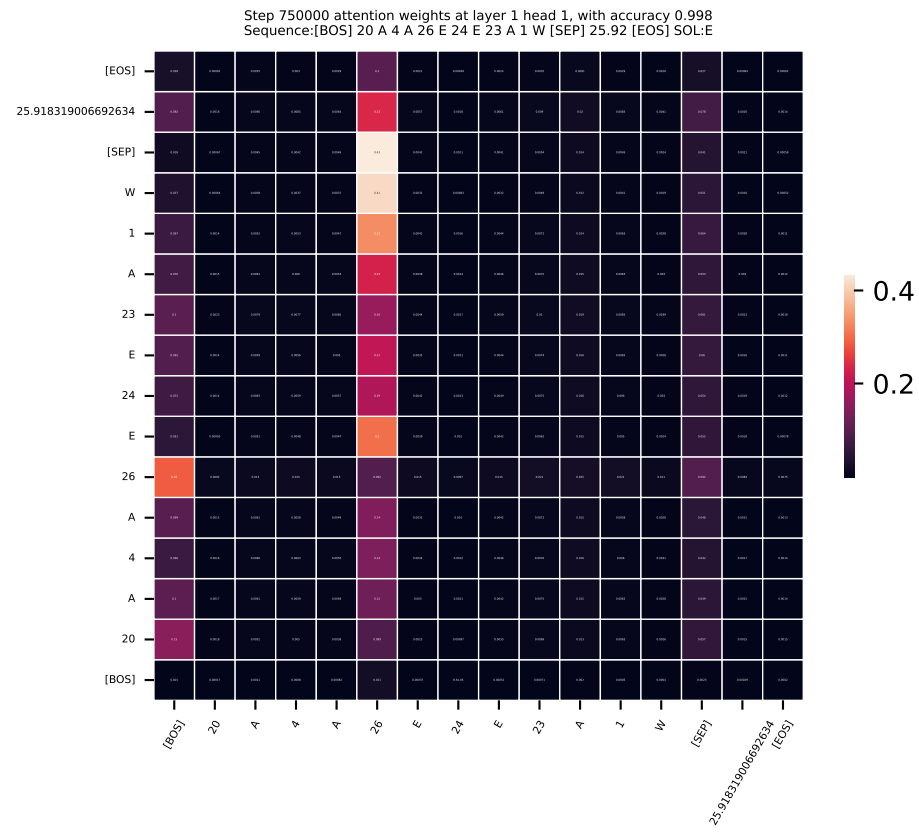

Step 750000 attention weights at layer 0 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.4
0.2

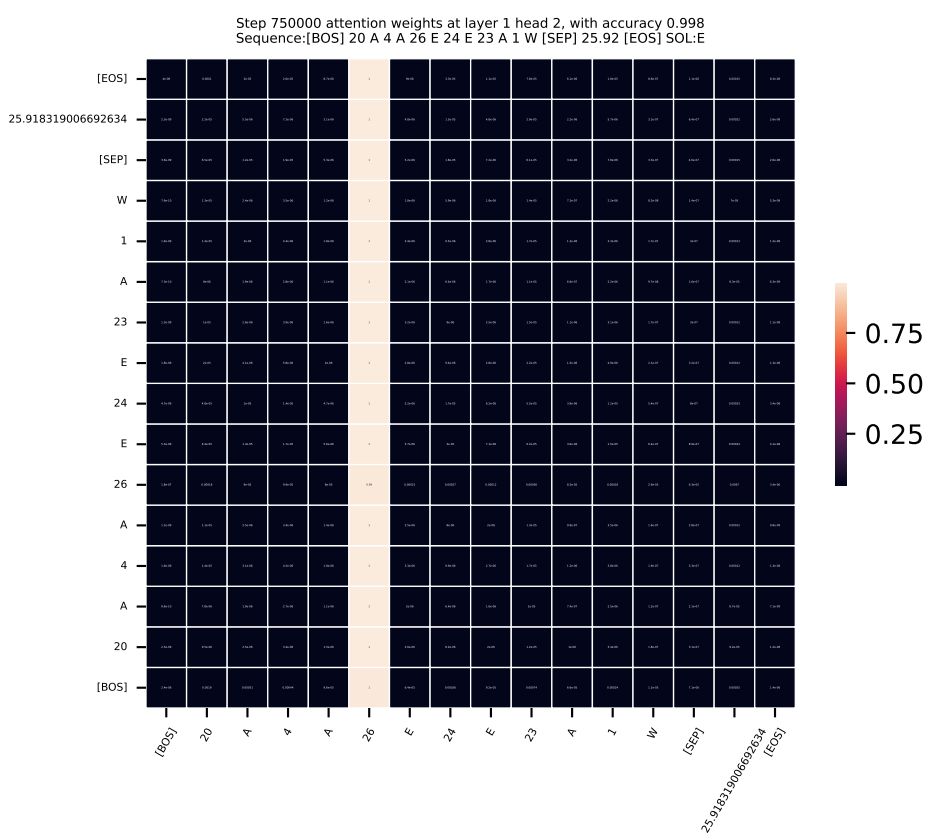

Step 750000 attention weights at layer 0 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

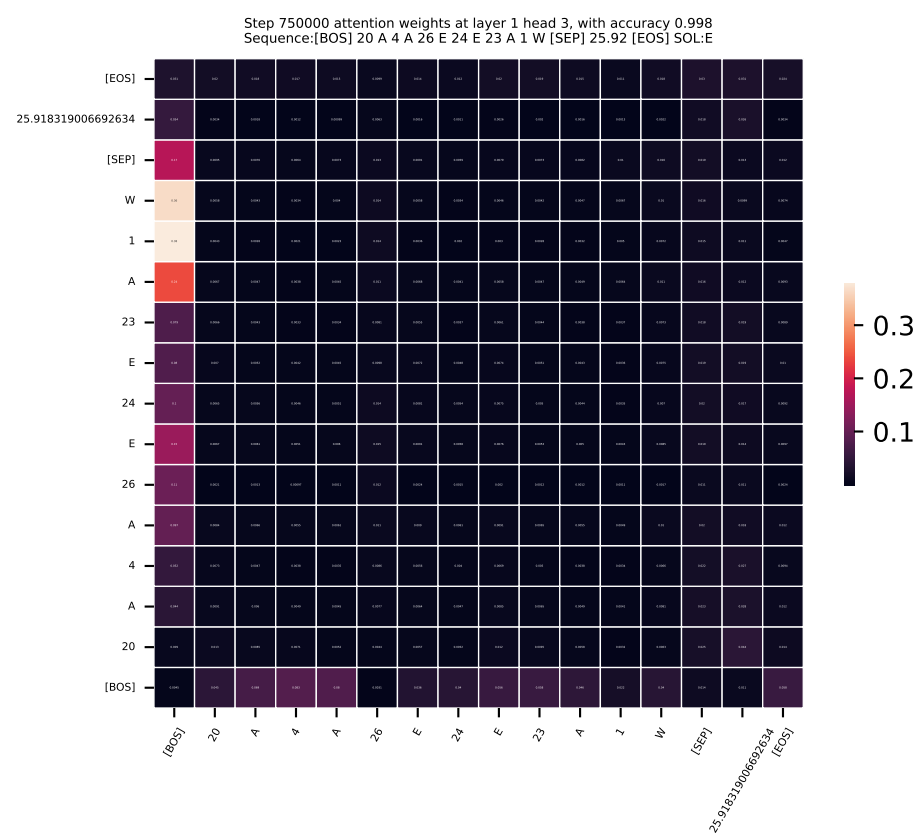

Step 750000 attention weights at layer 0 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

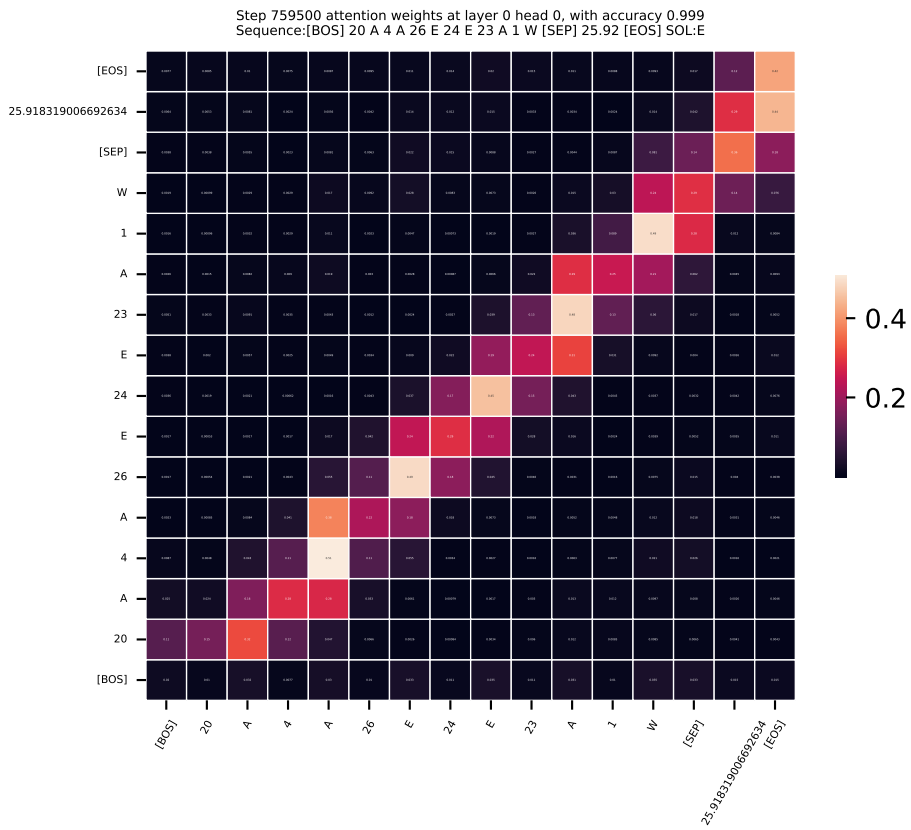

Step 750000 attention weights at layer 0 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

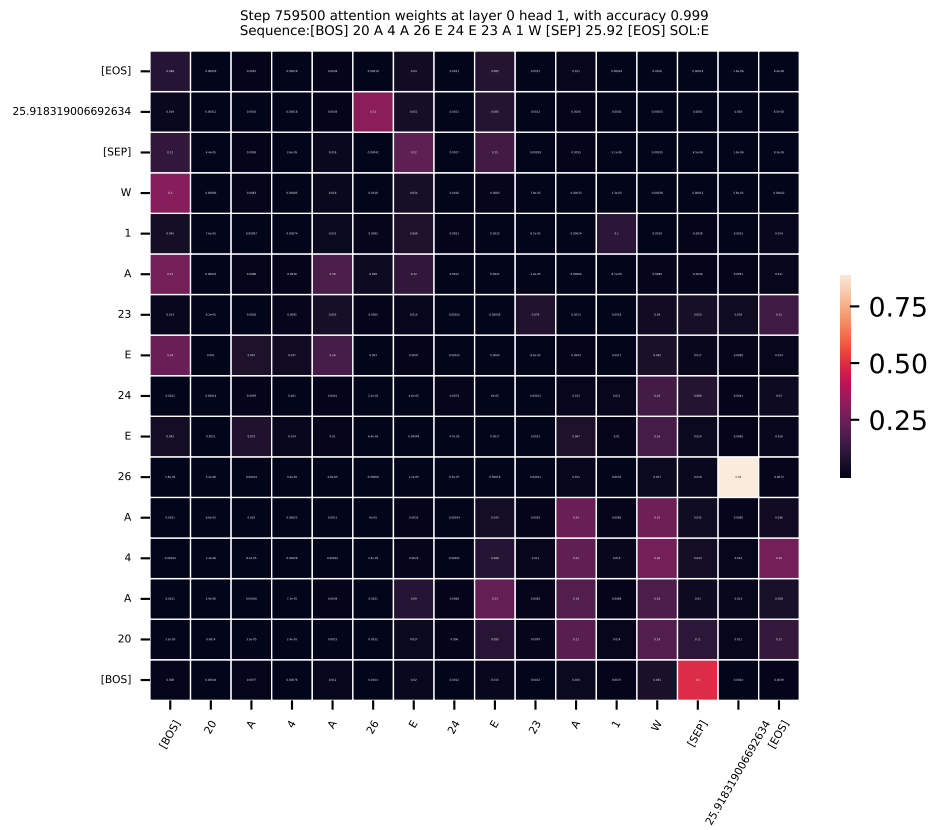

Step 750000 attention weights at layer 1 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25
0.00

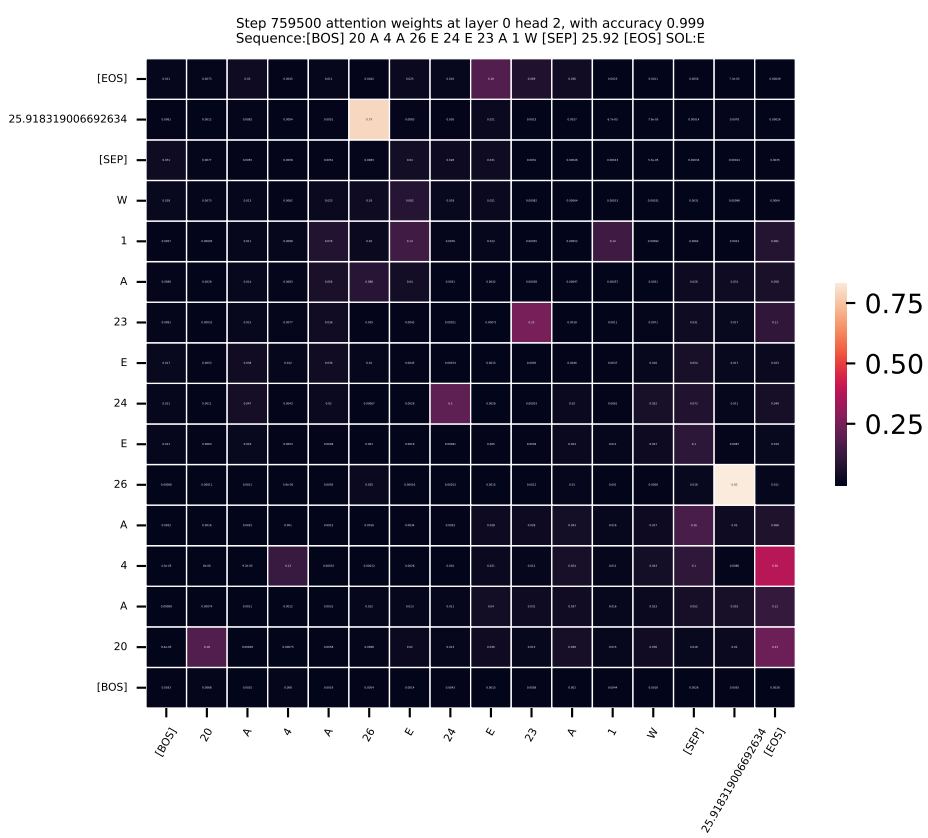

Step 750000 attention weights at layer 1 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.4
0.2

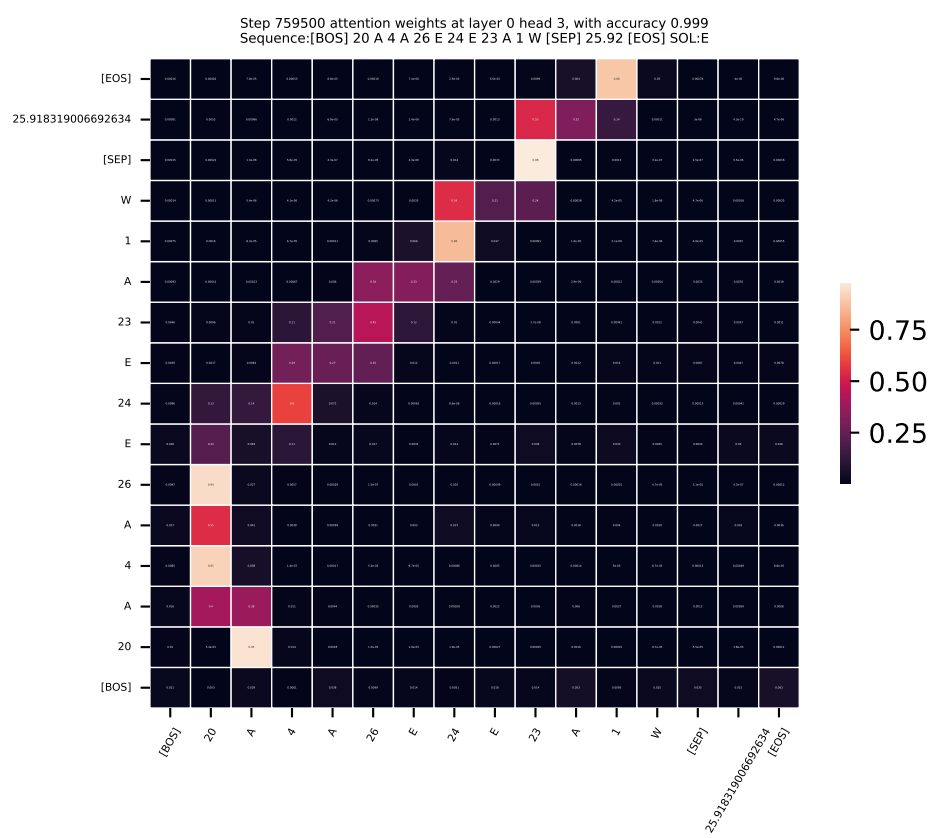

Step 750000 attention weights at layer 1 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

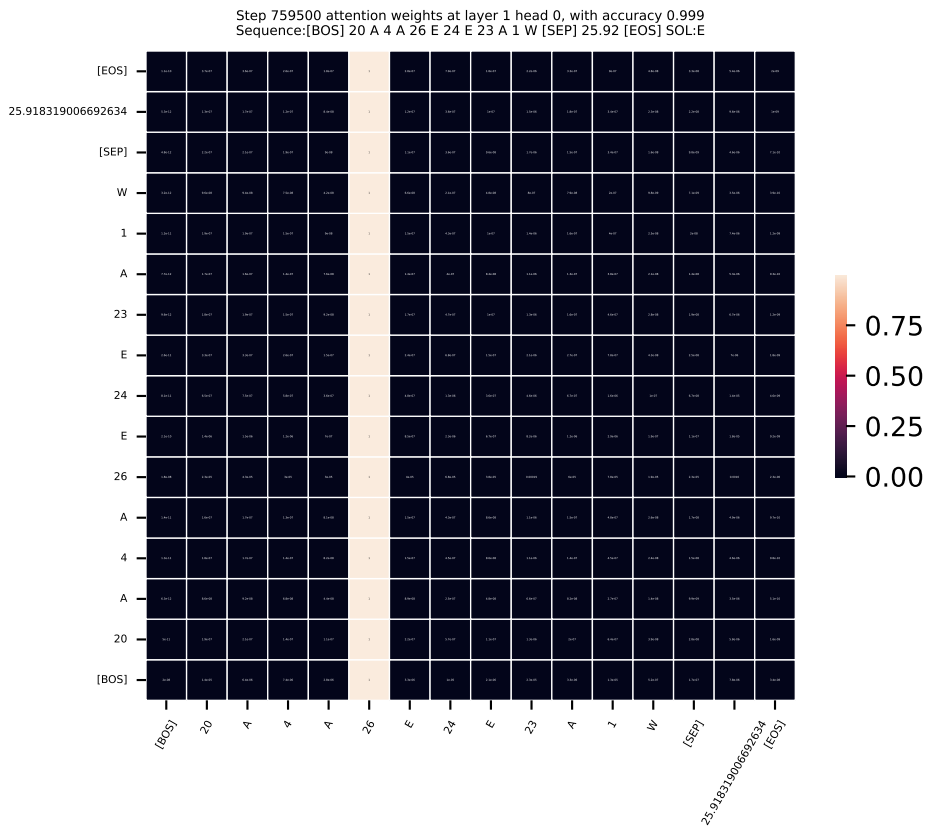

Step 750000 attention weights at layer 1 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.3
0.2
0.1

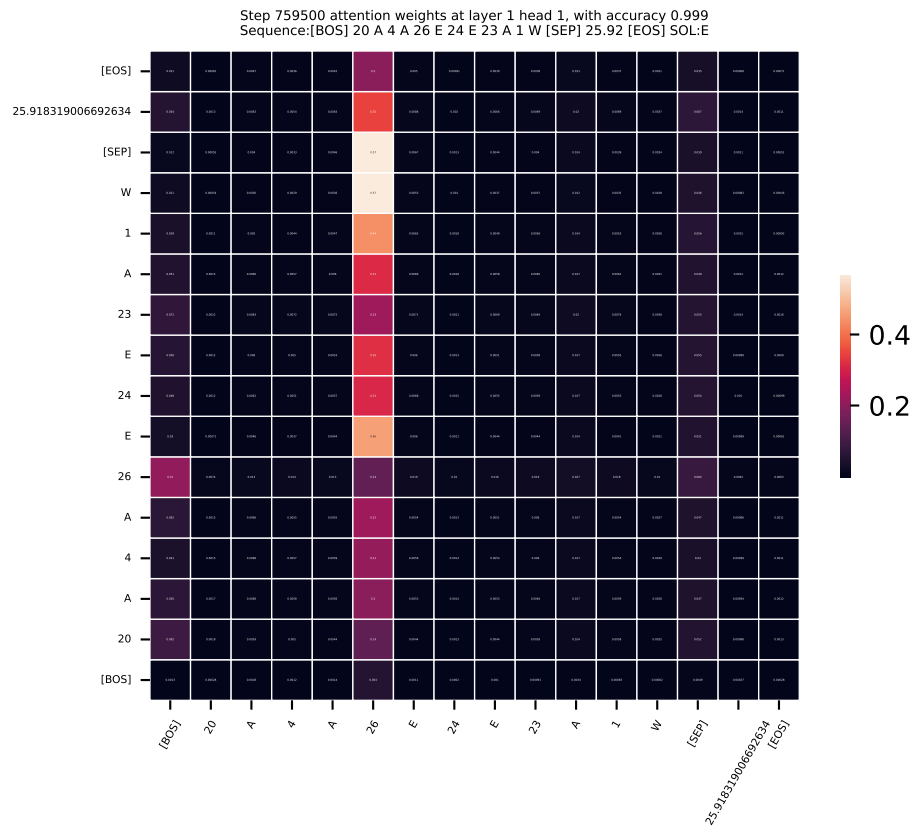

Step 759500 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

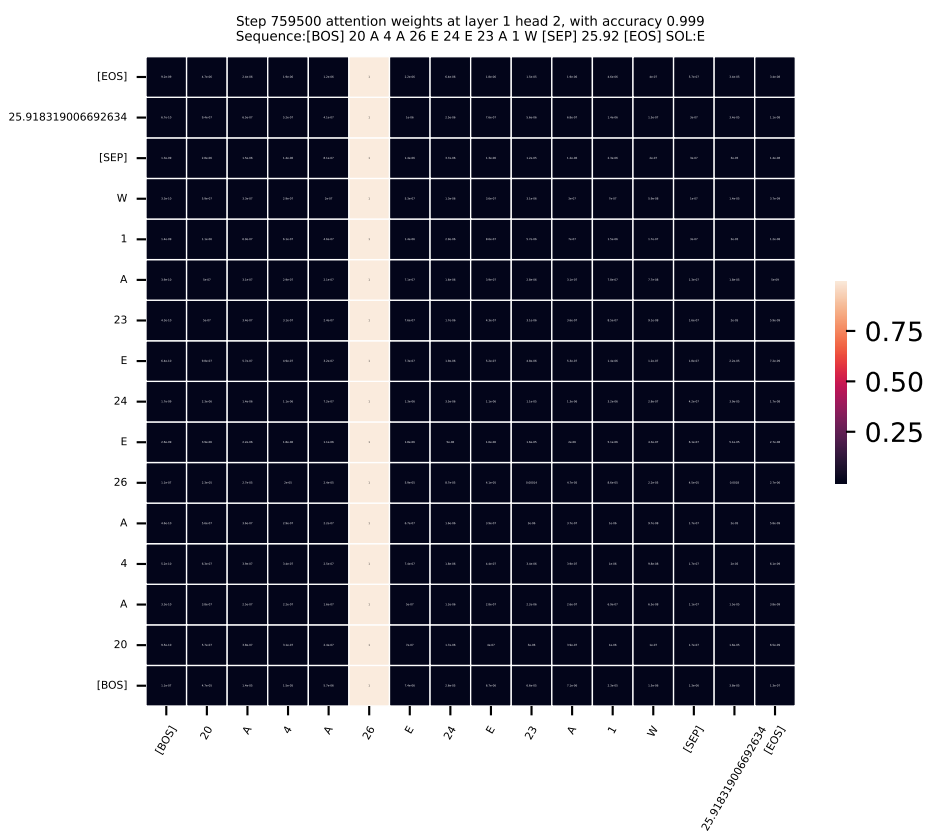

Step 759500 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

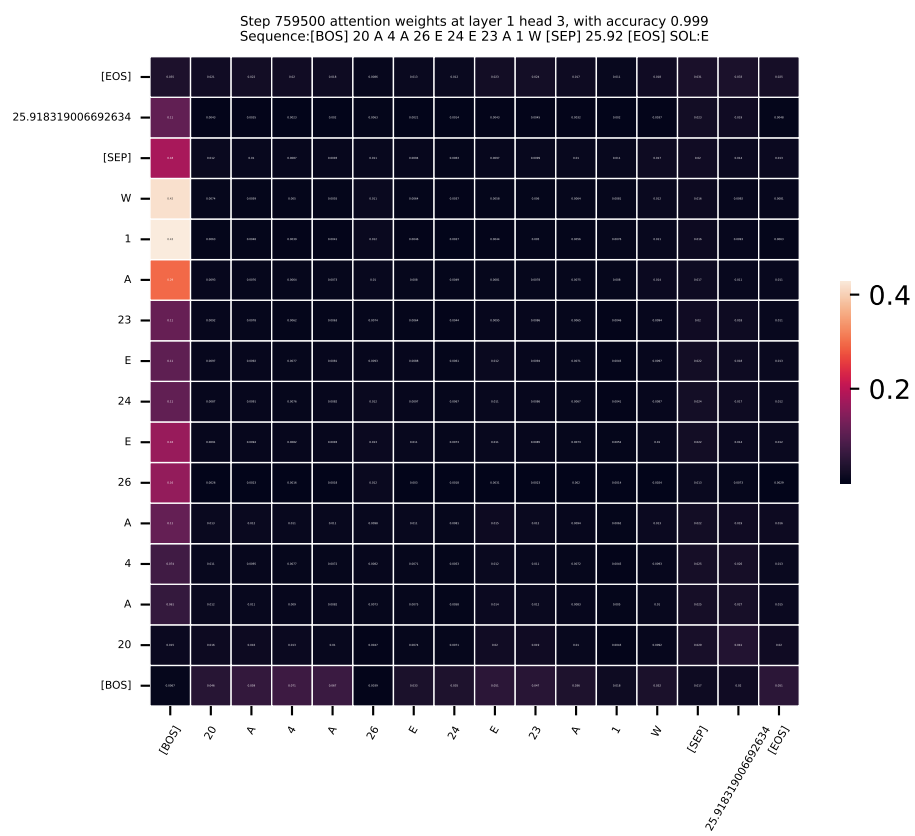

Step 759500 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

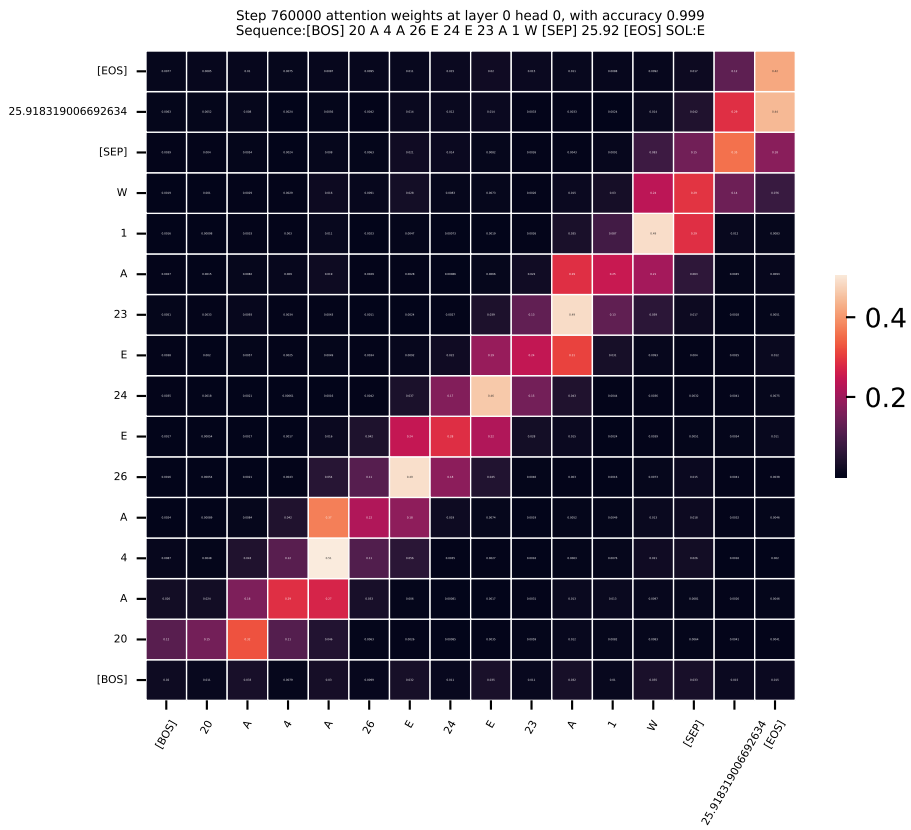

Step 759500 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

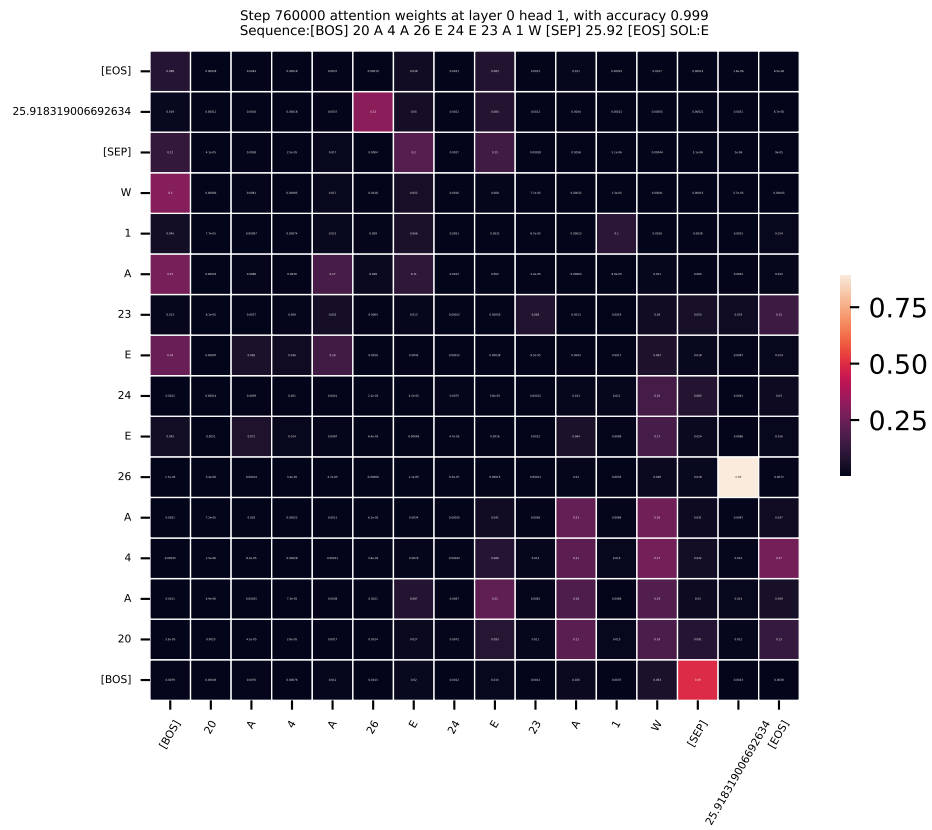

Step 759500 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

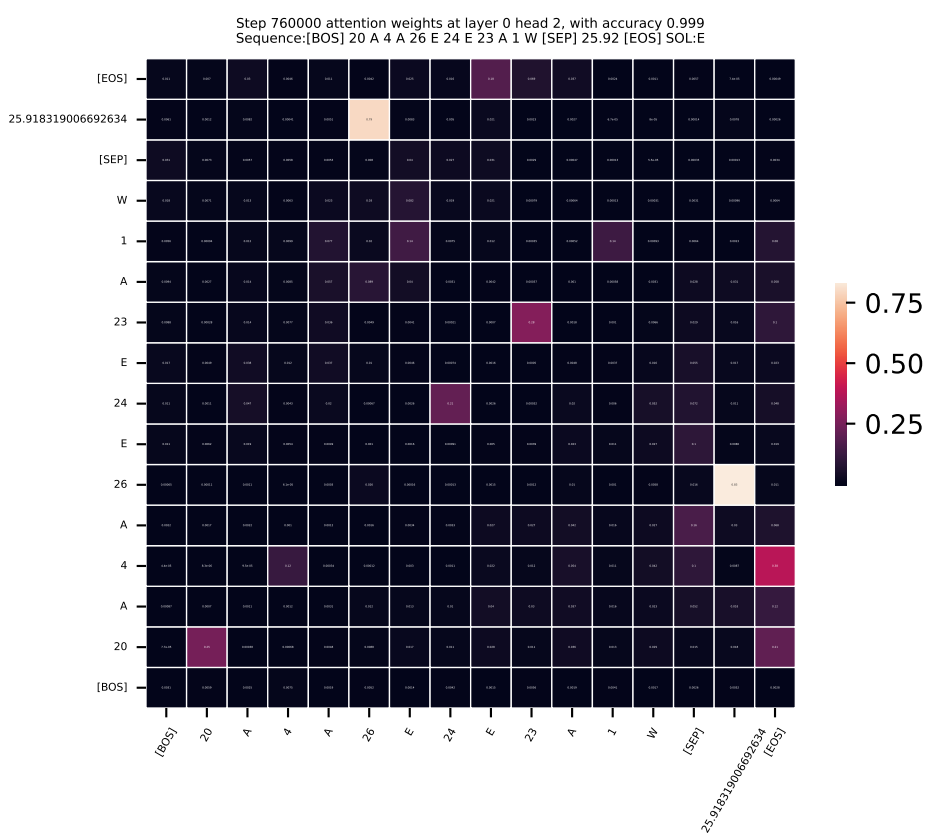

Step 759500 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

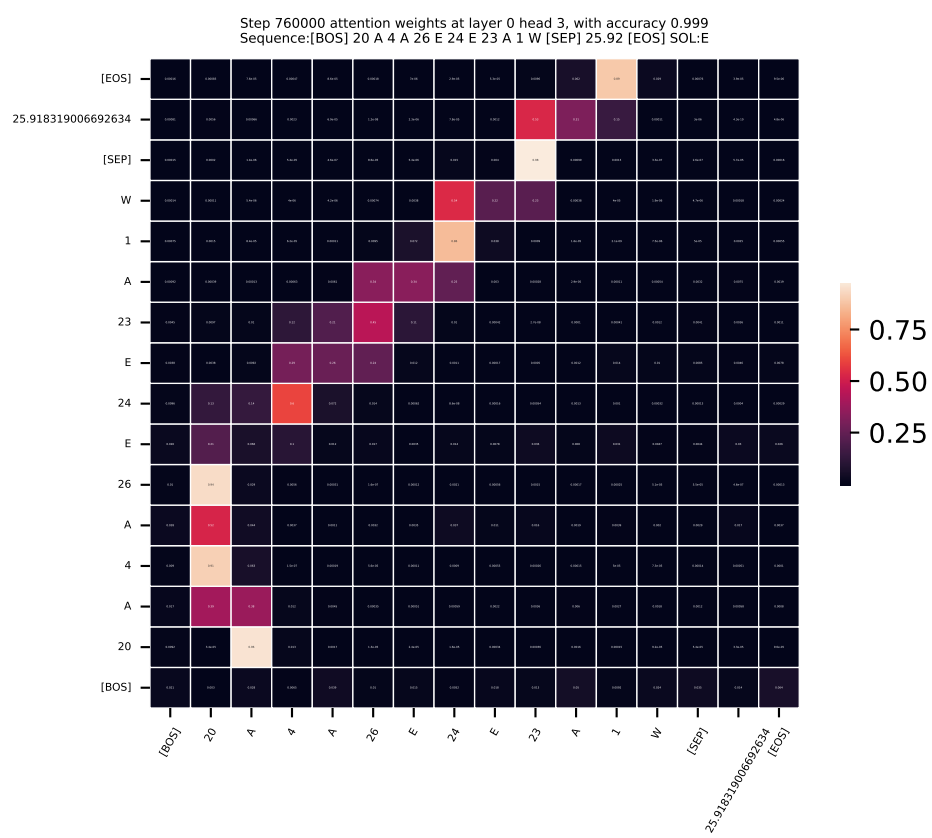

Step 759500 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

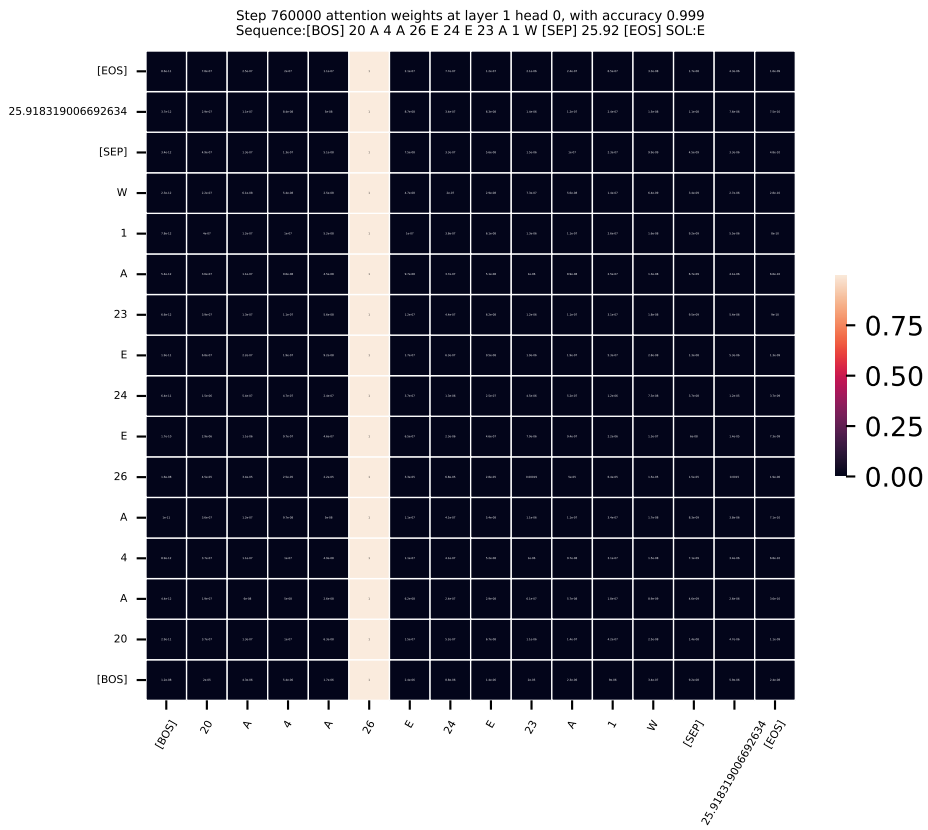

Step 759500 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

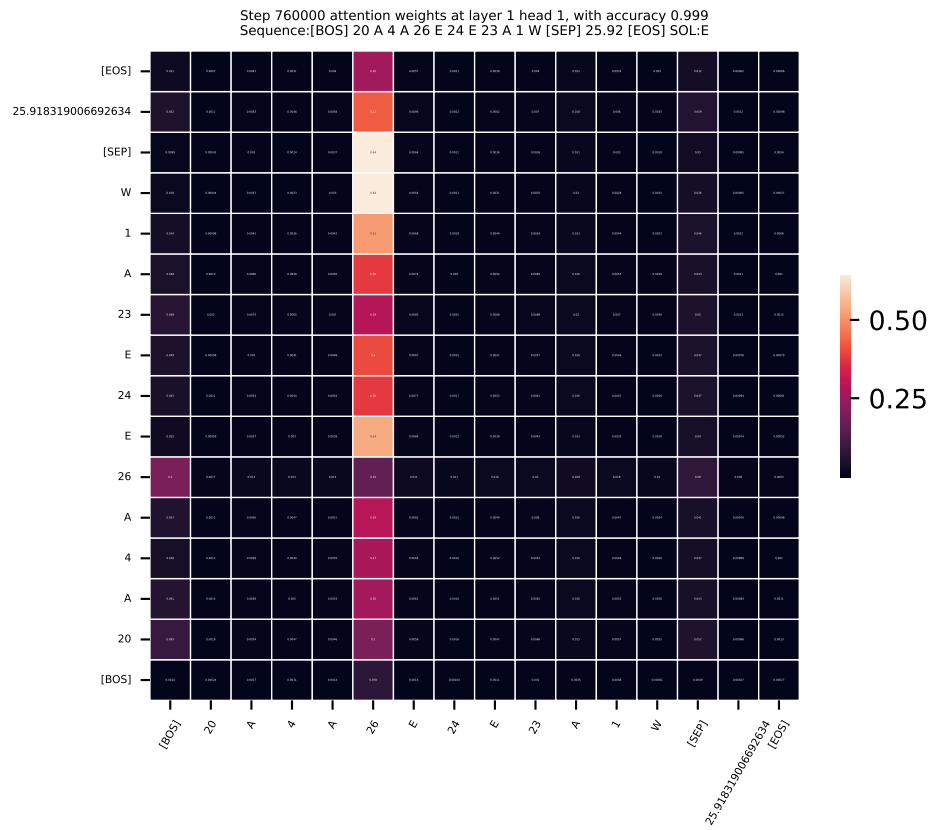

Step 760000 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

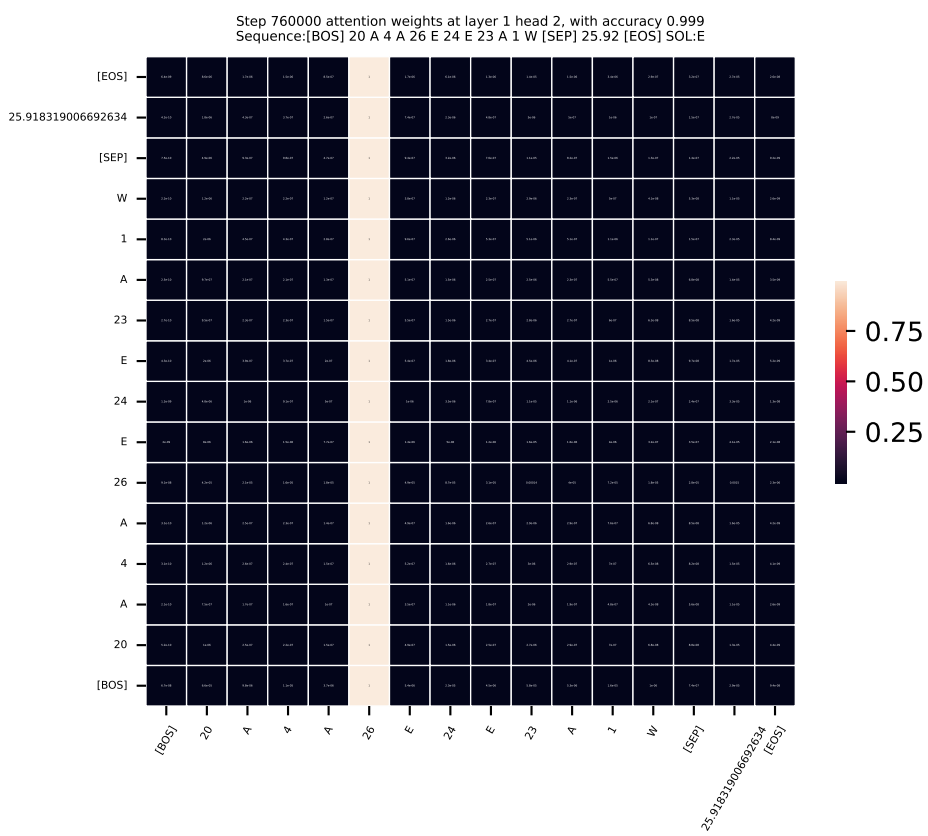

Step 760000 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

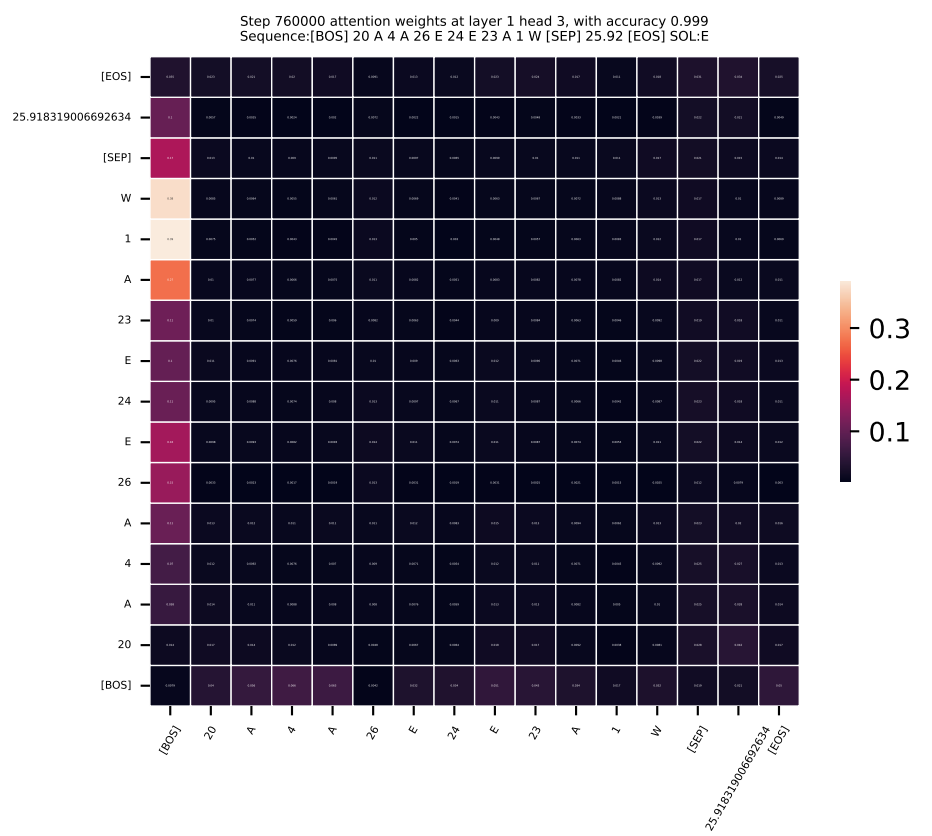

Step 760000 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

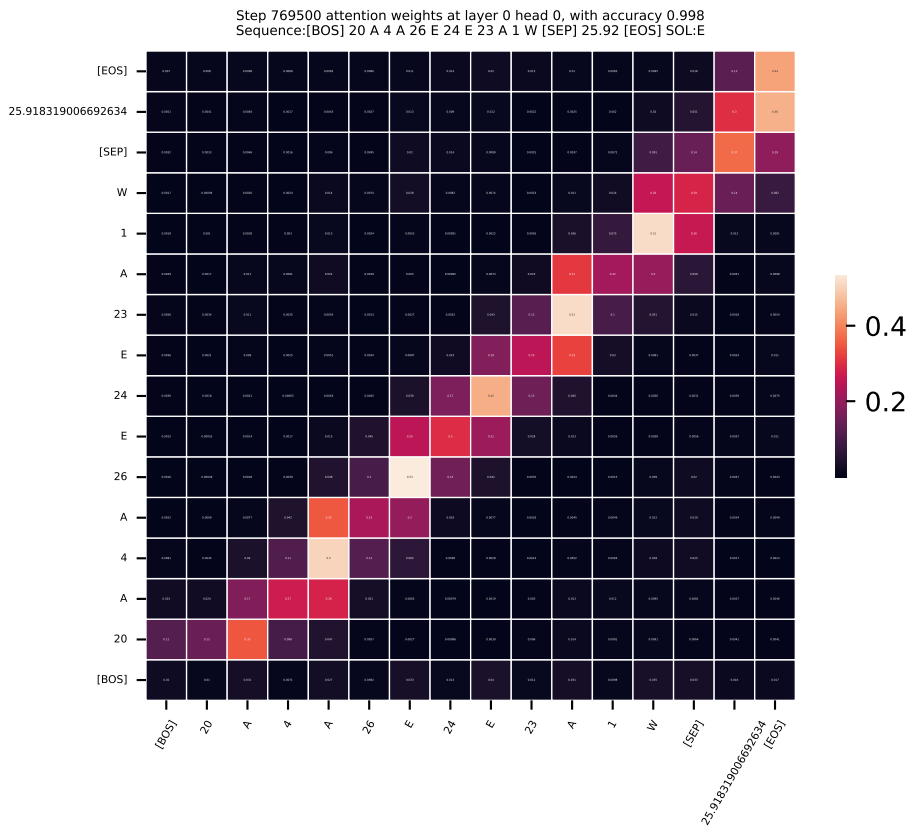

Step 760000 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

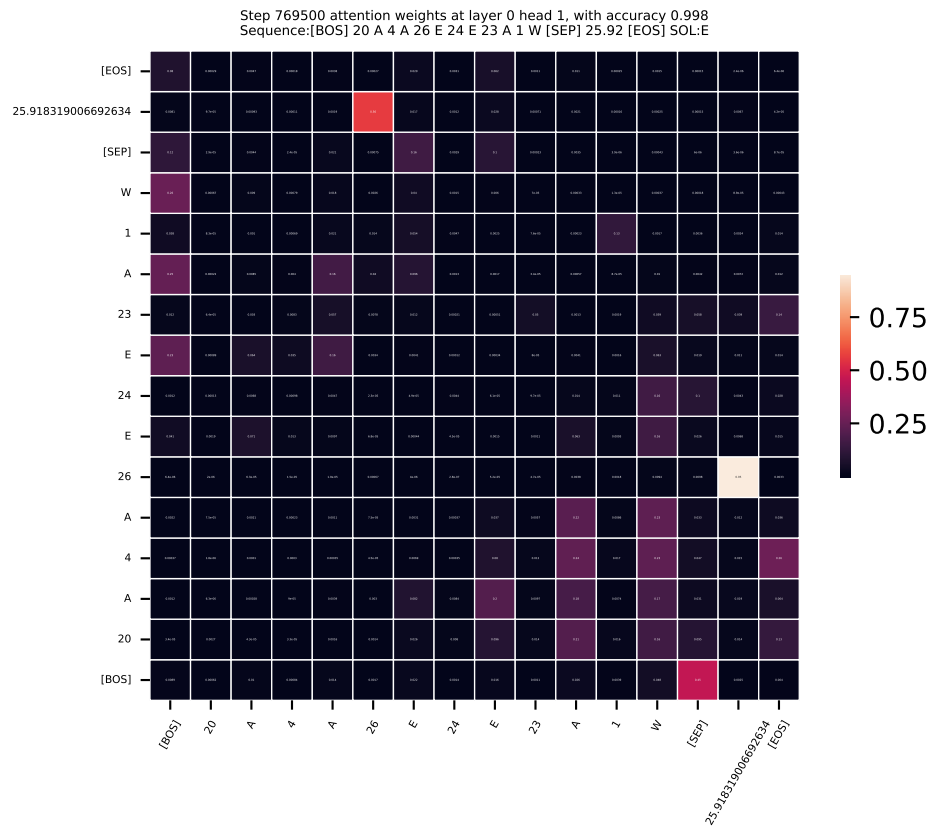

Step 760000 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

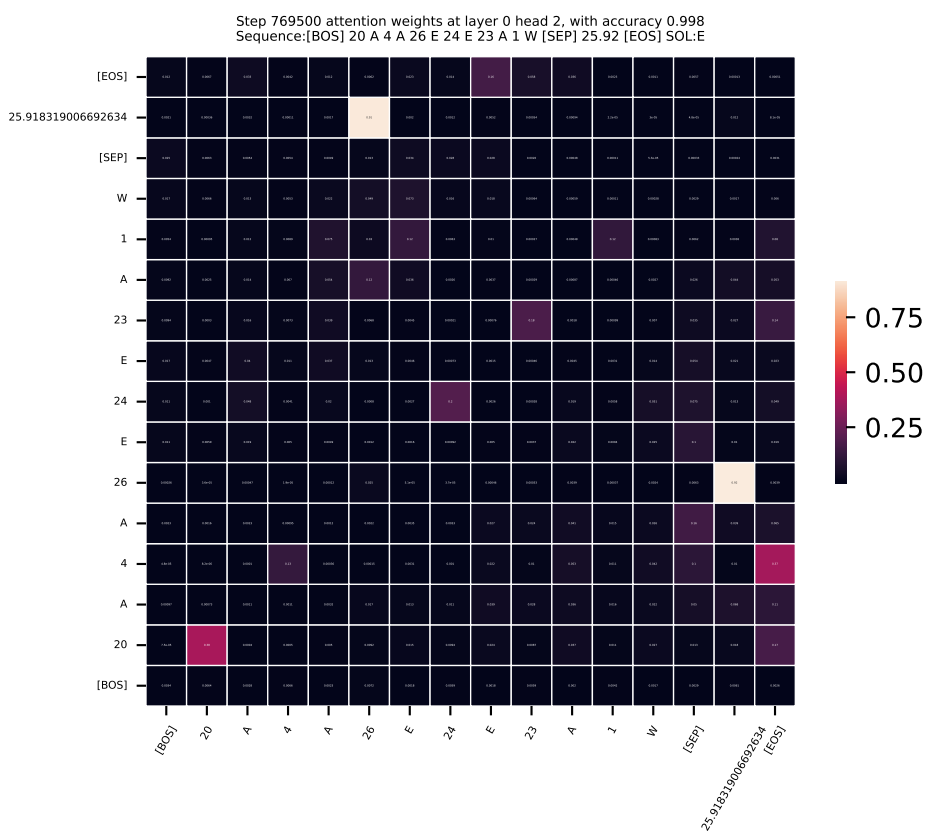

Step 760000 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

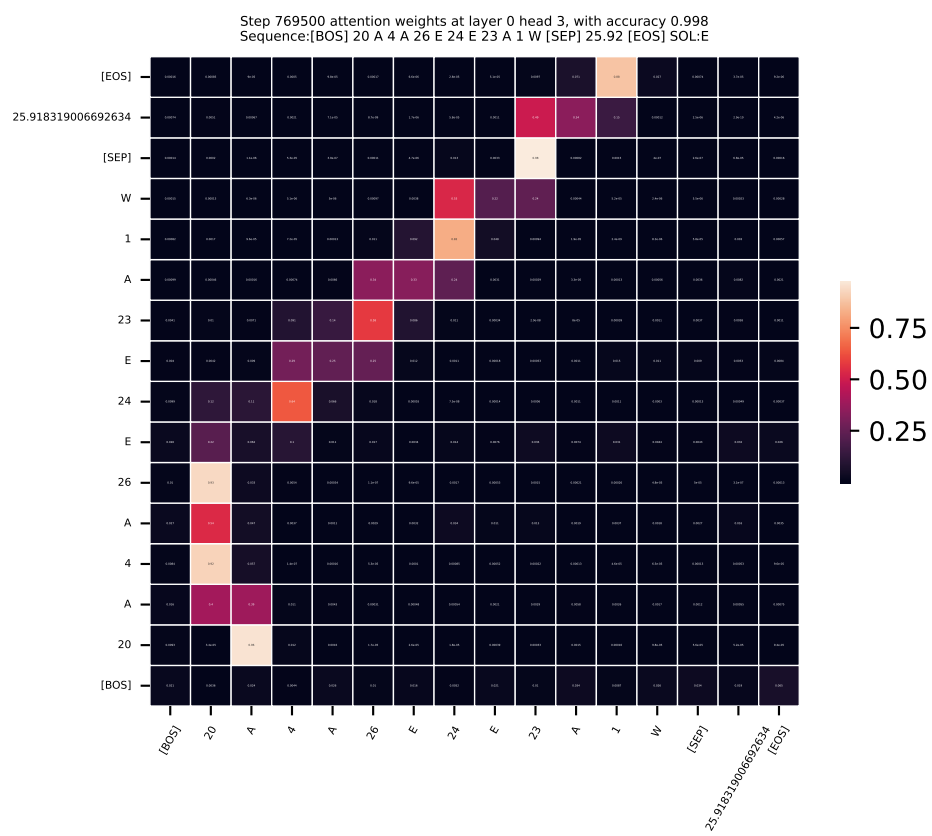

Step 760000 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

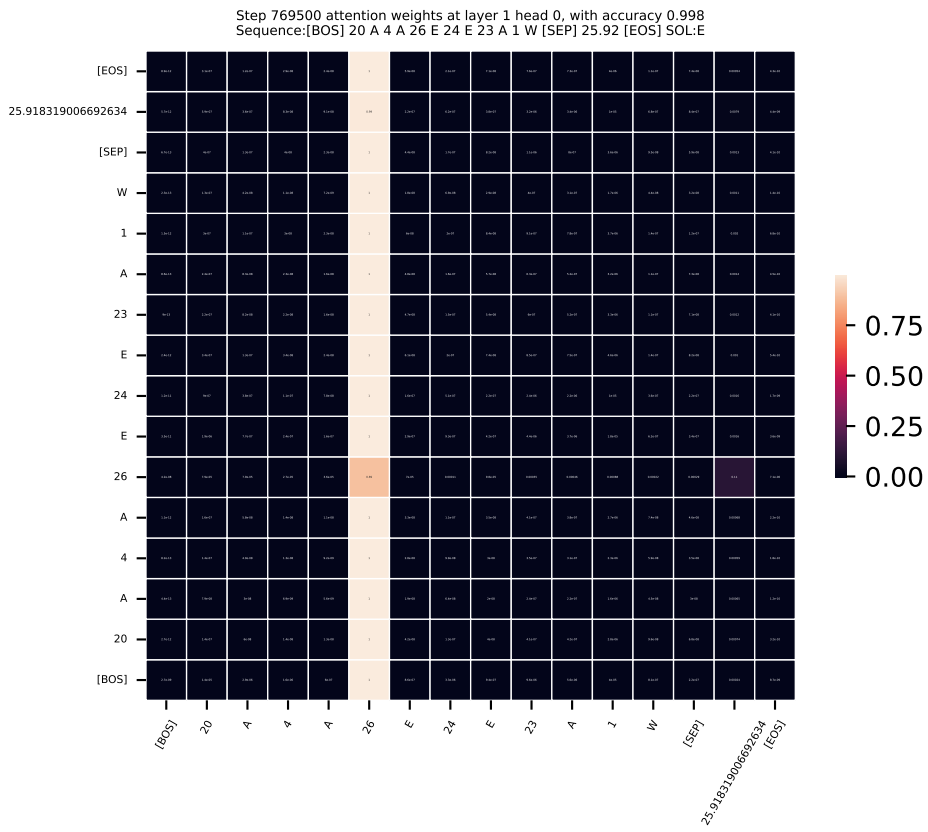

Step 760000 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

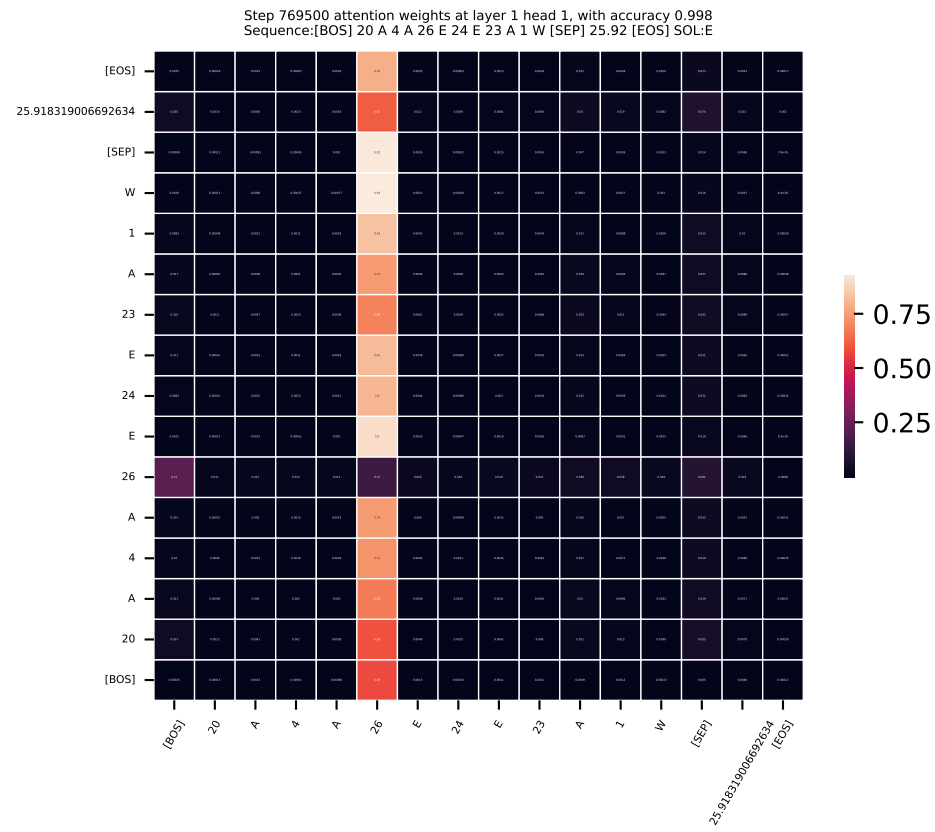

Step 769500 attention weights at layer 0 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.4
0.2

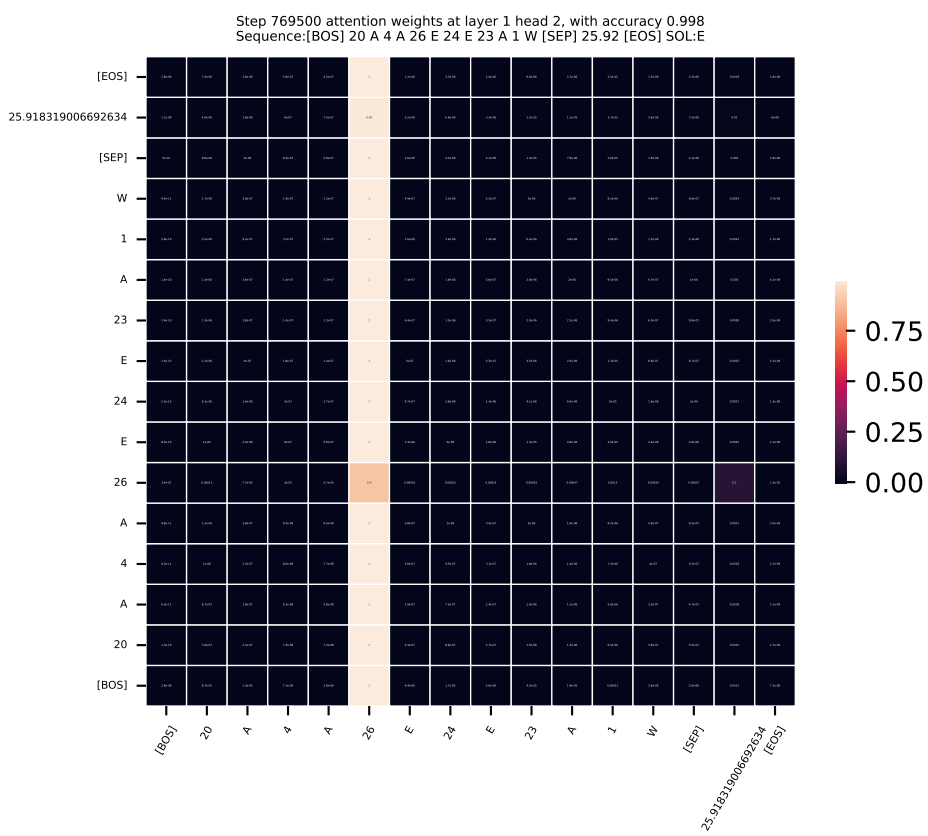

Step 769500 attention weights at layer 0 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

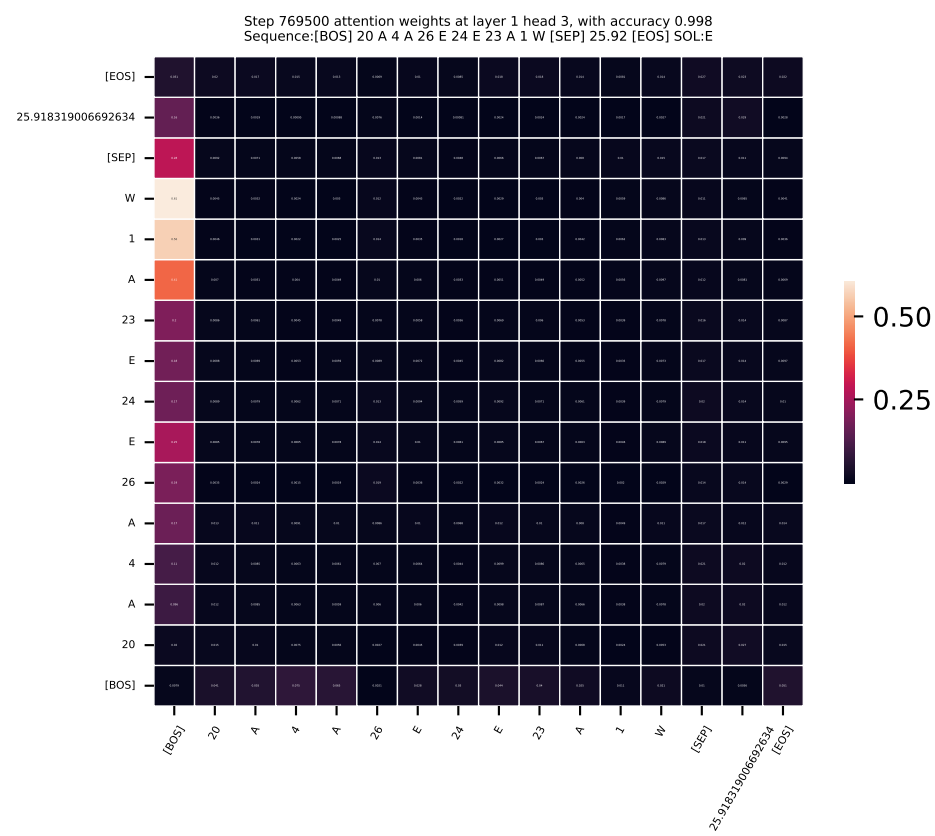

Step 769500 attention weights at layer 0 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

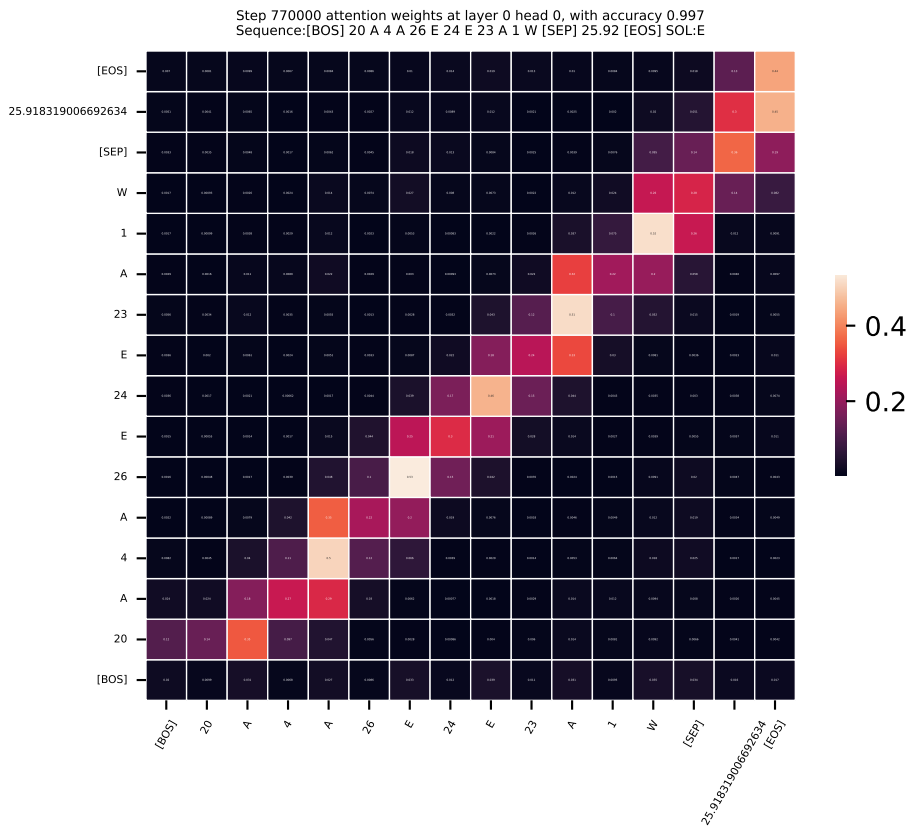

Step 769500 attention weights at layer 0 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

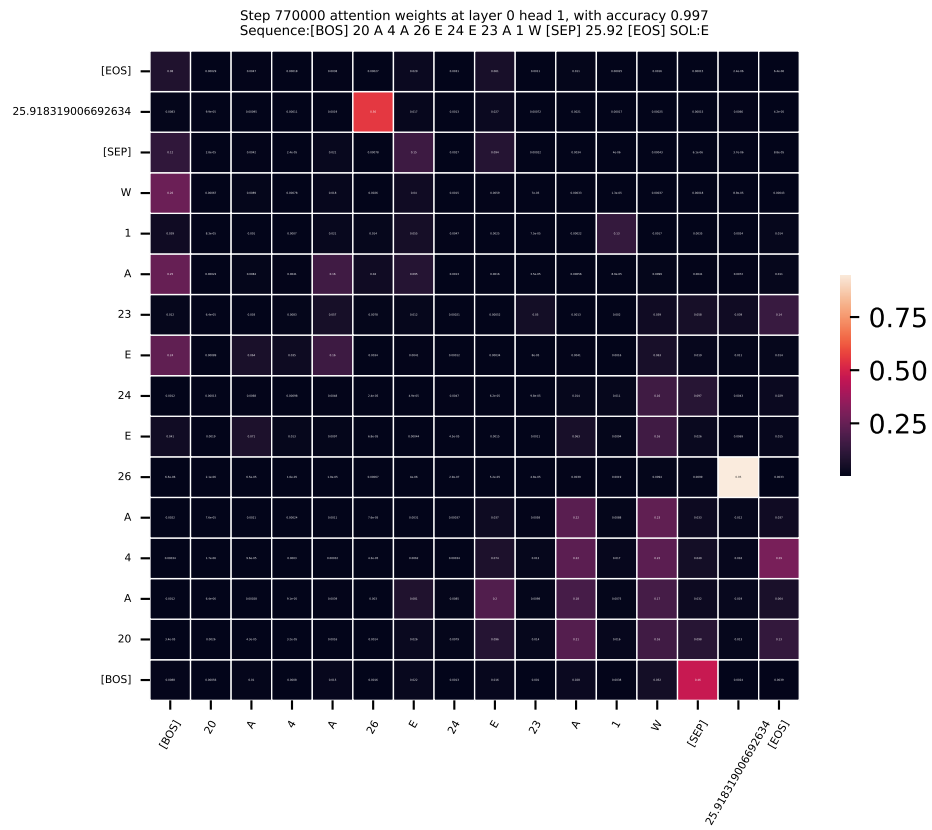

Step 769500 attention weights at layer 1 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

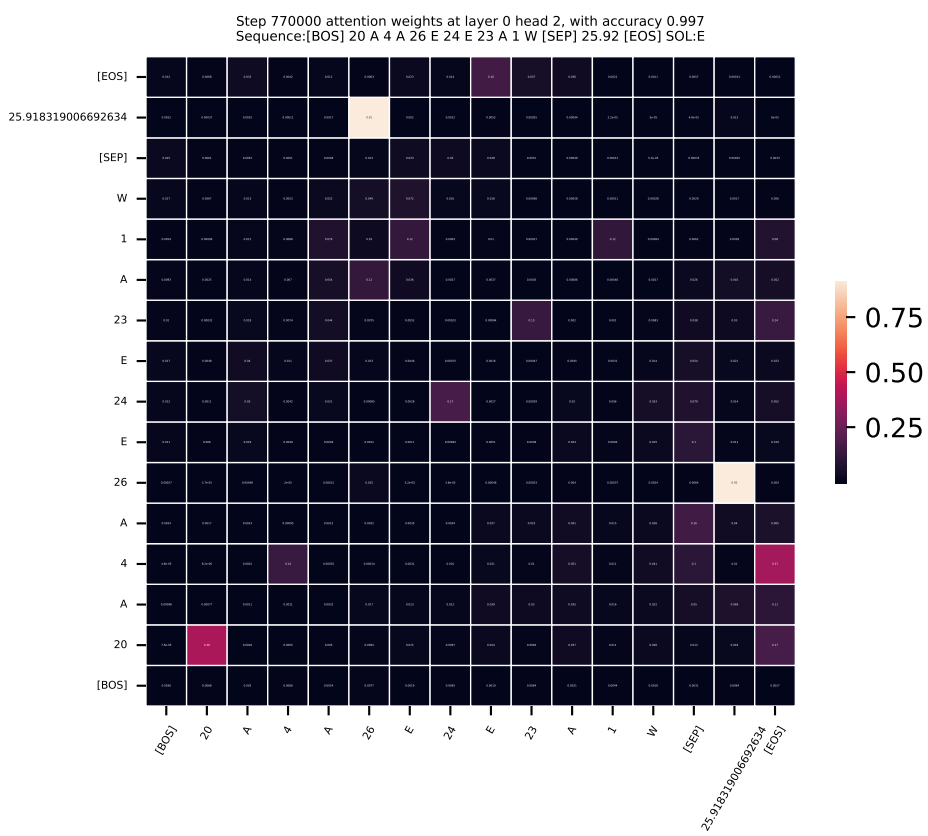

Step 769500 attention weights at layer 1 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

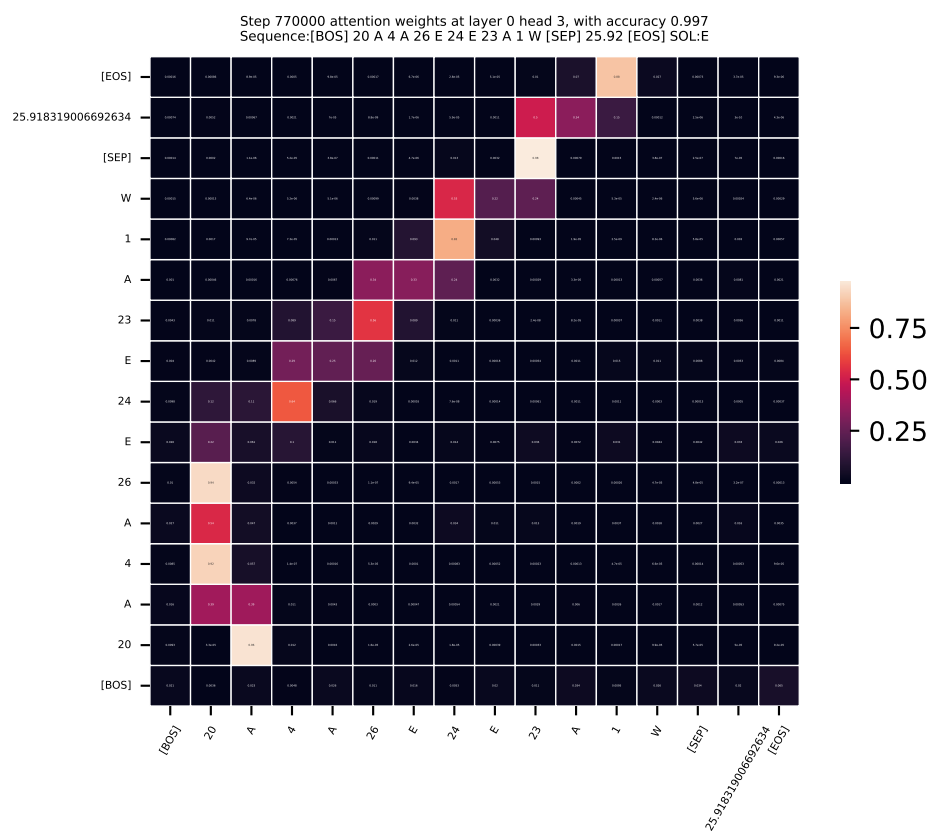

Step 769500 attention weights at layer 1 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

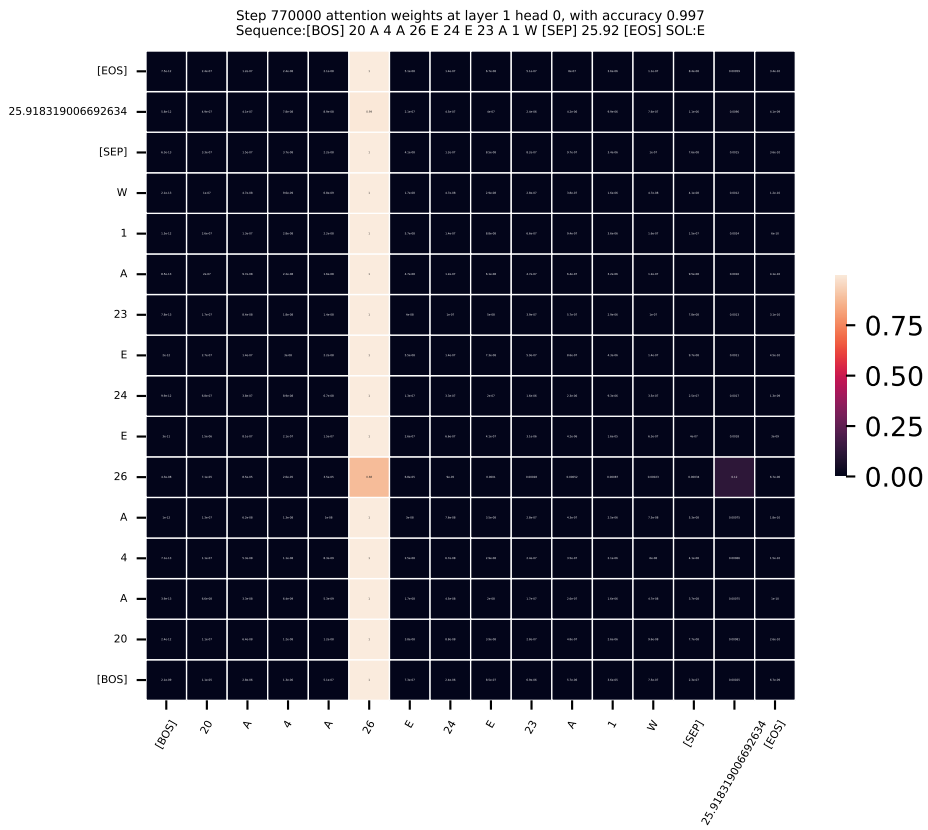

Step 769500 attention weights at layer 1 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

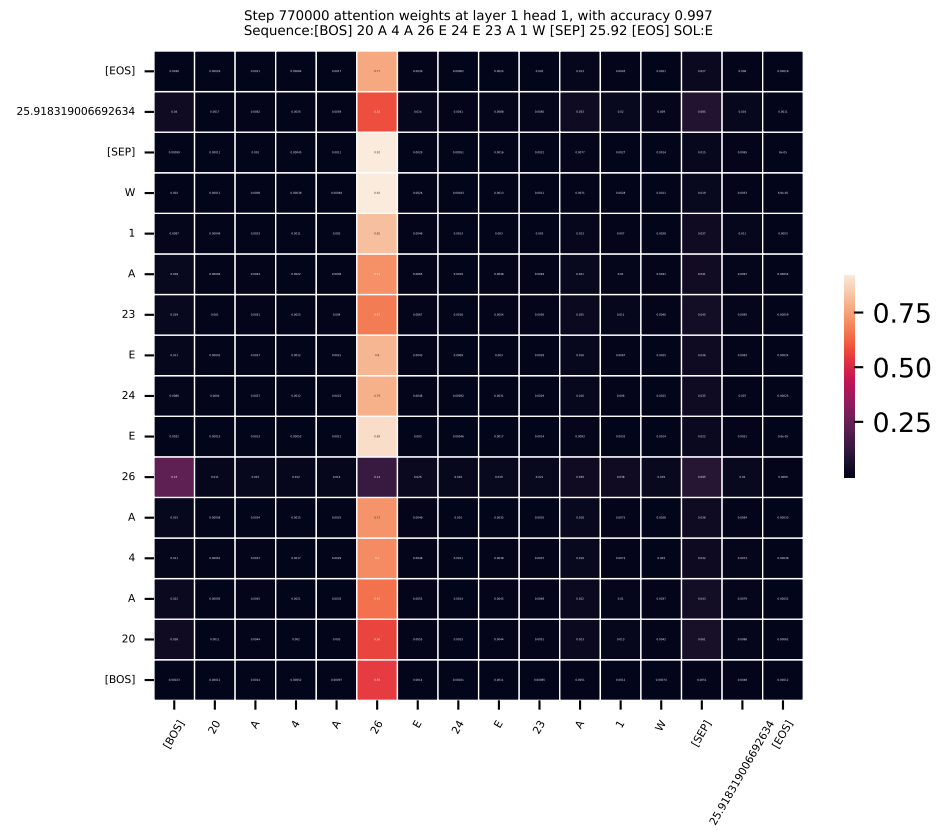

Step 770000 attention weights at layer 0 head 0, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

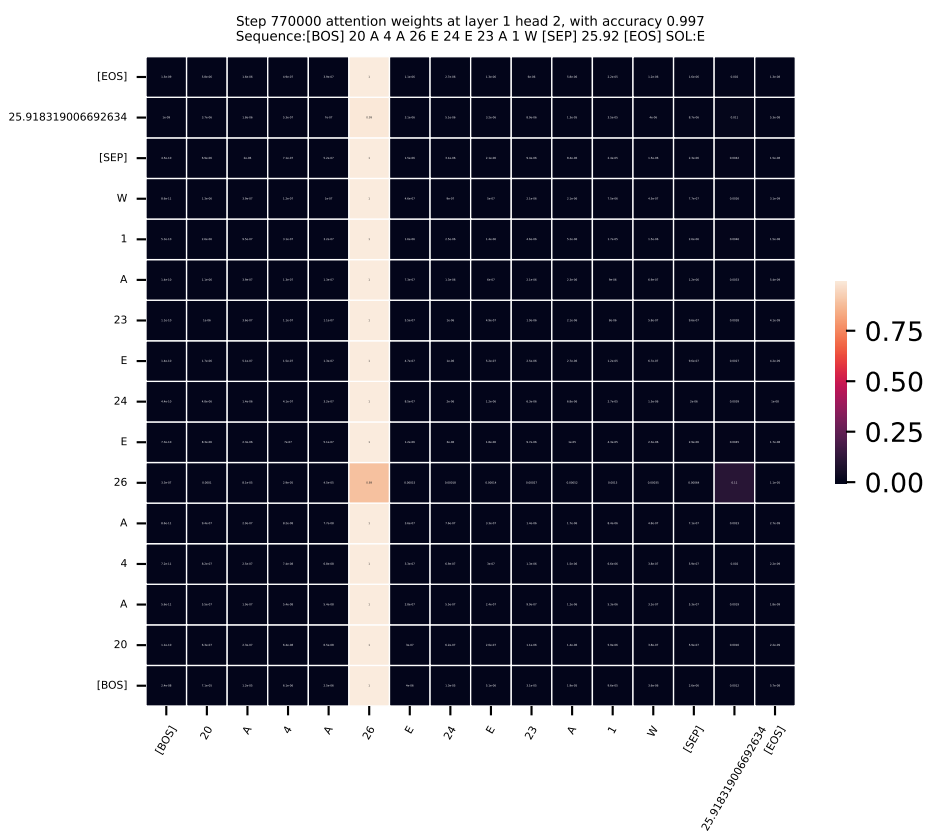

Step 770000 attention weights at layer 0 head 1, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

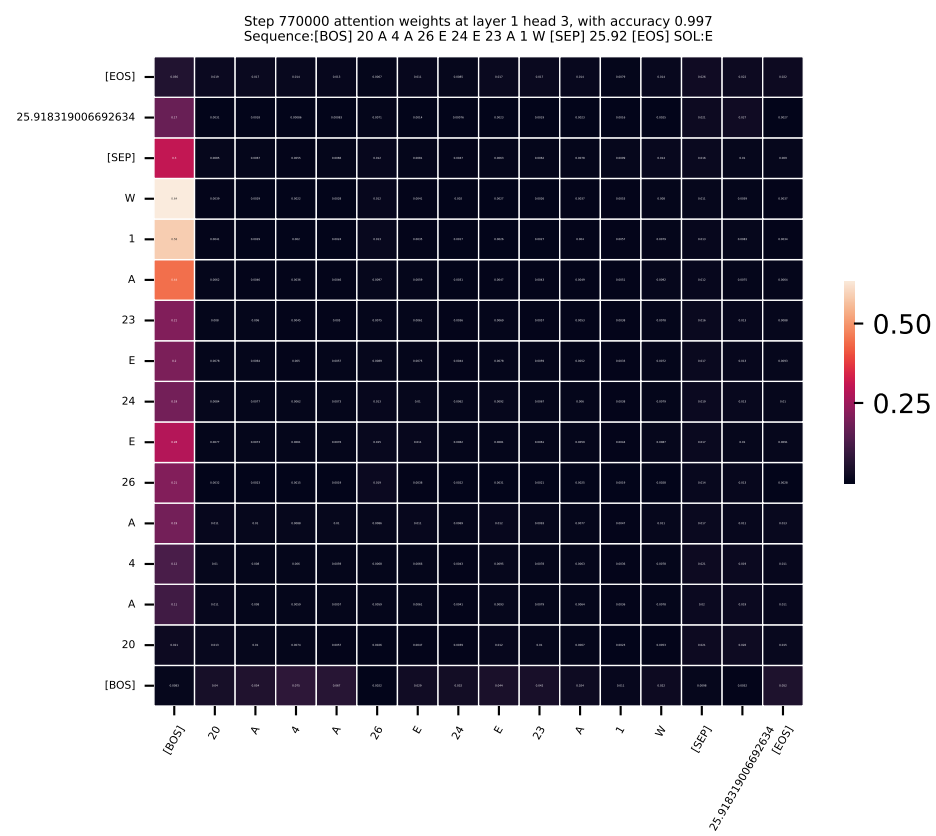

Step 770000 attention weights at layer 0 head 2, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

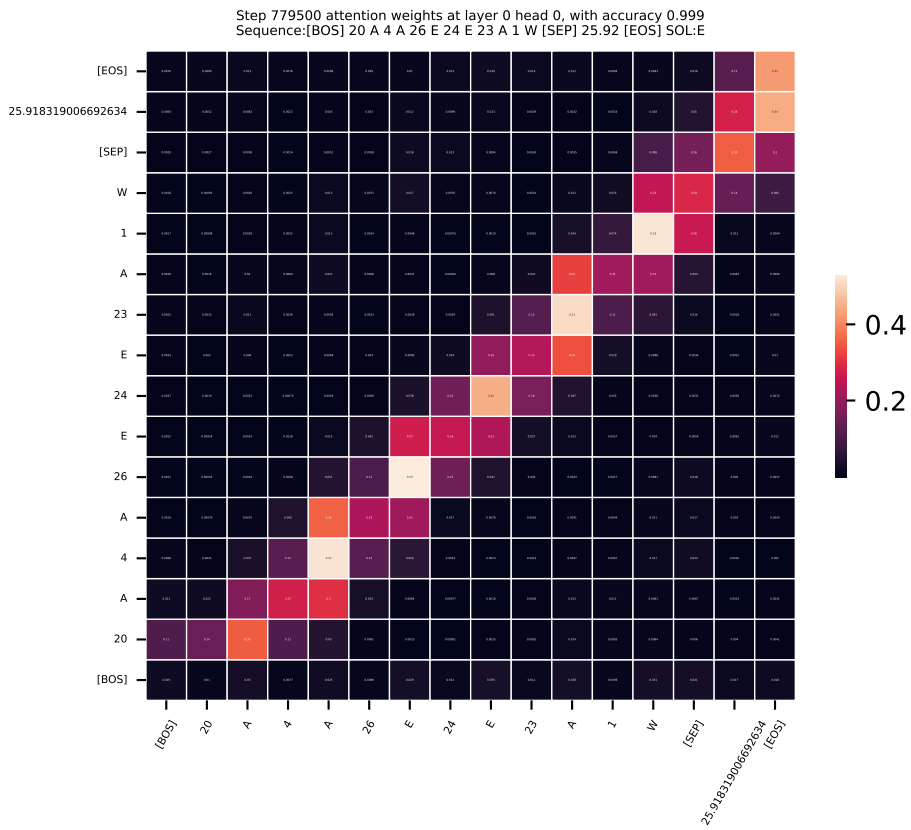

Step 770000 attention weights at layer 0 head 3, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

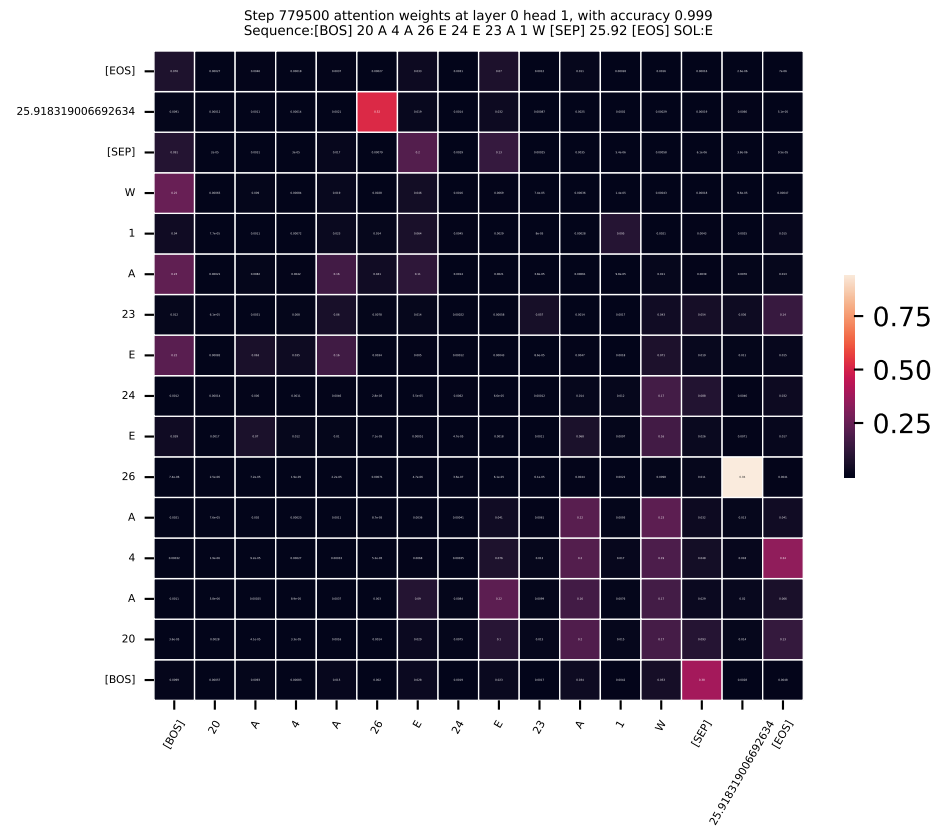

Step 770000 attention weights at layer 1 head 0, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

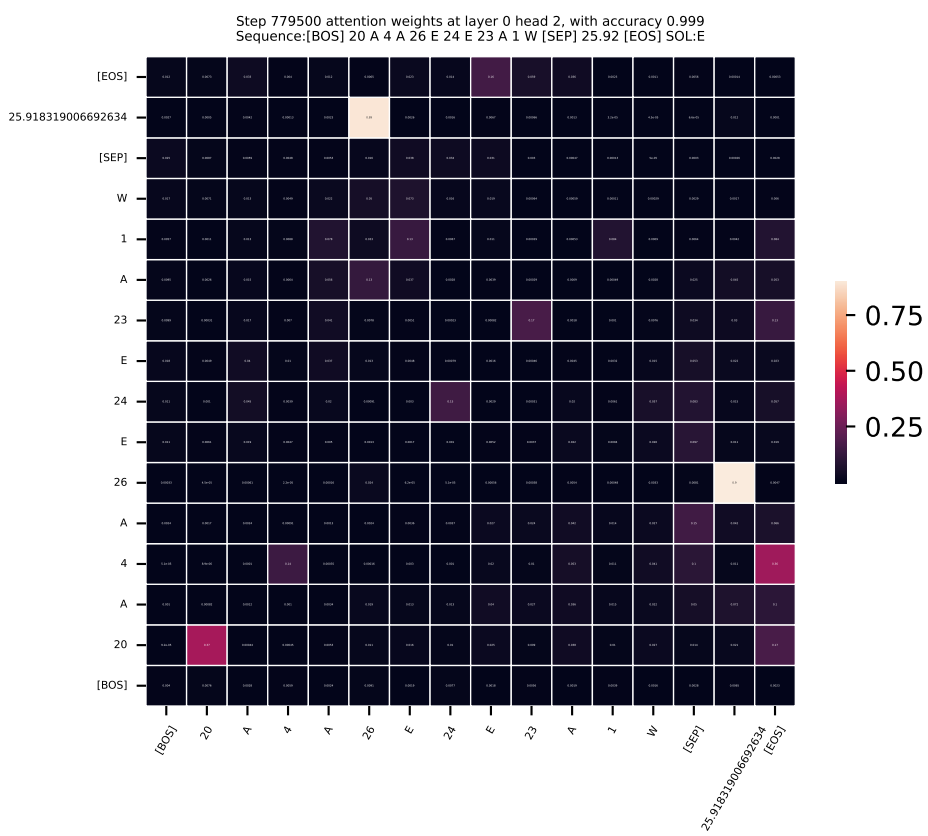

Step 770000 attention weights at layer 1 head 1, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

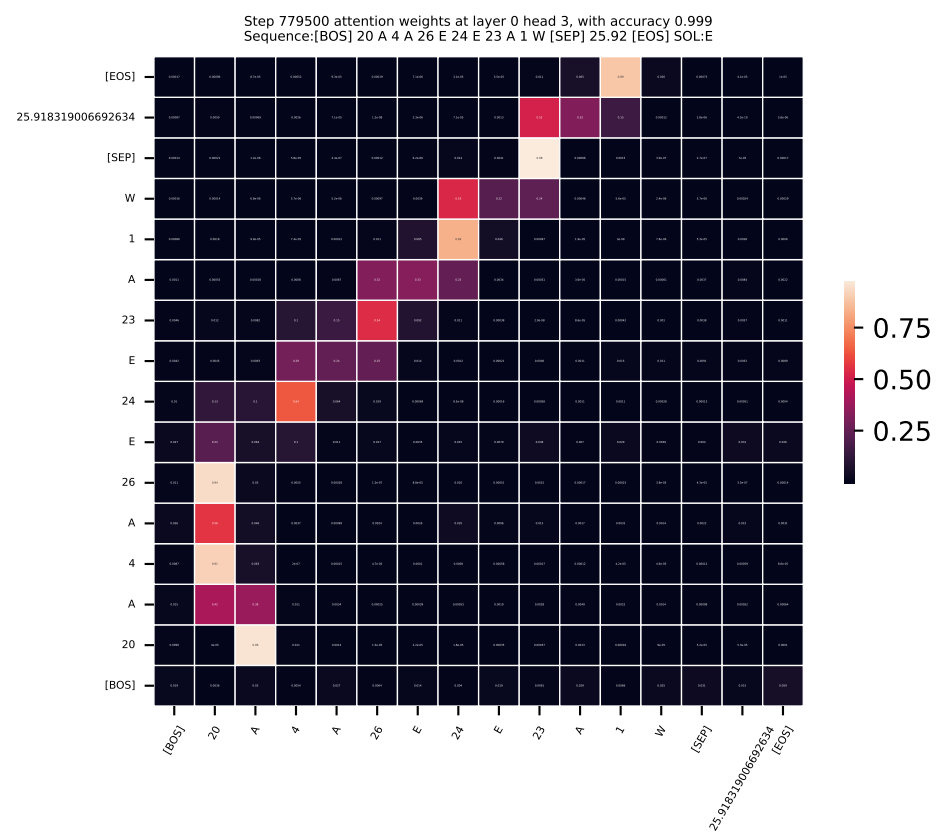

Step 770000 attention weights at layer 1 head 2, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

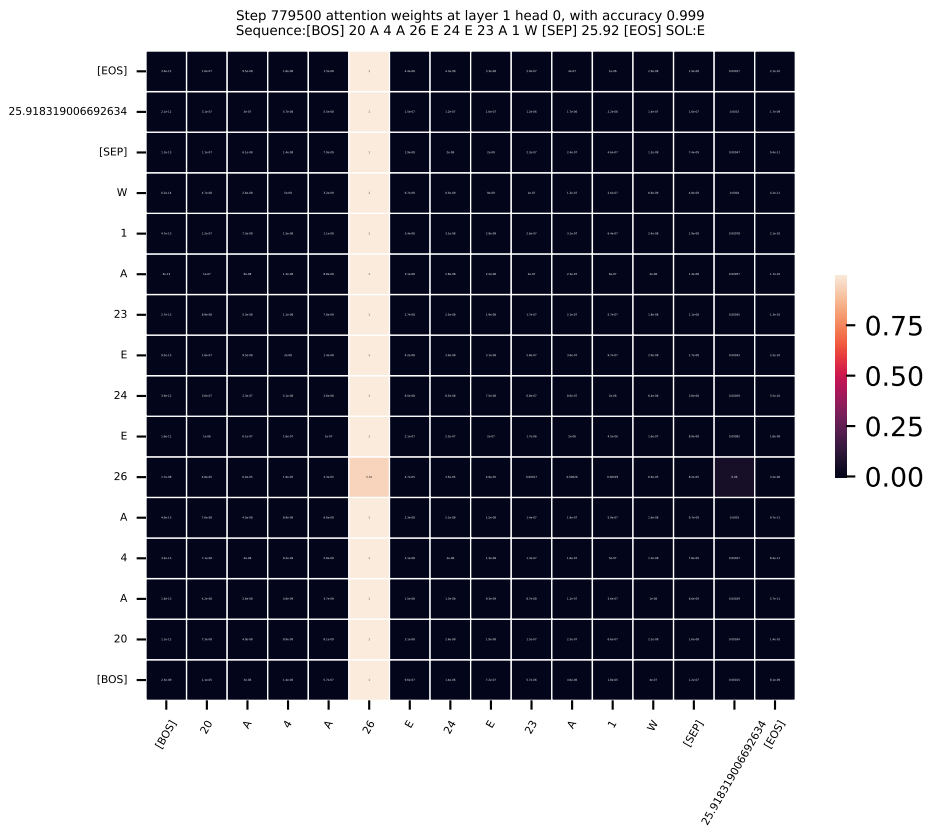

Step 770000 attention weights at layer 1 head 3, with accuracy 0.997
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

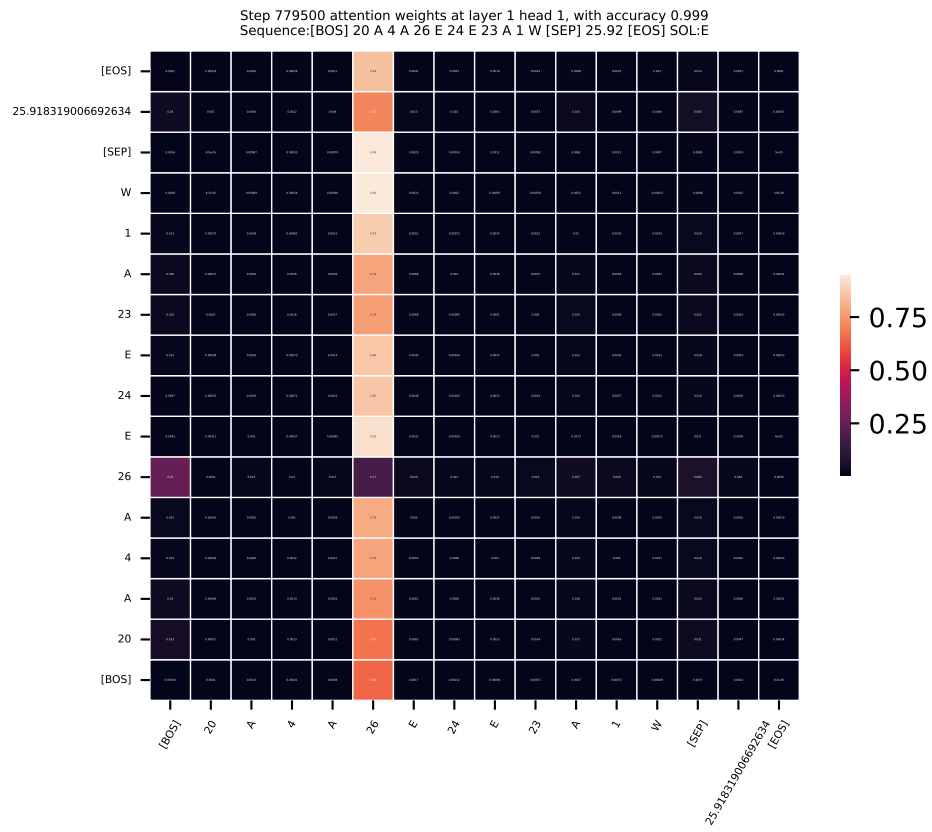

Step 779500 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

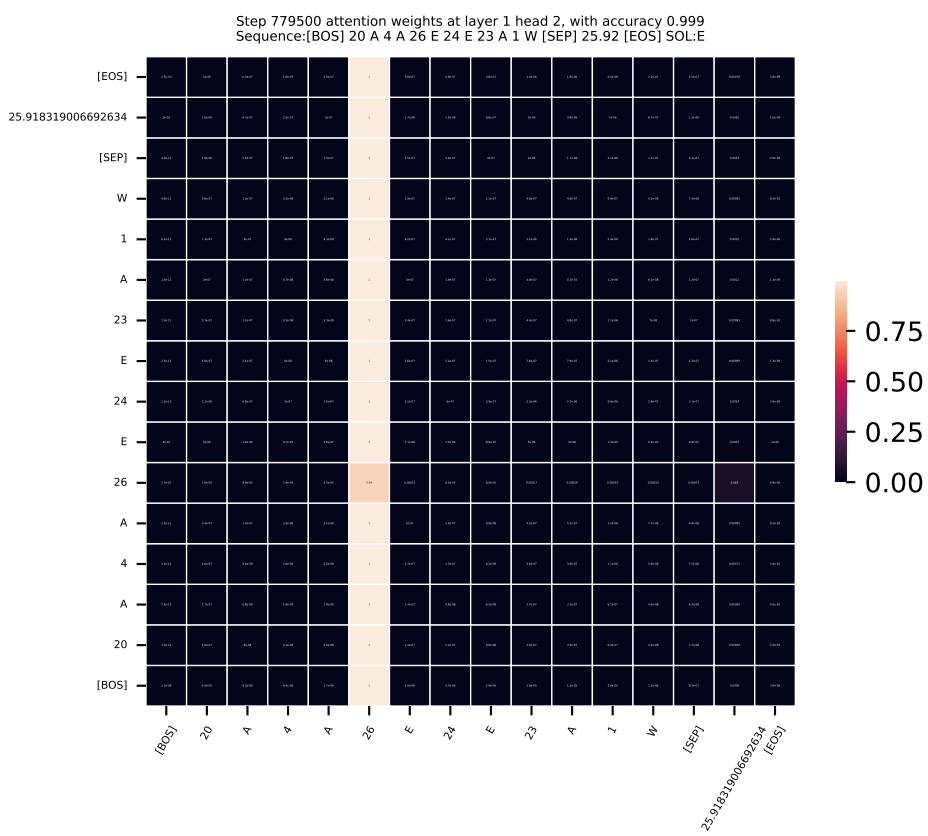

Step 779500 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

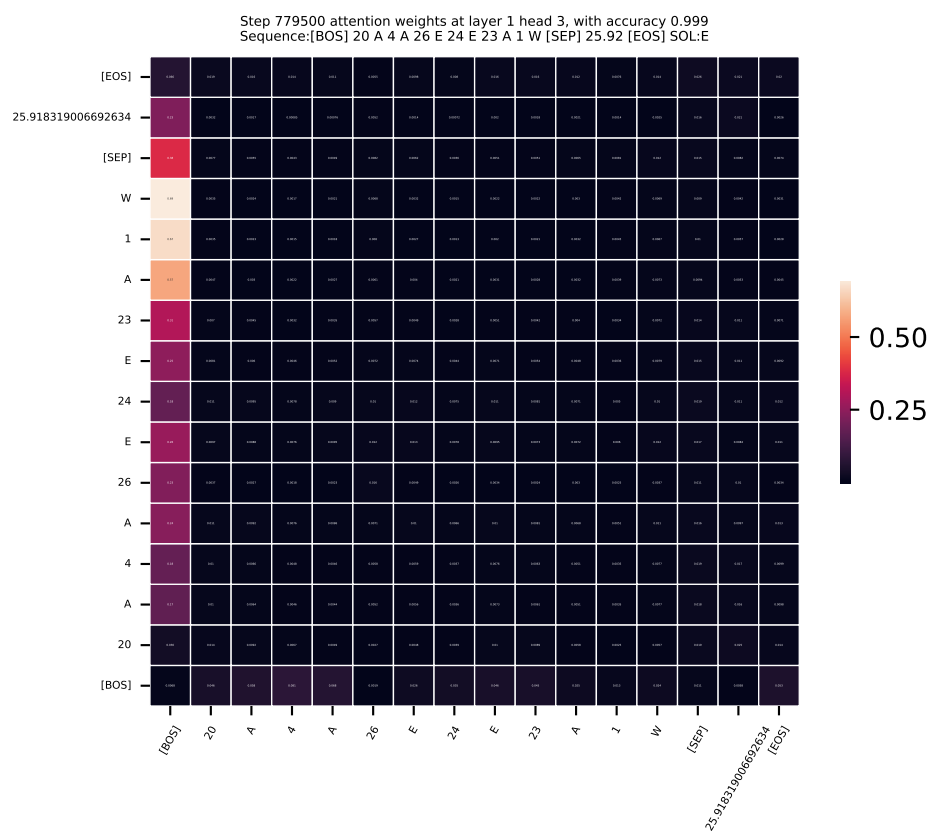

Step 779500 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

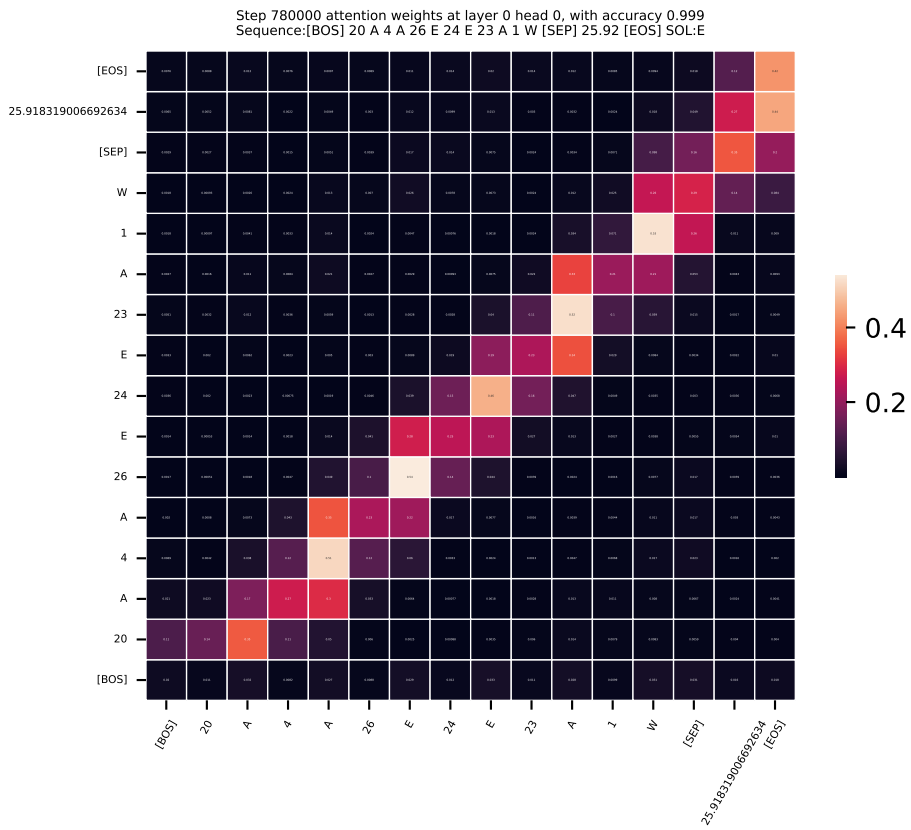

Step 779500 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

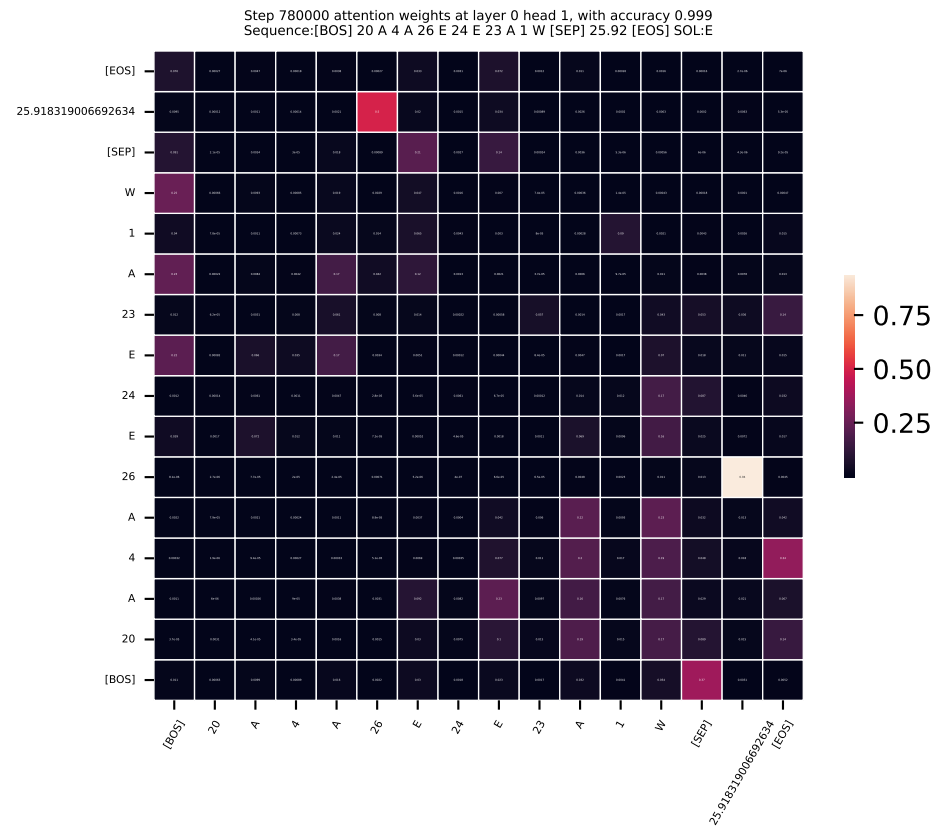

Step 779500 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

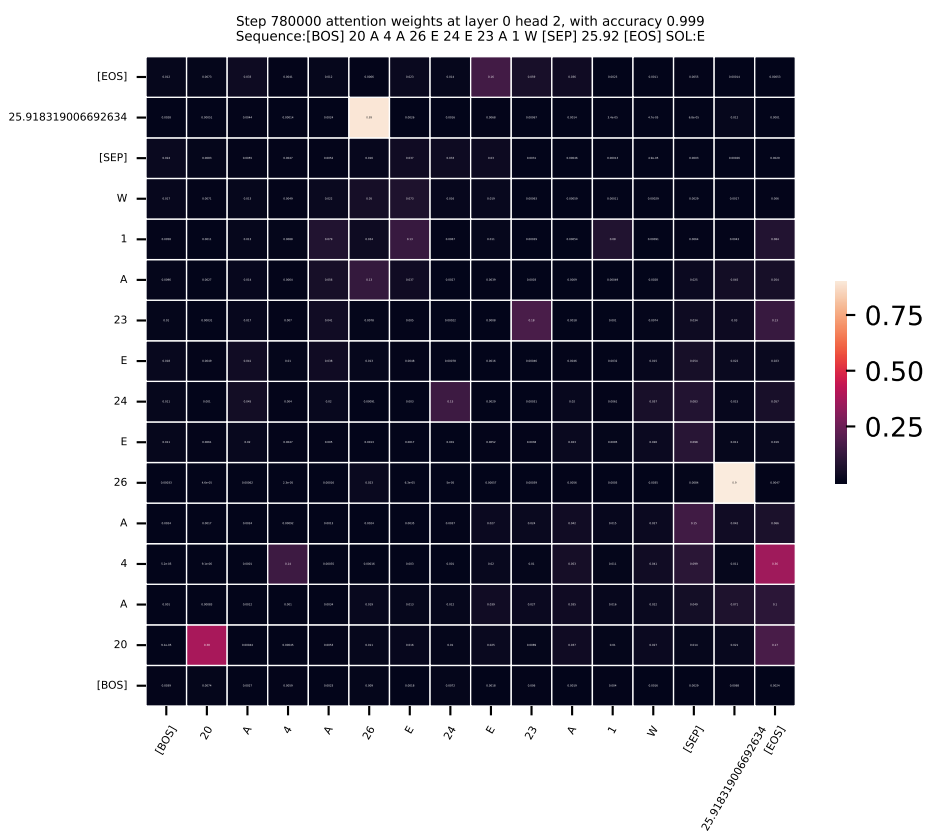

Step 779500 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

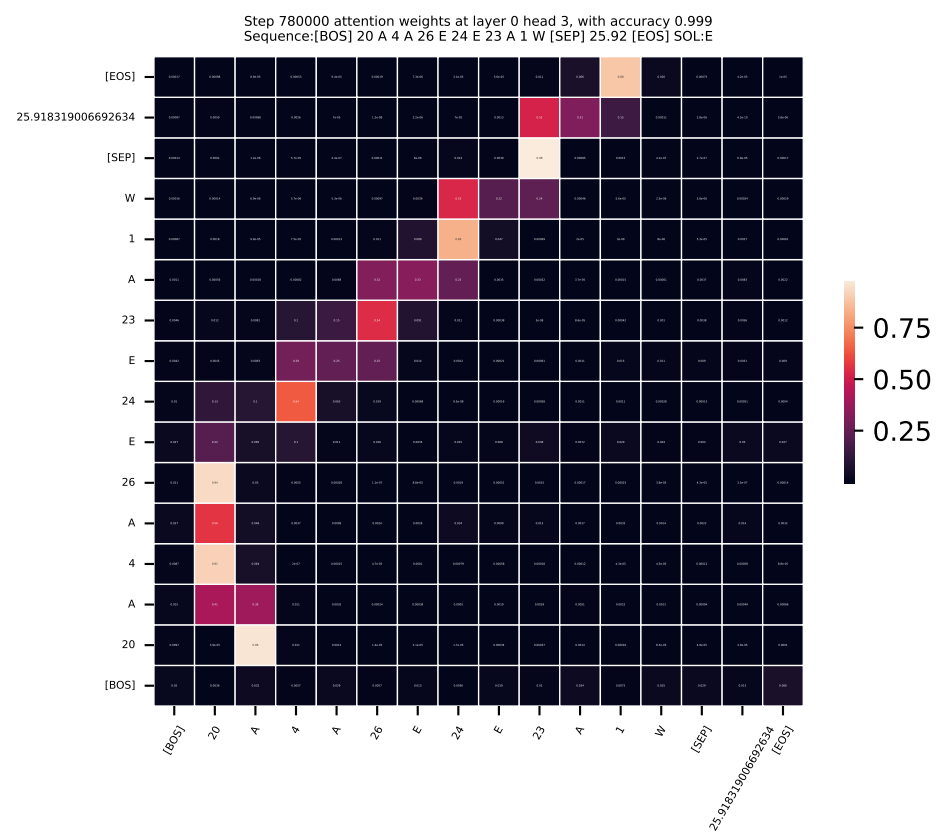

Step 779500 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

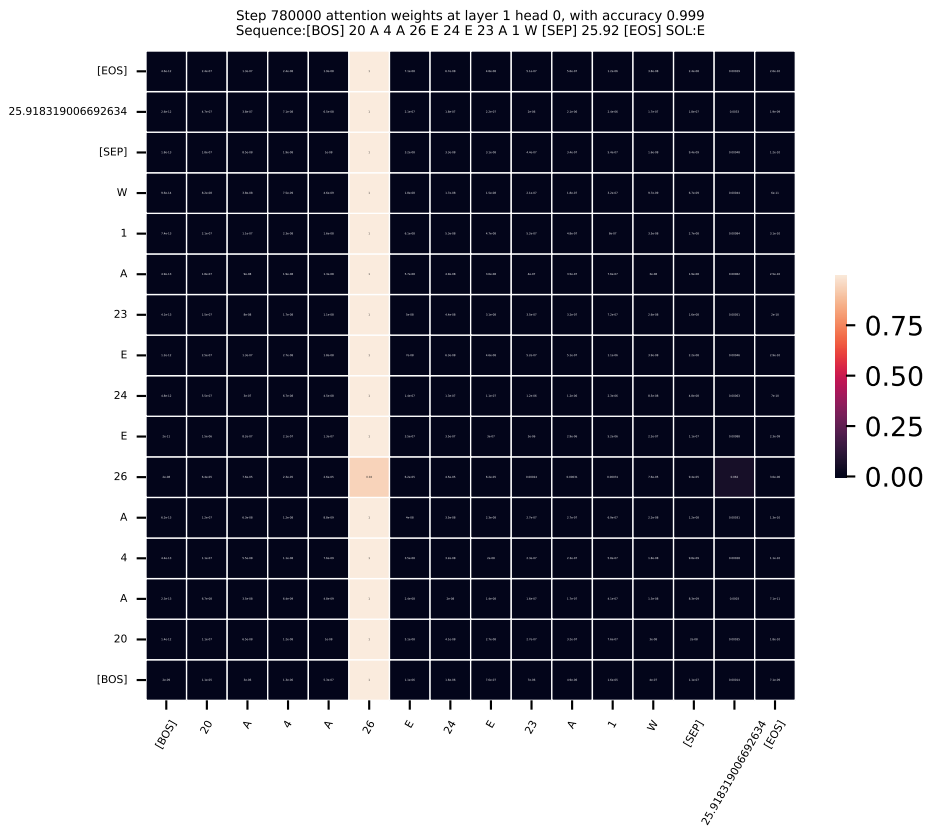

Step 779500 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

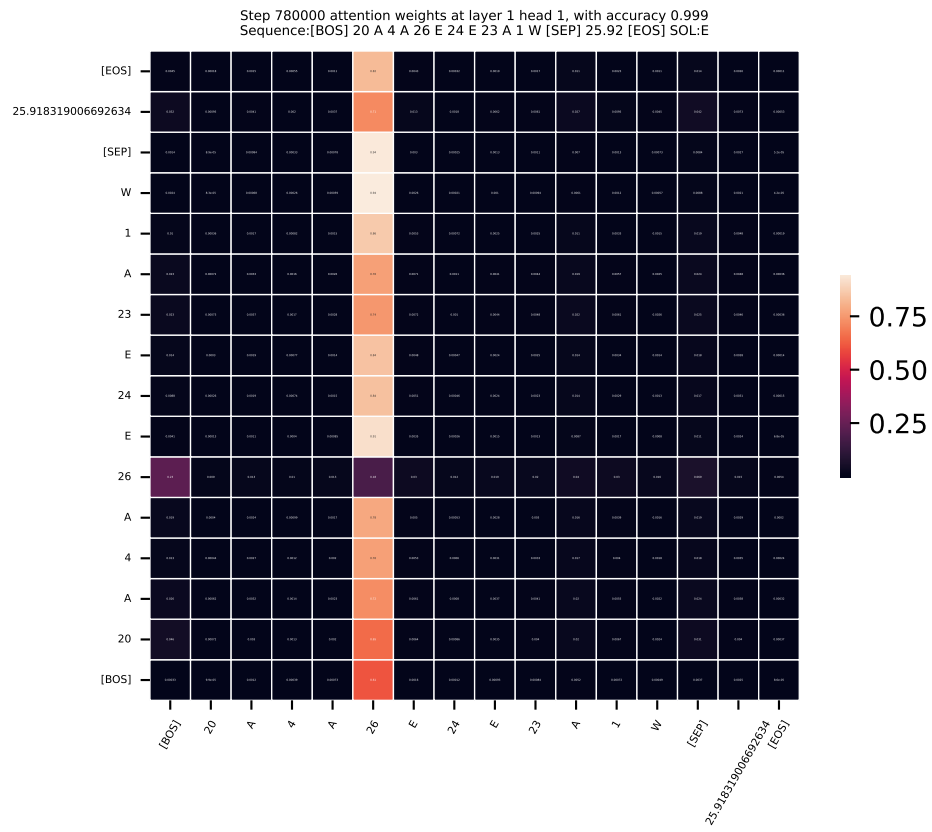

Step 780000 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

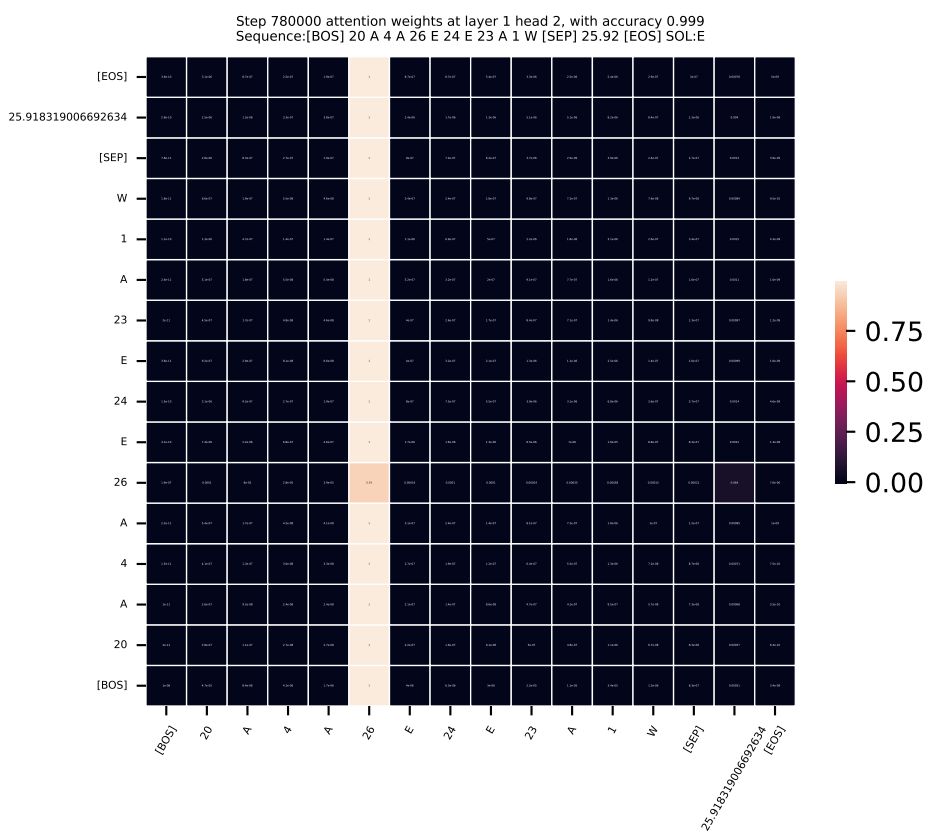

Step 780000 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

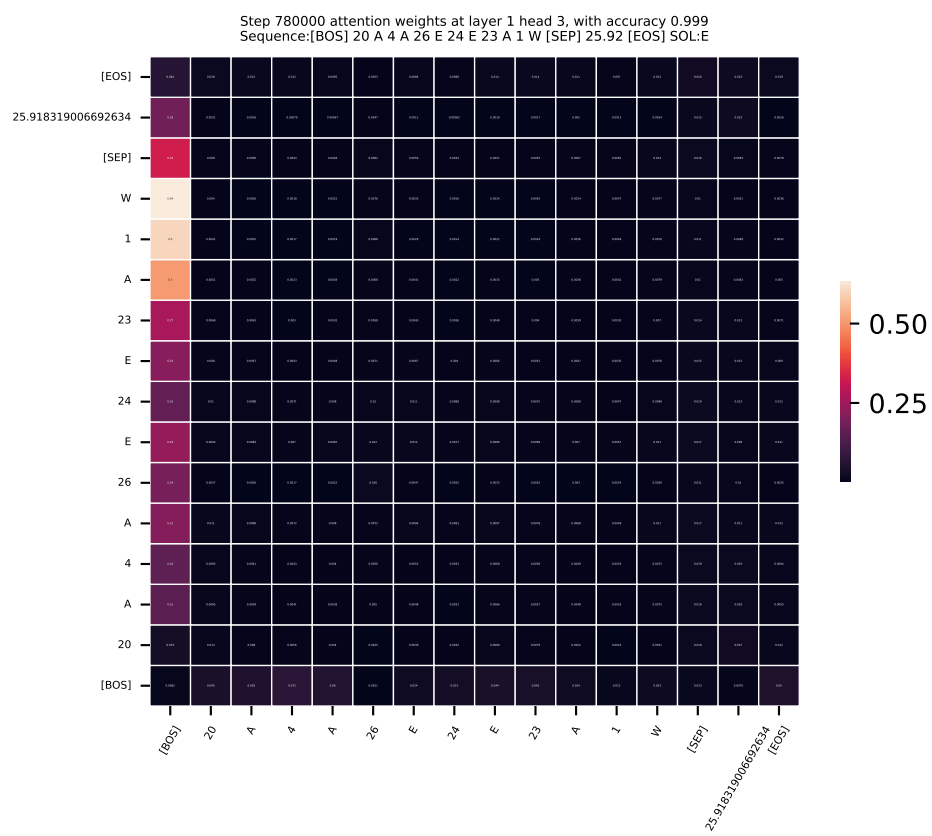

Step 780000 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

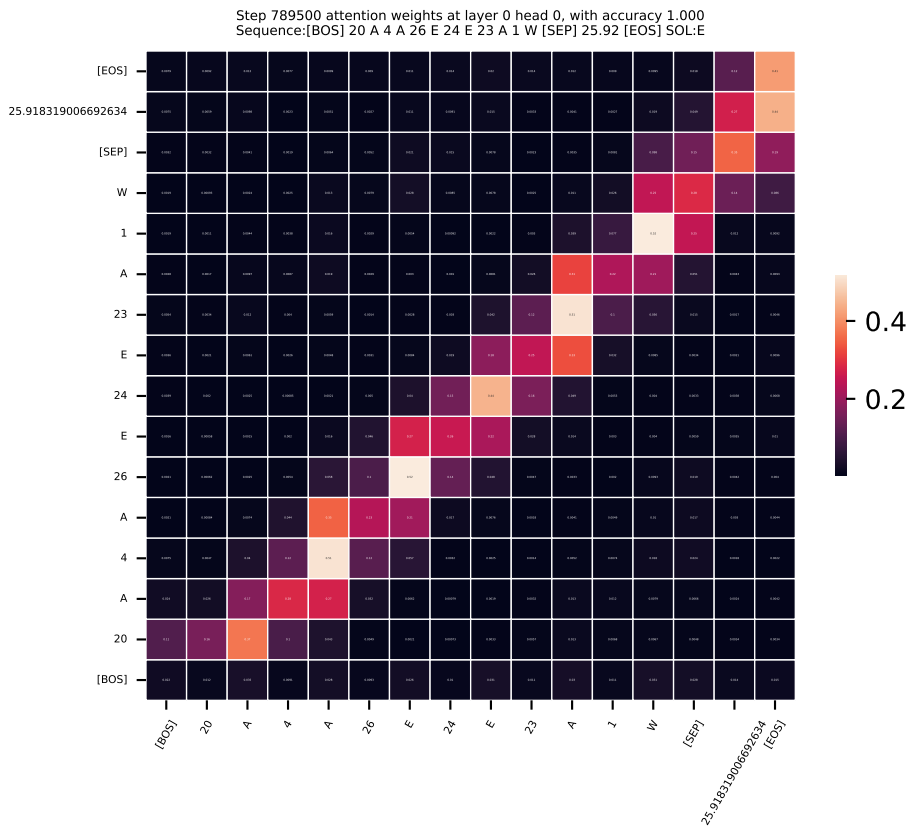

Step 780000 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

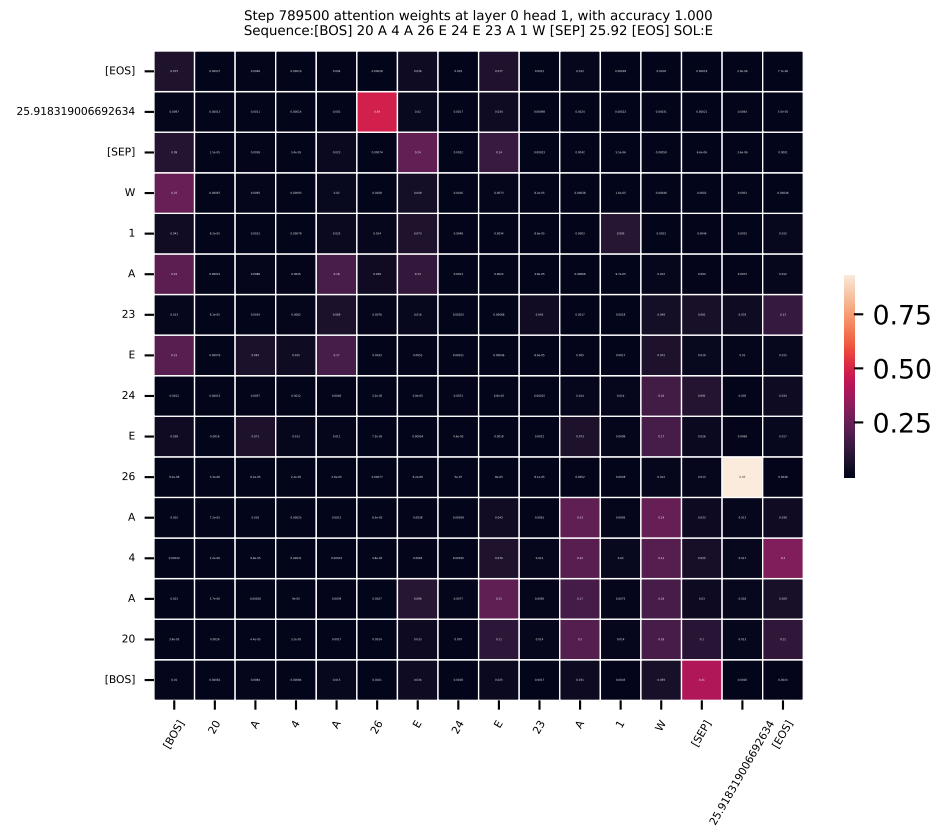

Step 780000 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25
0.00

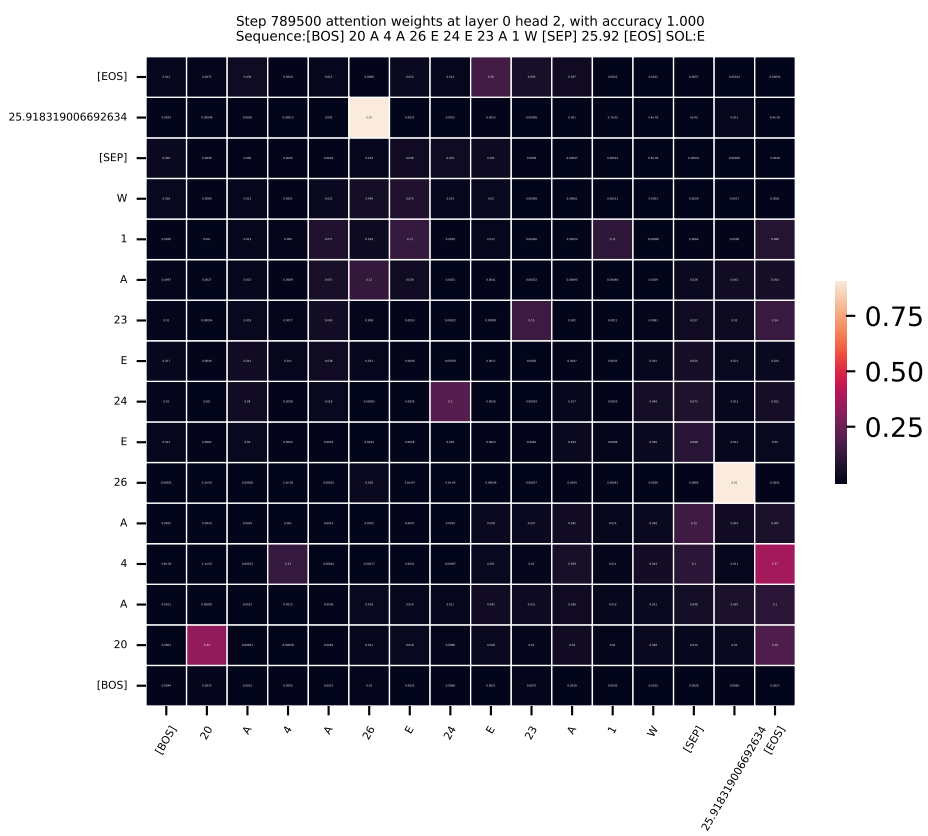

Step 780000 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

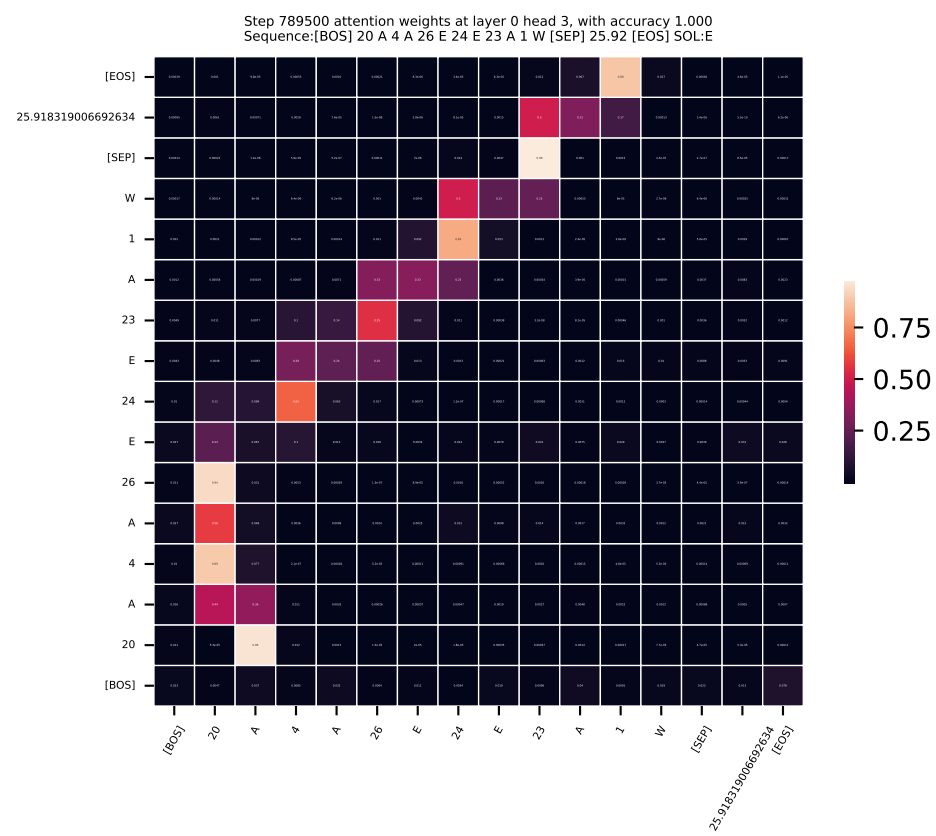

Step 780000 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25
0.00

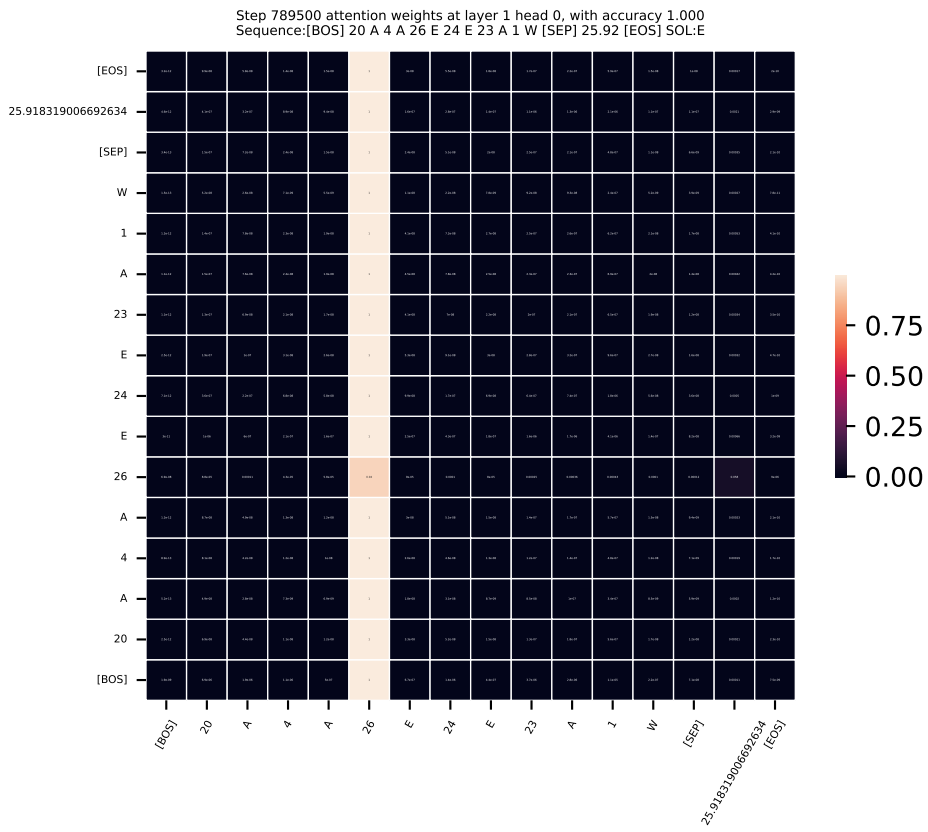

Step 780000 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.50
0.25

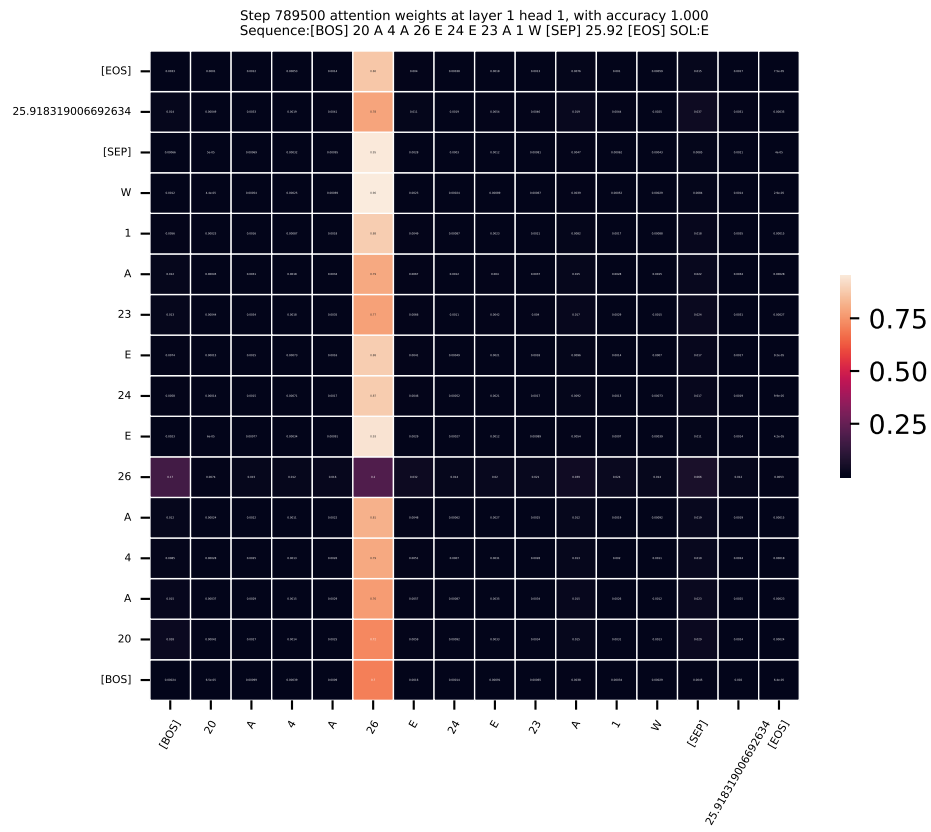

Step 789500 attention weights at layer 0 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

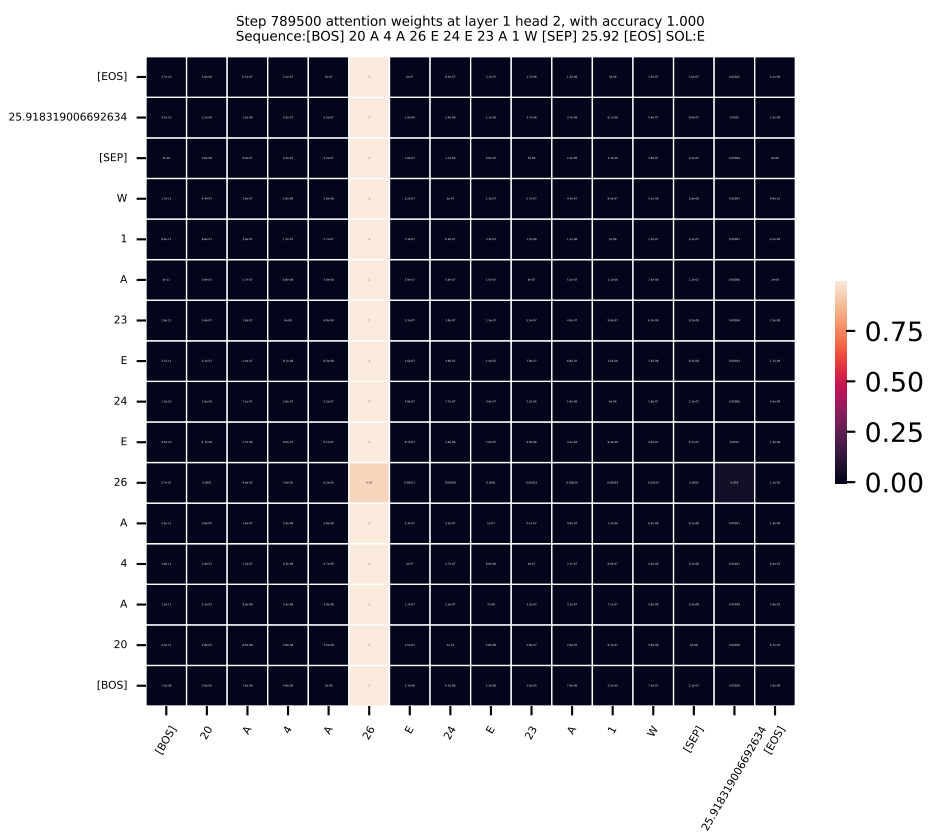

Step 789500 attention weights at layer 0 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

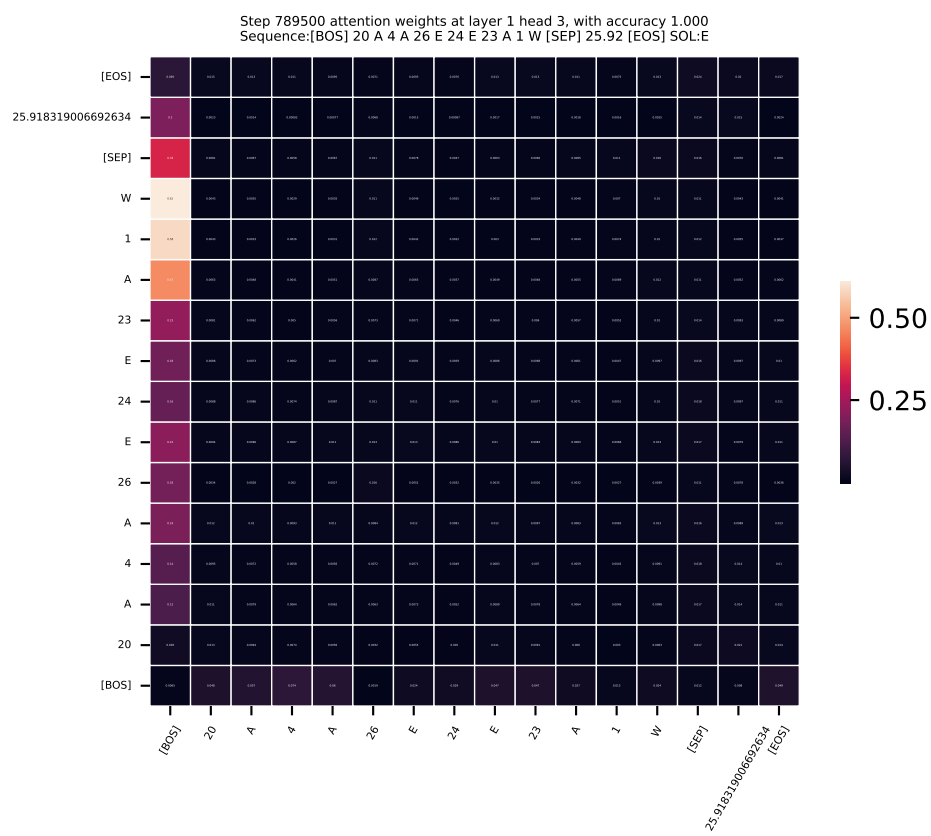

Step 789500 attention weights at layer 0 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

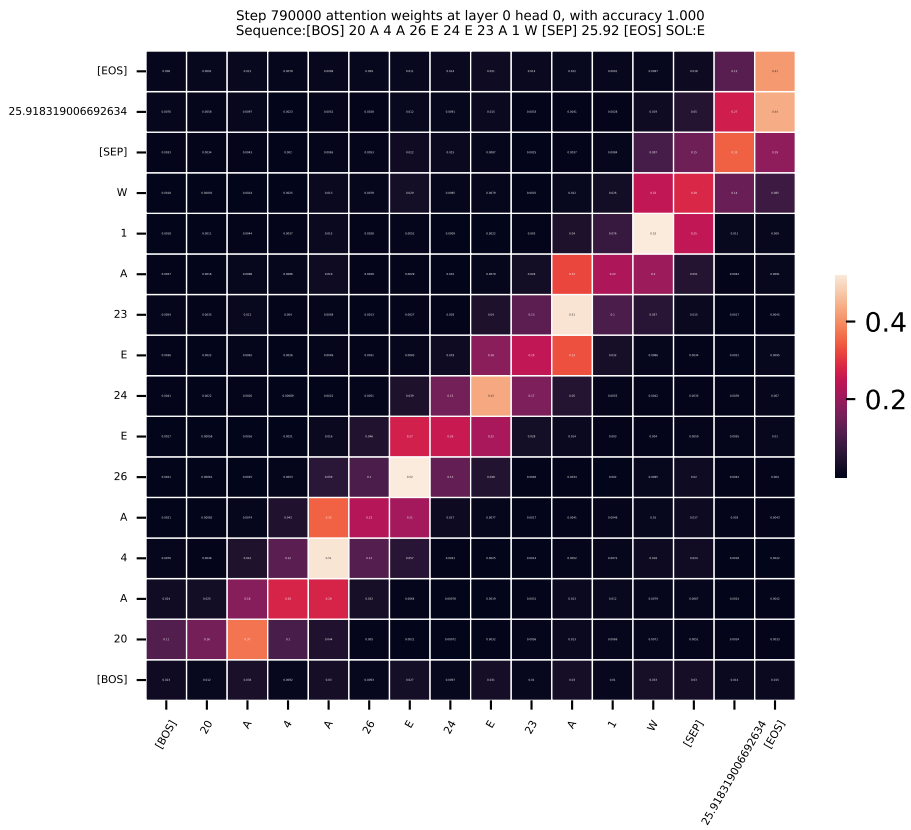

Step 789500 attention weights at layer 0 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

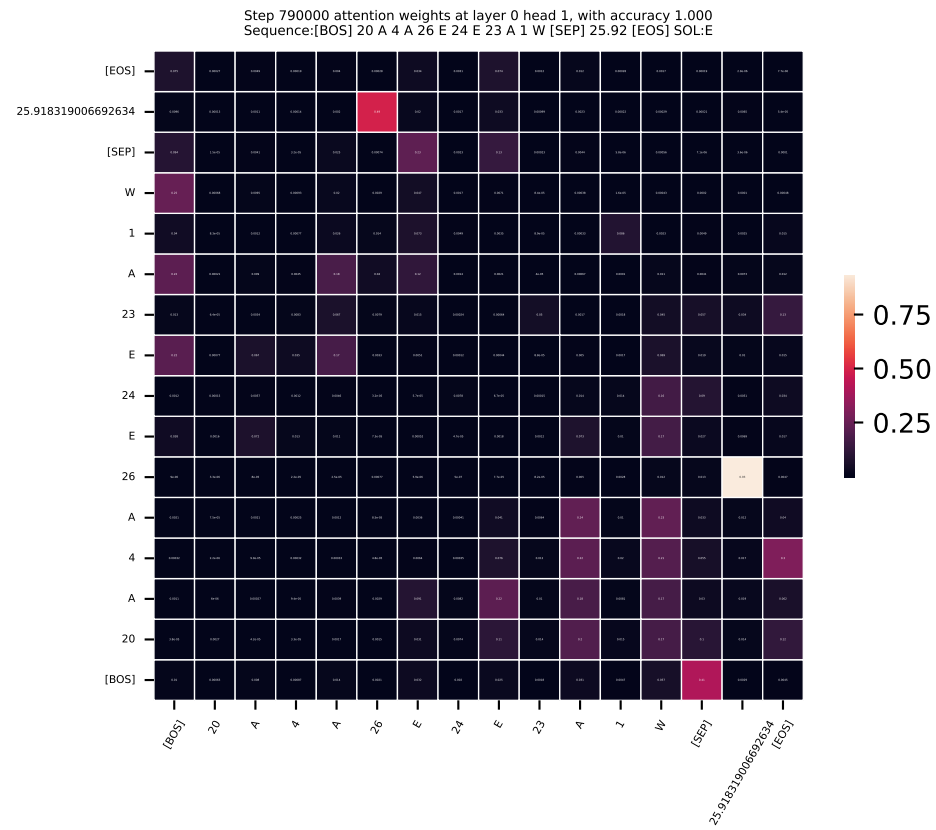

Step 789500 attention weights at layer 1 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

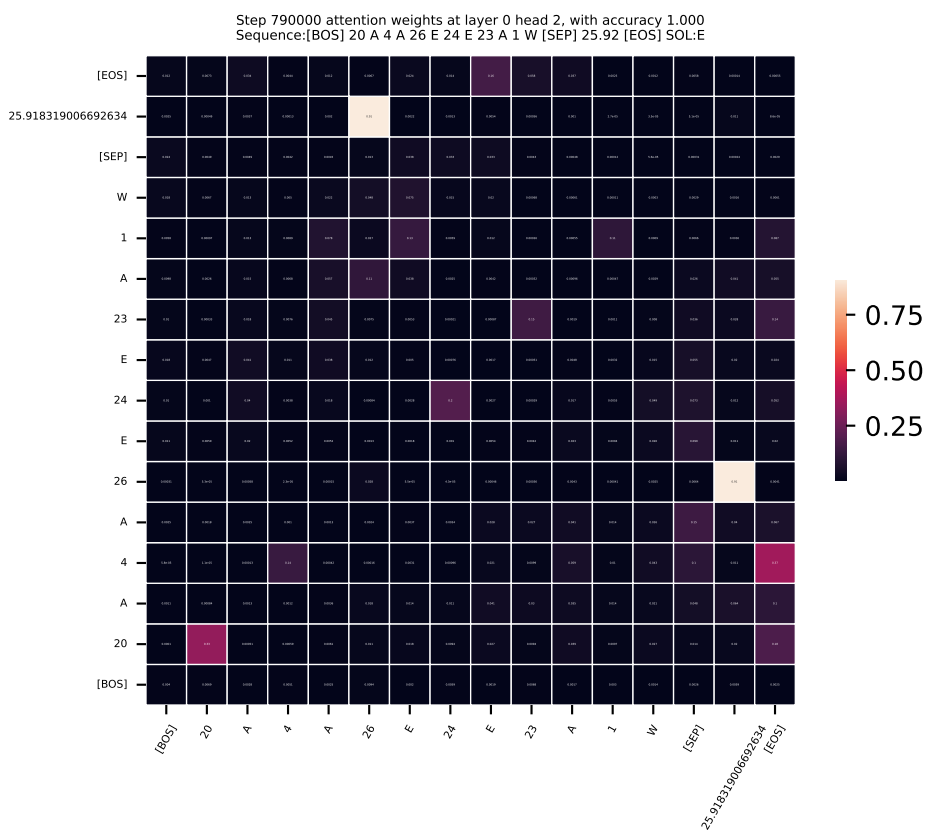

Step 789500 attention weights at layer 1 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

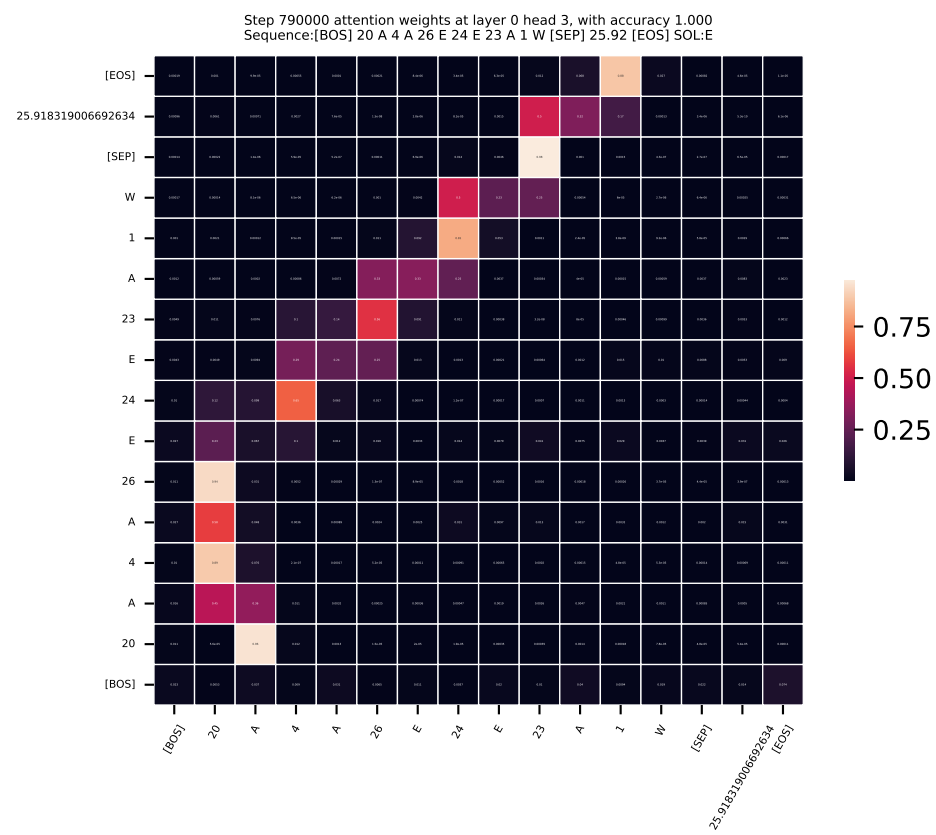

Step 789500 attention weights at layer 1 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

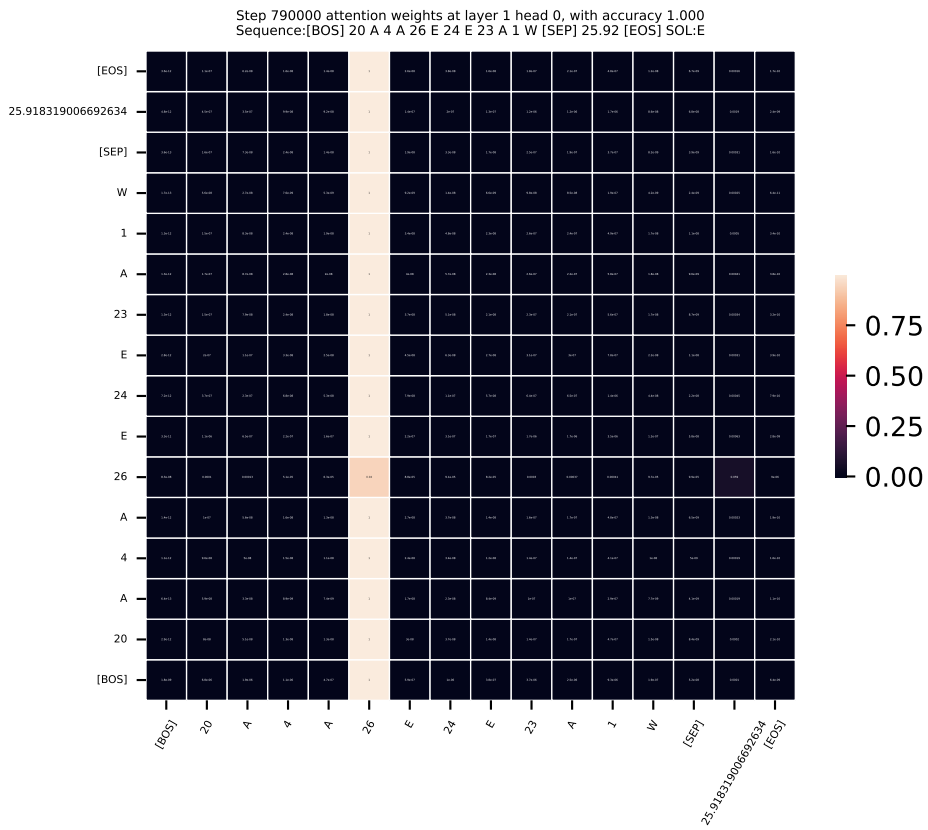

Step 789500 attention weights at layer 1 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

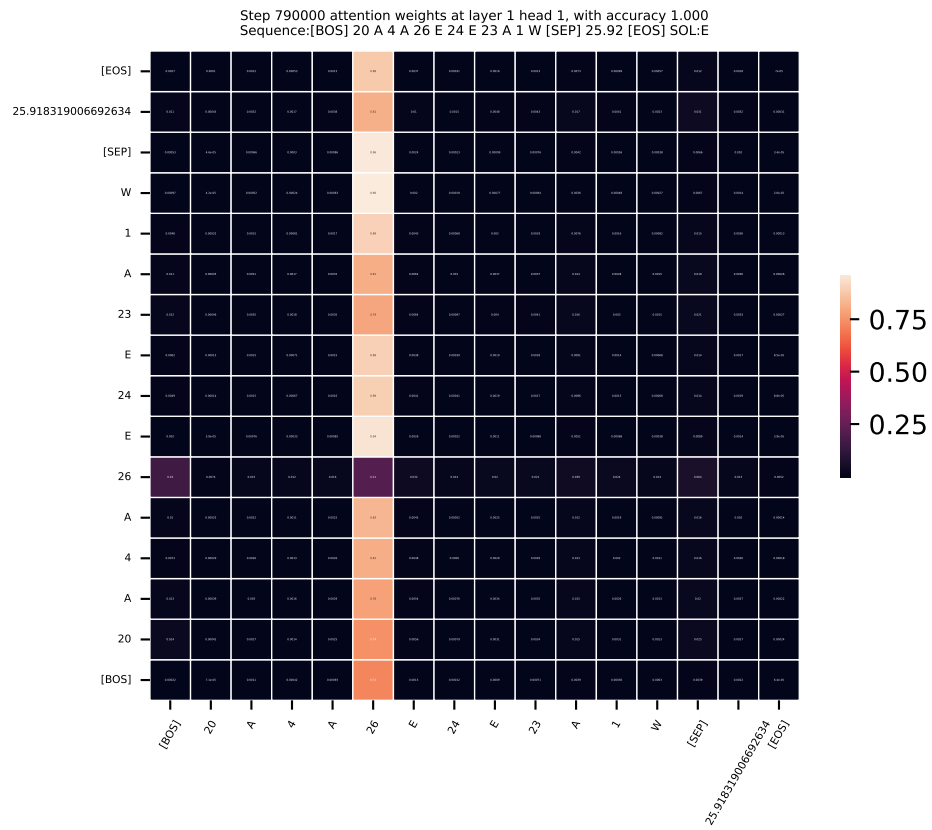

Step 790000 attention weights at layer 0 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

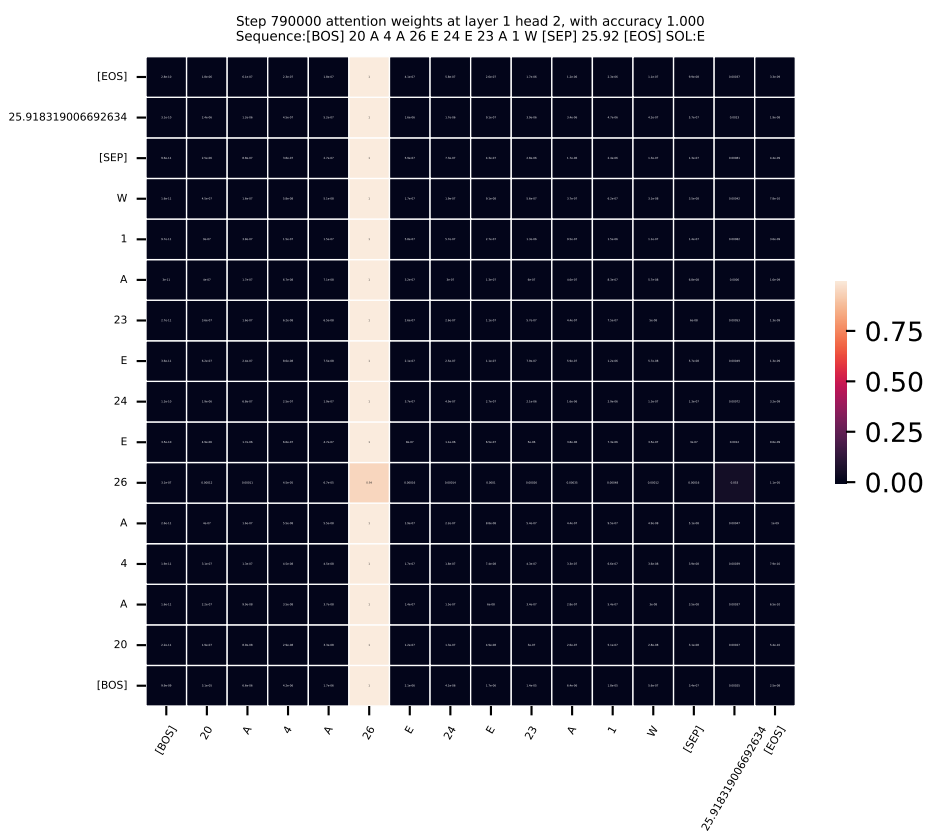

Step 790000 attention weights at layer 0 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

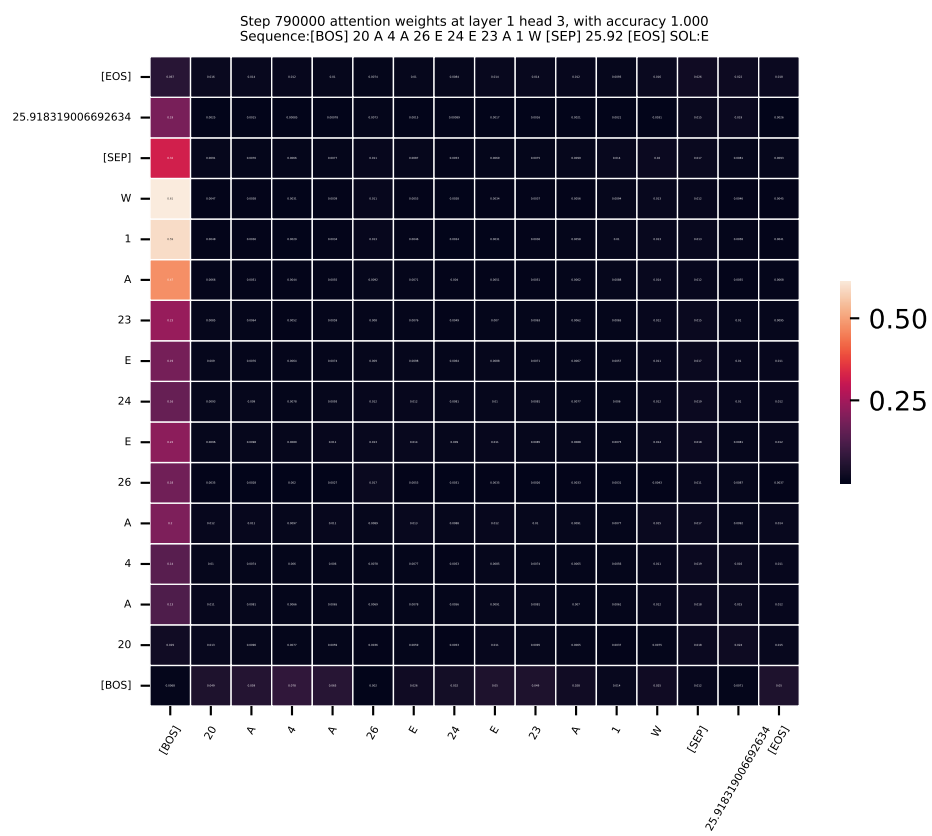

Step 790000 attention weights at layer 0 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

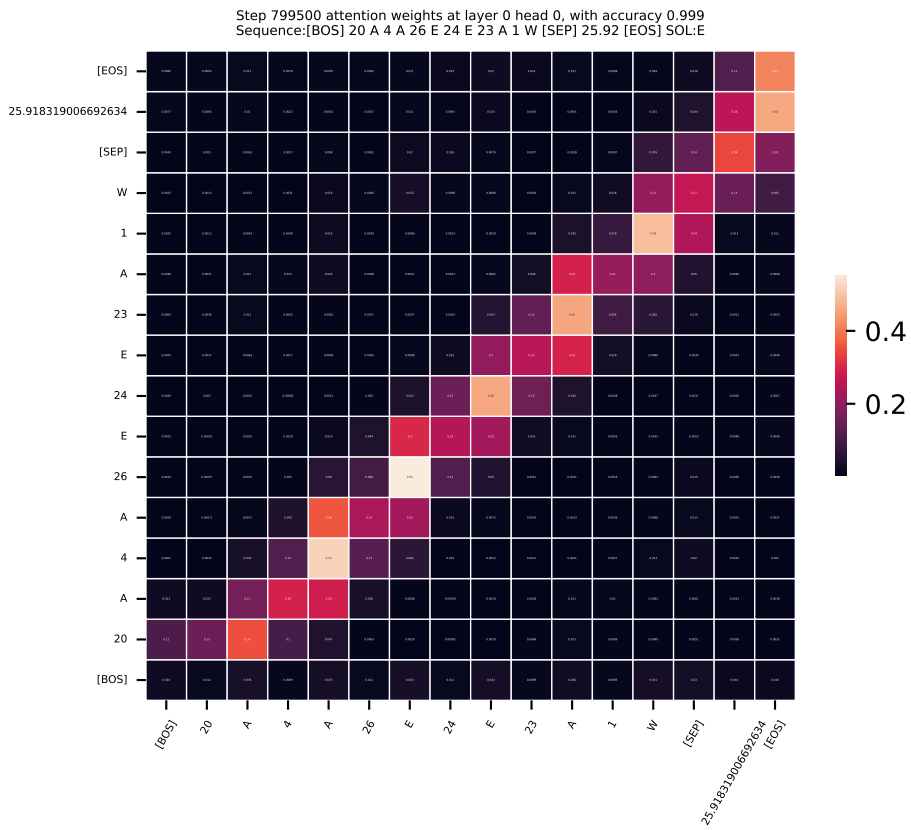

Step 790000 attention weights at layer 0 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

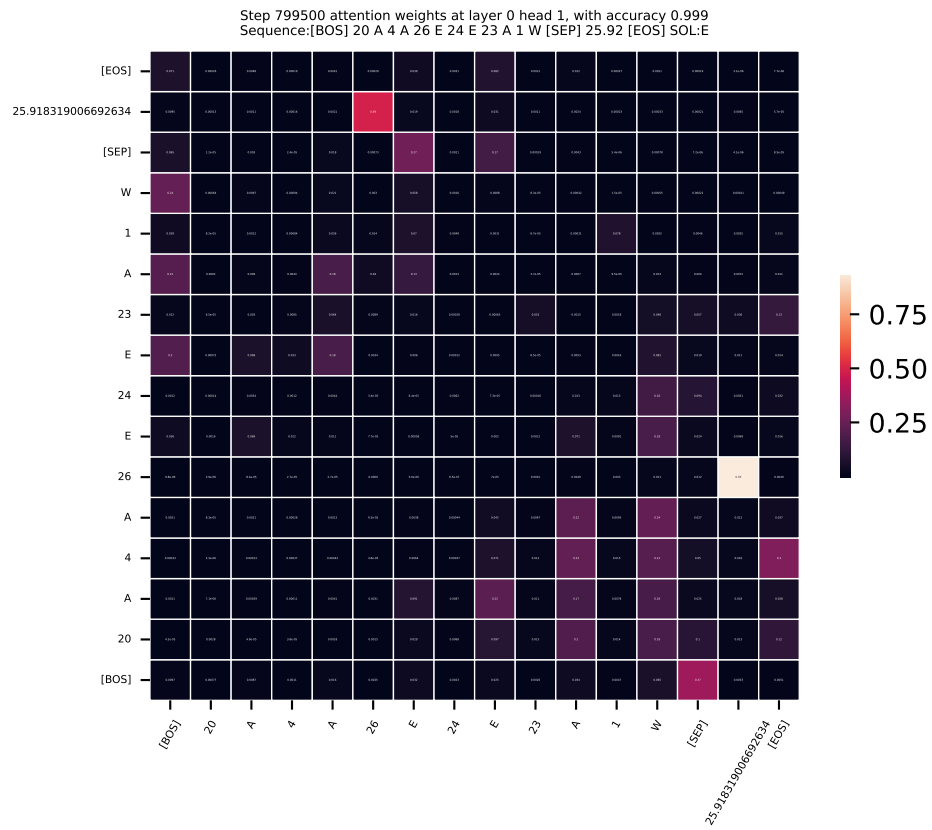

Step 790000 attention weights at layer 1 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25
0.00

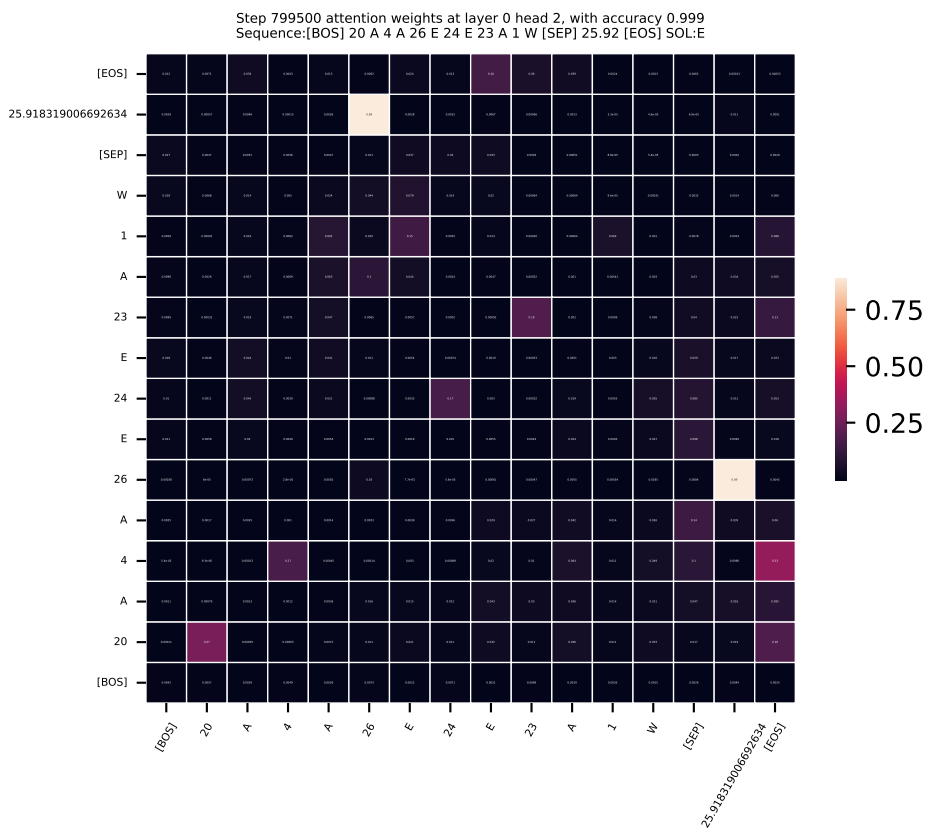

Step 790000 attention weights at layer 1 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

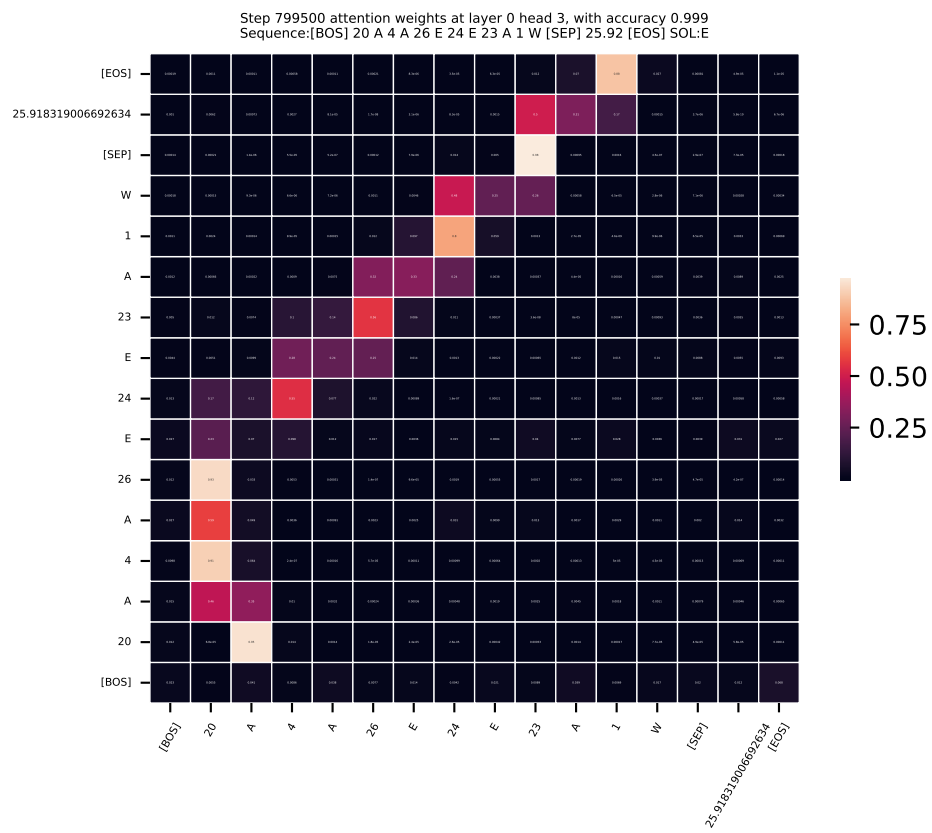

Step 790000 attention weights at layer 1 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25
0.00

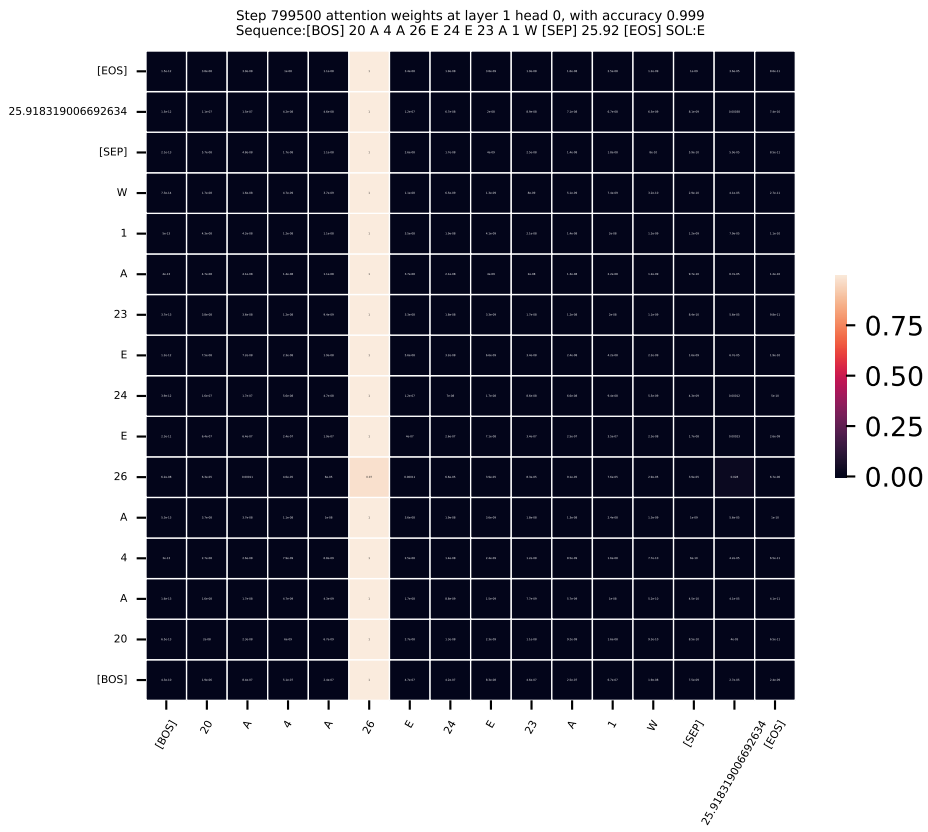

Step 790000 attention weights at layer 1 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.50
0.25

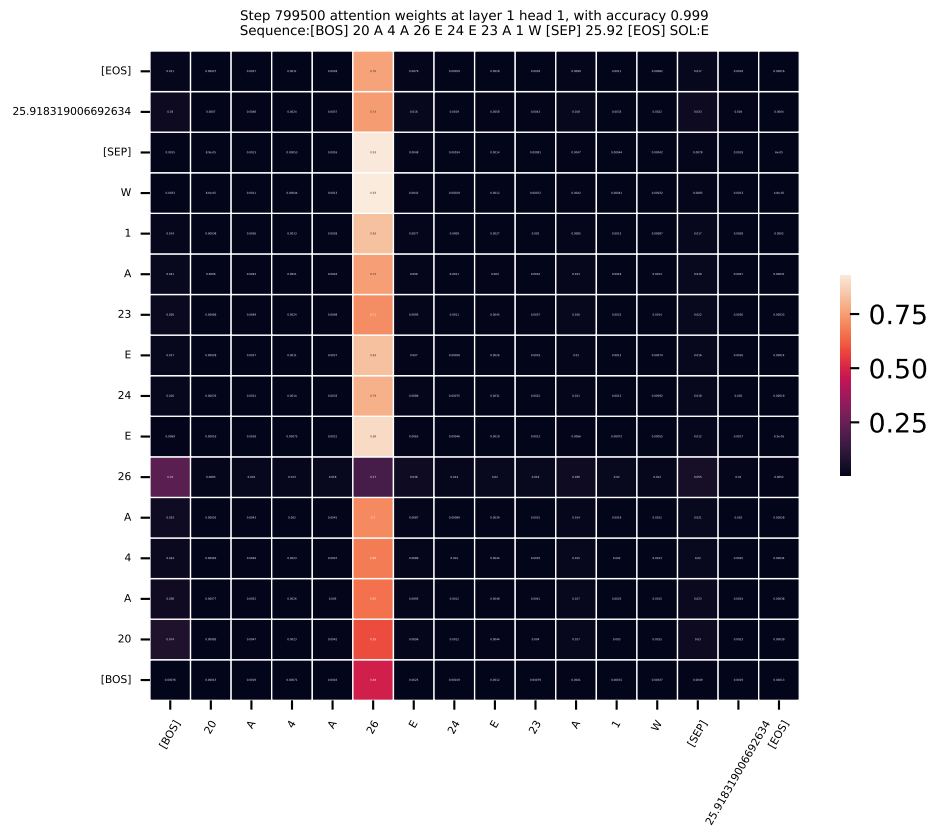

Step 799500 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

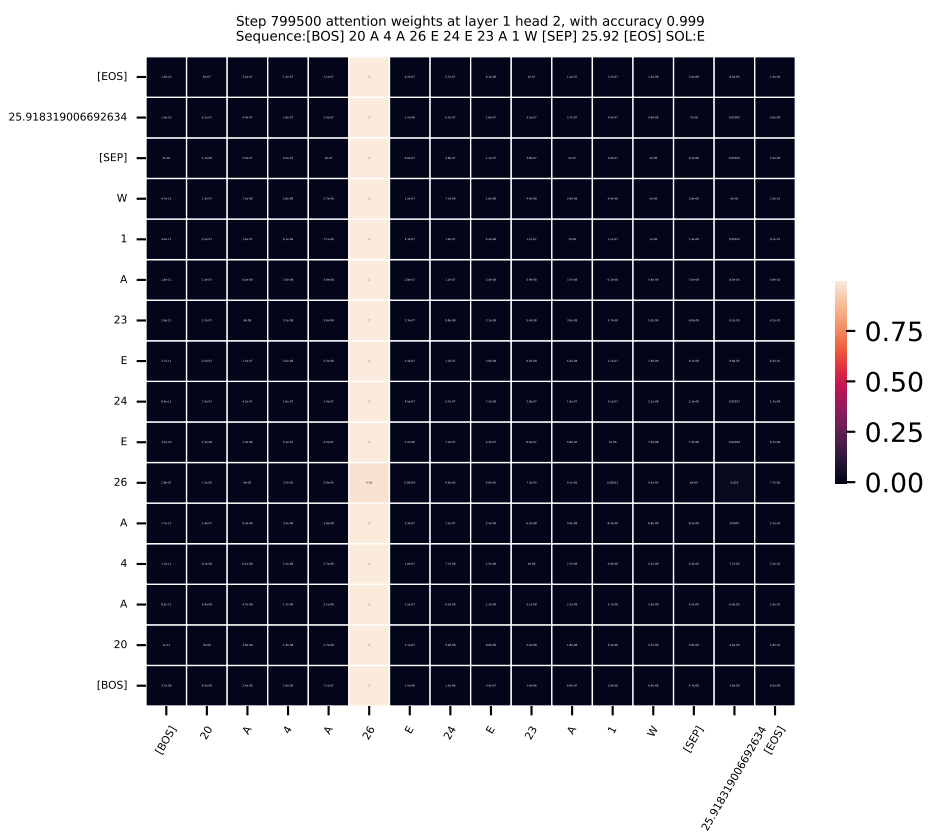

Step 799500 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

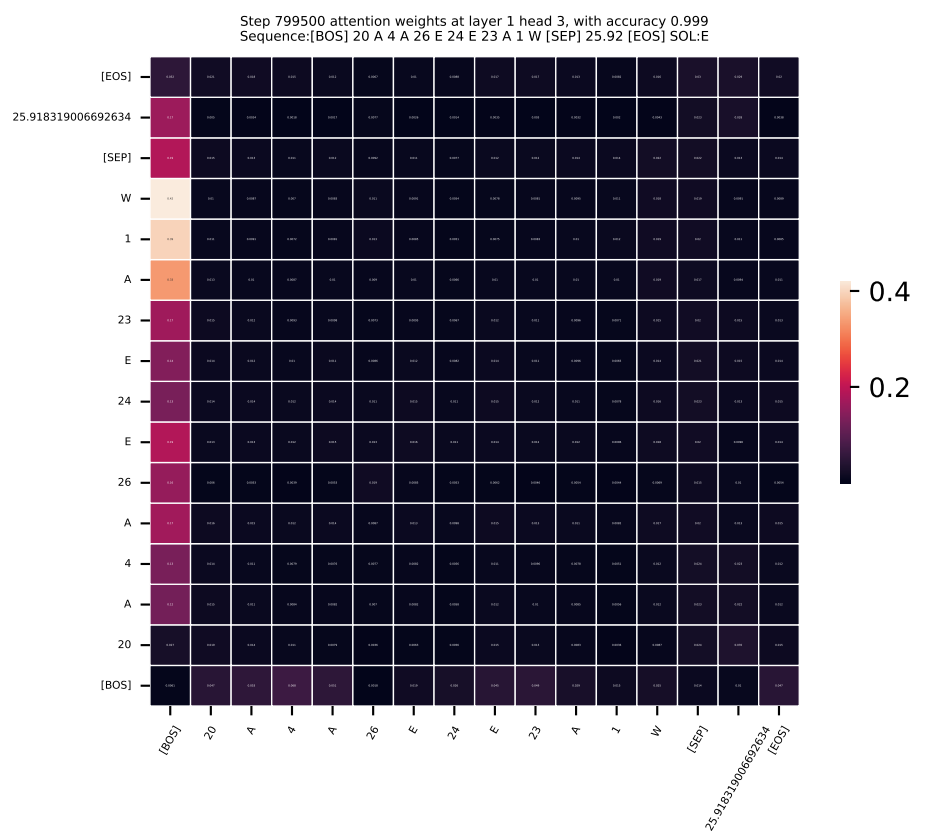

Step 799500 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

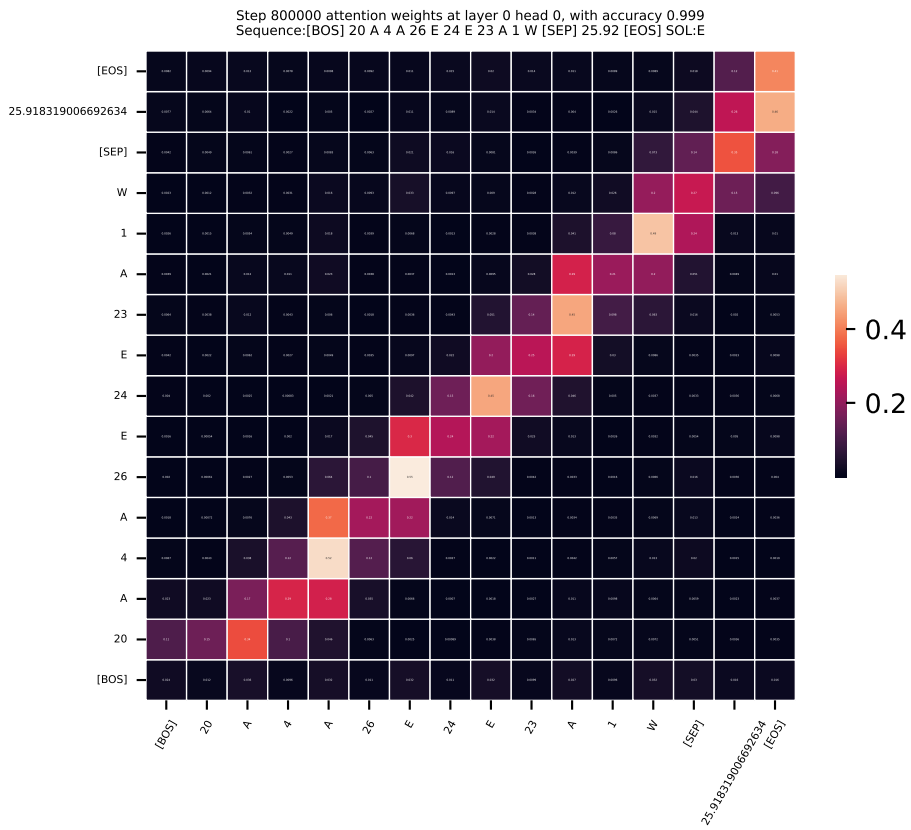

Step 799500 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

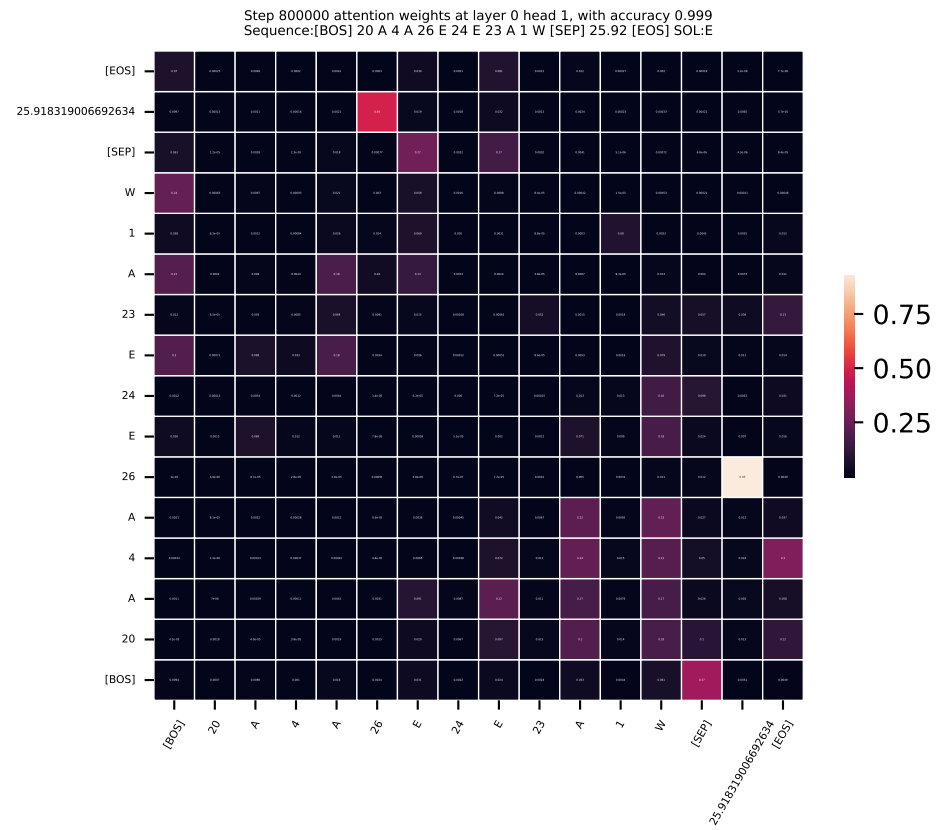

Step 799500 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

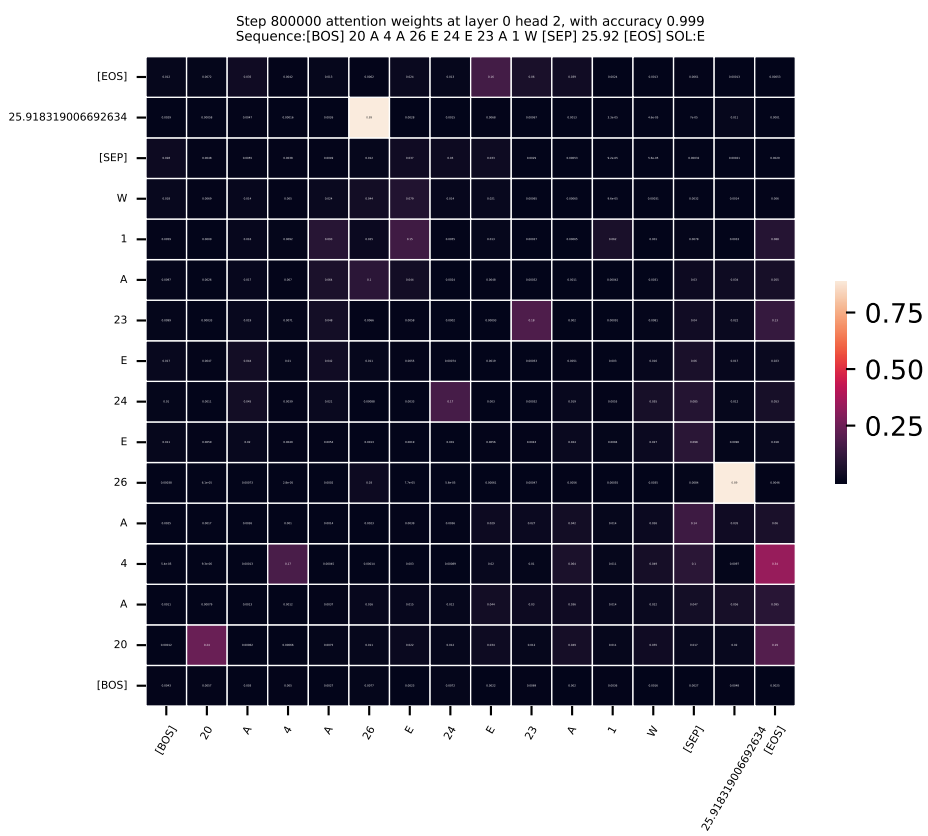

Step 799500 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

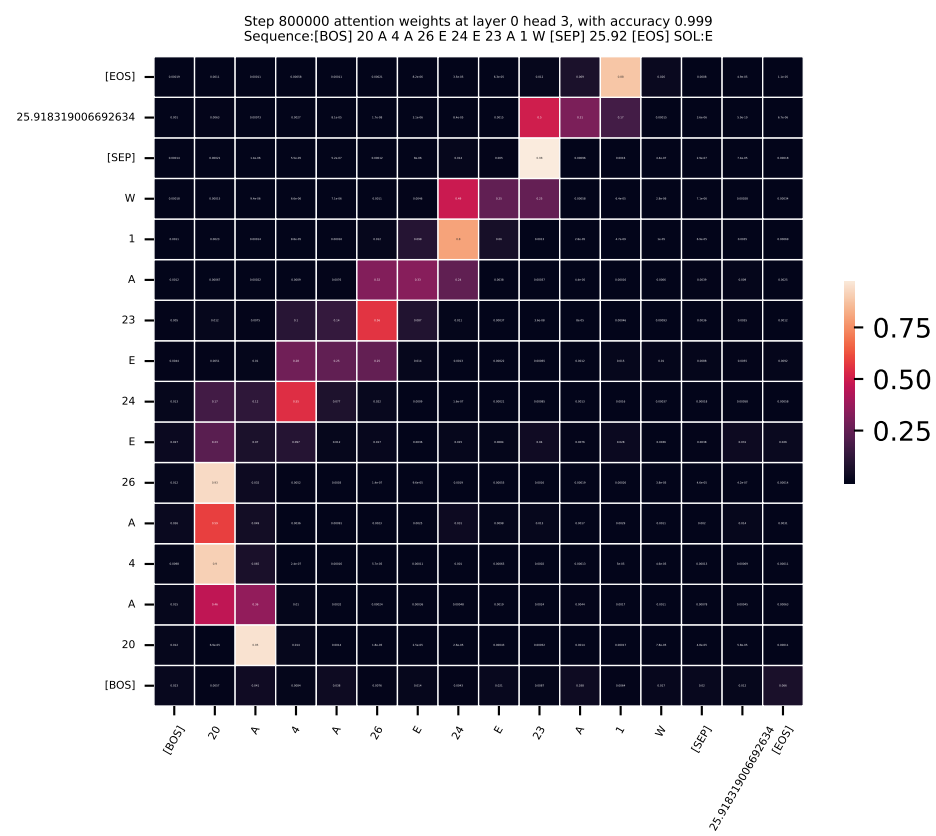

Step 799500 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

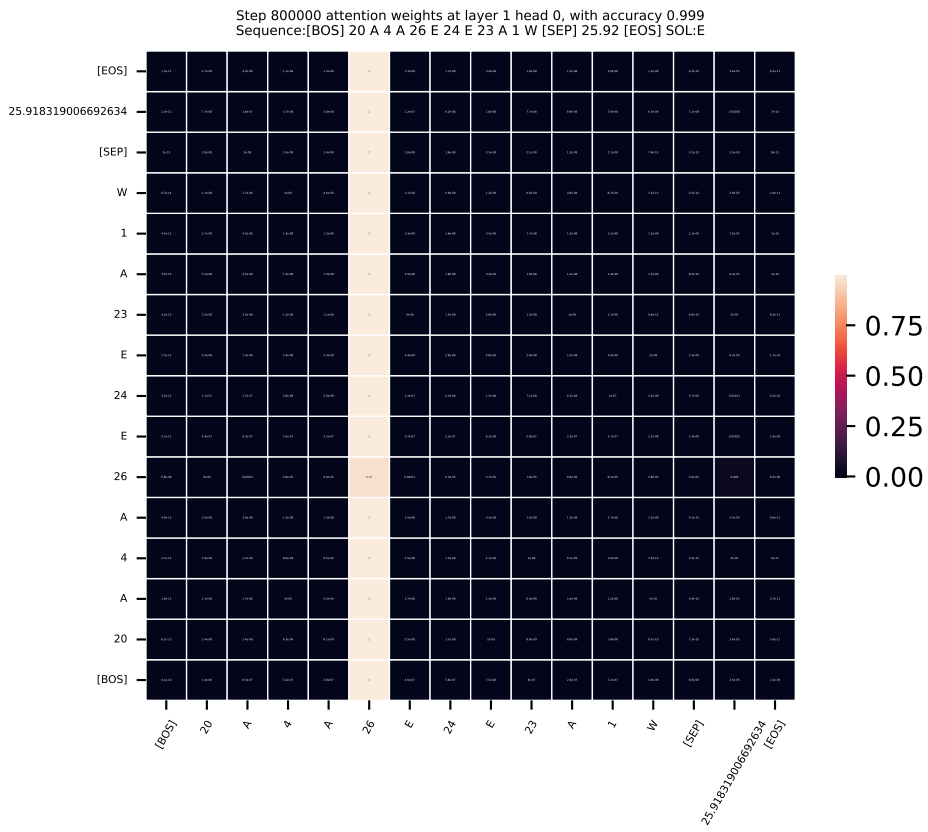

Step 799500 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

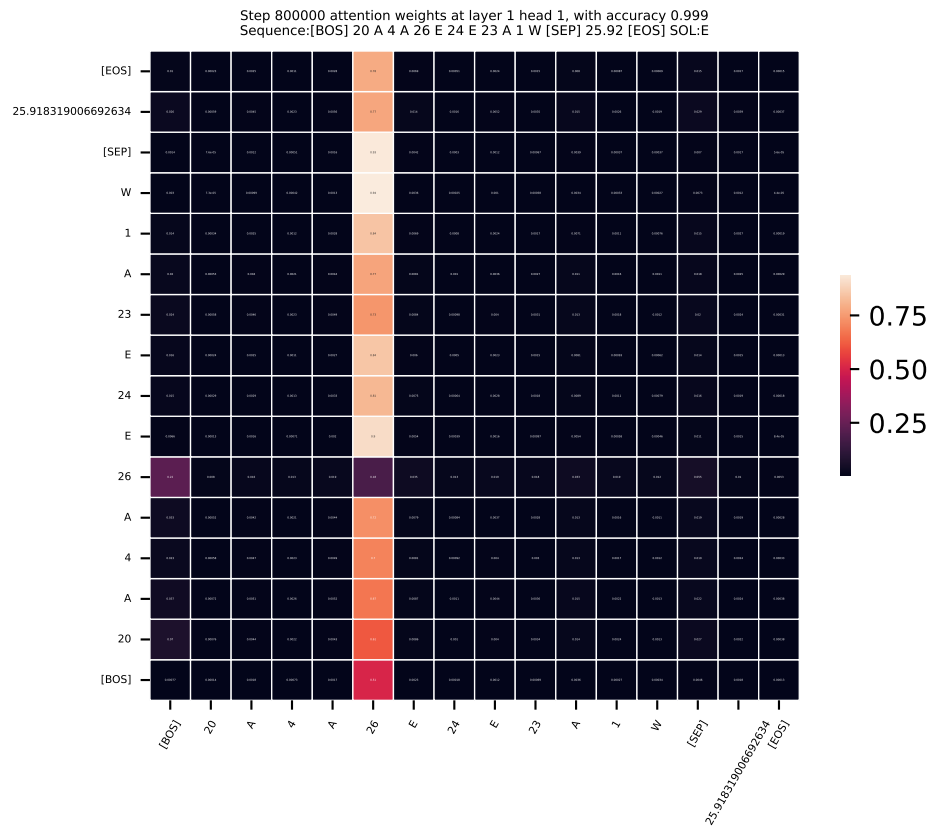

Step 800000 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

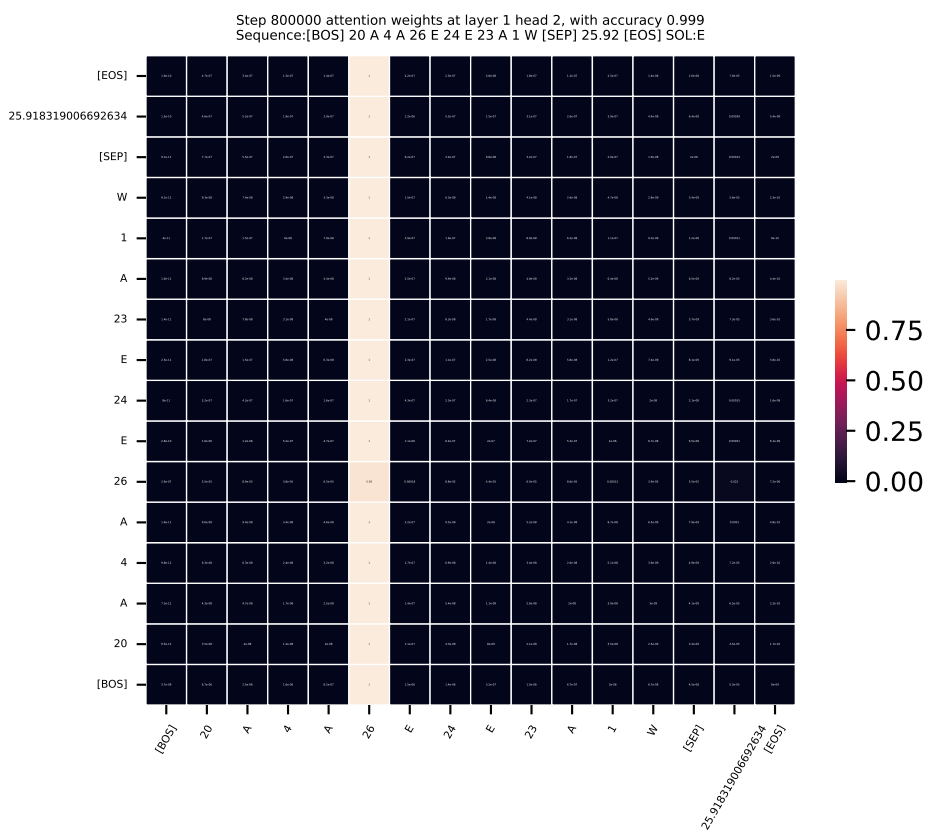

Step 800000 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

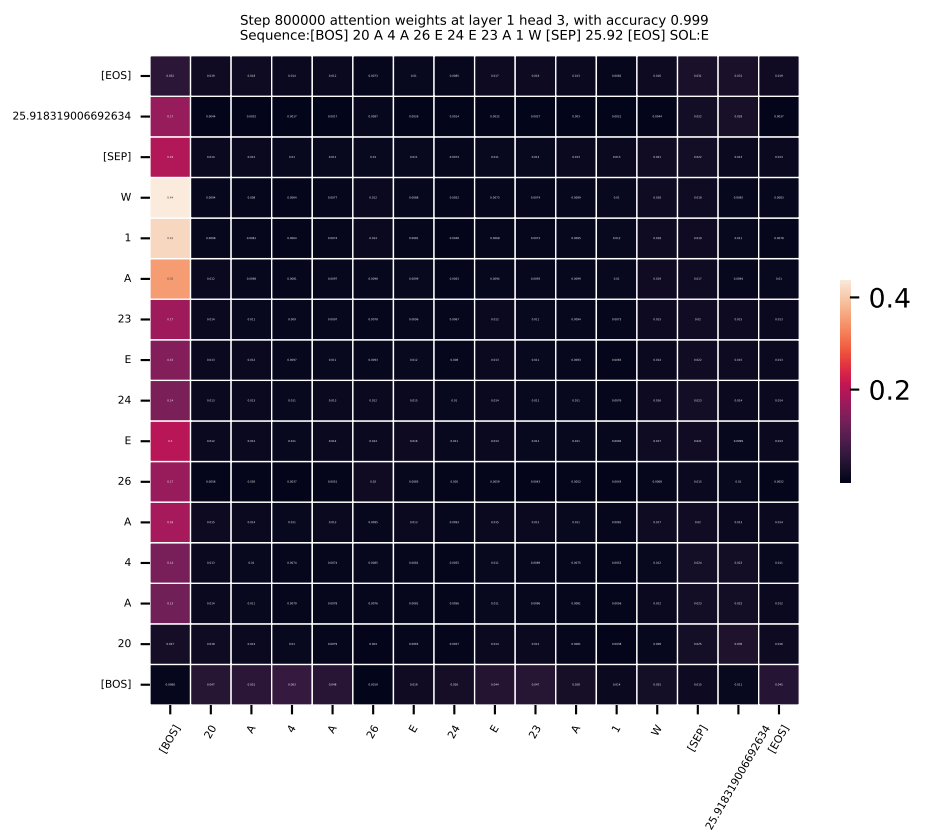

Step 800000 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

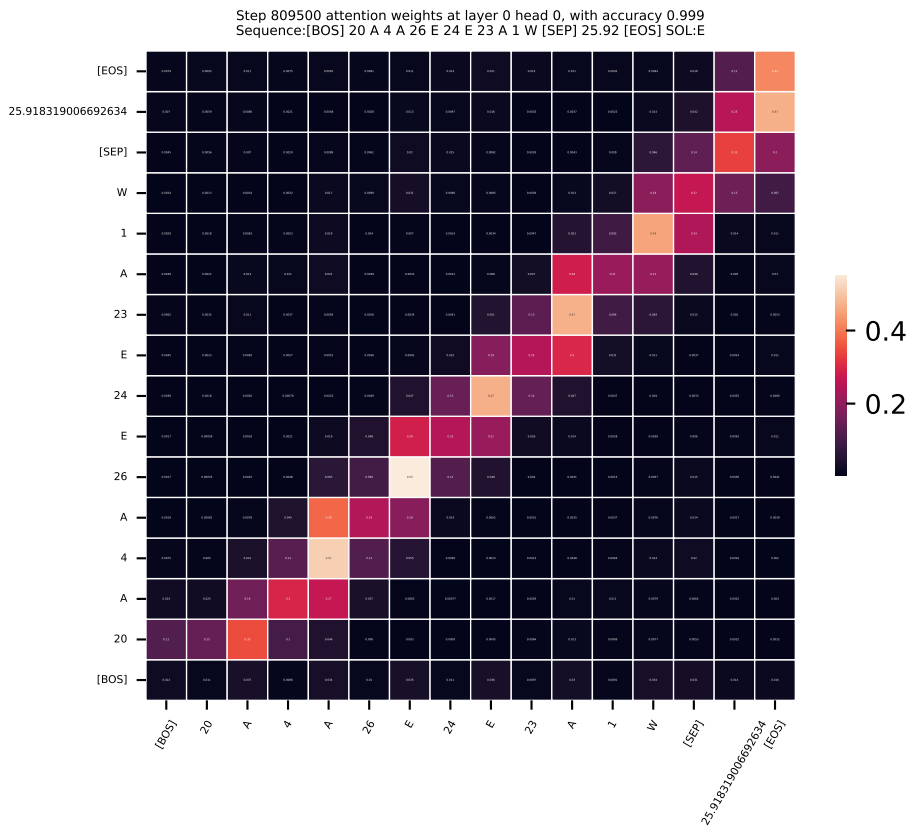

Step 800000 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

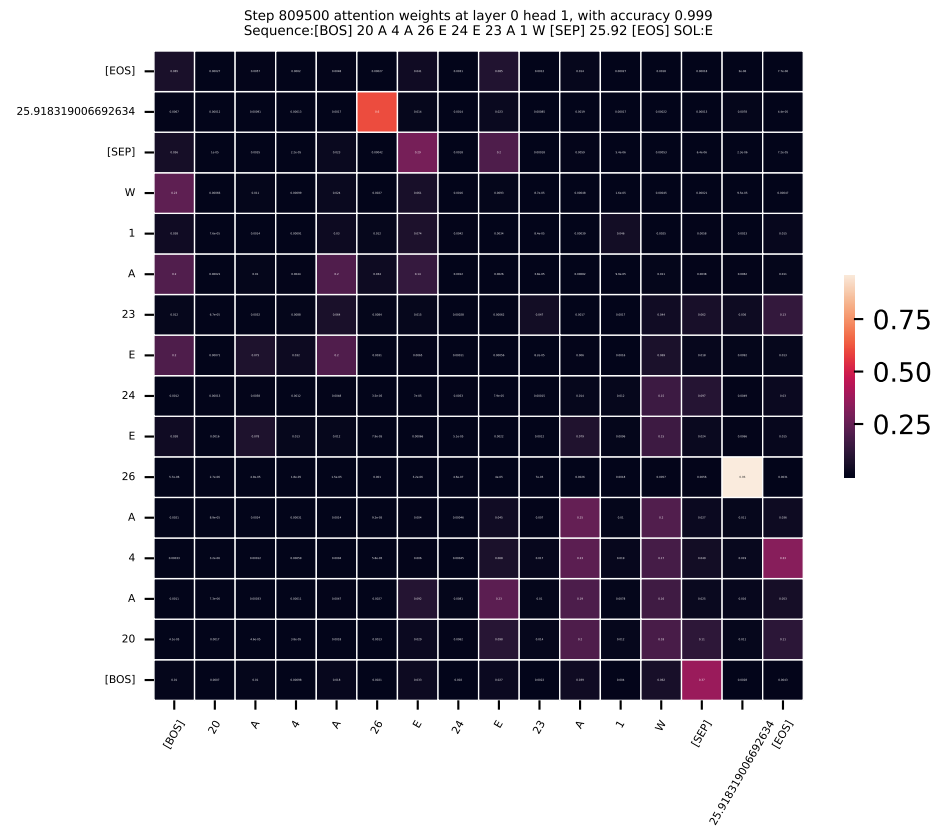

Step 800000 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

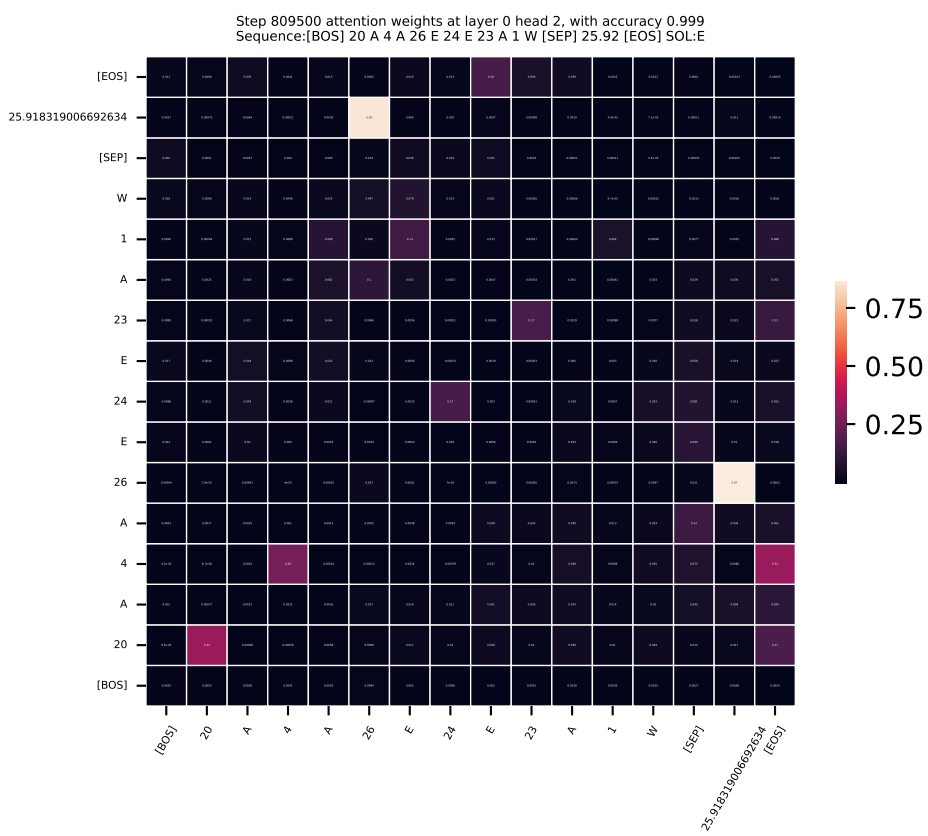

Step 800000 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

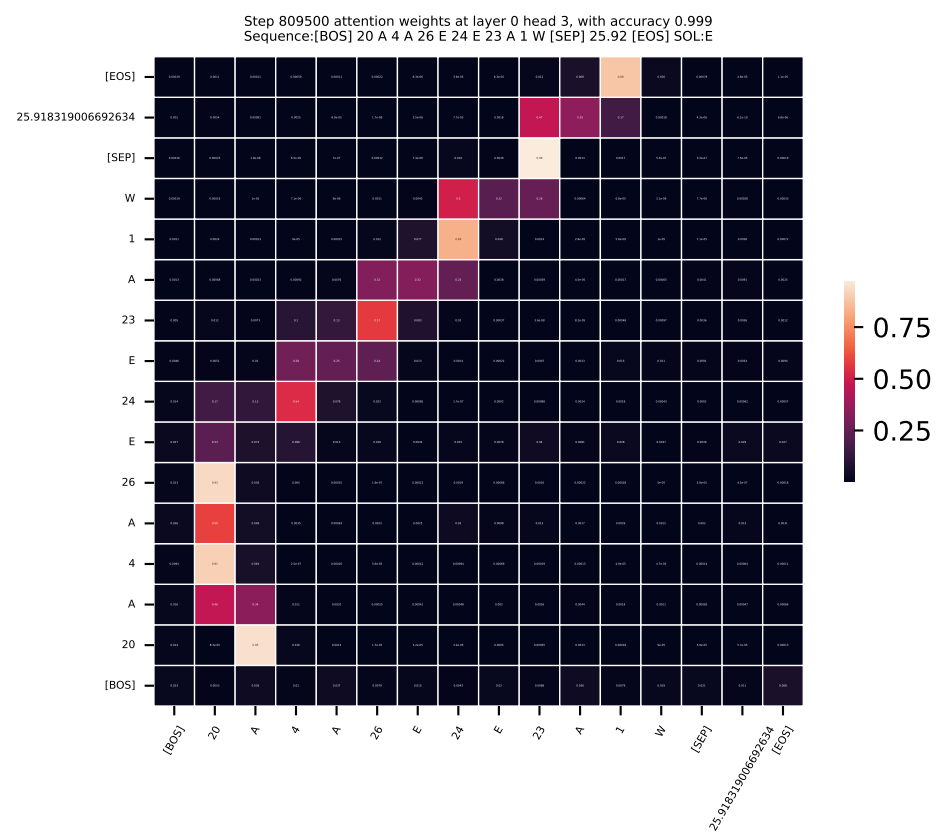

Step 800000 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

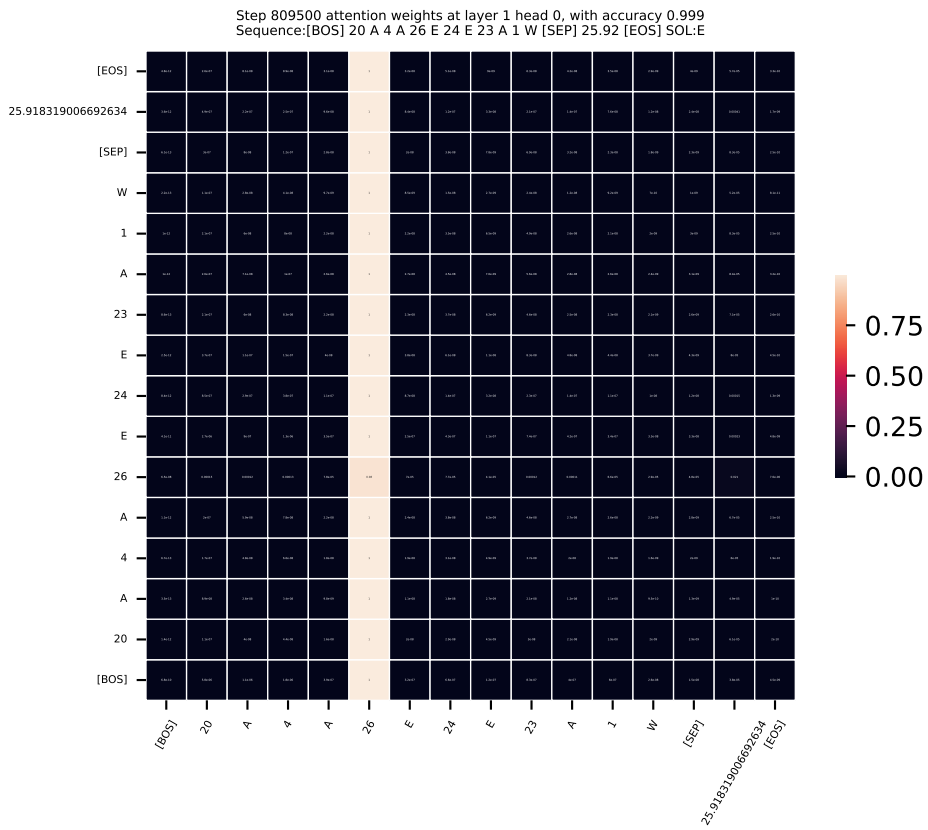

Step 800000 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

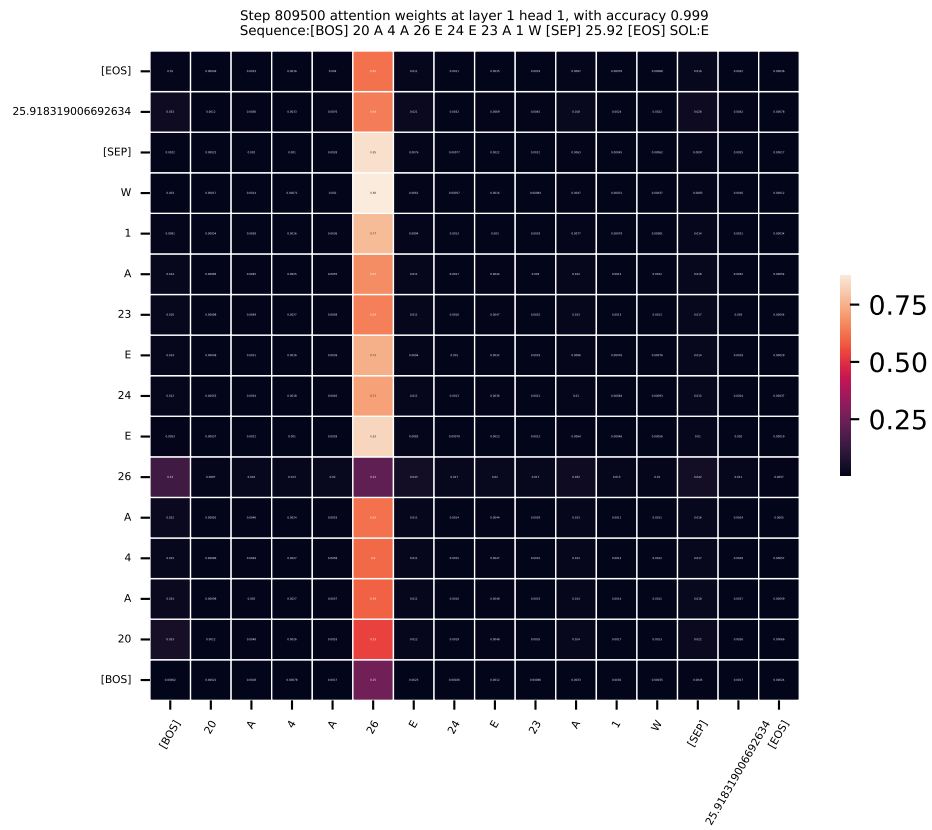

Step 809500 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

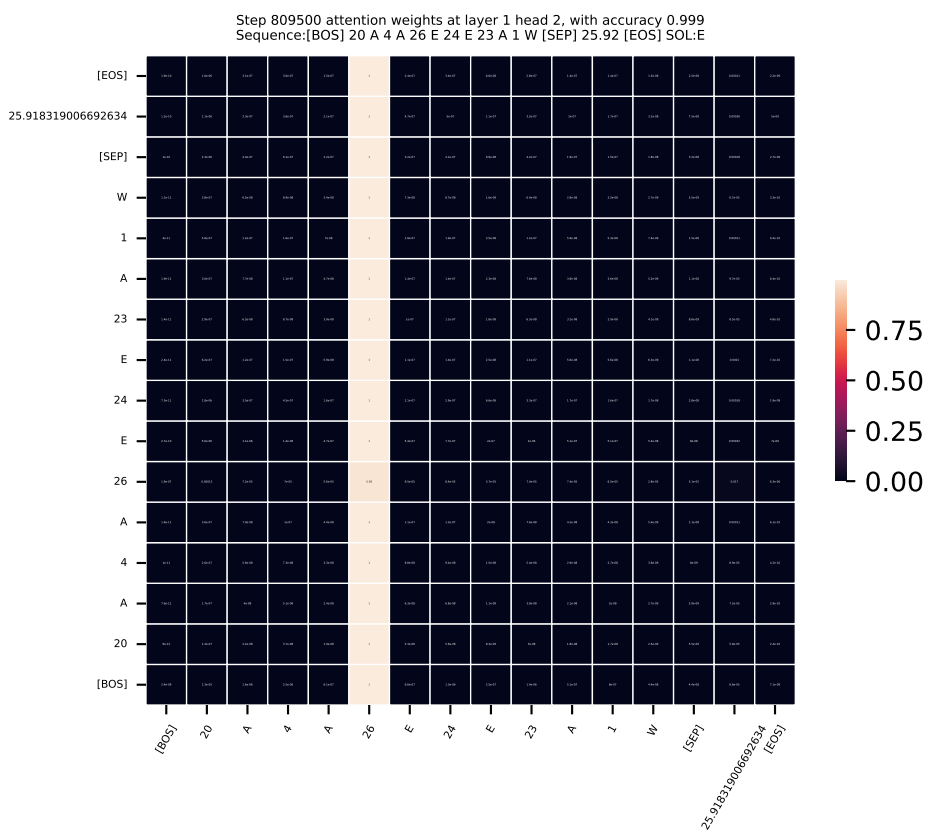

Step 809500 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

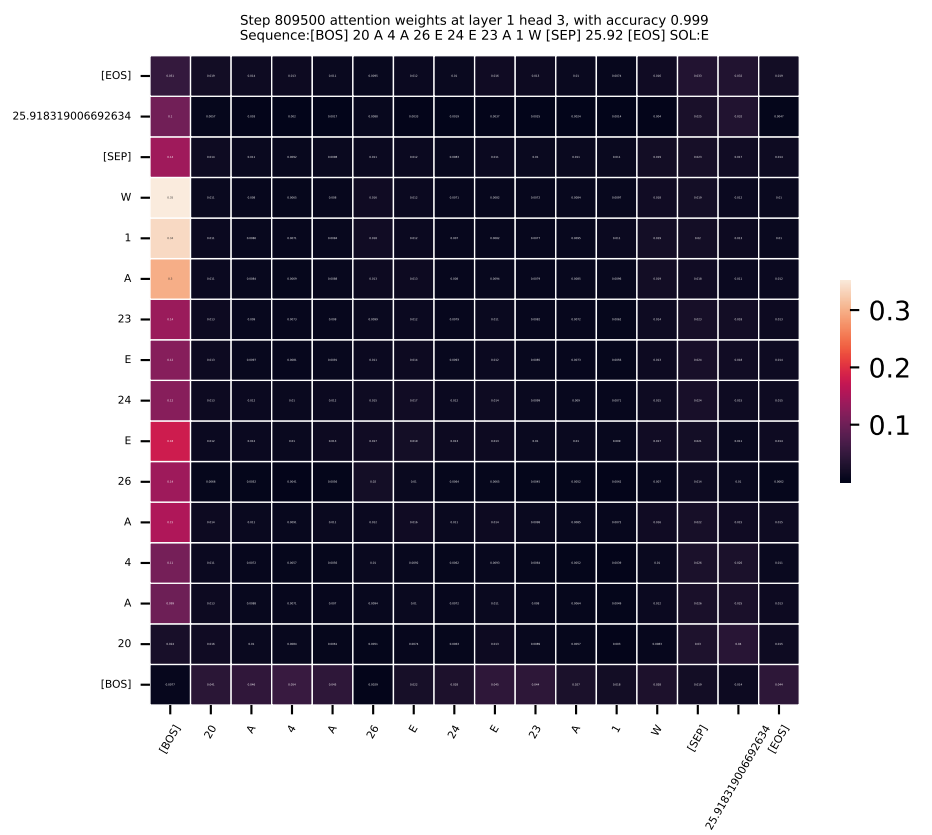

Step 809500 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

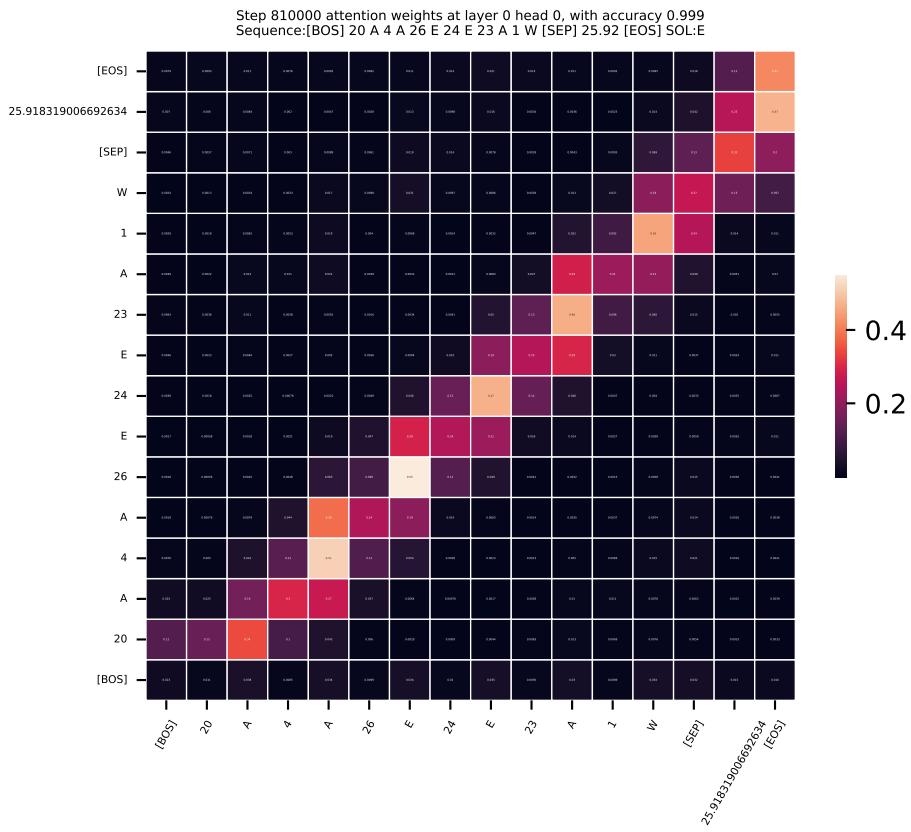

Step 809500 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

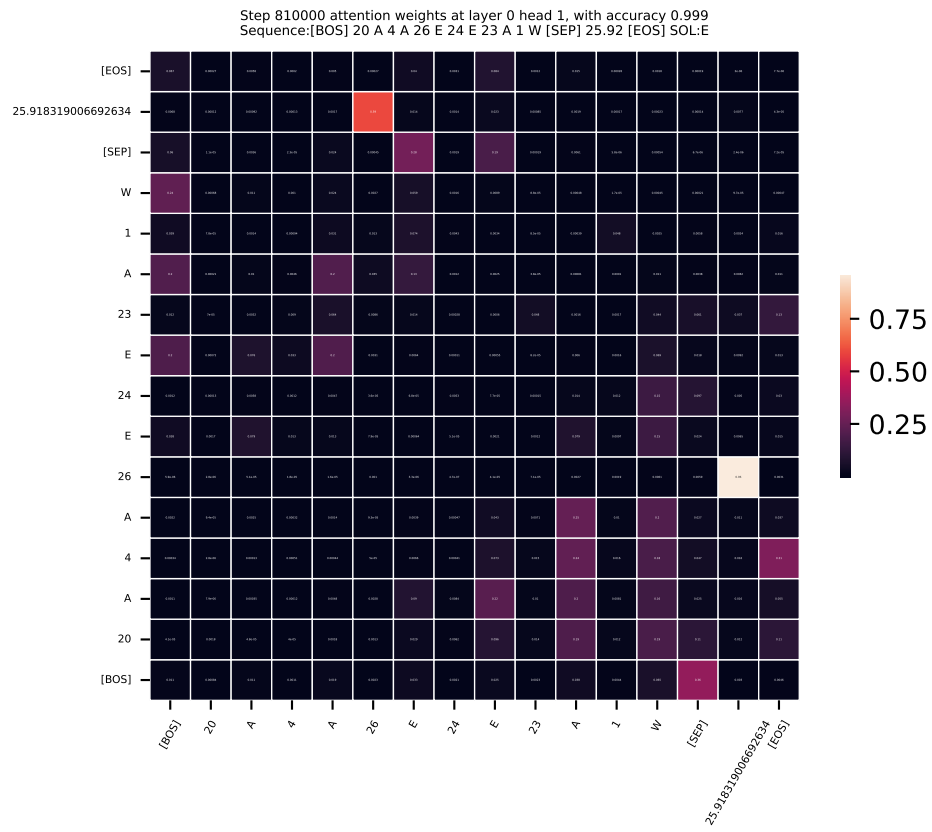

Step 809500 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

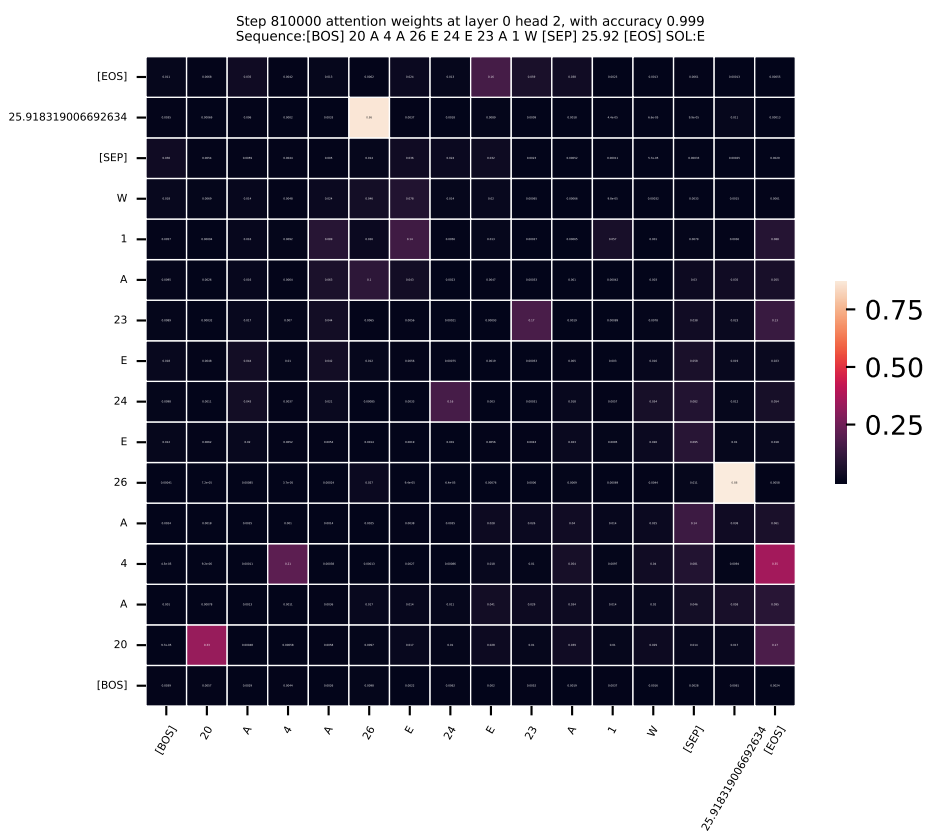

Step 809500 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

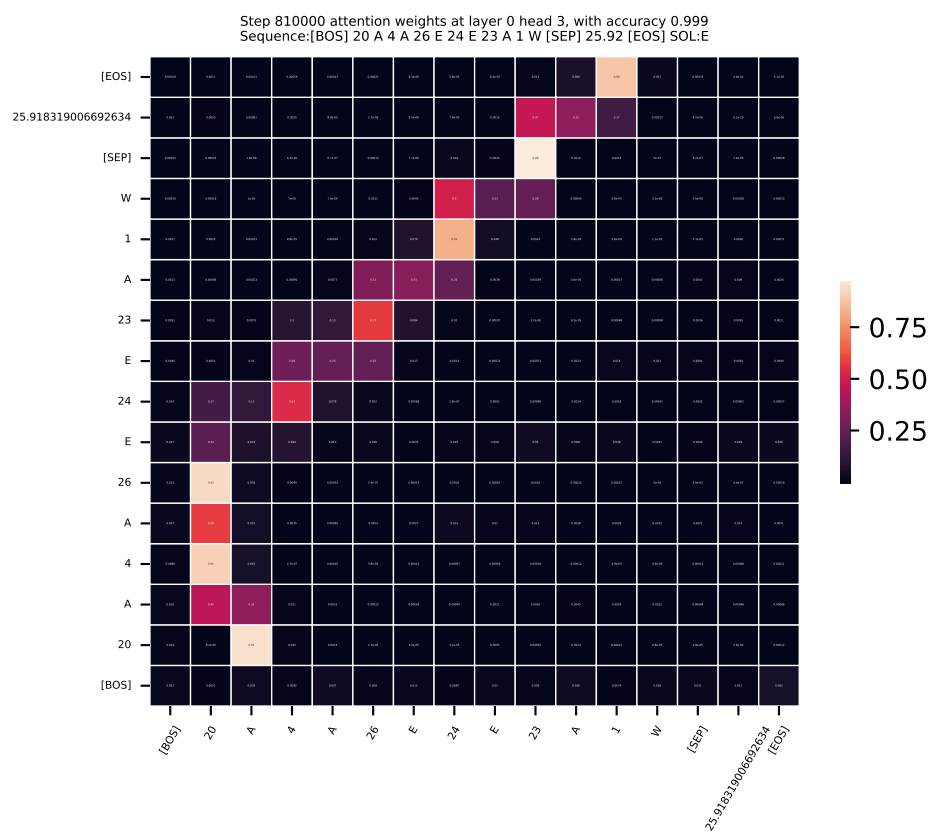

Step 809500 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

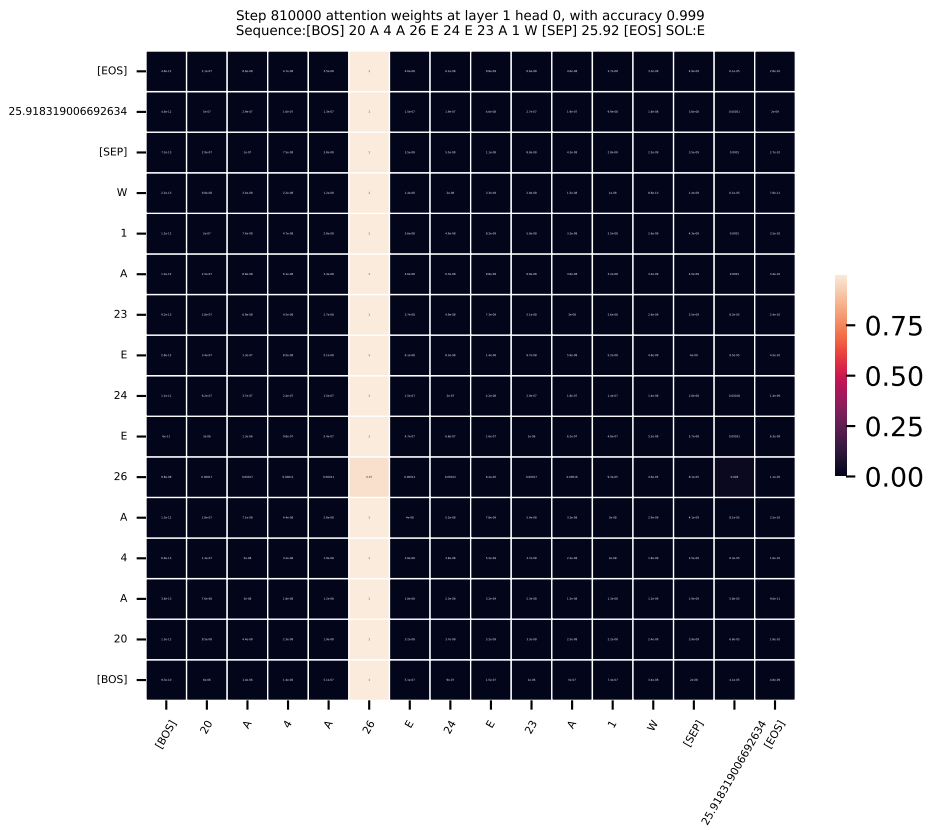

Step 809500 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

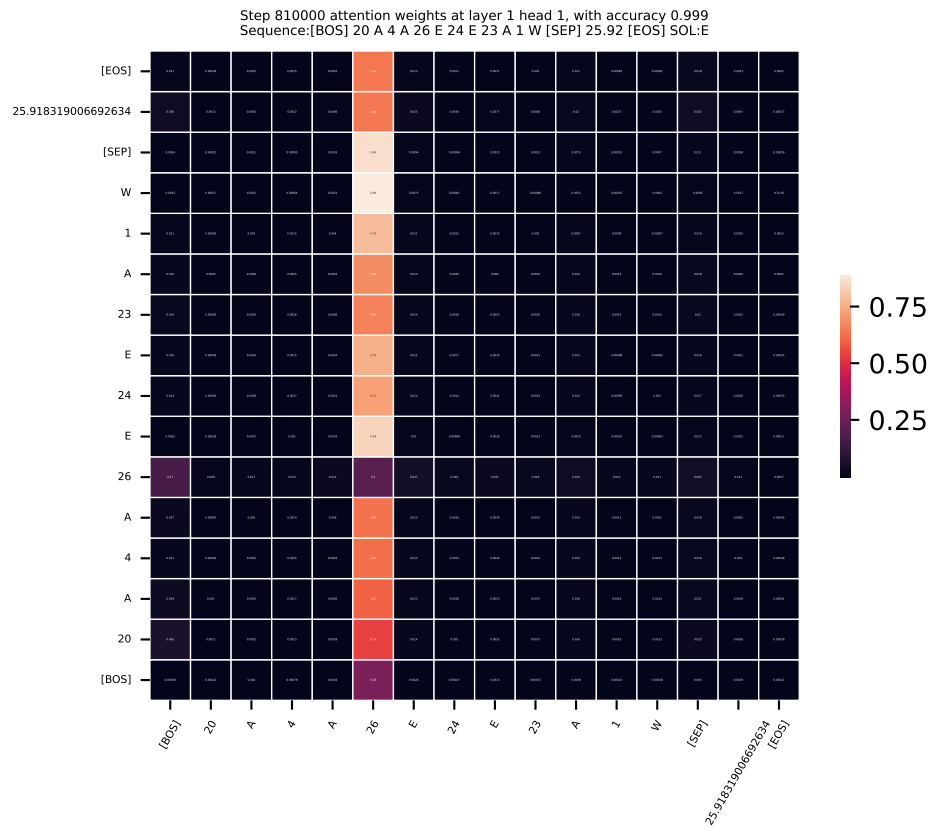

Step 810000 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.4
0.2

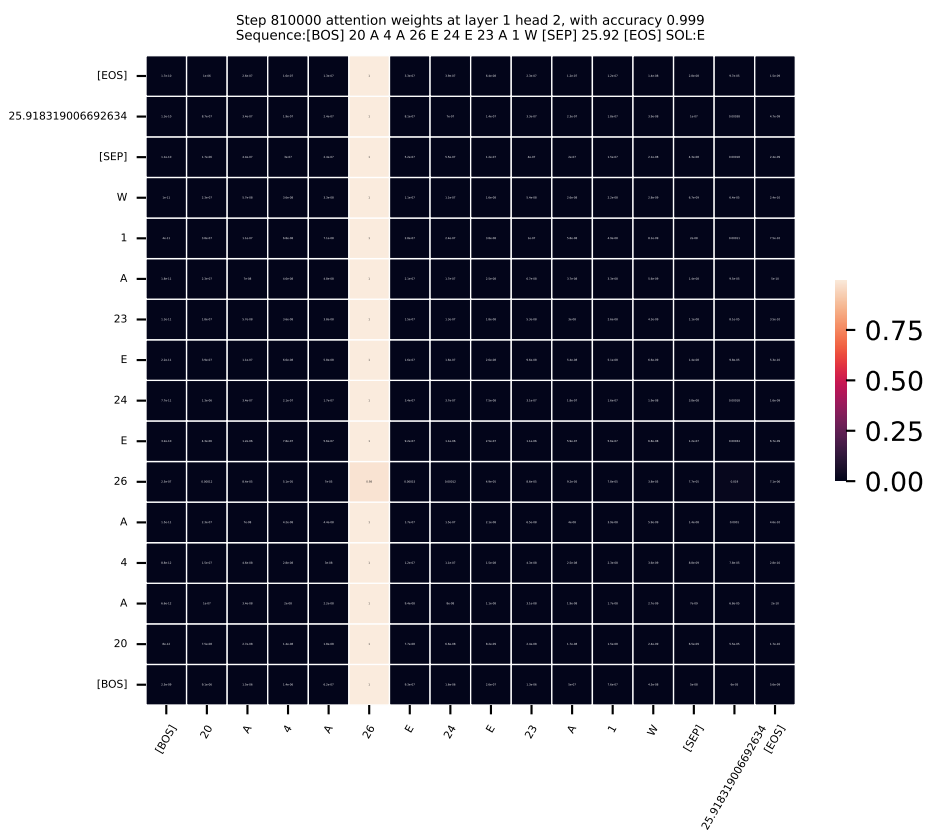

Step 810000 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

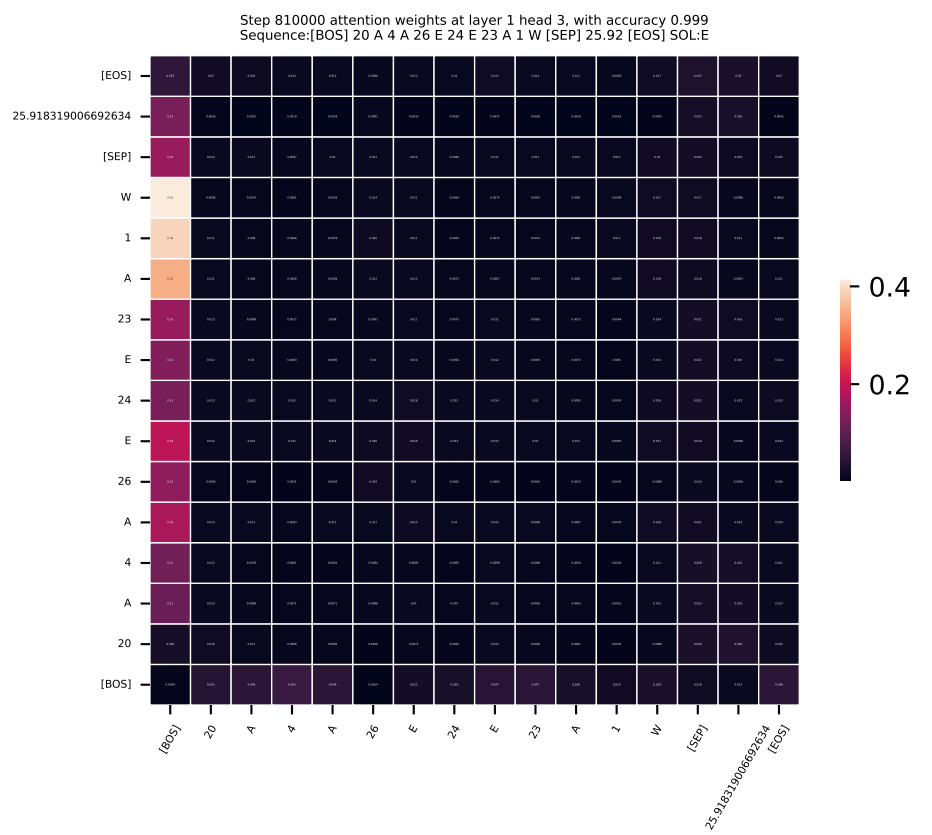

Step 810000 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

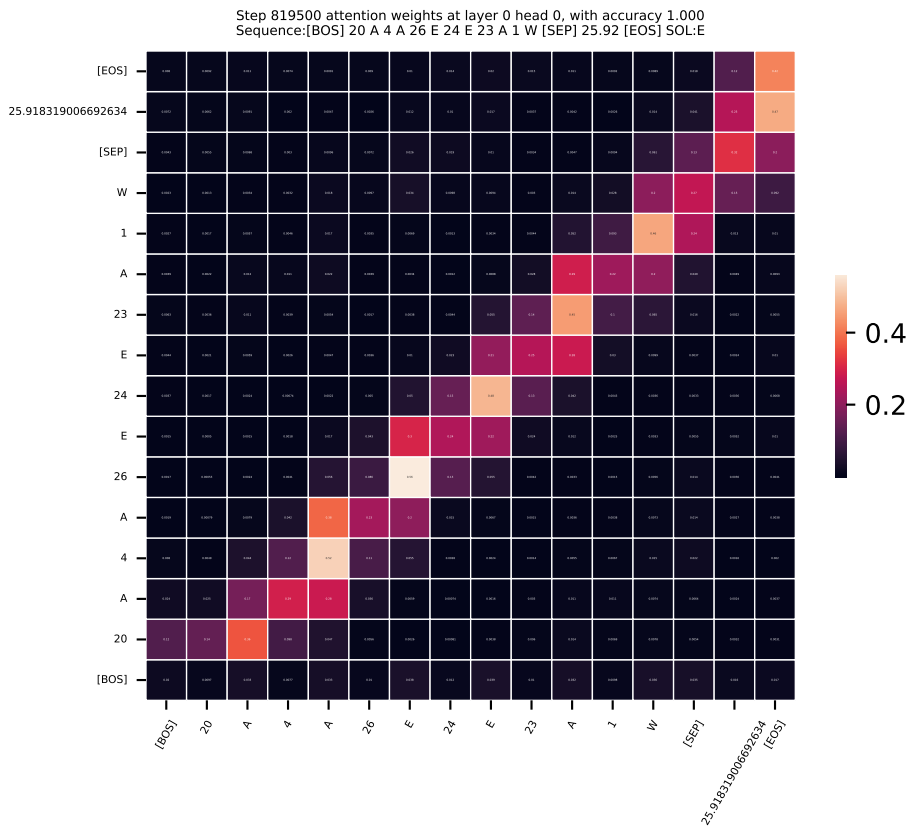

Step 810000 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

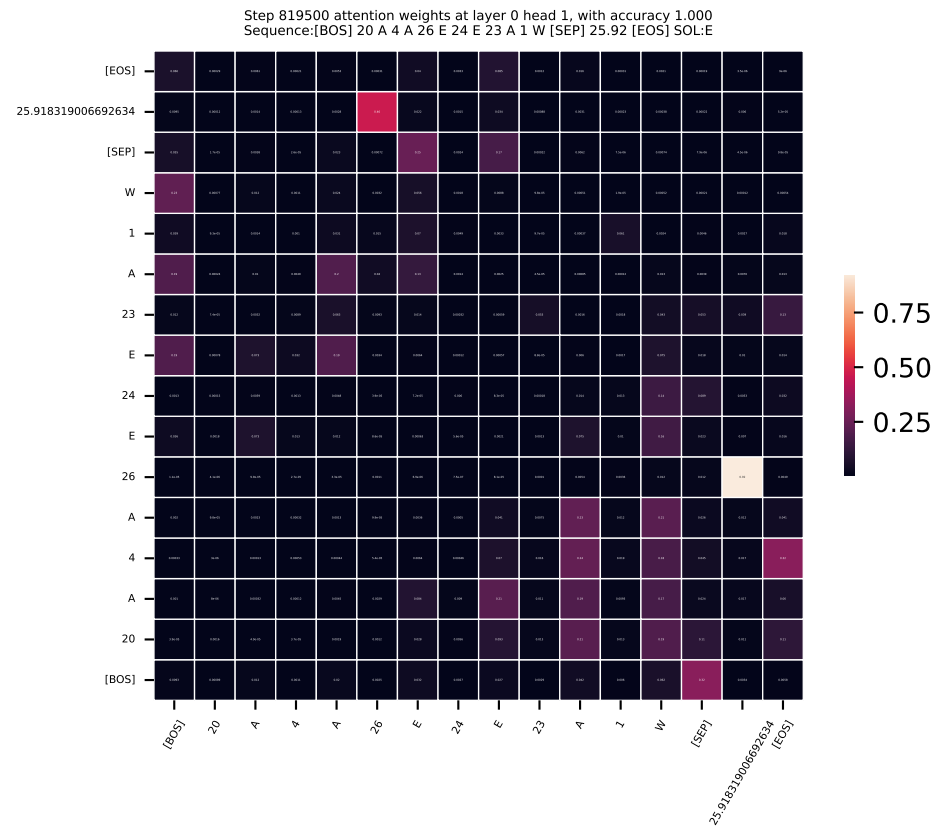

Step 810000 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

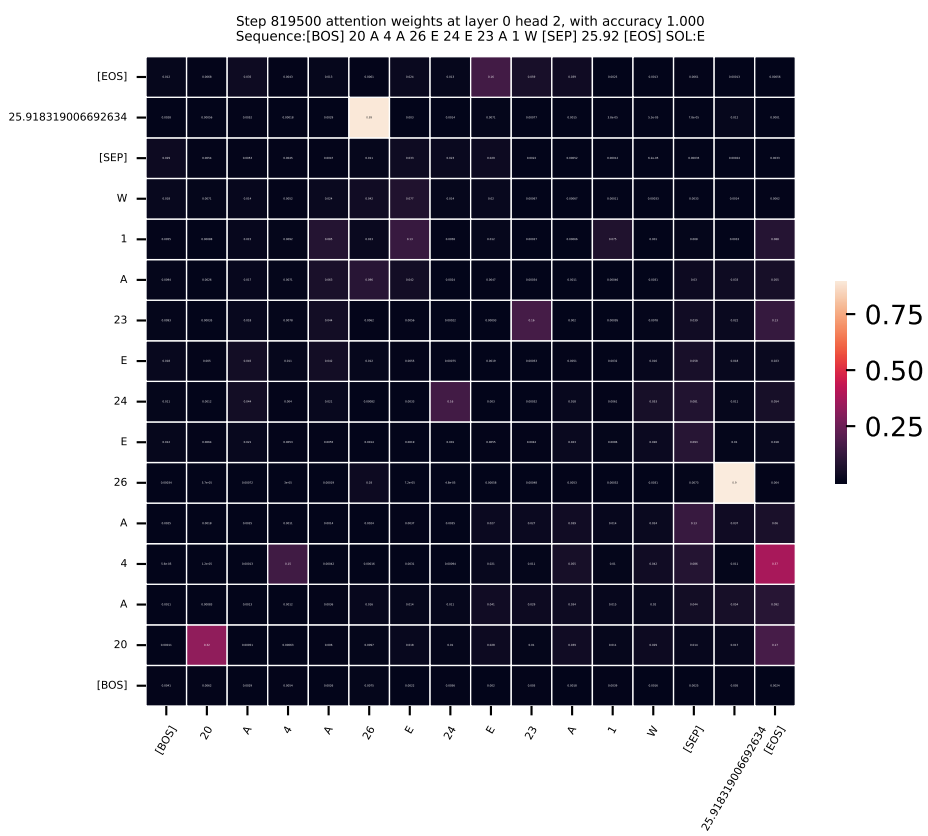

Step 810000 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

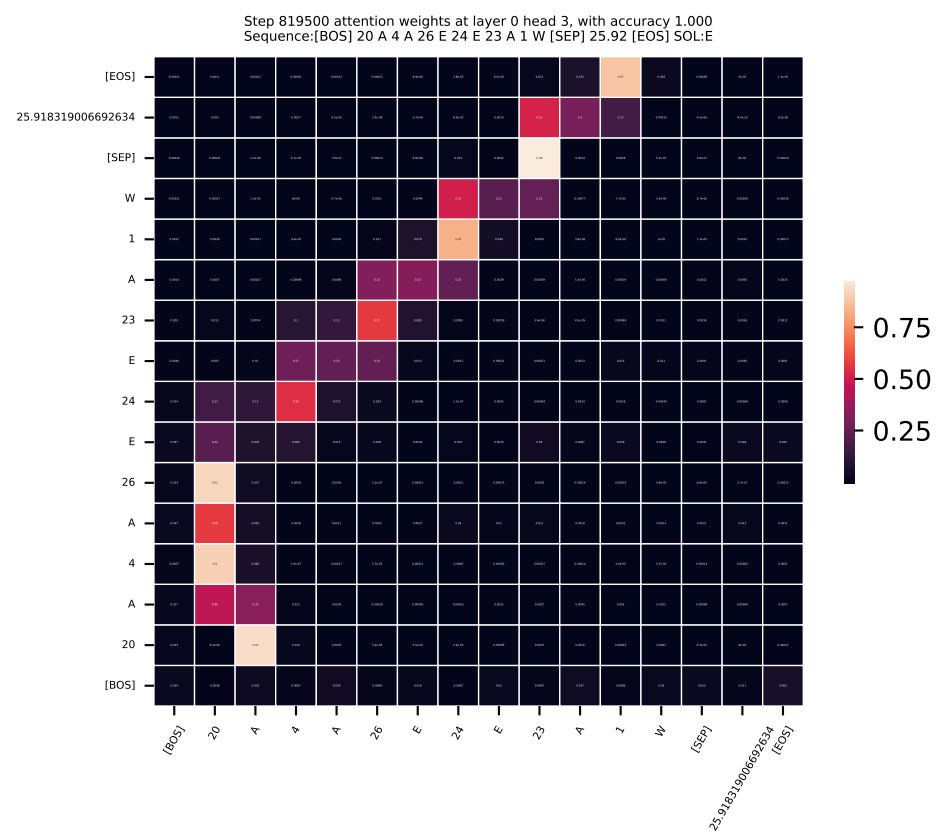

Step 810000 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

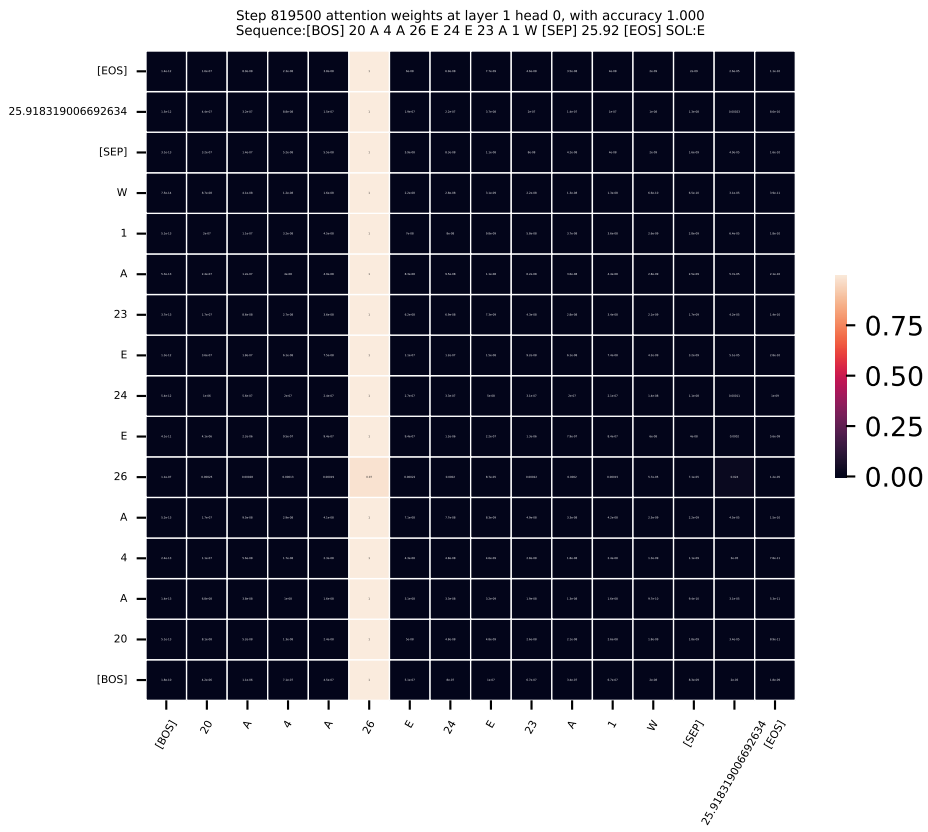

Step 810000 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

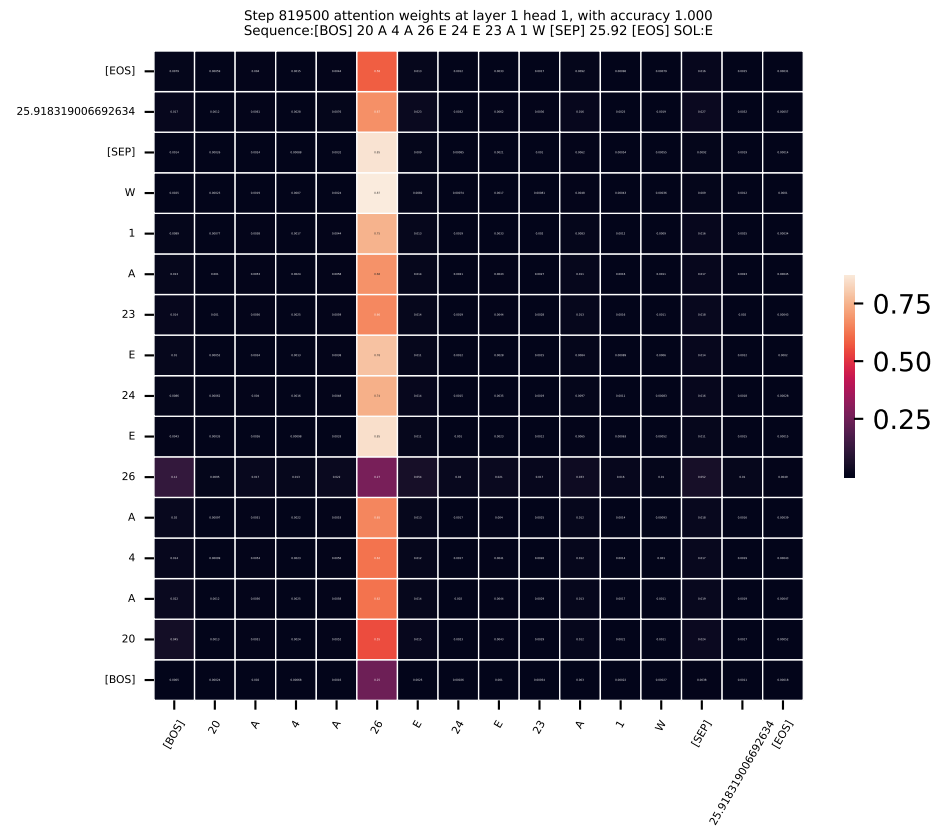

Step 819500 attention weights at layer 0 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

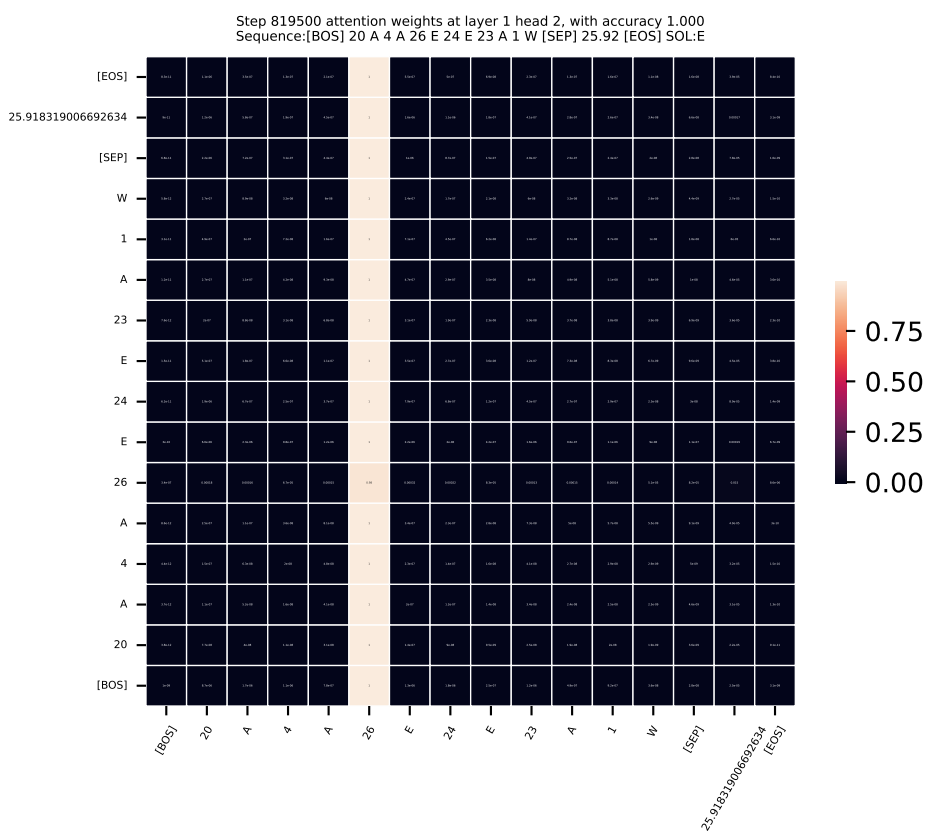

Step 819500 attention weights at layer 0 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

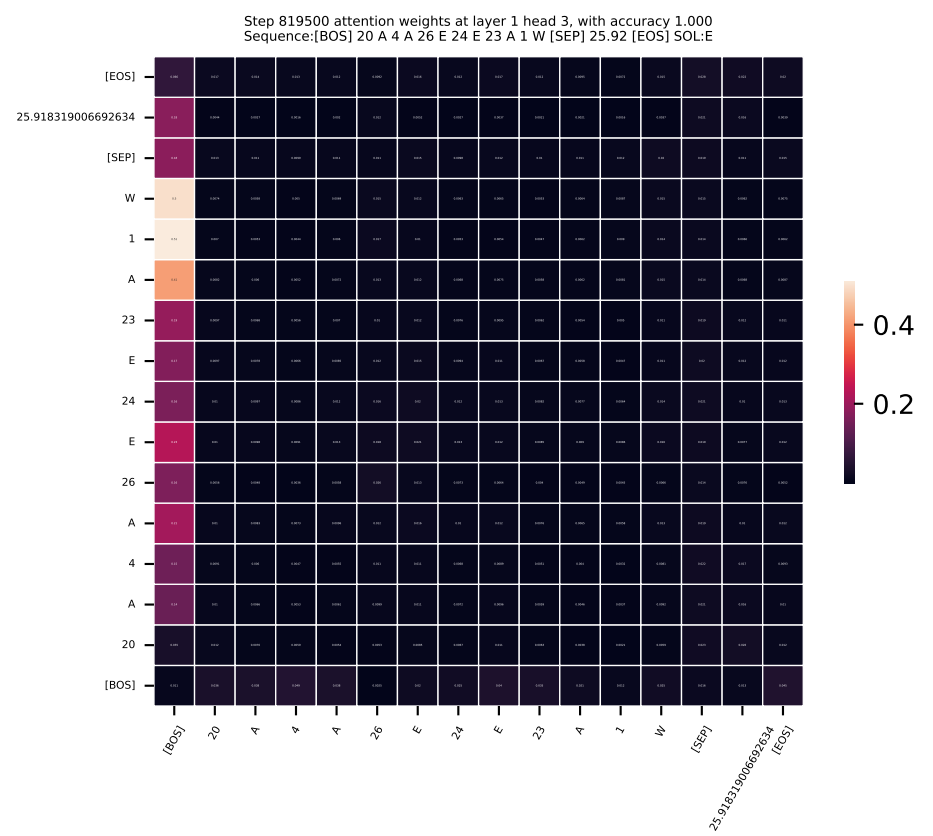

Step 819500 attention weights at layer 0 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

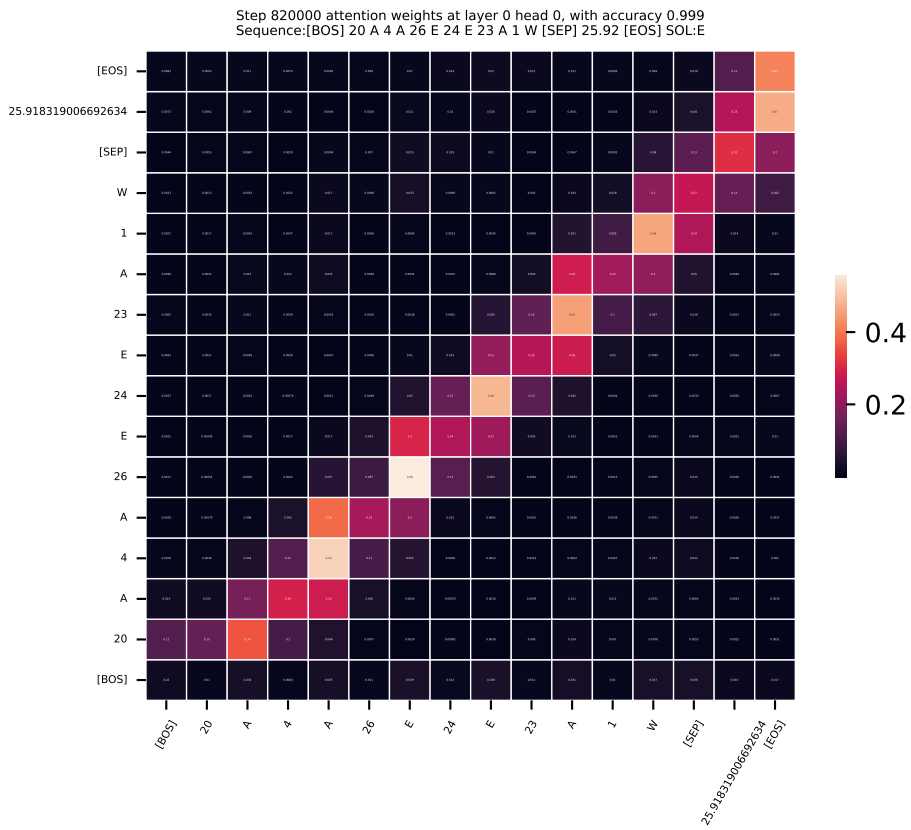

Step 819500 attention weights at layer 0 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

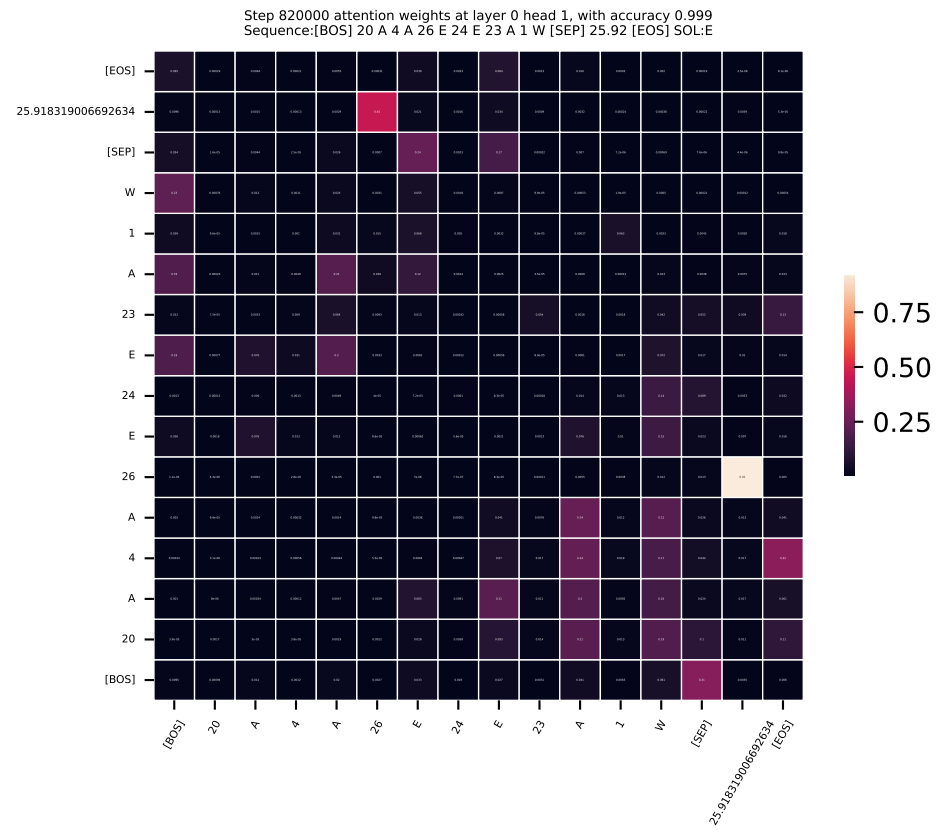

Step 819500 attention weights at layer 1 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

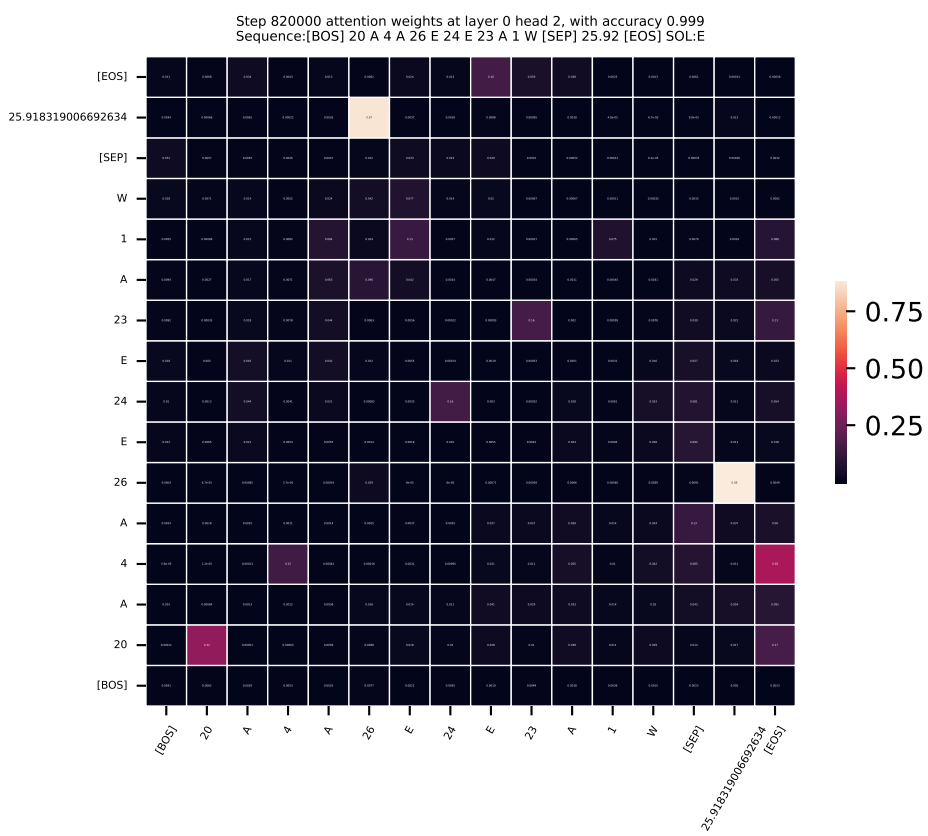

Step 819500 attention weights at layer 1 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

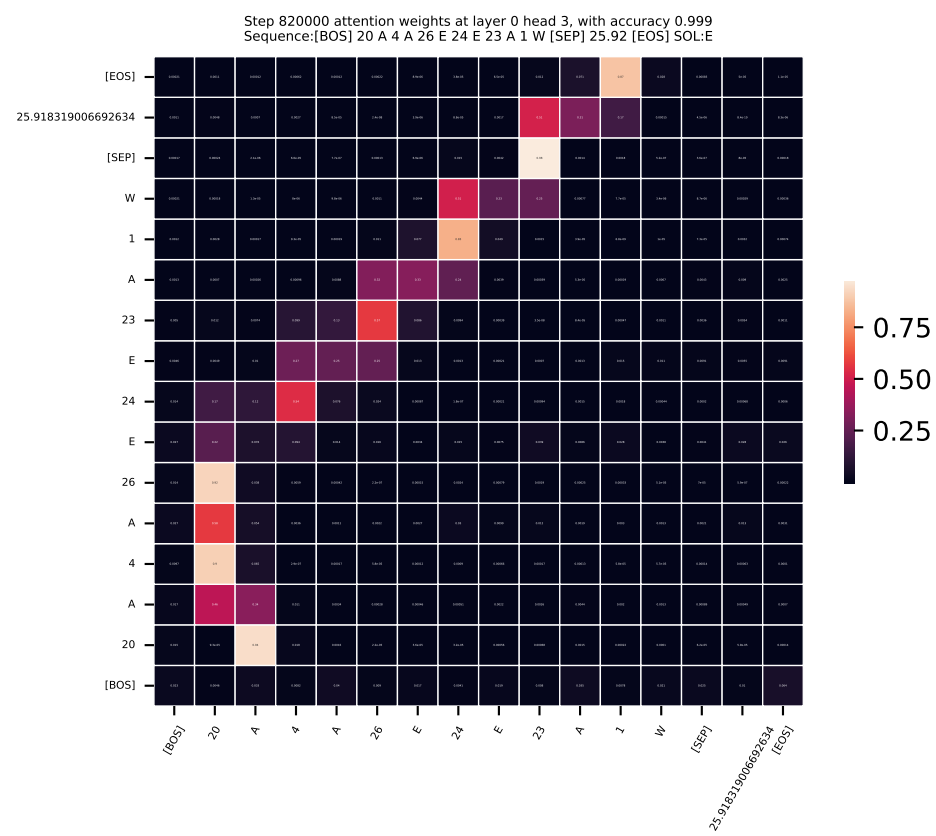

Step 819500 attention weights at layer 1 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

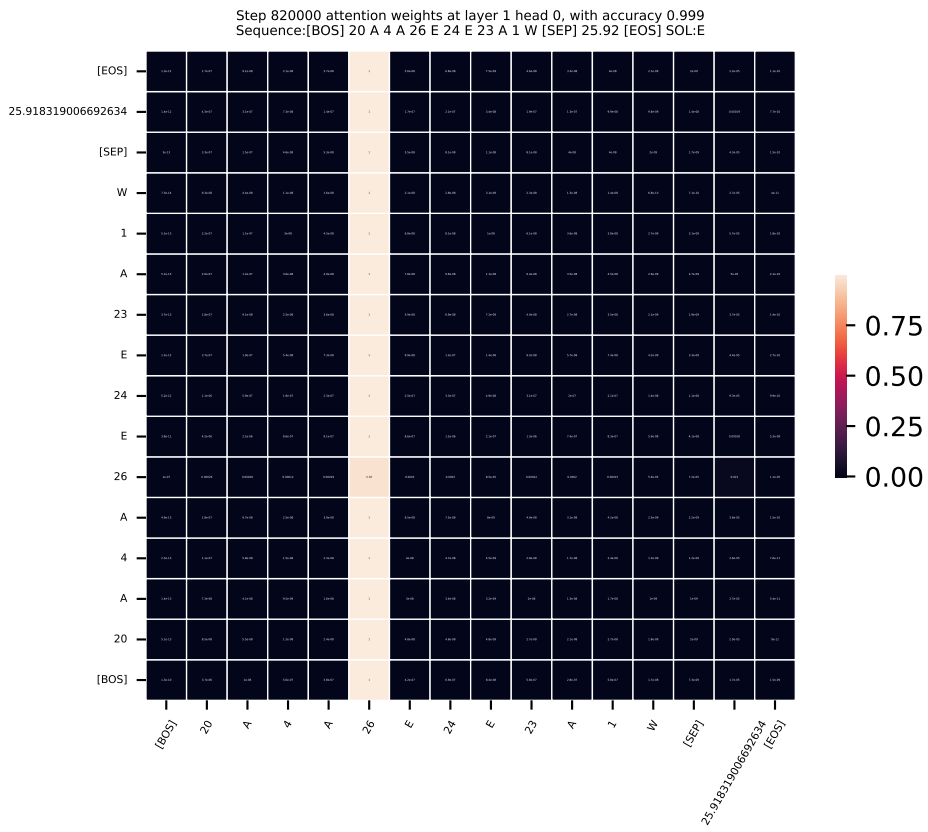

Step 819500 attention weights at layer 1 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

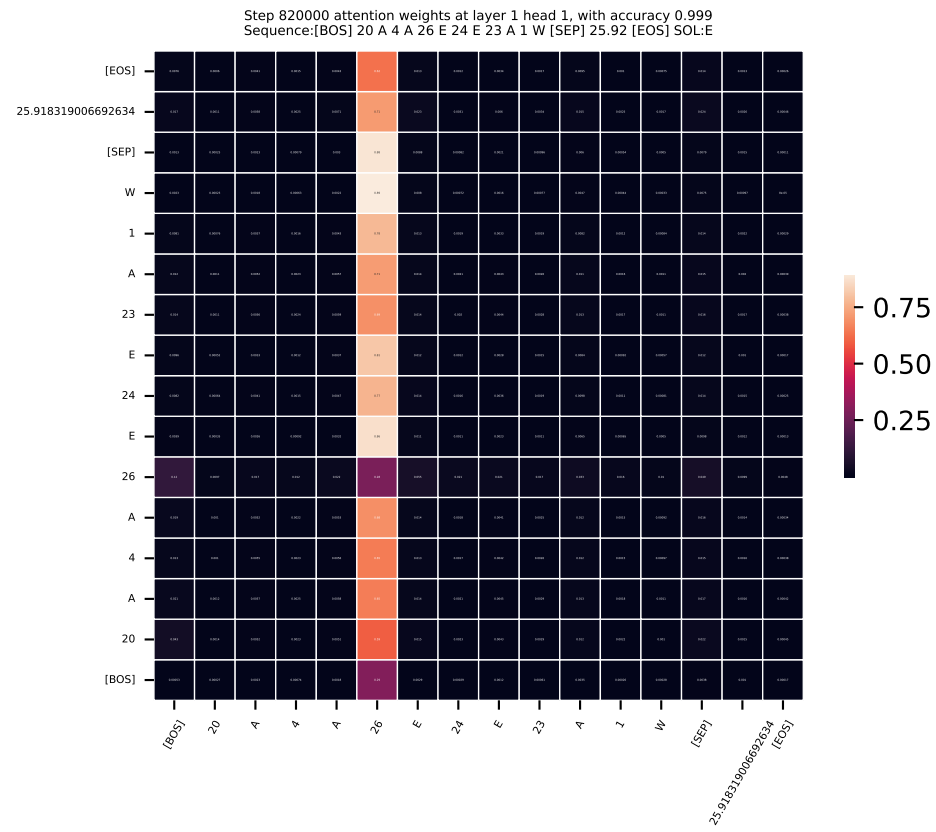

Step 820000 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

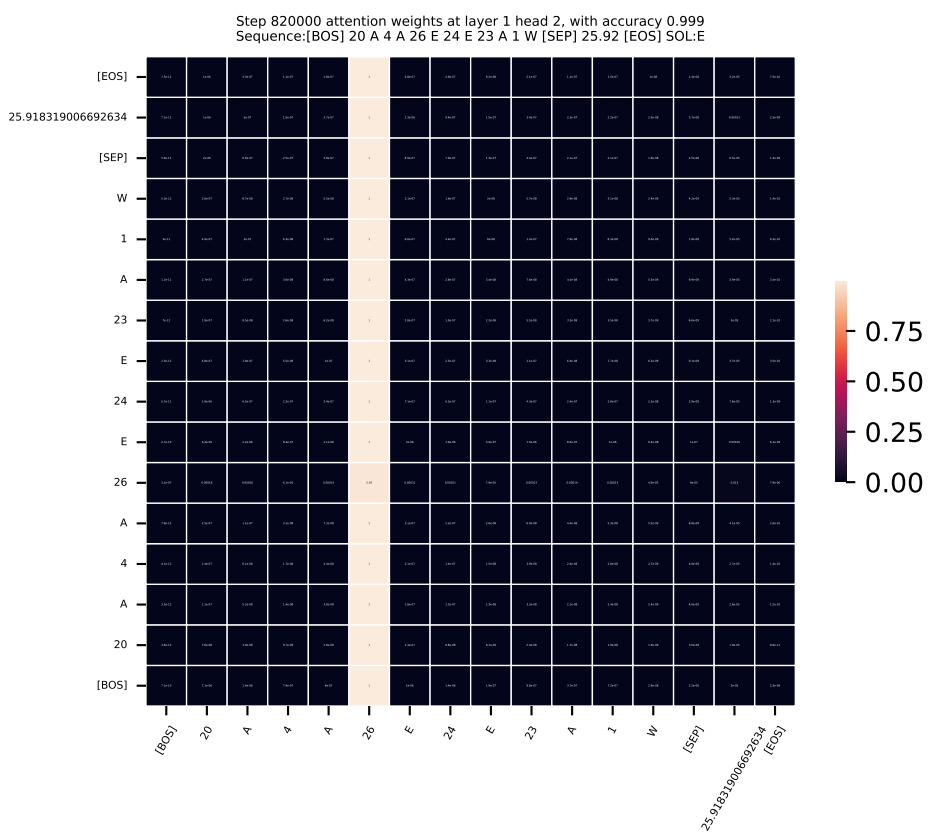

Step 820000 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

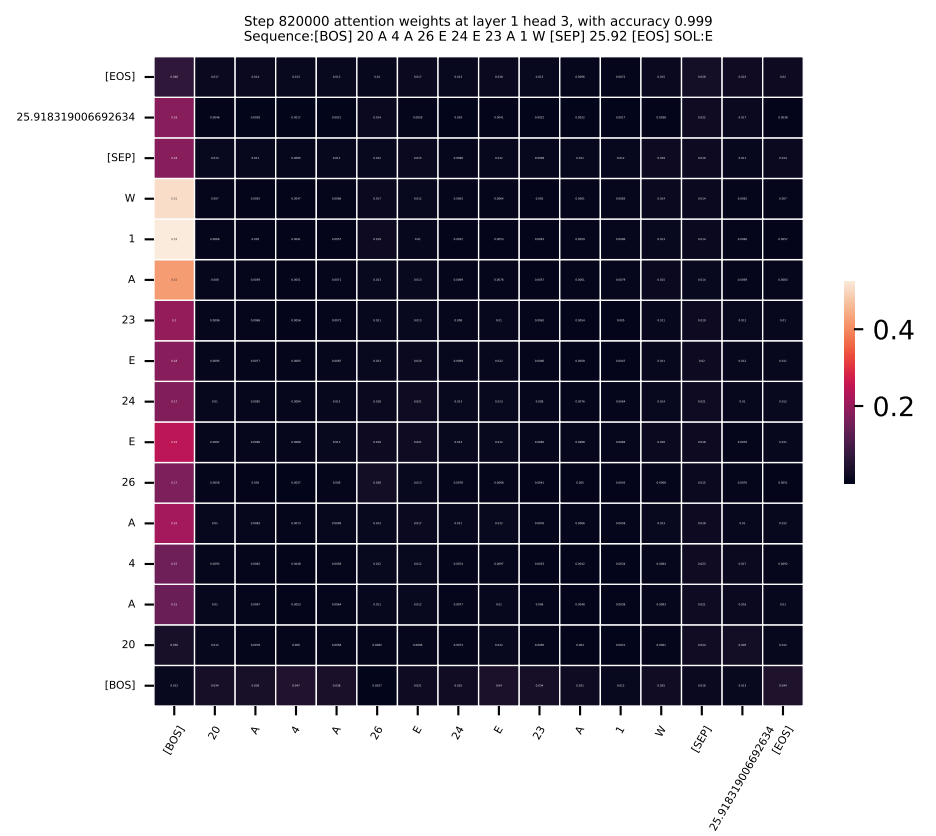

Step 820000 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

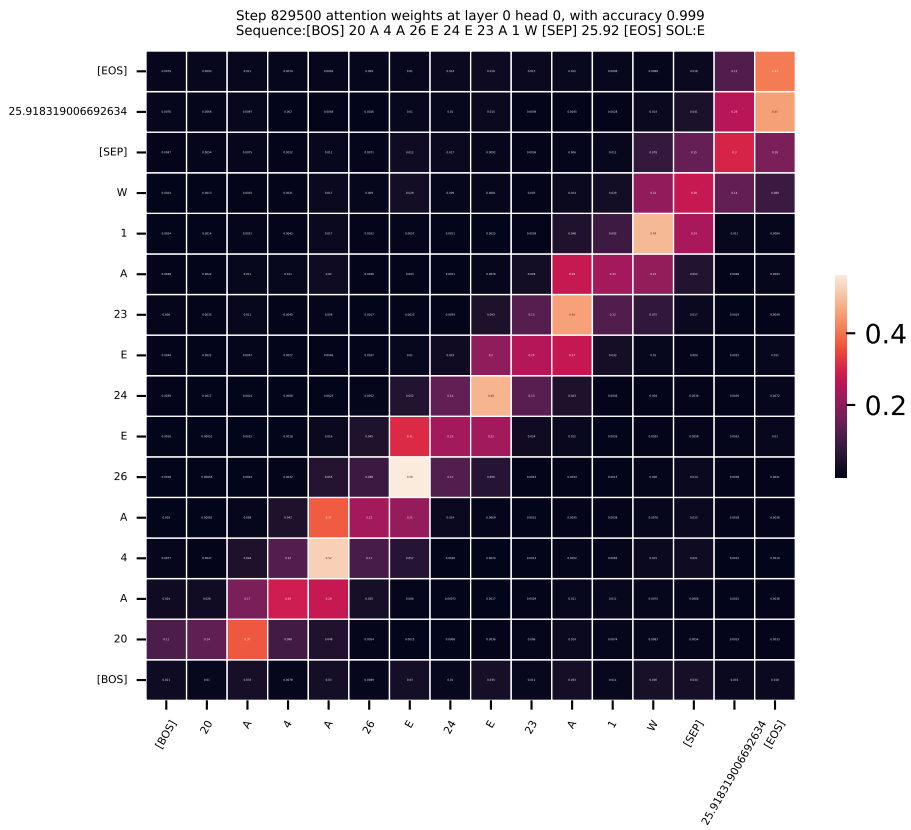

Step 820000 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

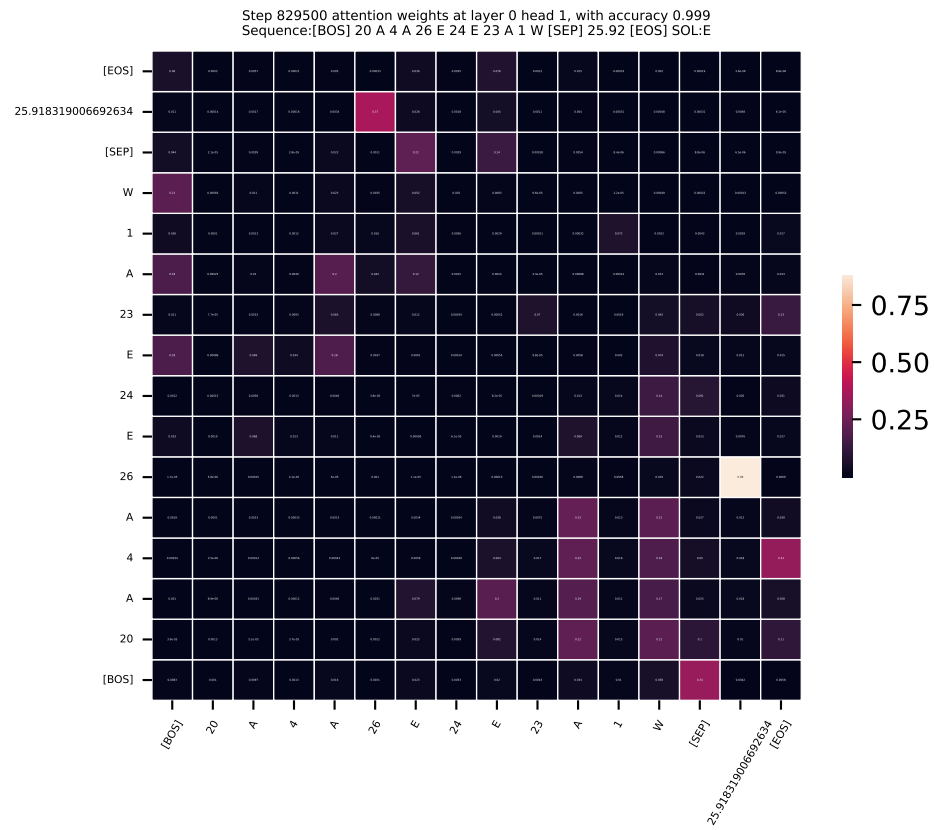

Step 820000 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25
0.00

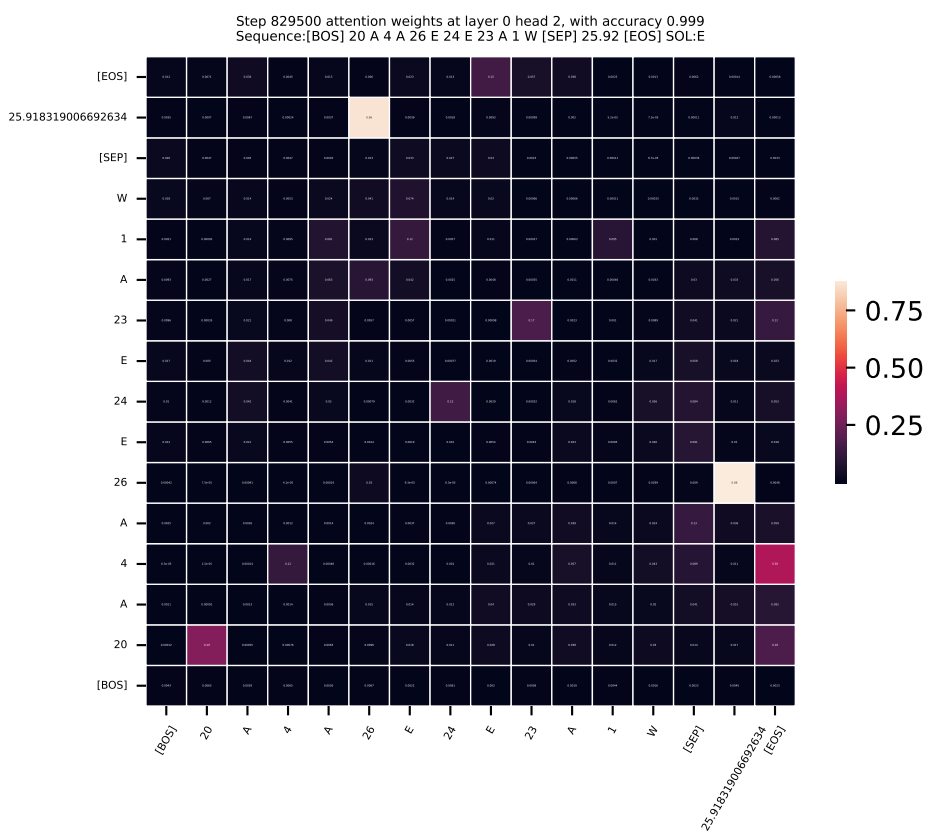

Step 820000 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25

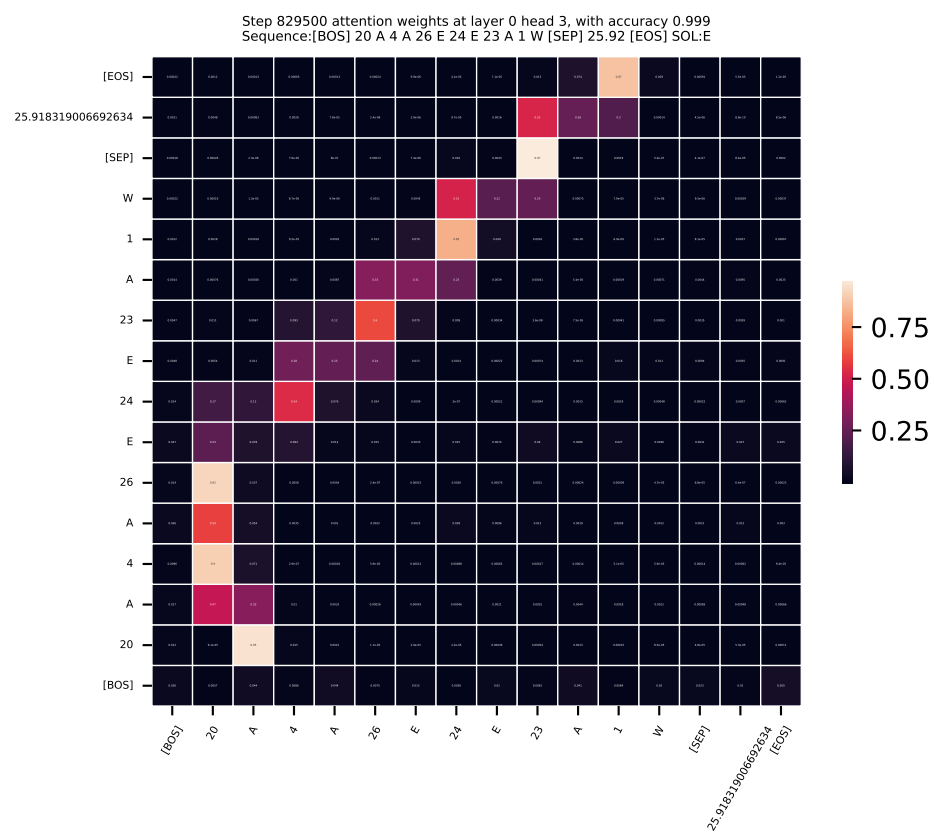

Step 820000 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.75
0.50
0.25
0.00

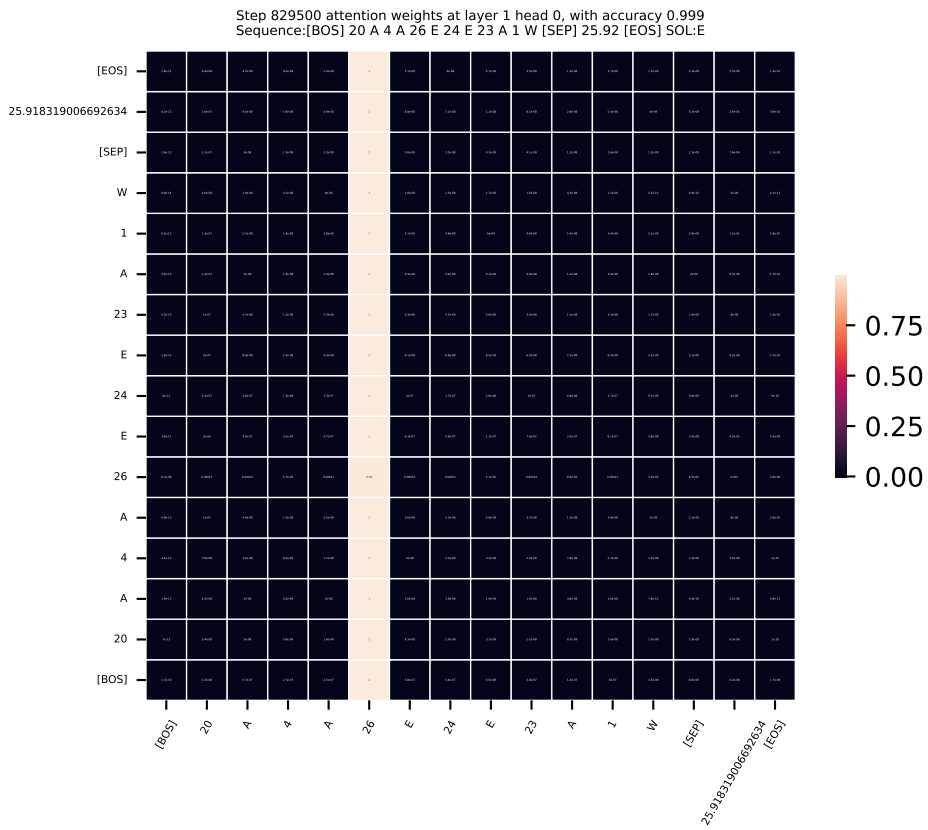

Step 820000 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS]
0.4
0.2

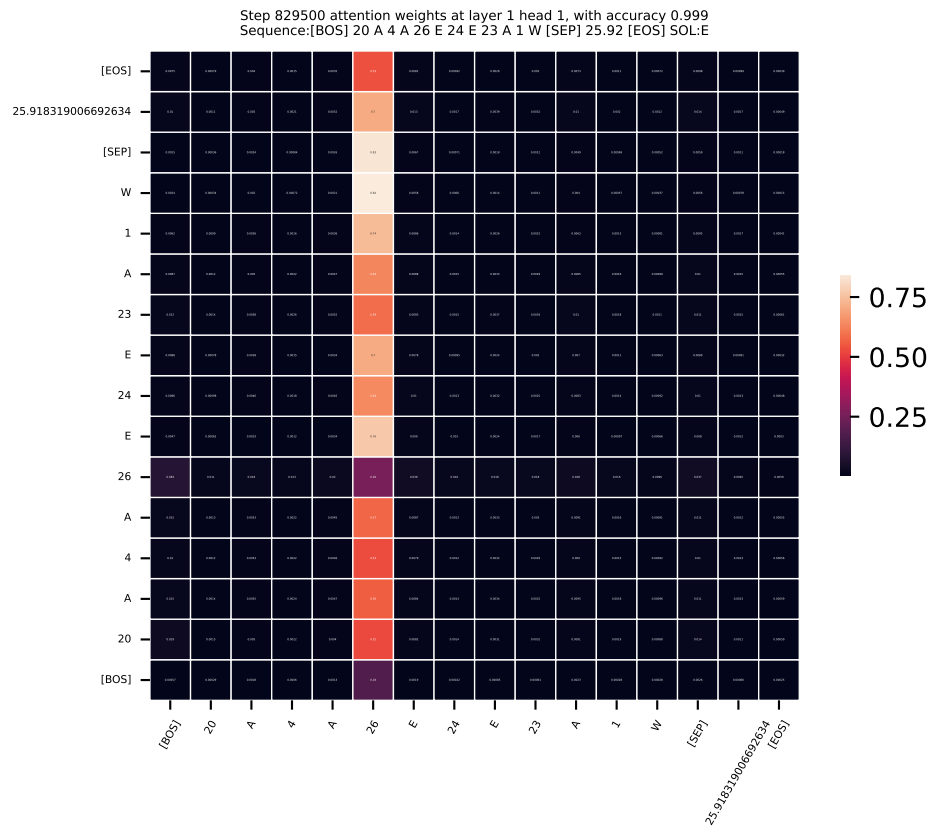

Step 829500 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

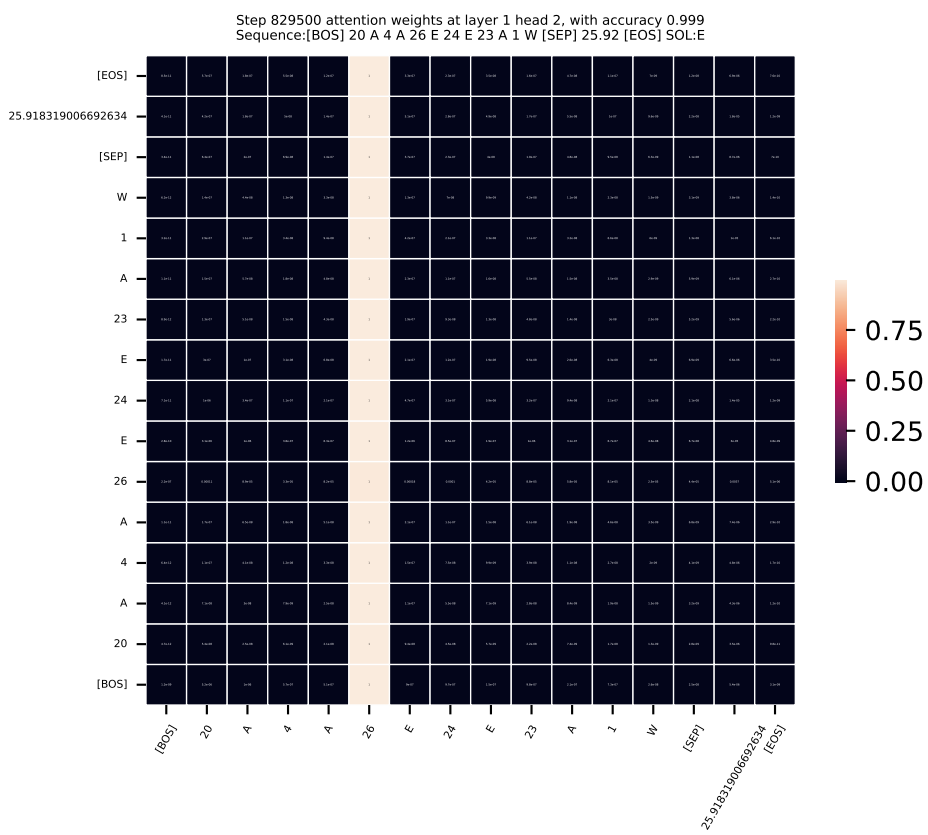

Step 829500 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

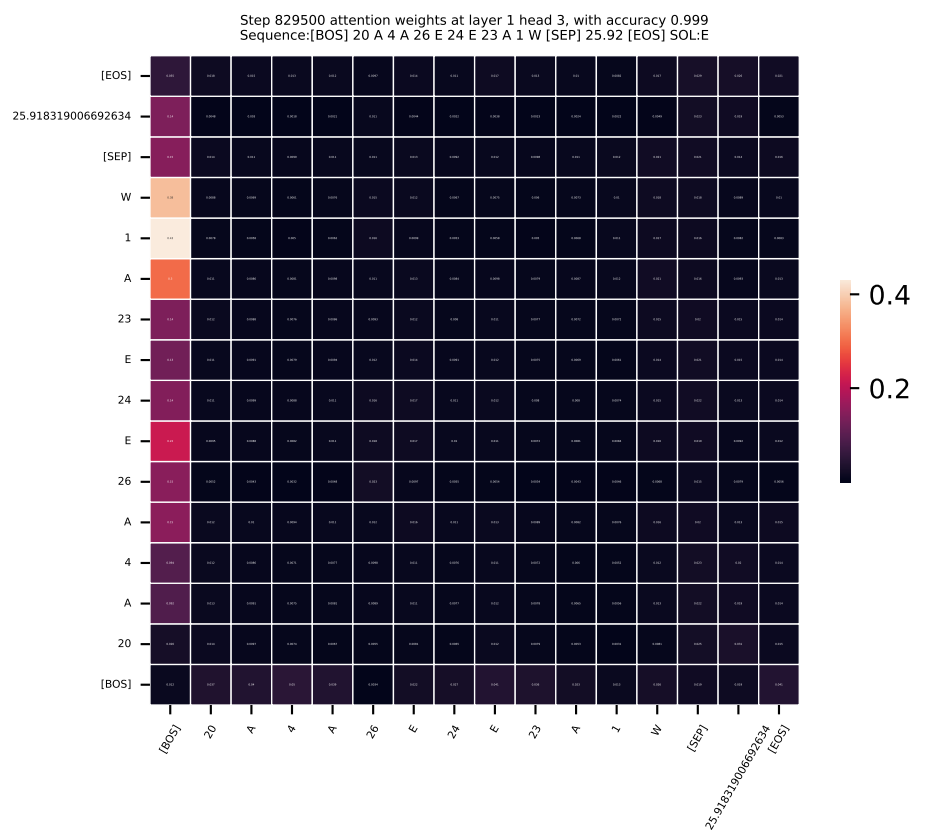

Step 829500 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

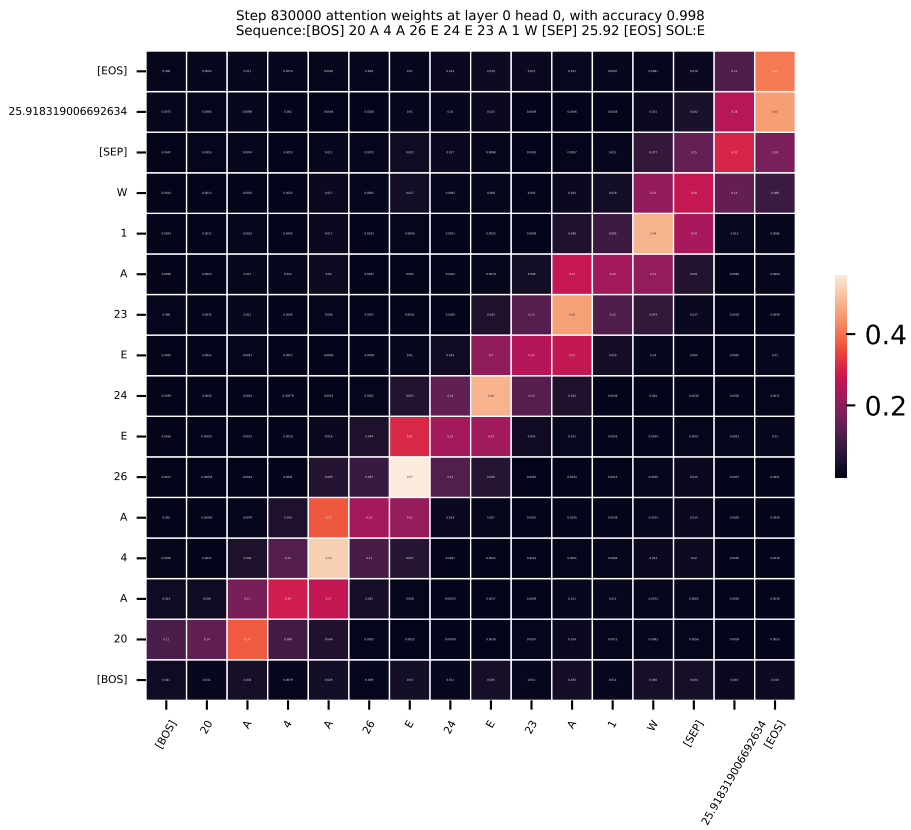

Step 829500 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

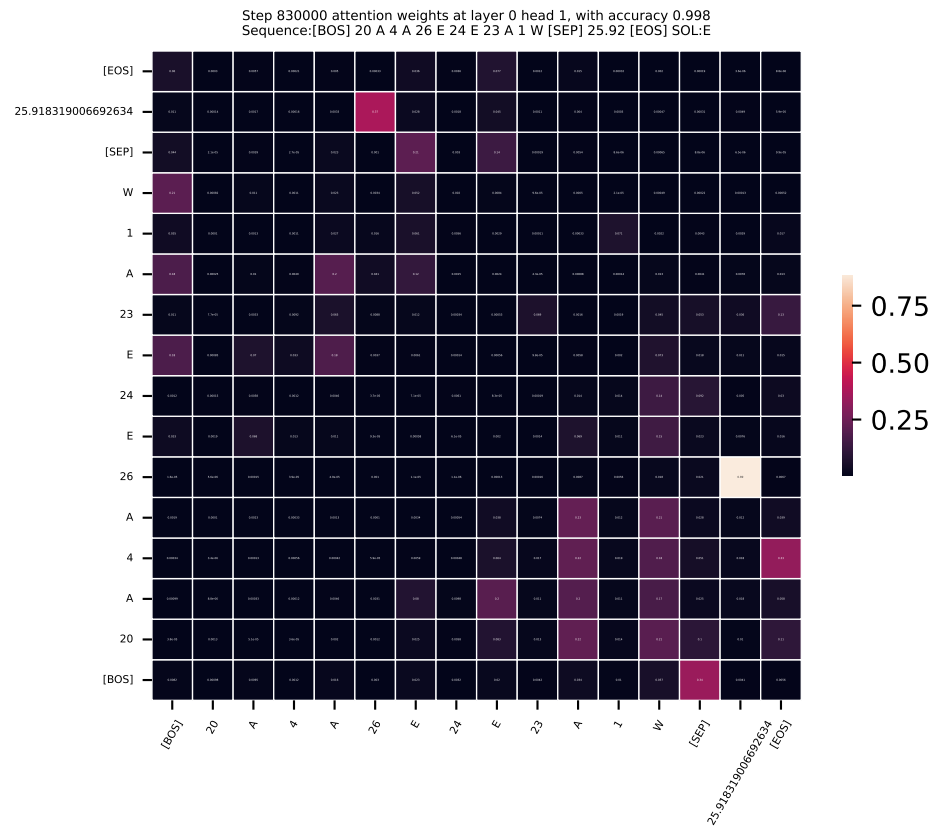

Step 829500 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

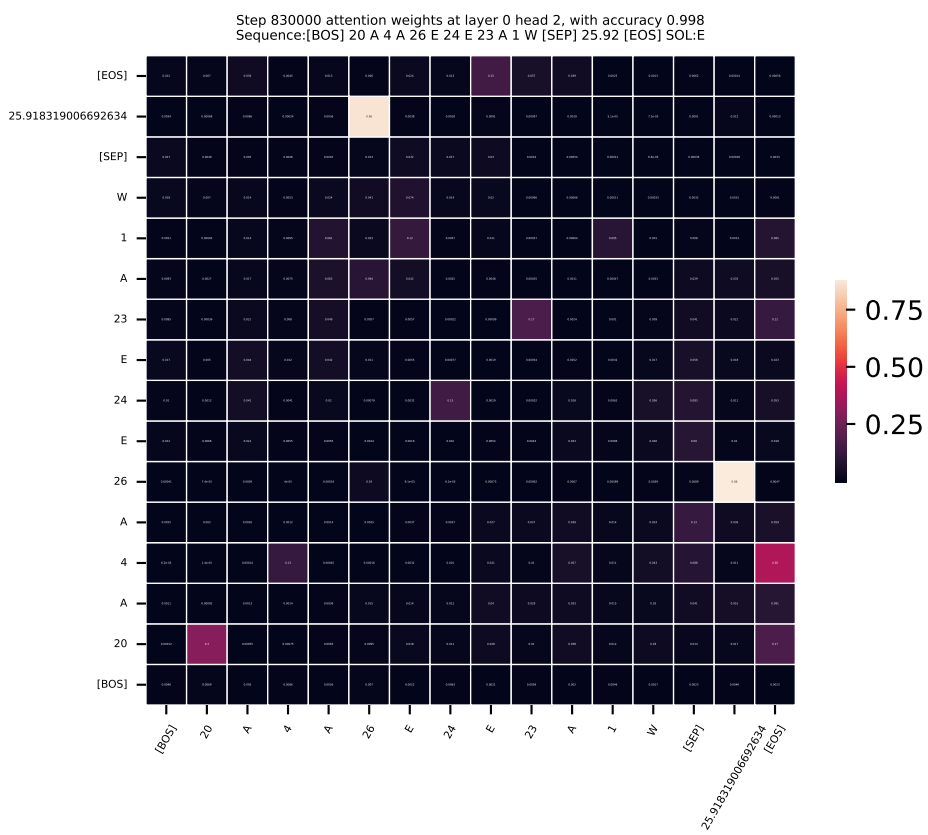

Step 829500 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

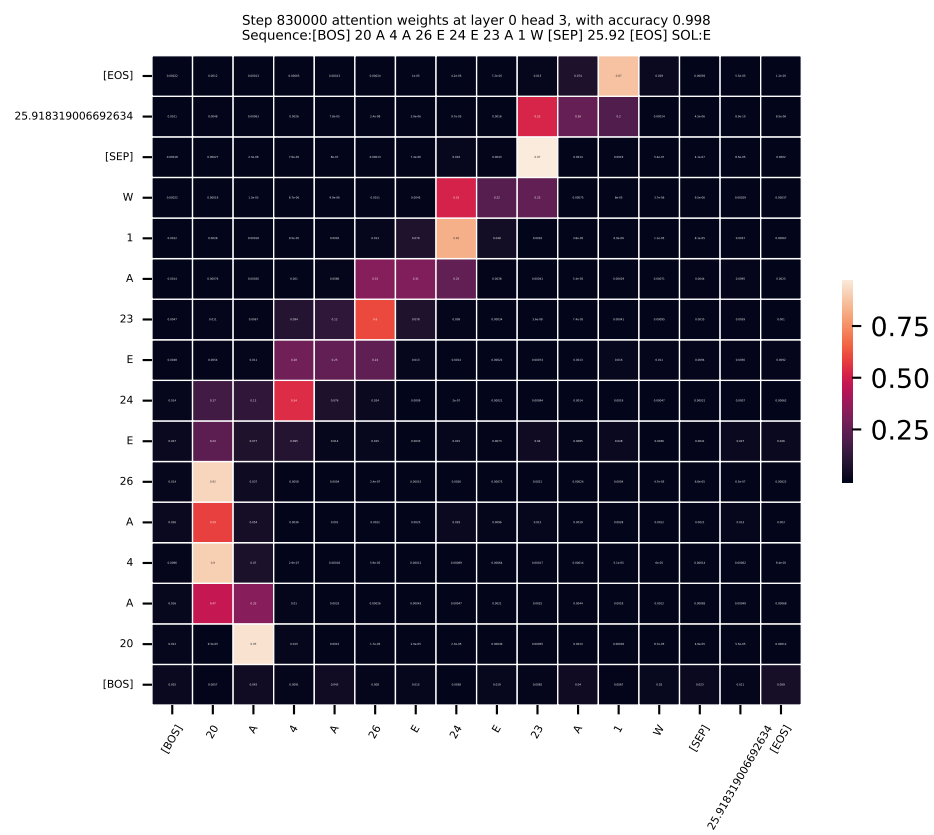

Step 829500 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

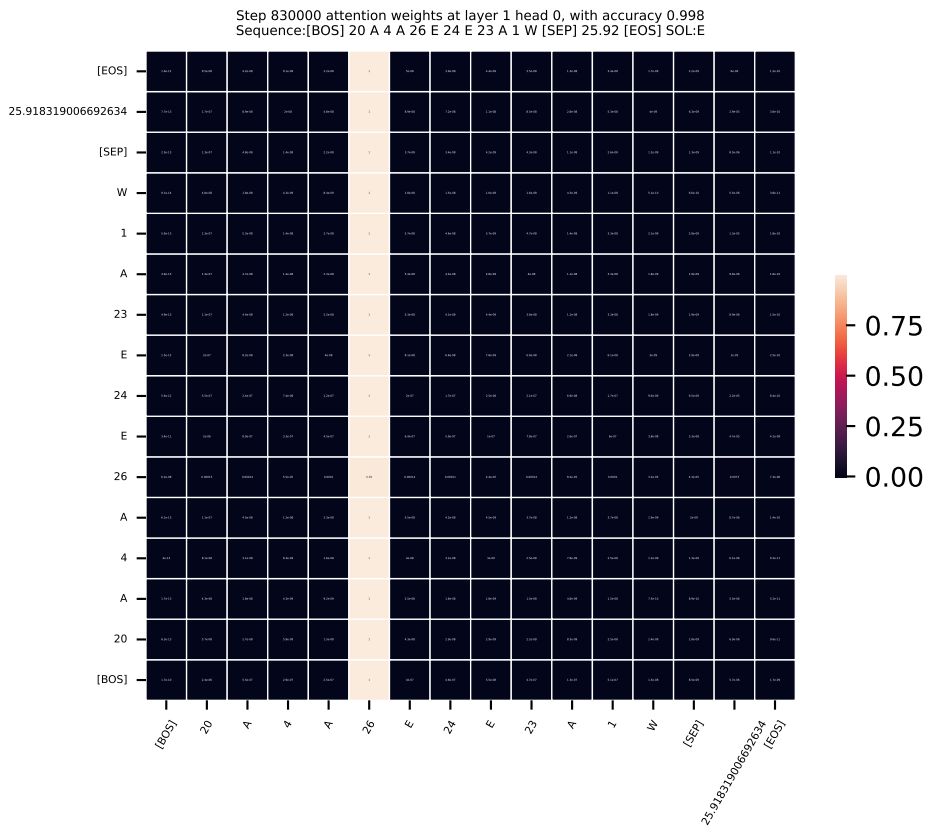

Step 829500 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

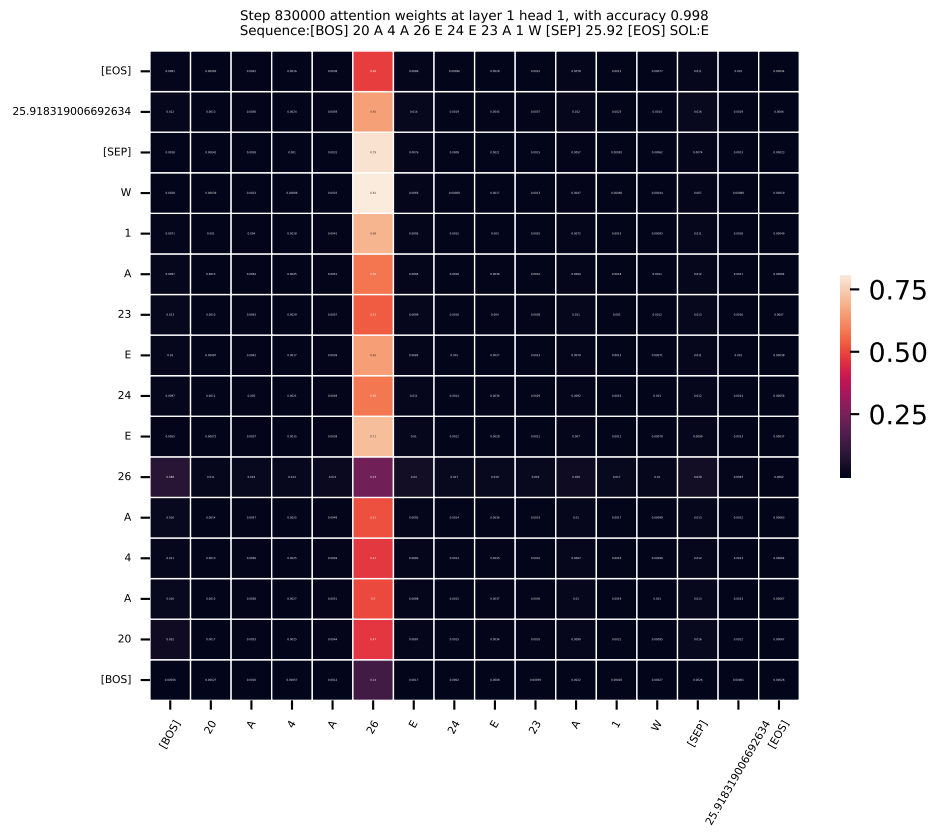

Step 830000 attention weights at layer 0 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

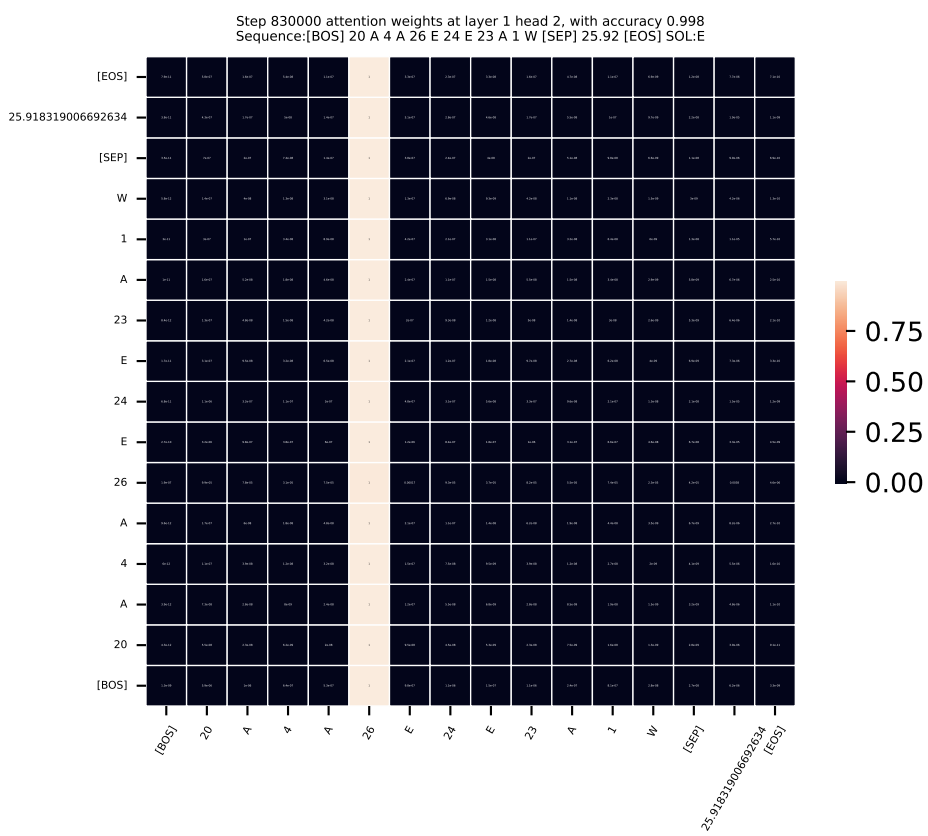

Step 830000 attention weights at layer 0 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

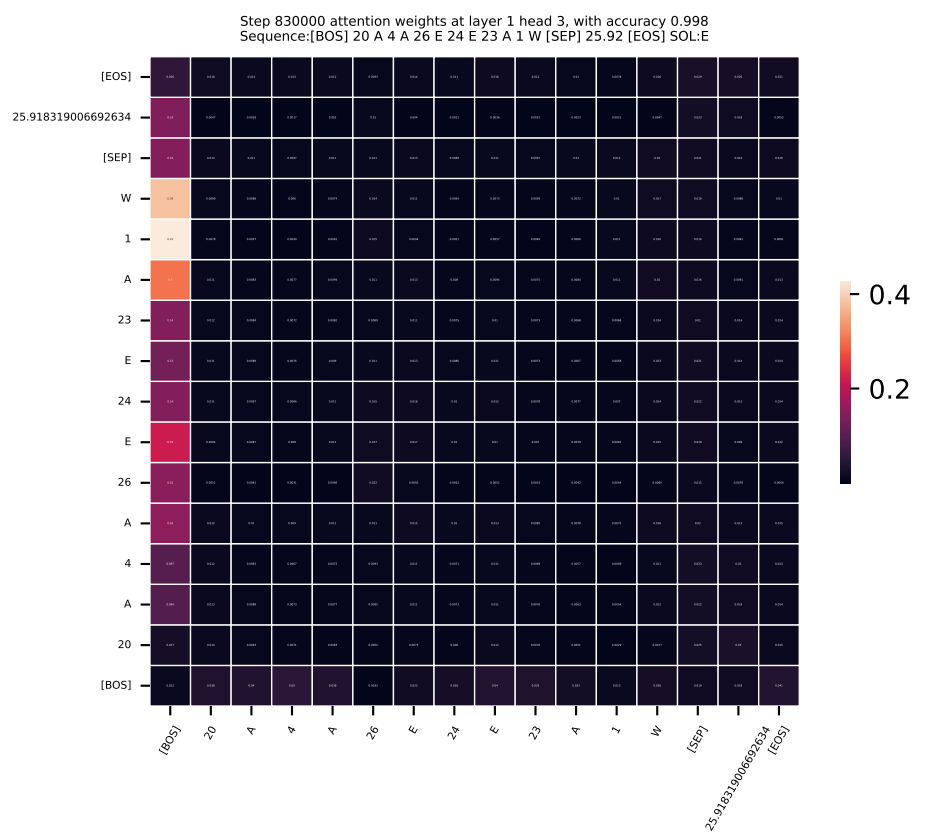

Step 830000 attention weights at layer 0 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

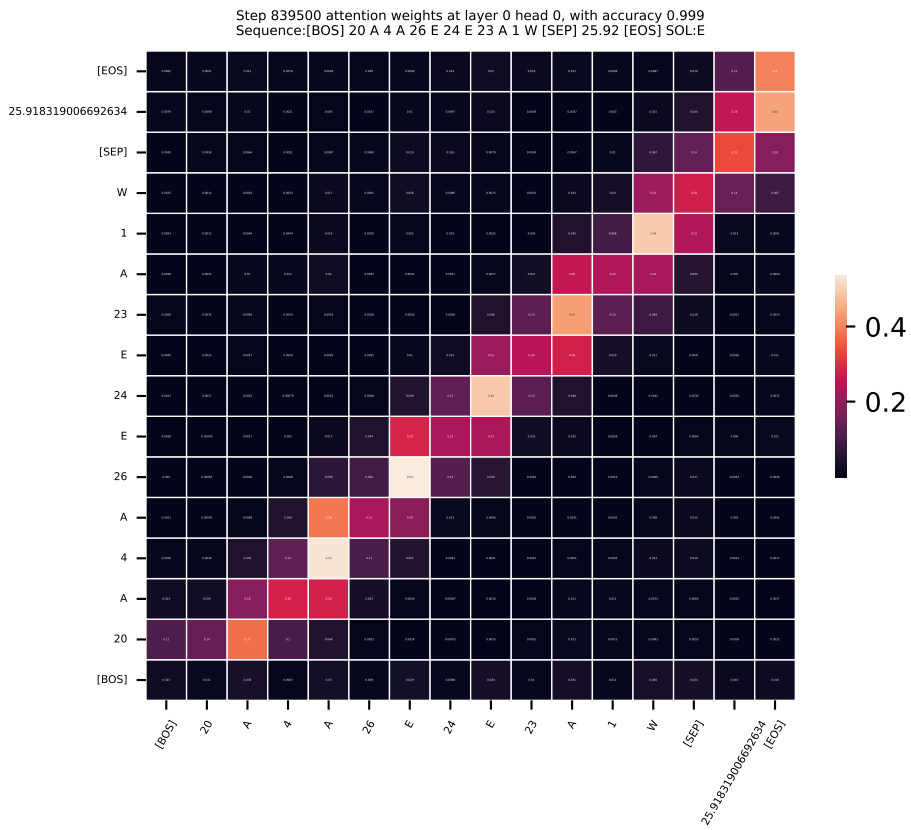

Step 830000 attention weights at layer 0 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

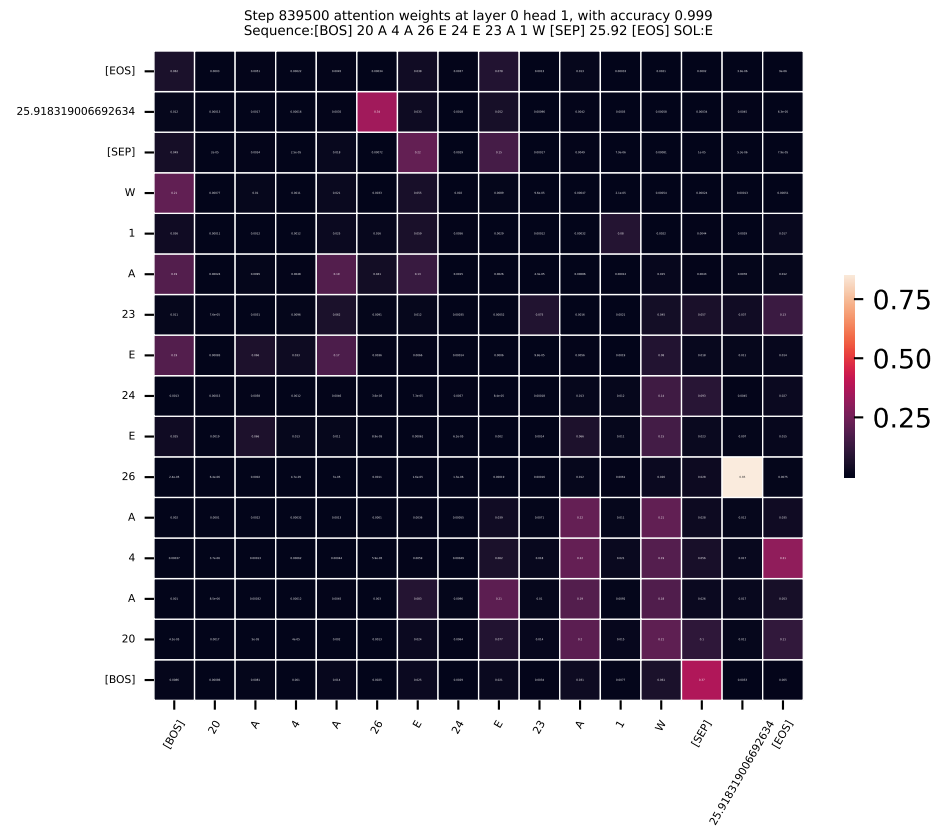

Step 830000 attention weights at layer 1 head 0, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

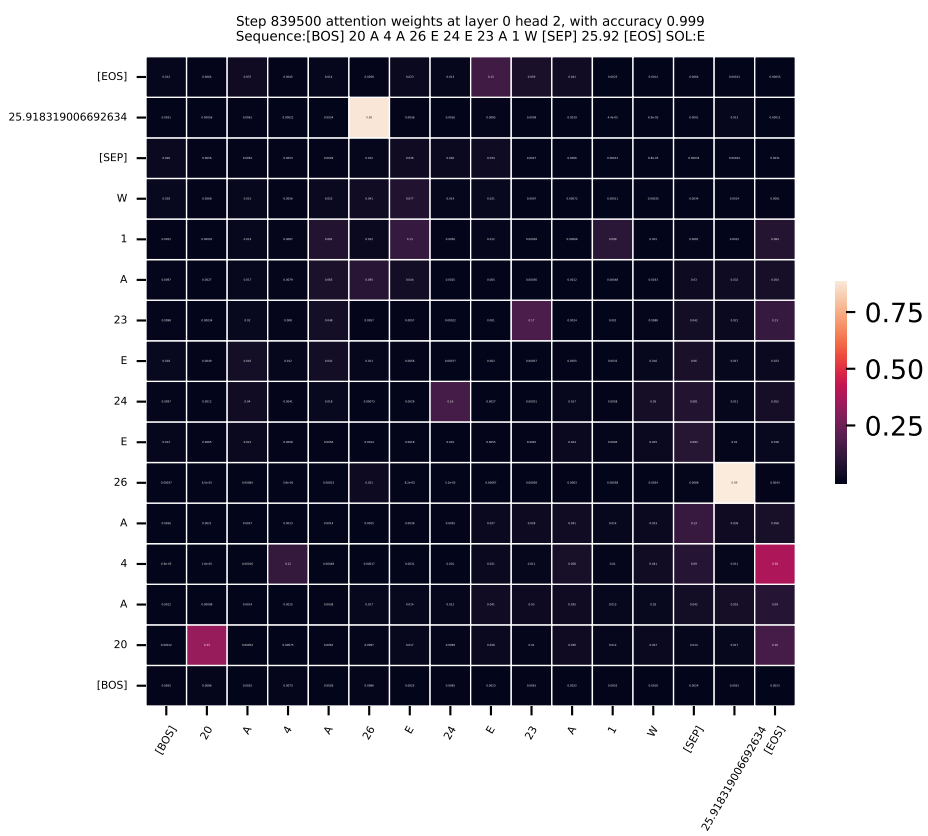

Step 830000 attention weights at layer 1 head 1, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

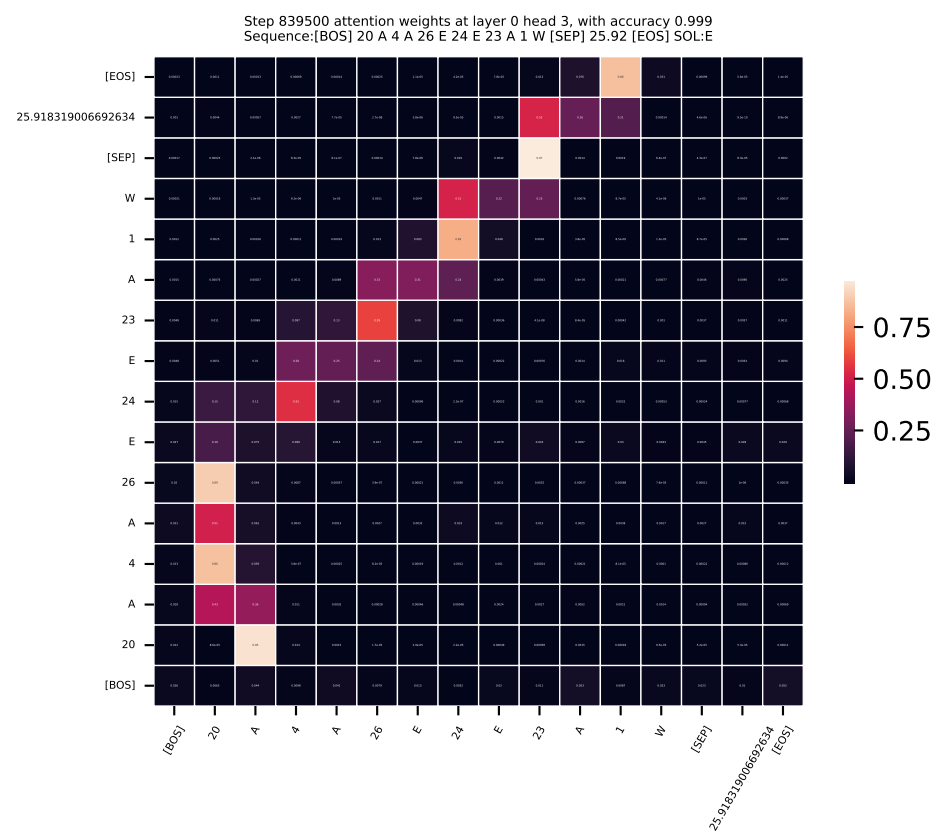

Step 830000 attention weights at layer 1 head 2, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

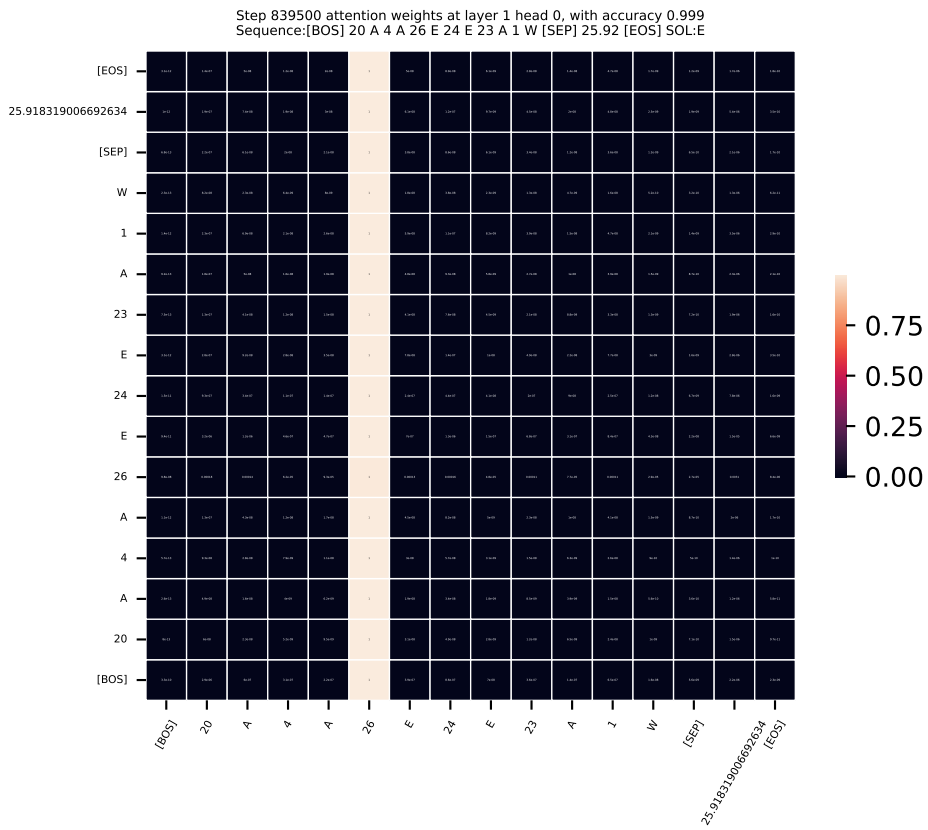

Step 830000 attention weights at layer 1 head 3, with accuracy 0.998
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

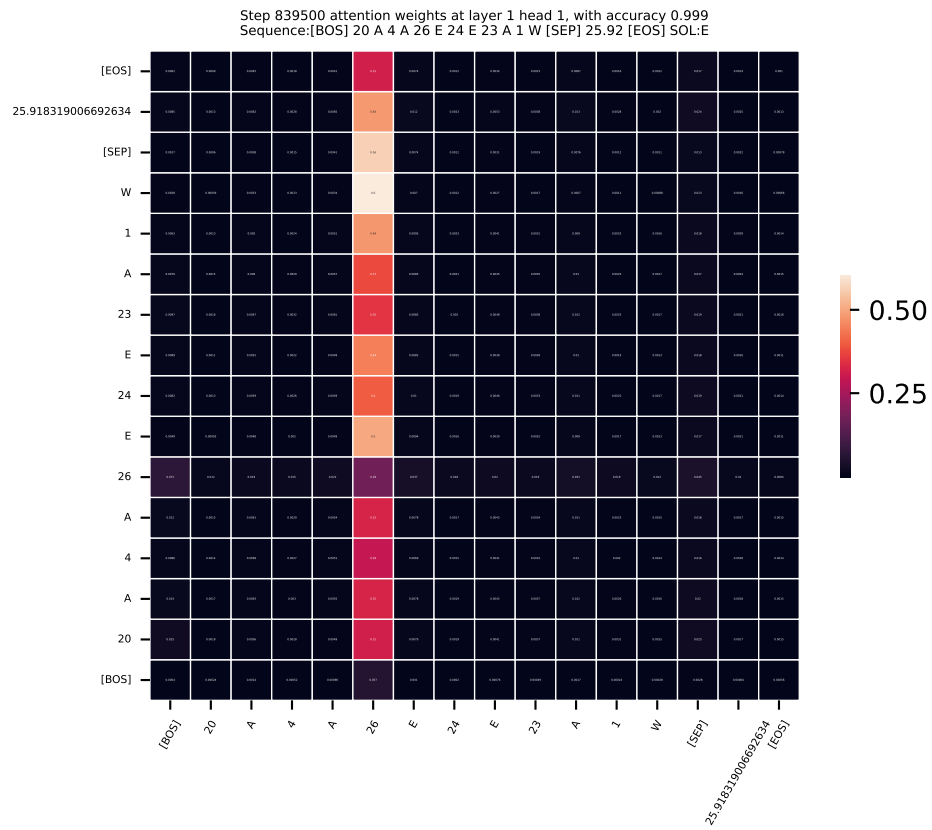

Step 839500 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

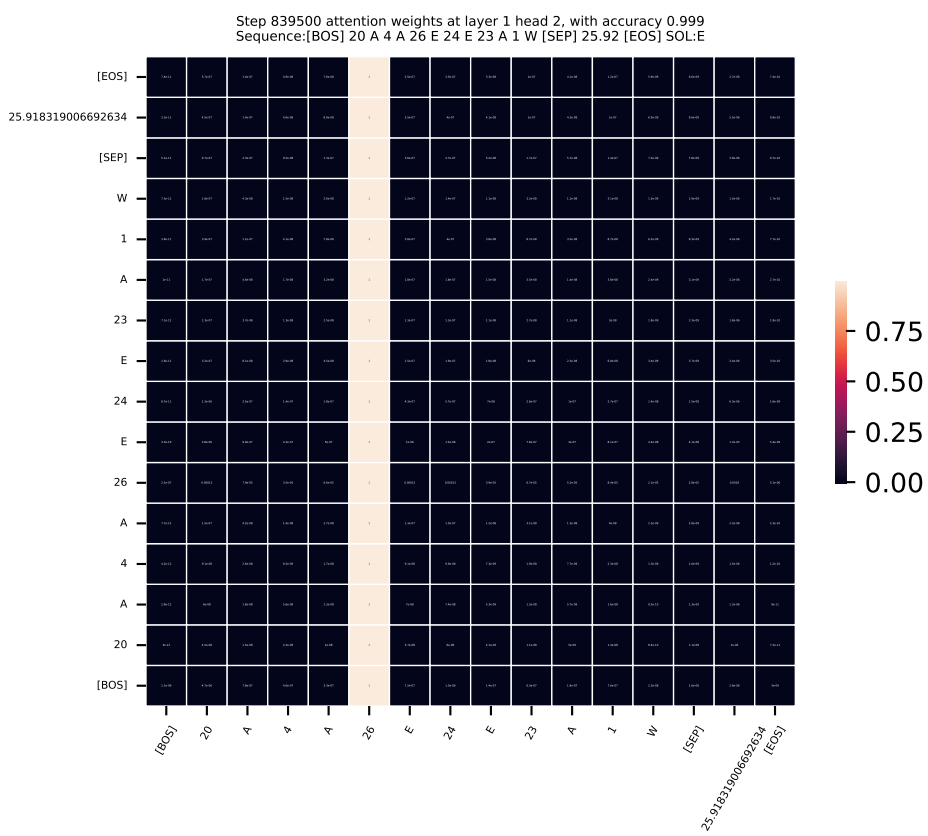

Step 839500 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

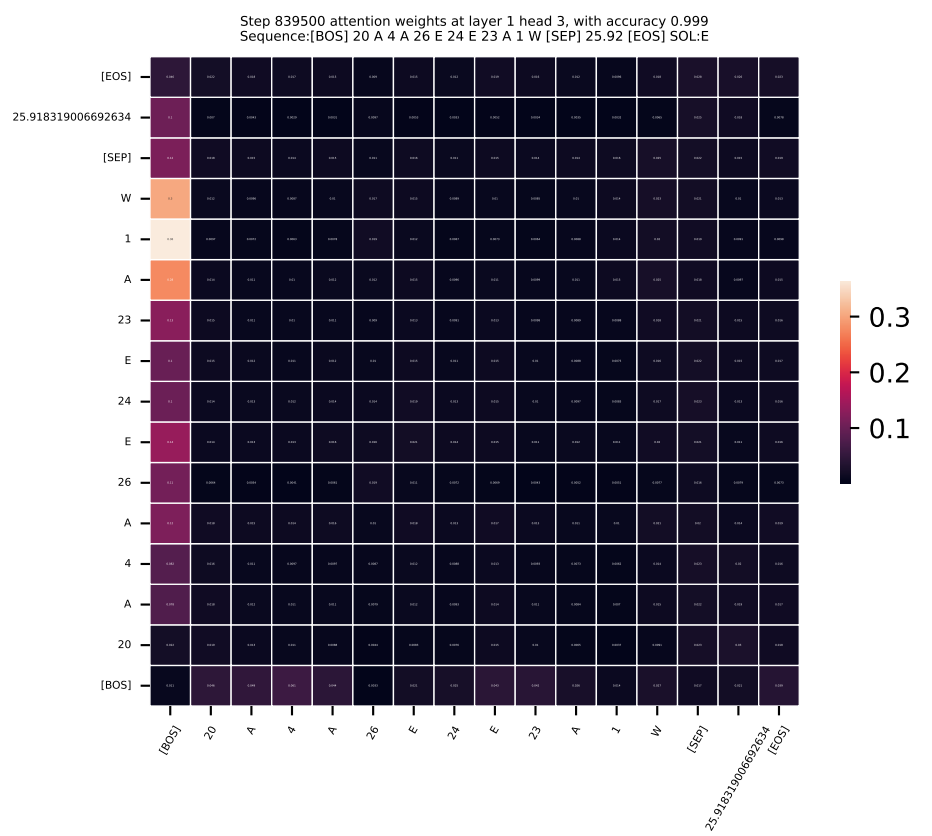

Step 839500 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

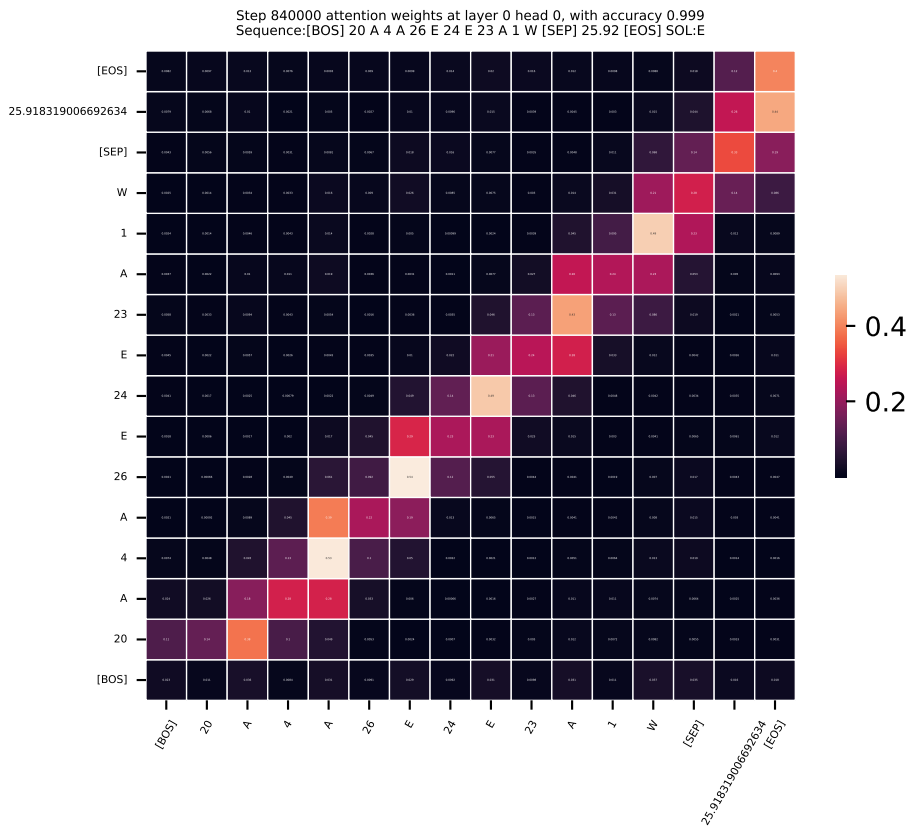

Step 839500 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

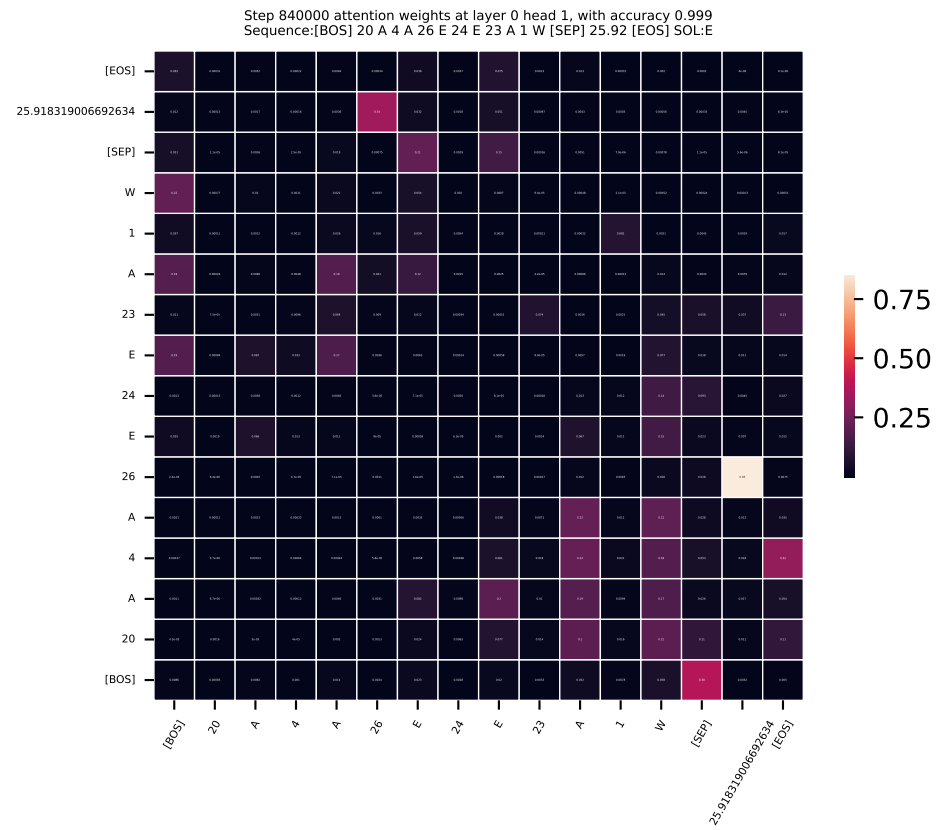

Step 839500 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

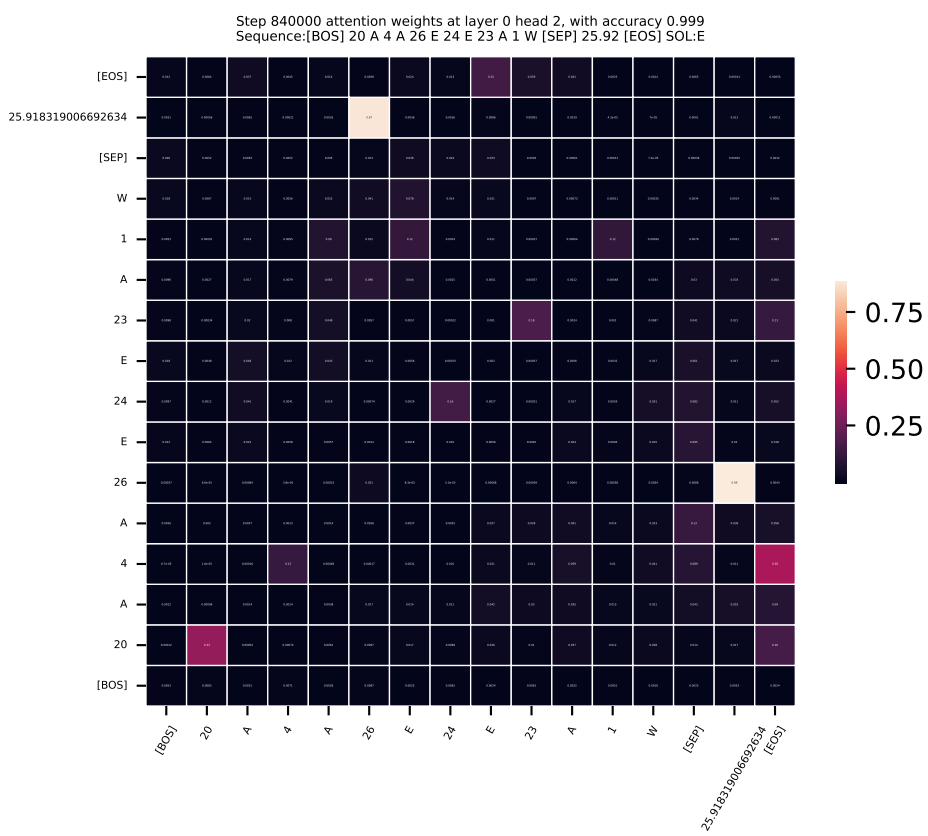

Step 839500 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

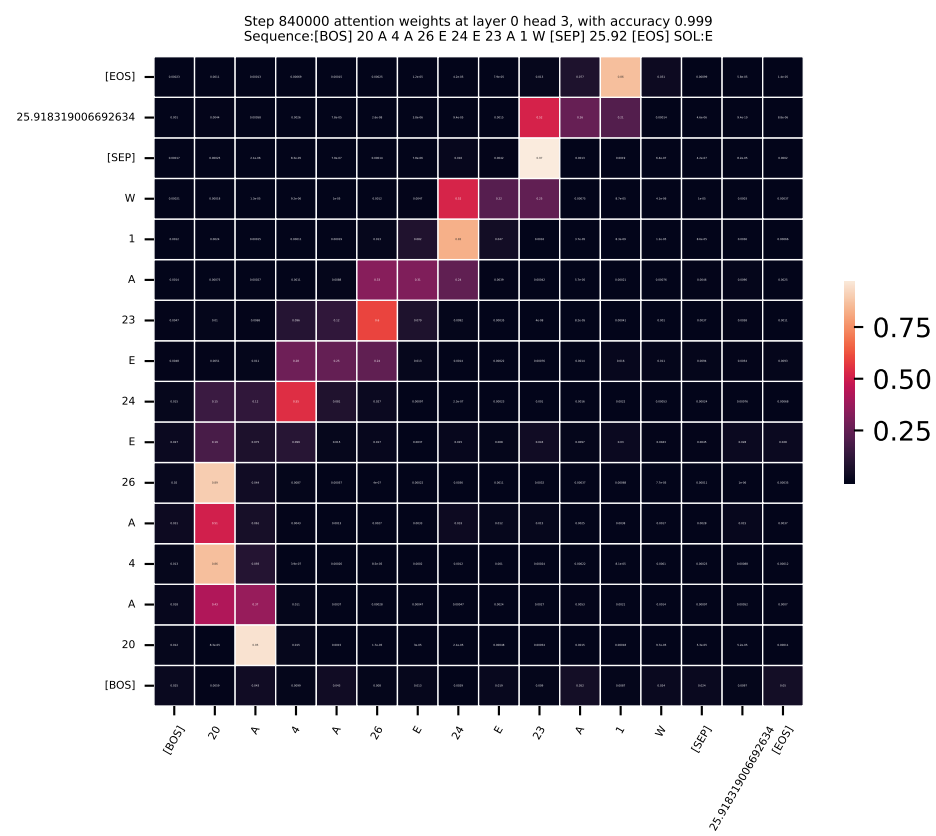

Step 839500 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

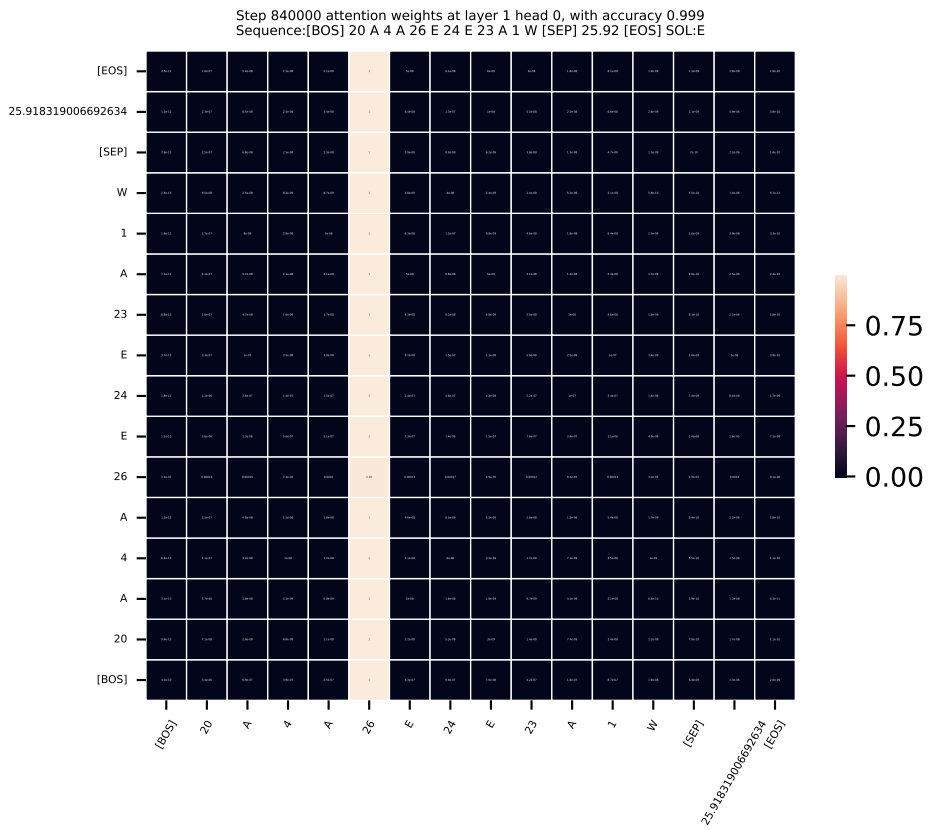

Step 839500 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

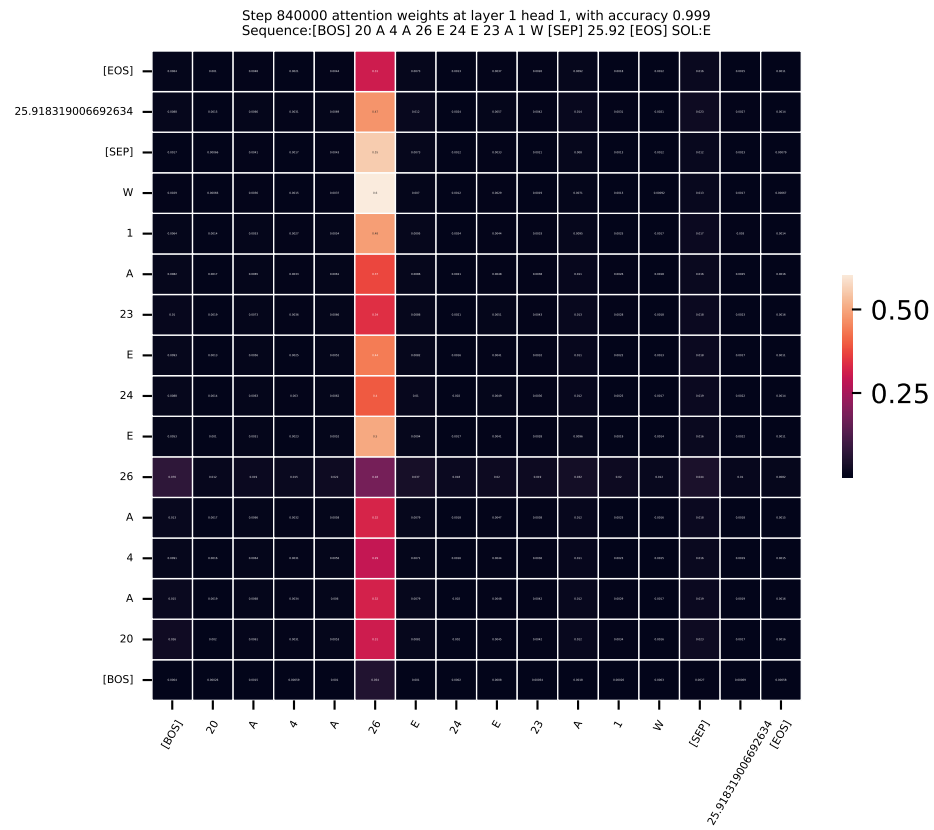

Step 840000 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

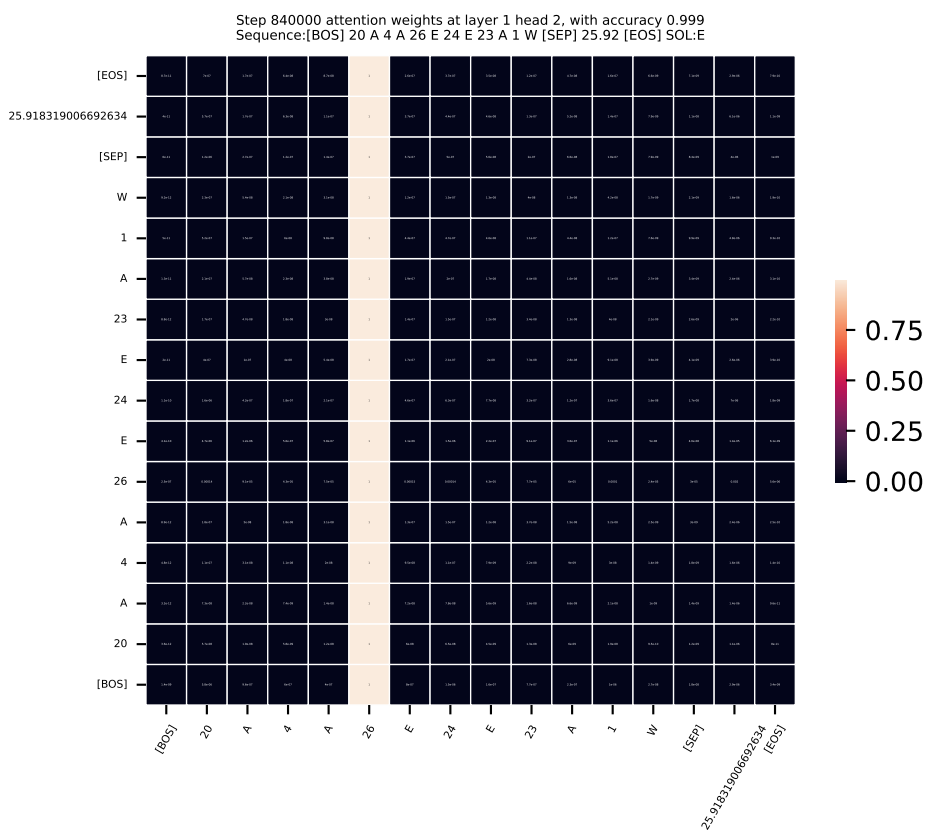

Step 840000 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

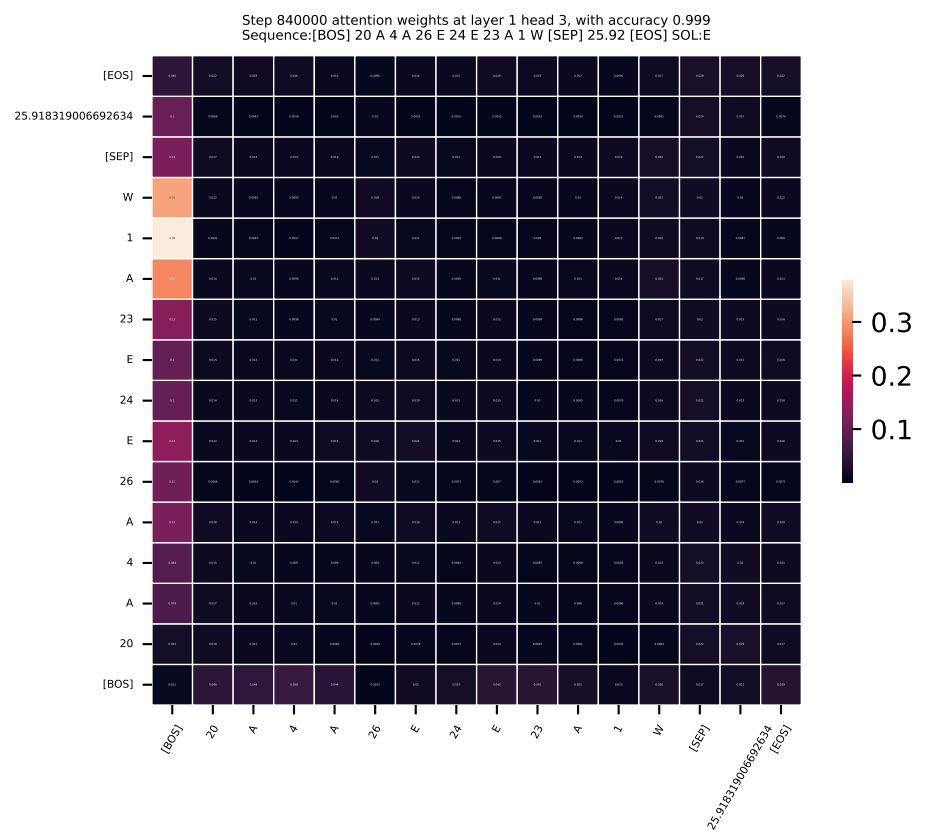

Step 840000 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

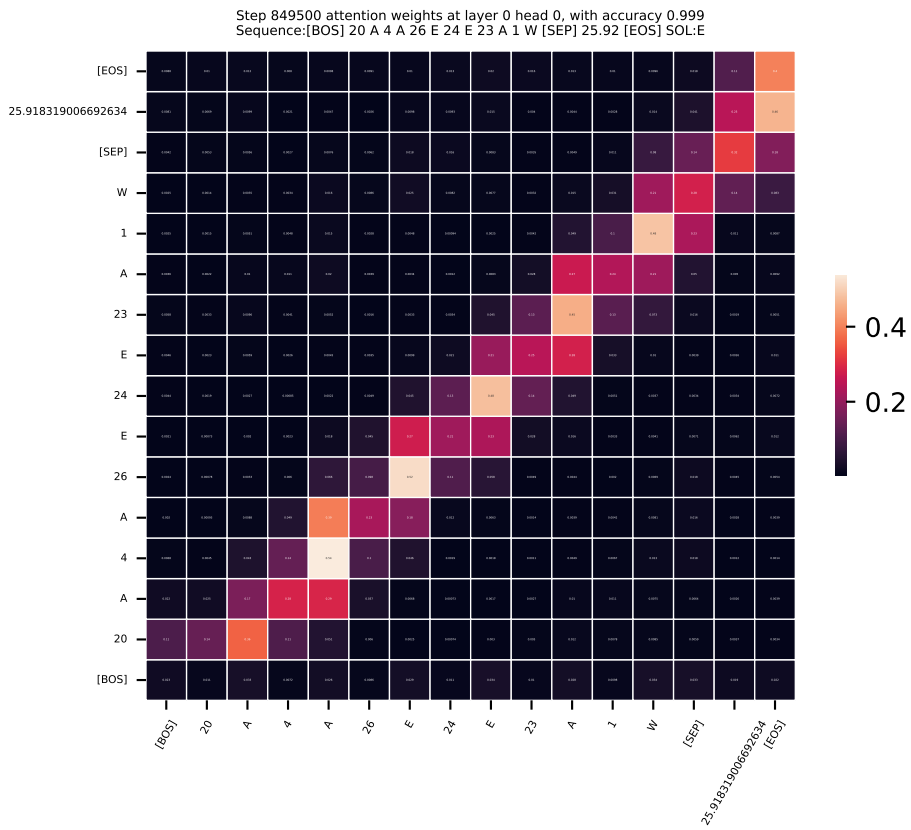

Step 840000 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

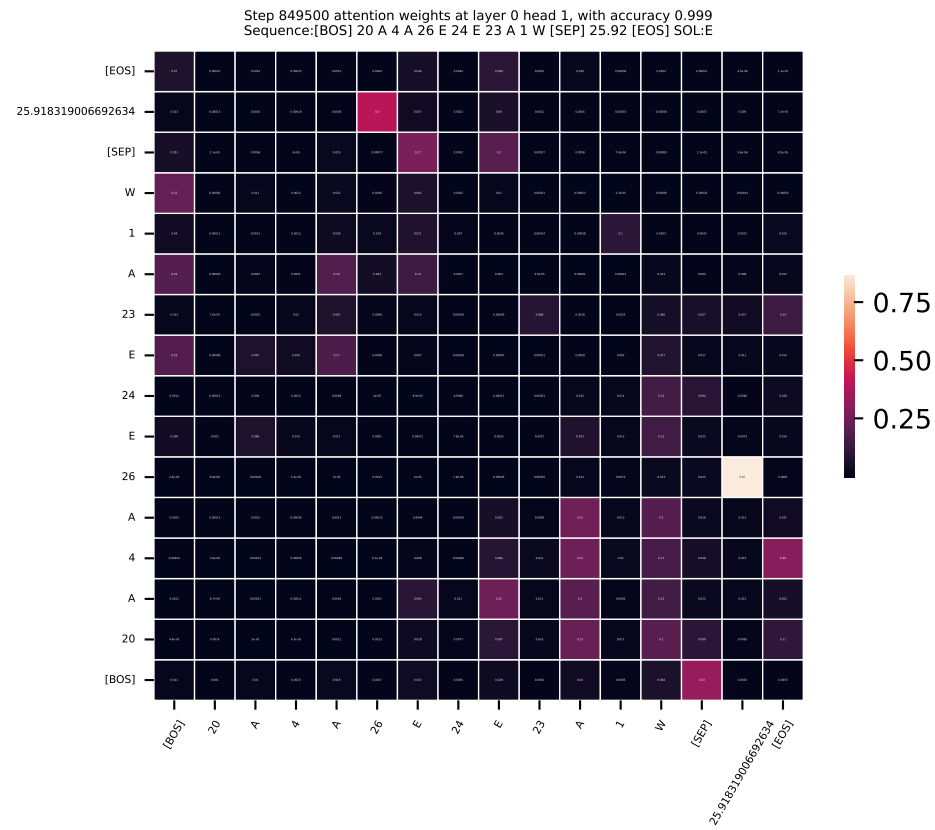

Step 840000 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

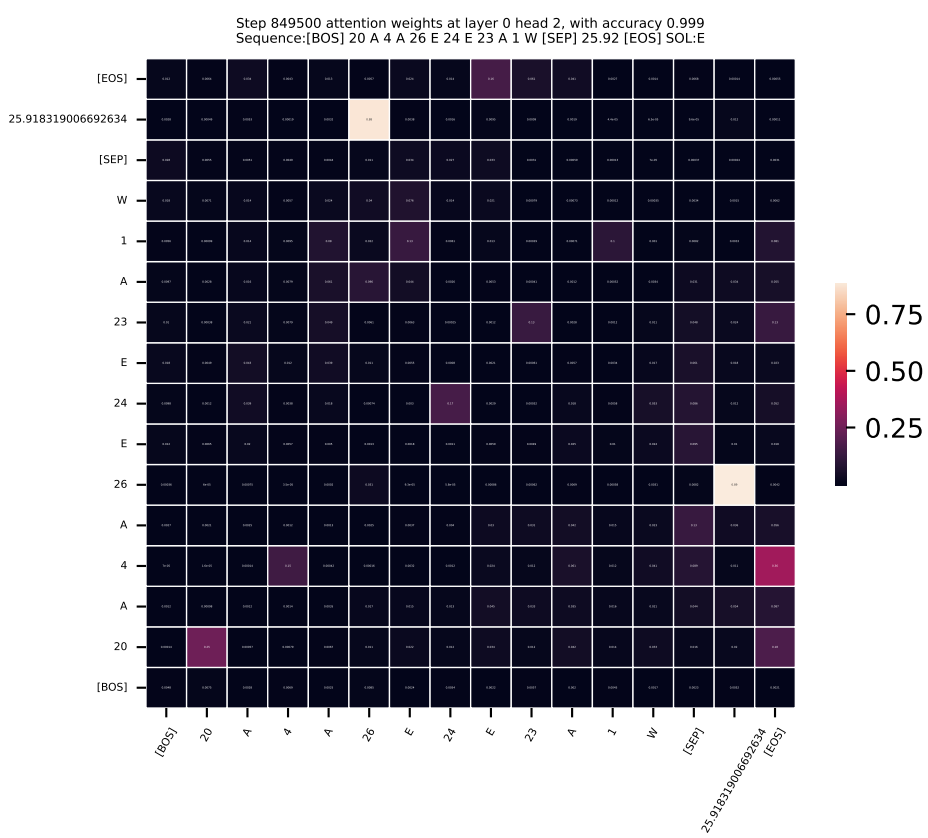

Step 840000 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

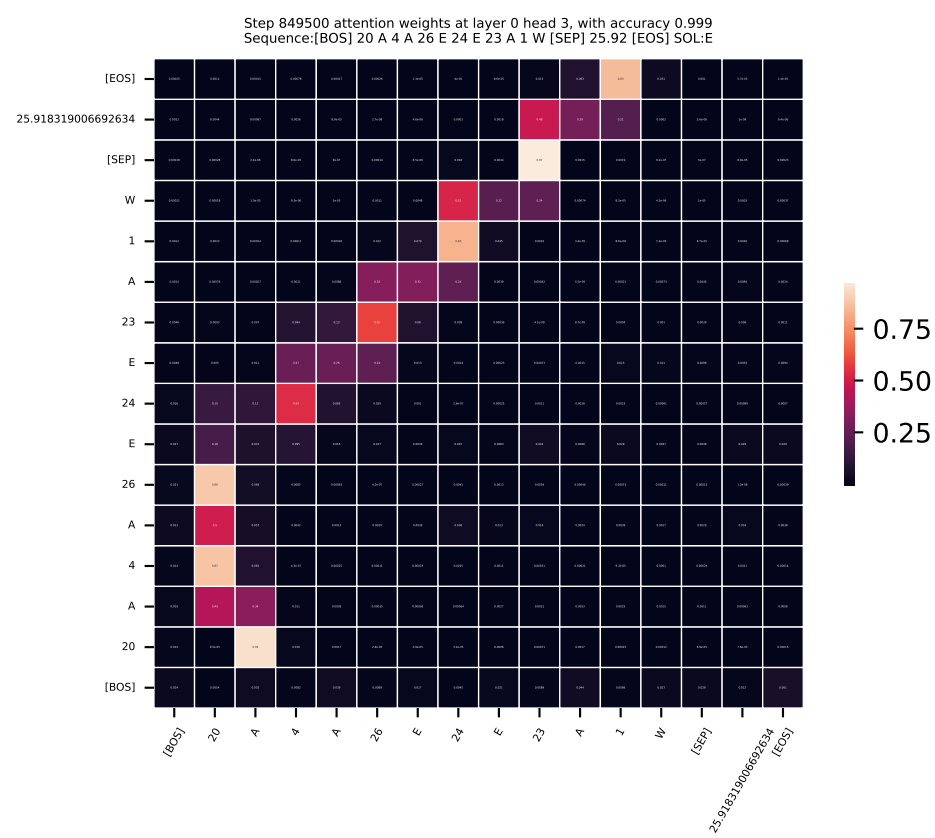

Step 840000 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

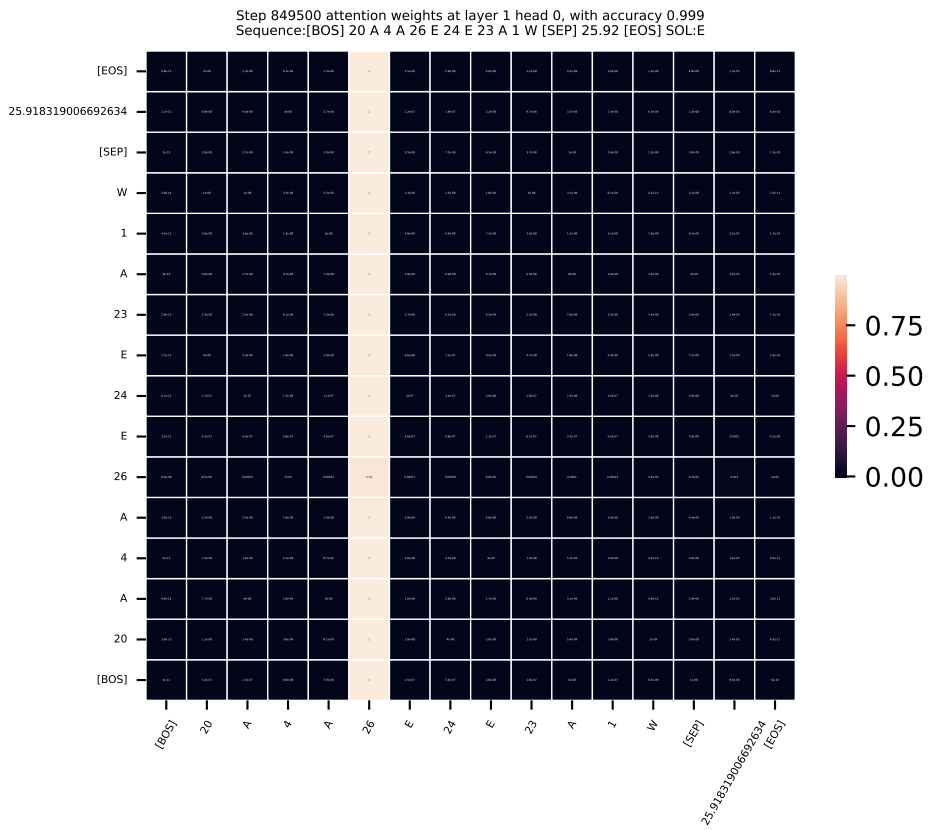

Step 840000 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

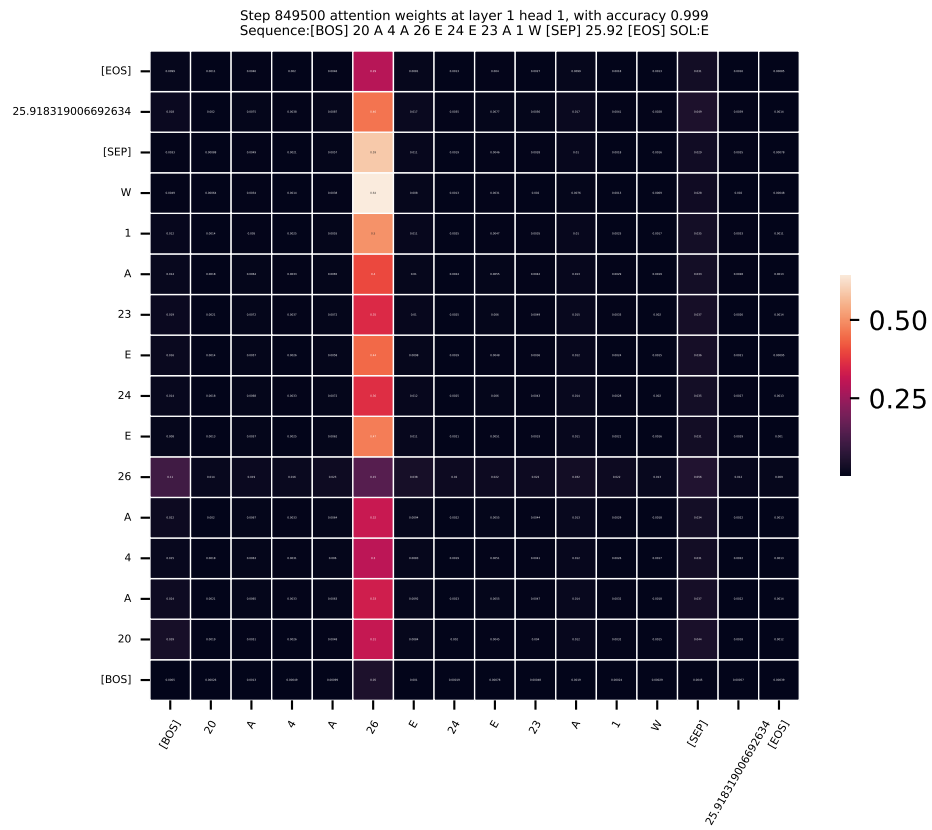

Step 849500 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

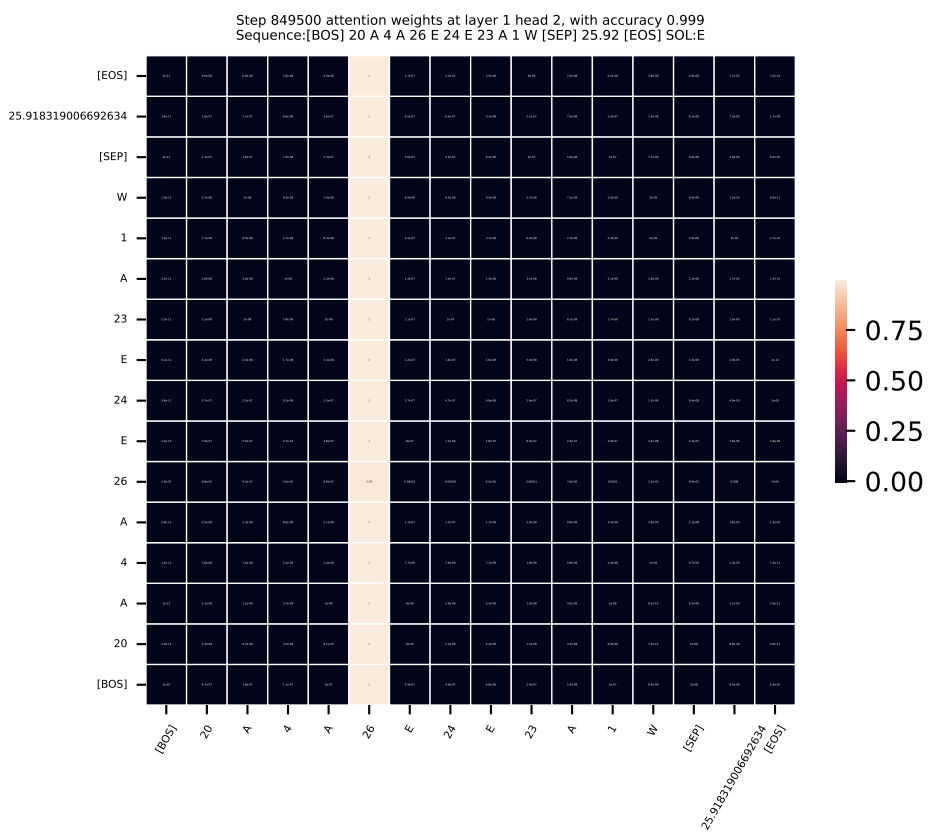

Step 849500 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

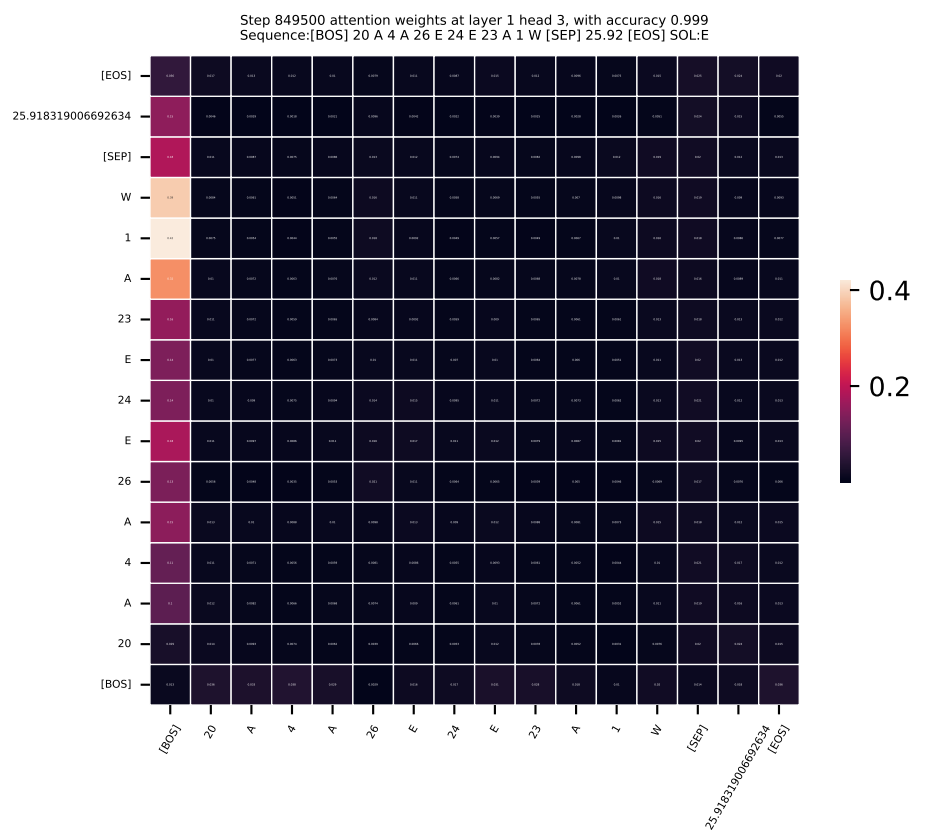

Step 849500 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

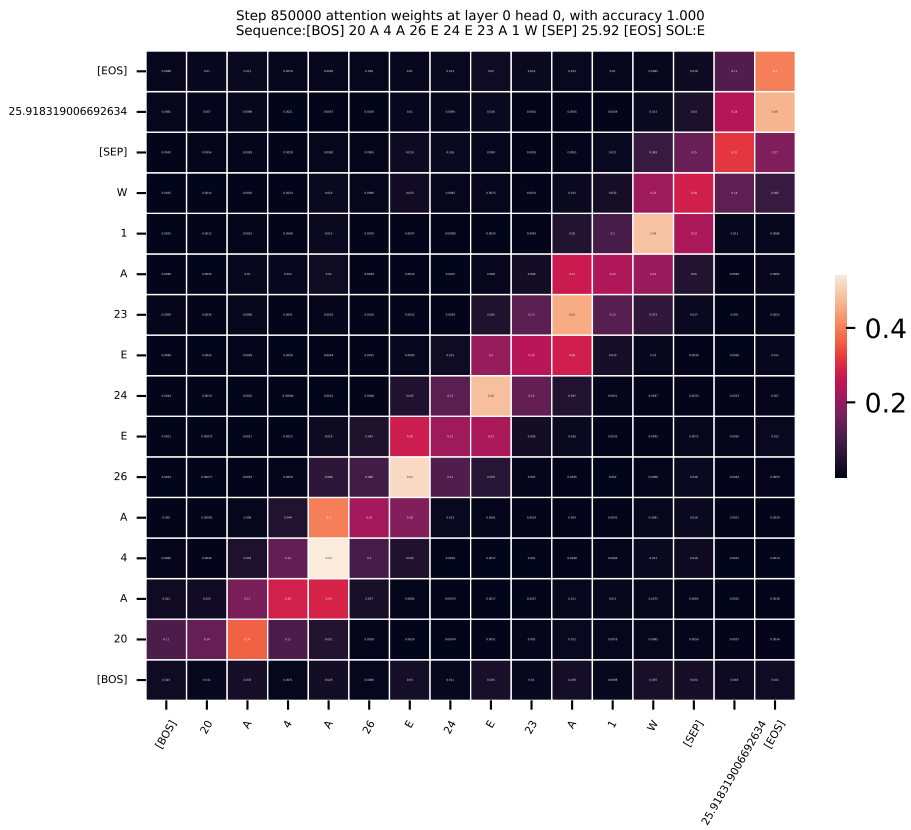

Step 849500 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

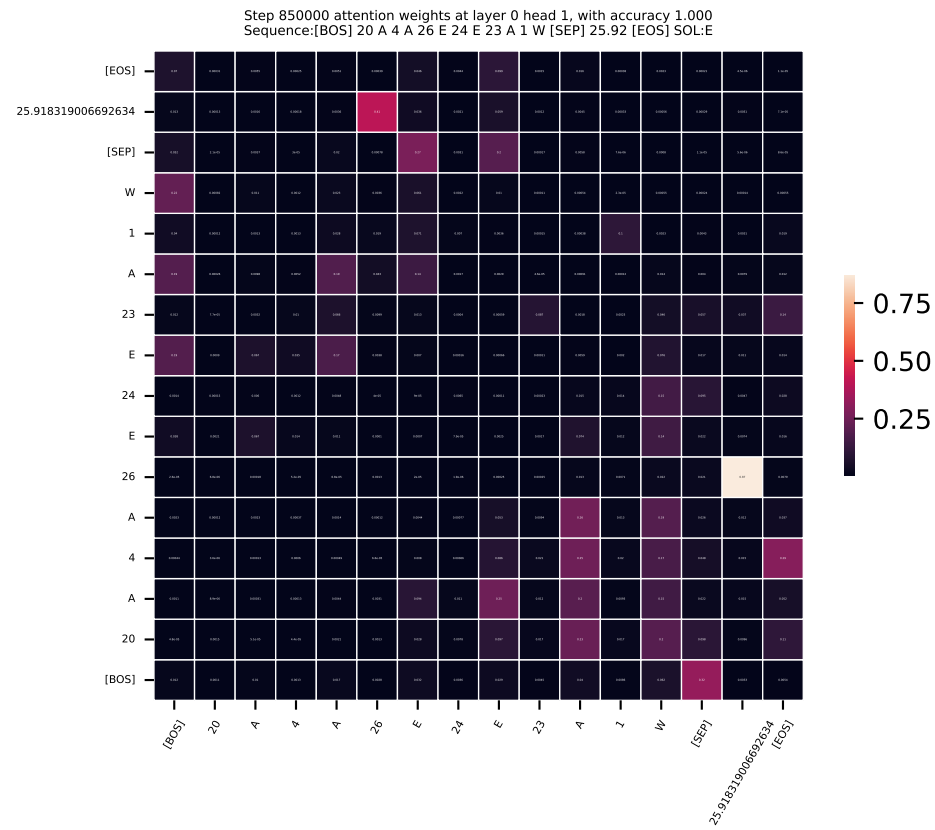

Step 849500 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

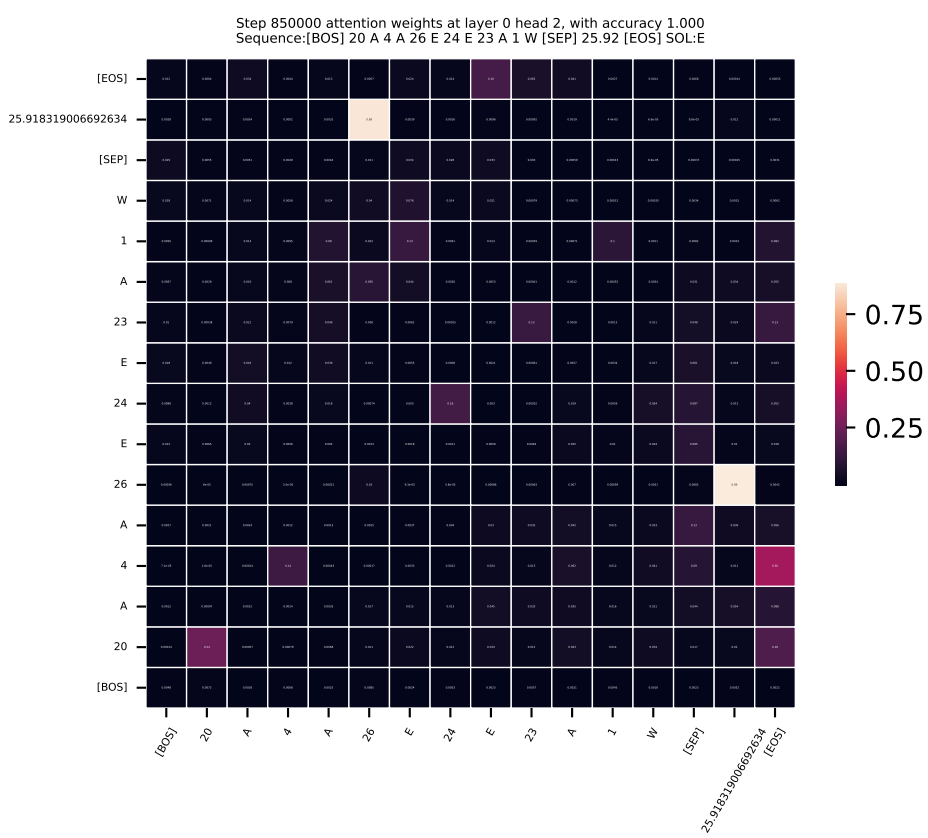

Step 849500 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

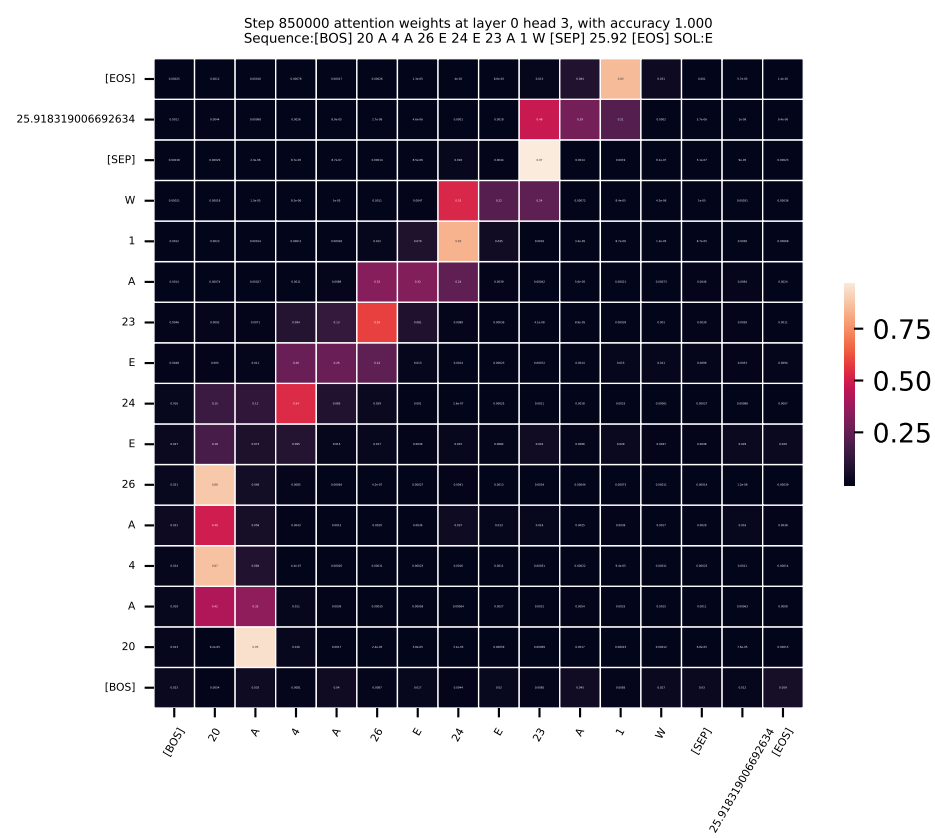

Step 849500 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

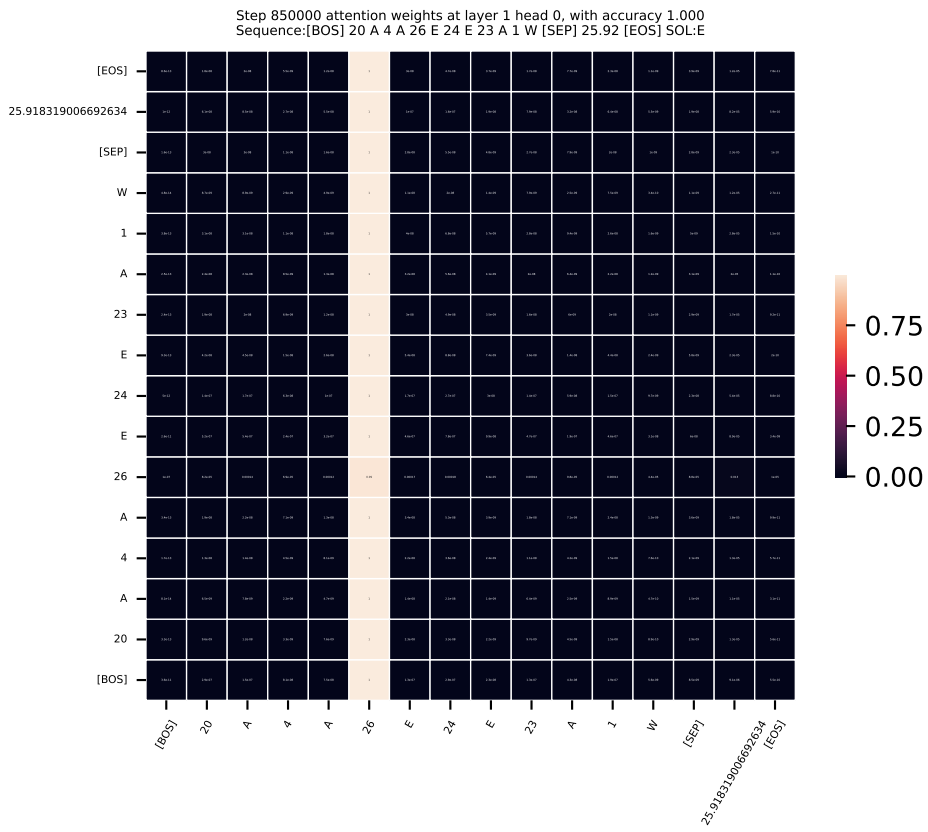

Step 849500 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

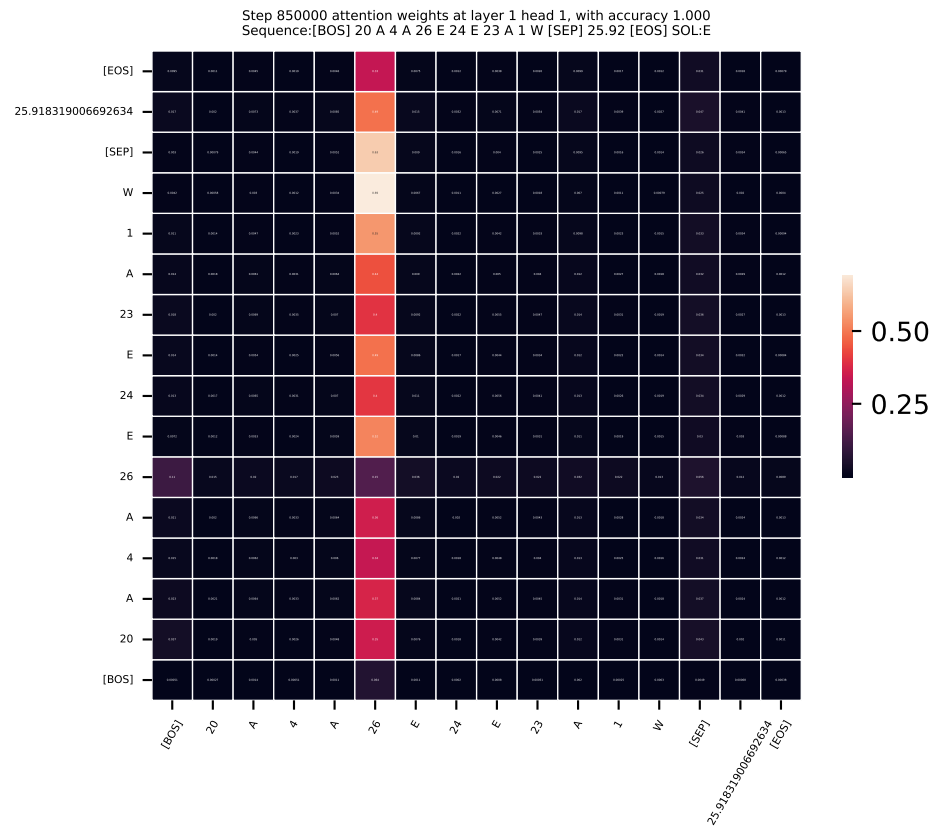

Step 850000 attention weights at layer 0 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

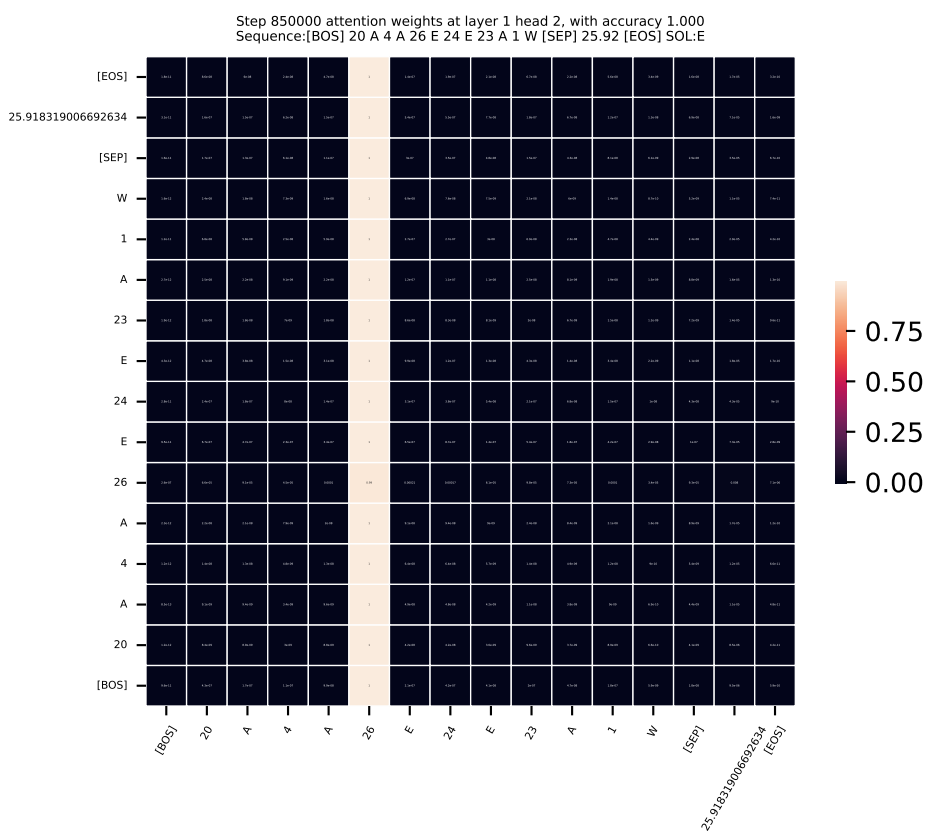

Step 850000 attention weights at layer 0 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

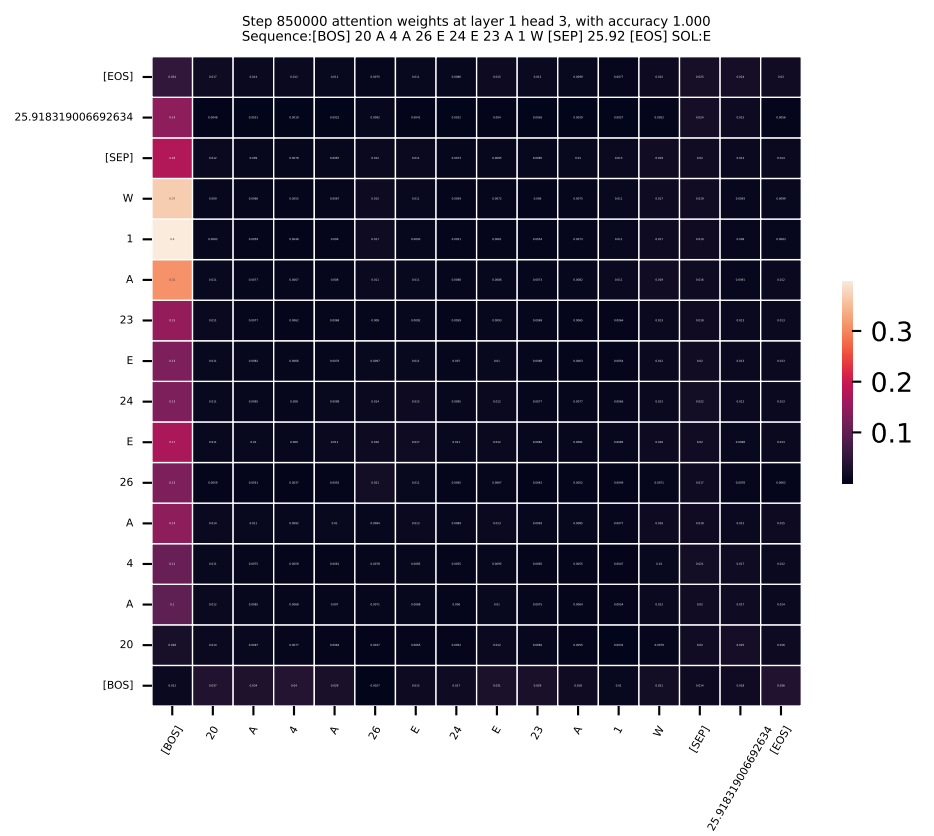

Step 850000 attention weights at layer 0 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

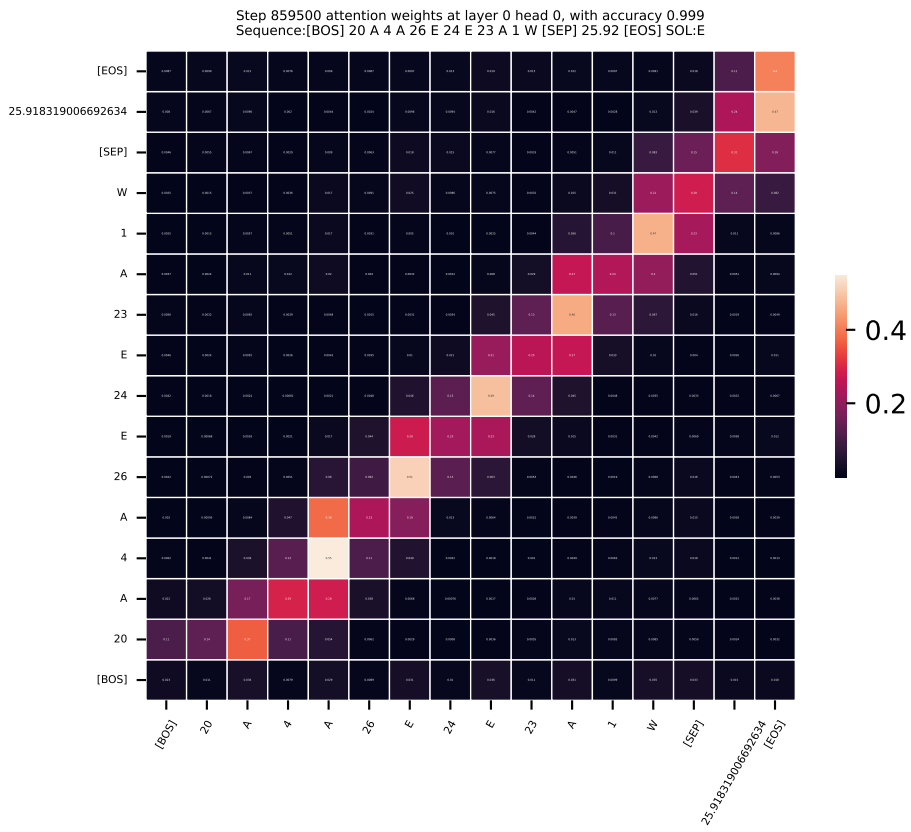

Step 850000 attention weights at layer 0 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

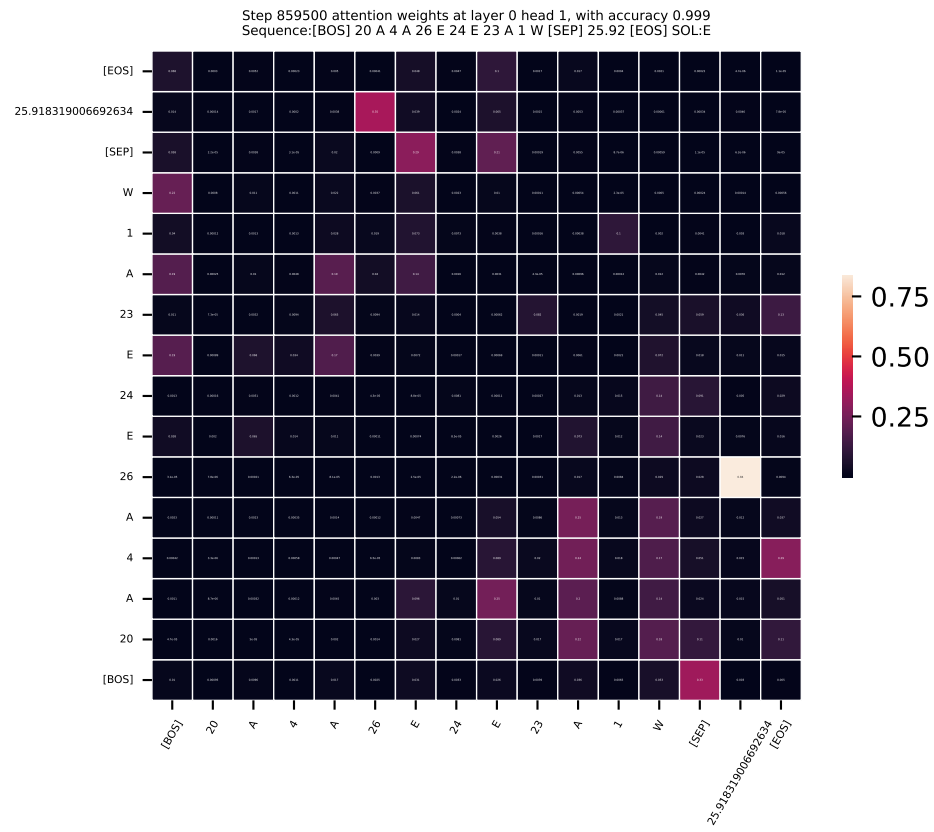

Step 850000 attention weights at layer 1 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

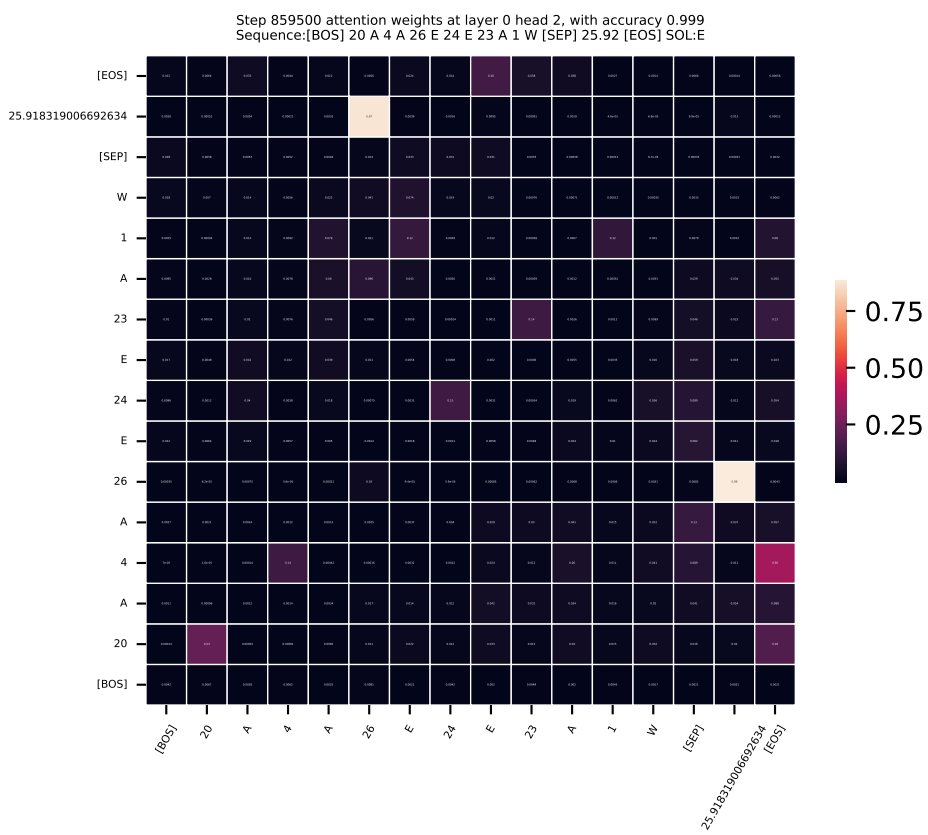

Step 850000 attention weights at layer 1 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

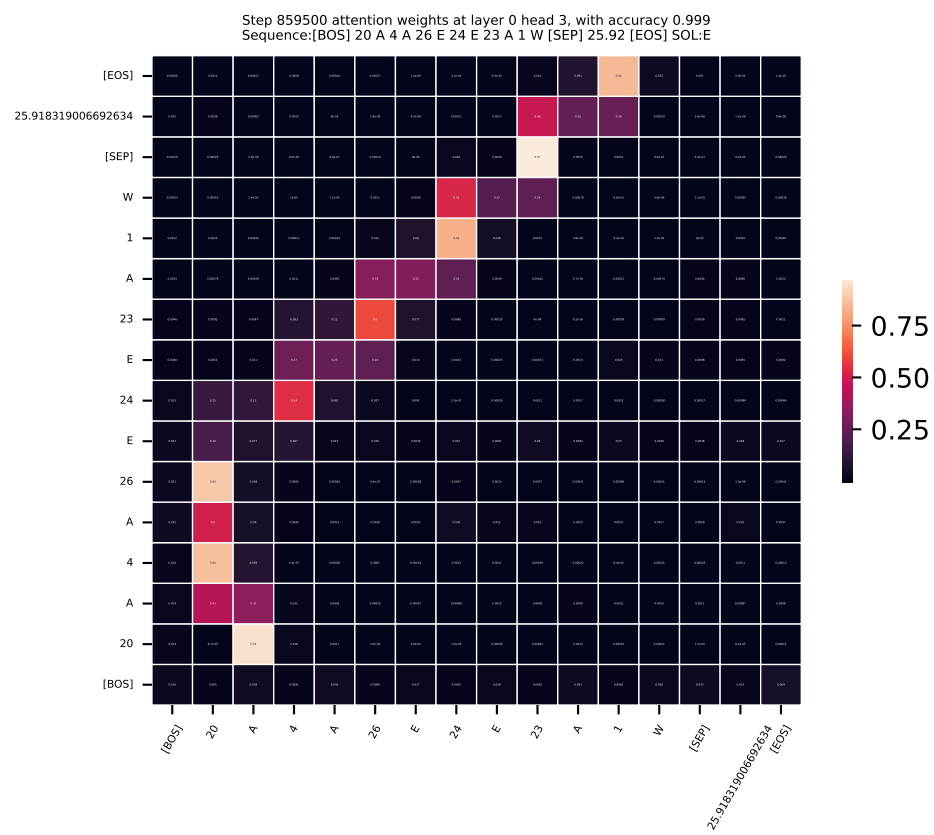

Step 850000 attention weights at layer 1 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

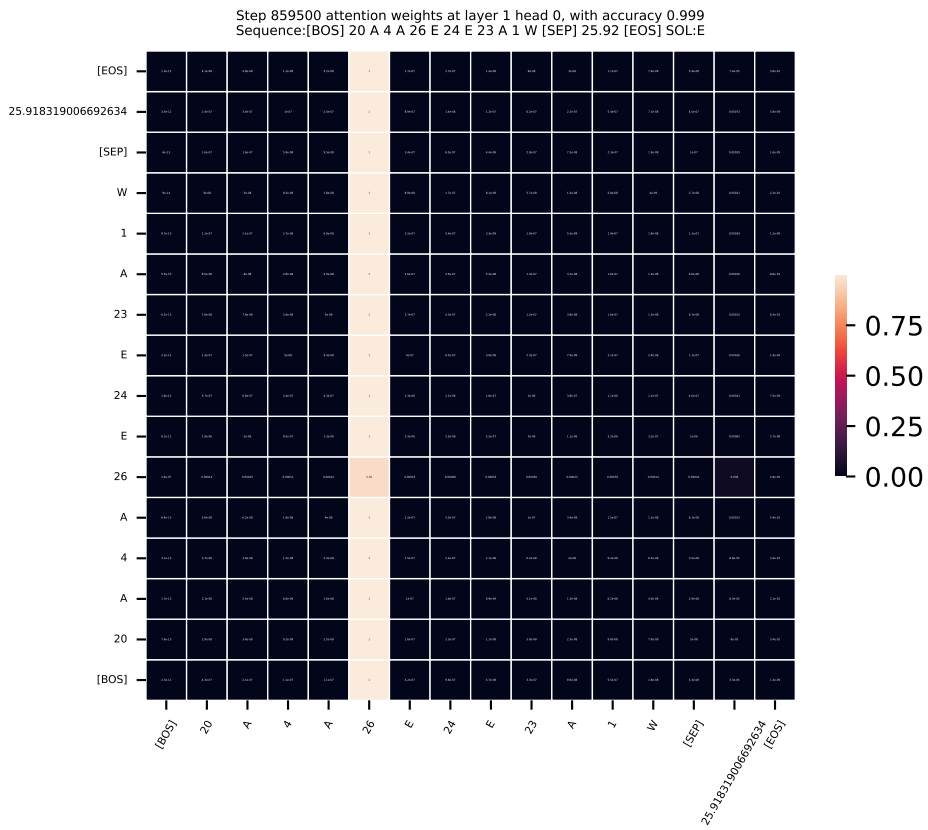

Step 850000 attention weights at layer 1 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

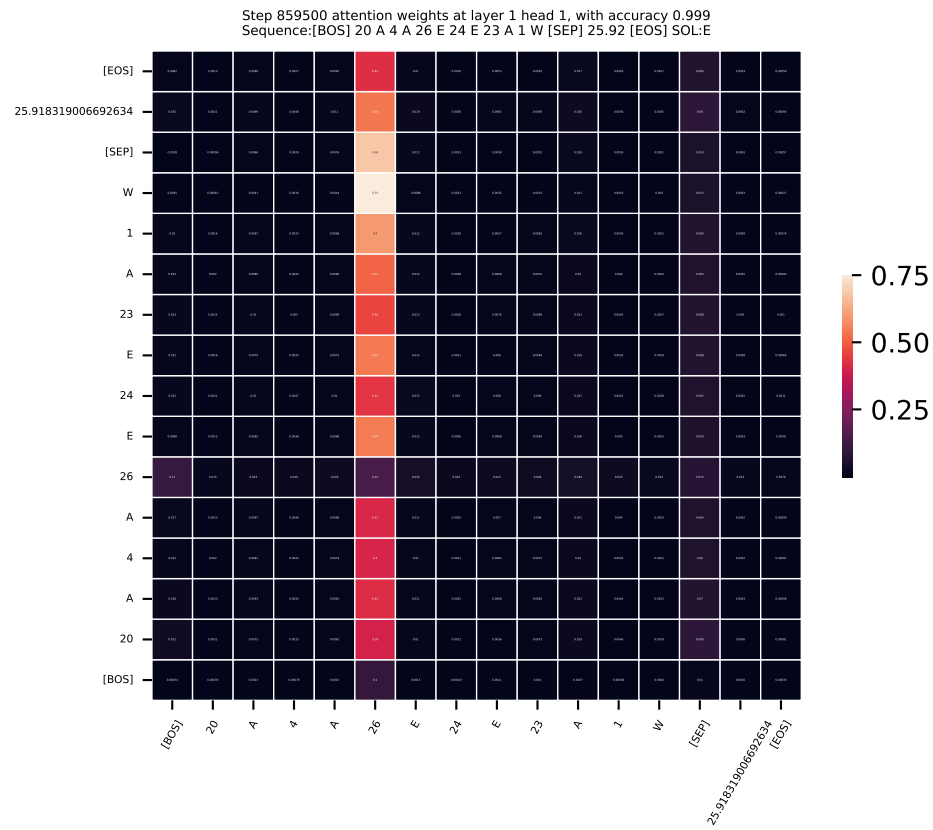

Step 859500 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

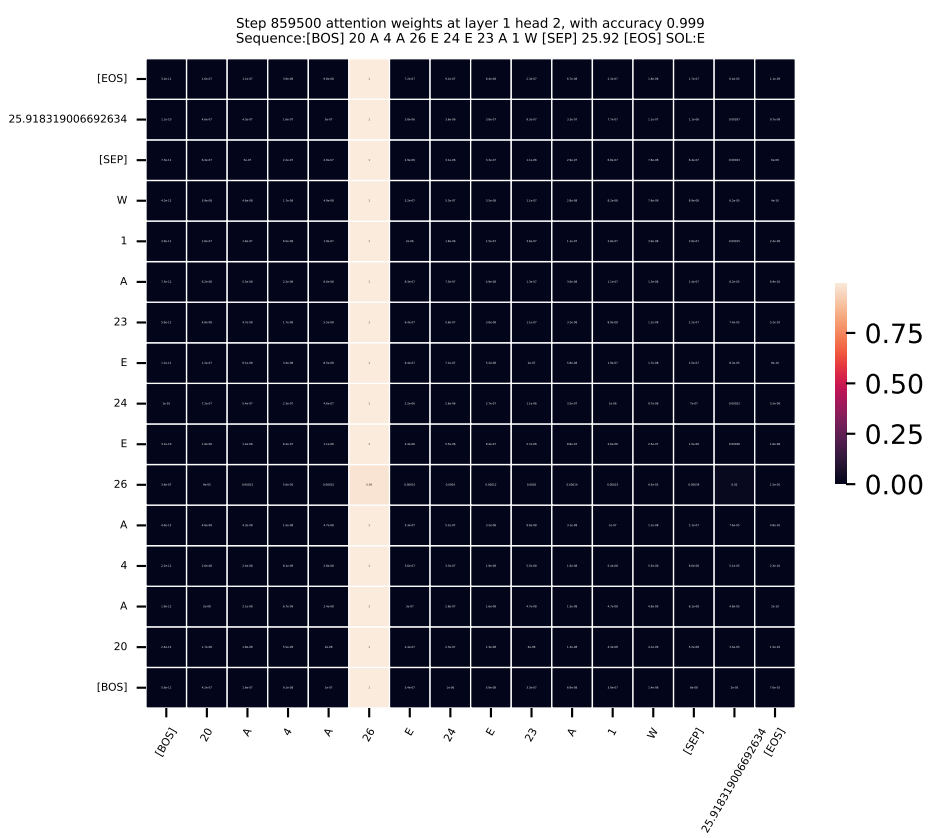

Step 859500 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

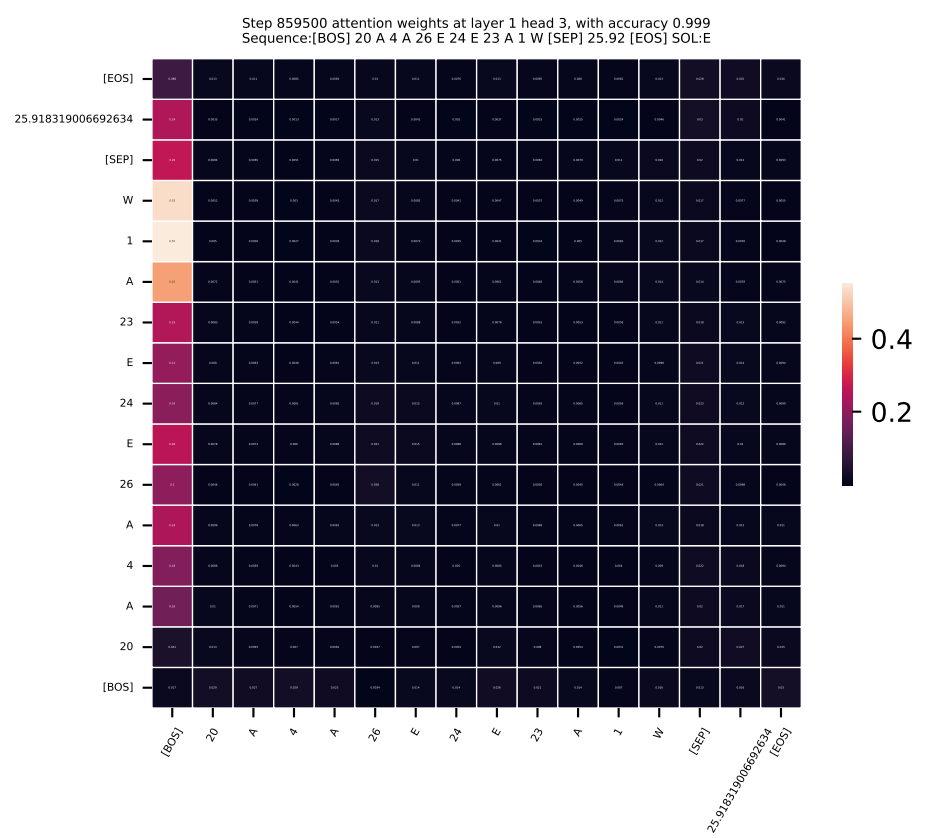

Step 859500 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

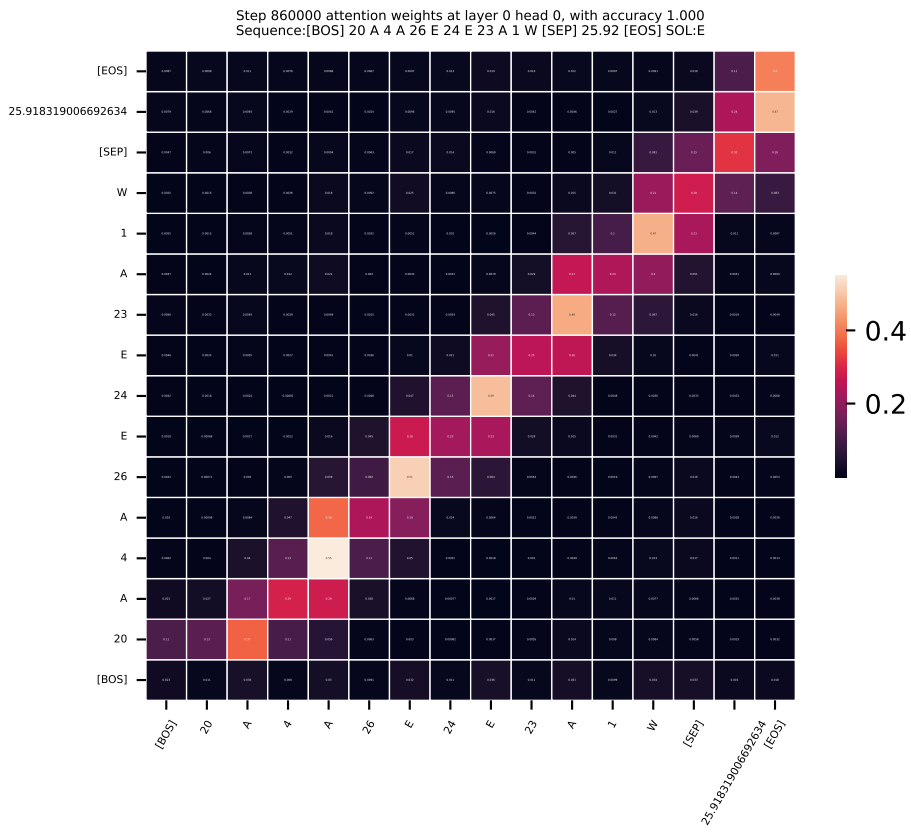

Step 859500 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

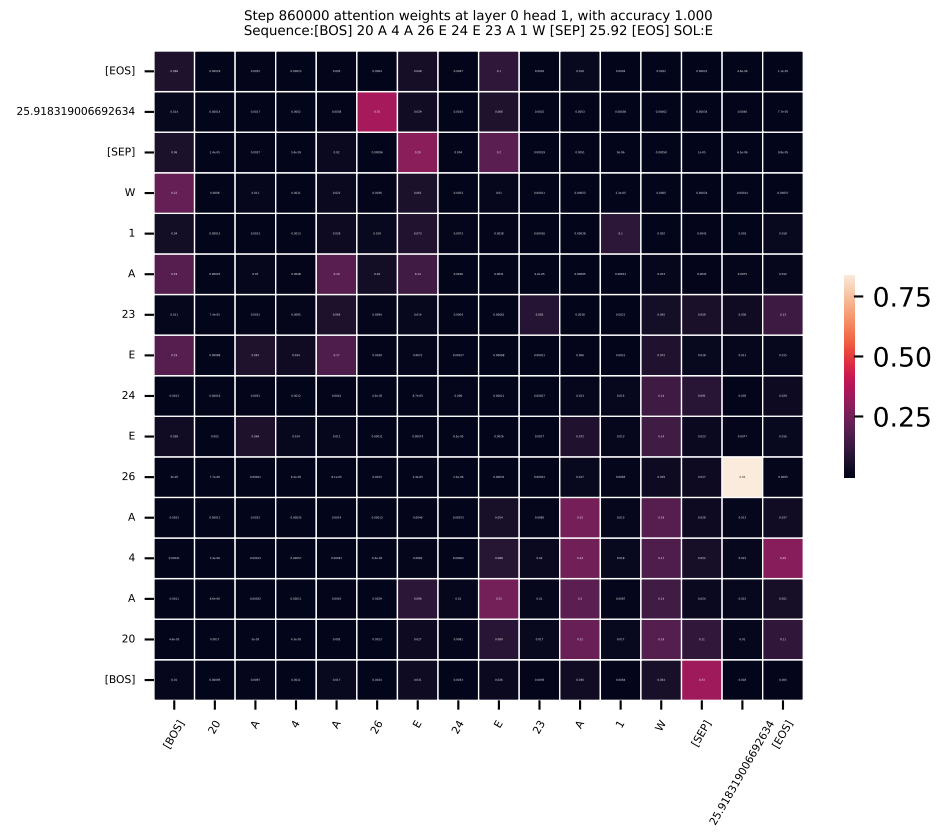

Step 859500 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

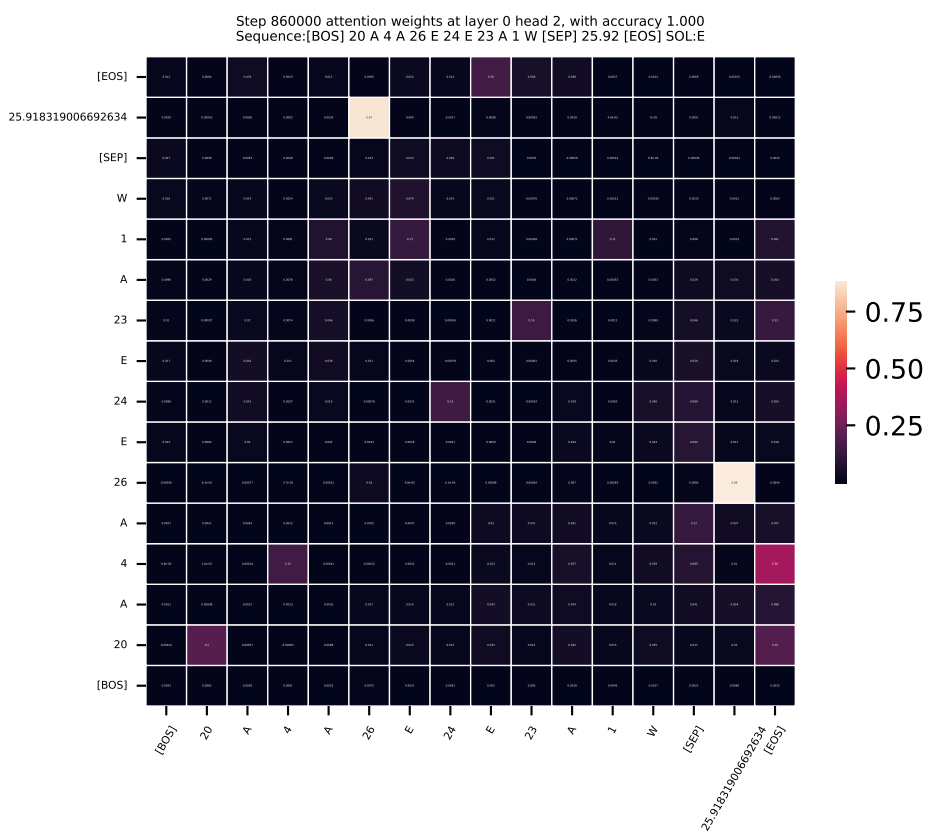

Step 859500 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

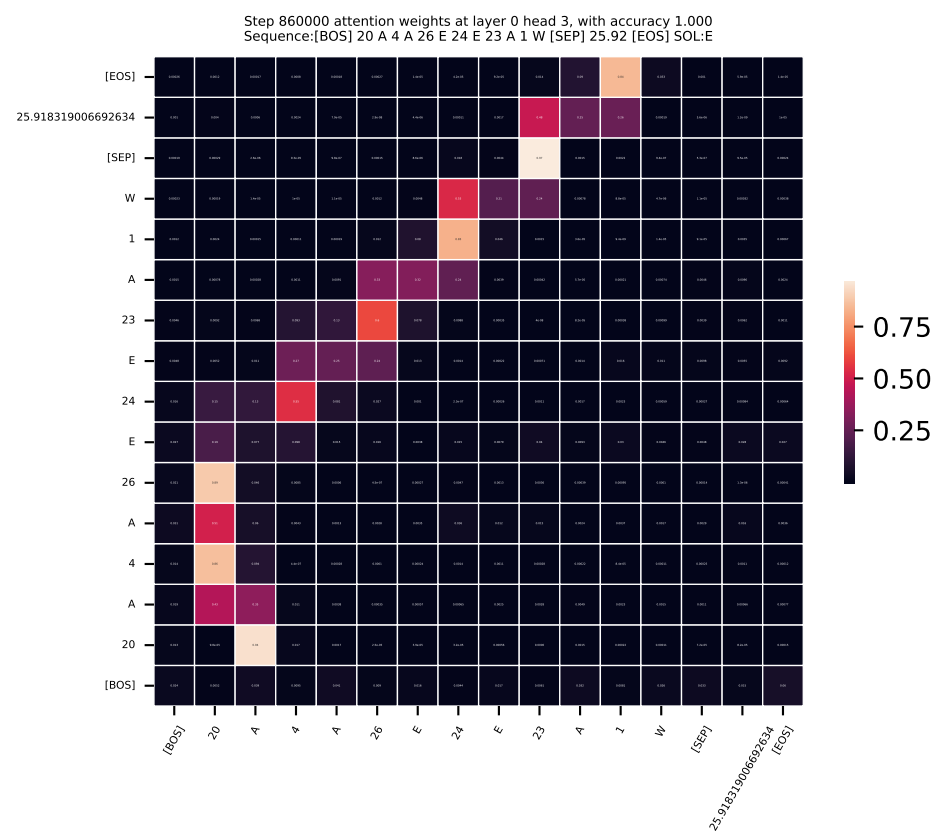

Step 859500 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

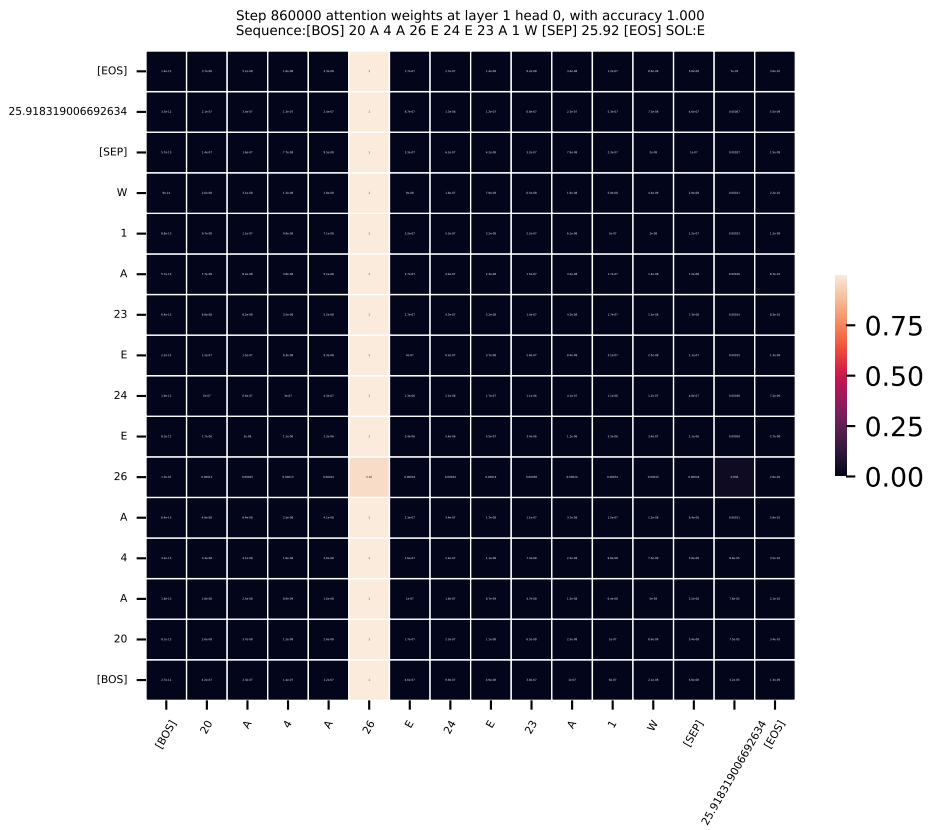

Step 859500 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

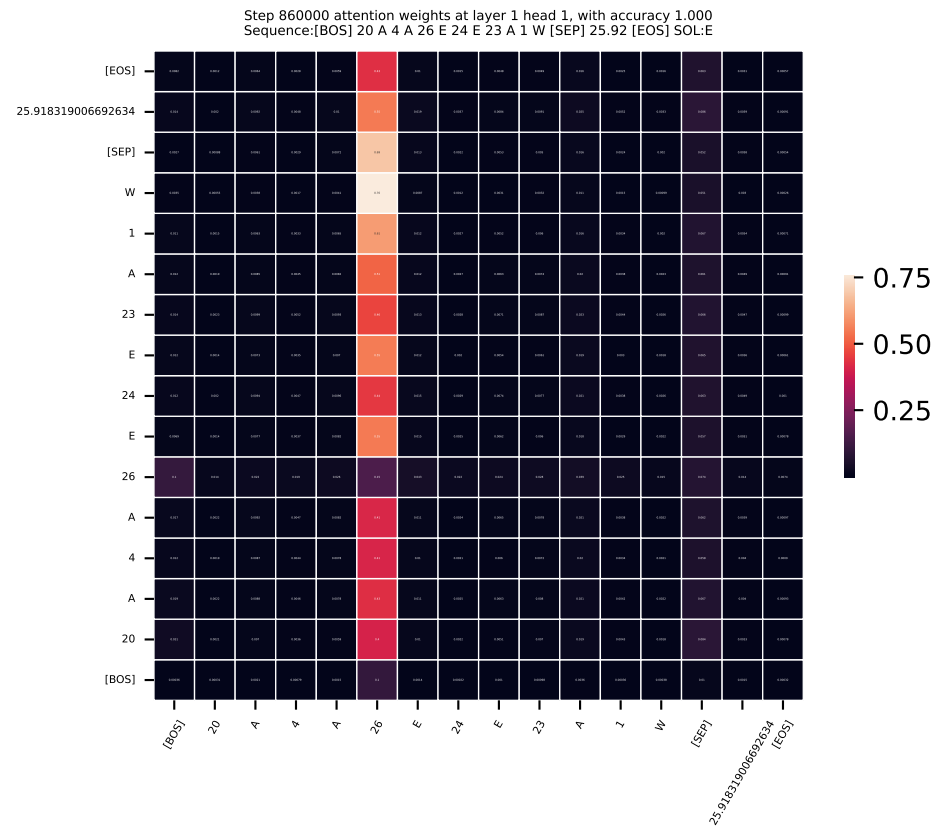

Step 860000 attention weights at layer 0 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

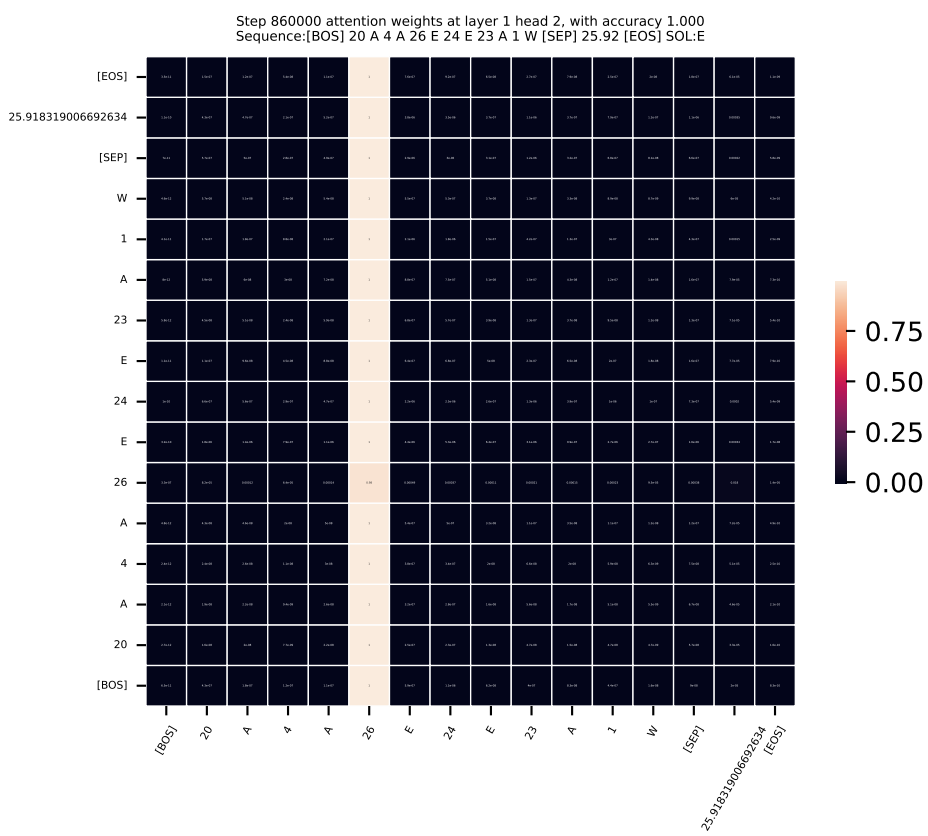

Step 860000 attention weights at layer 0 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

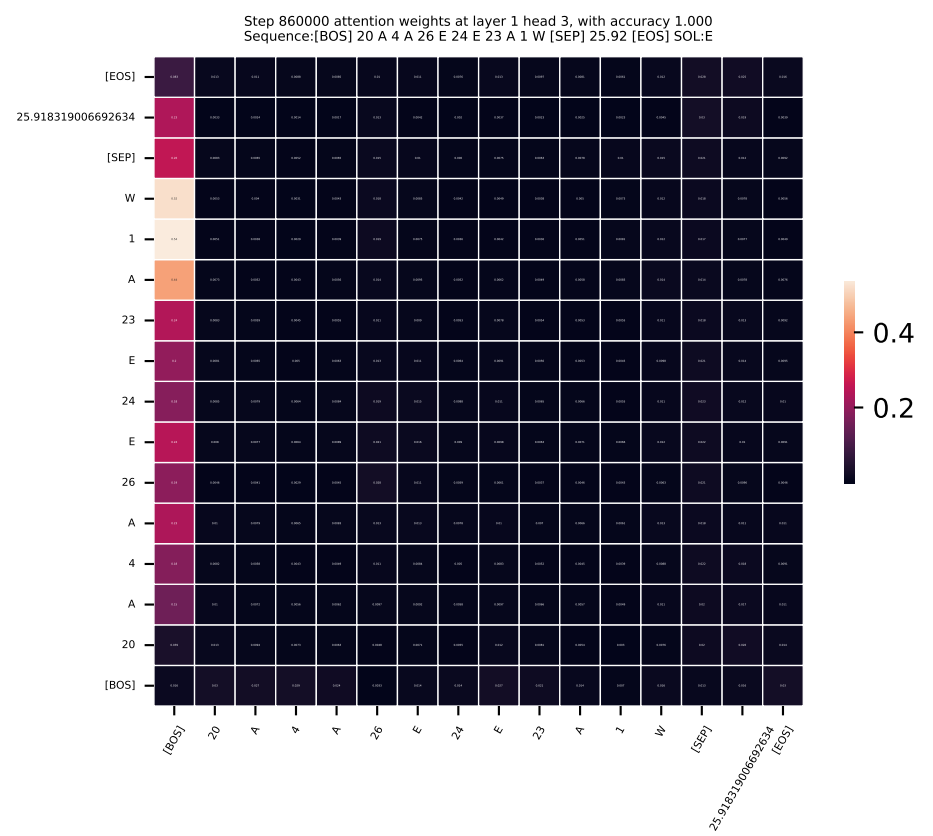

Step 860000 attention weights at layer 0 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

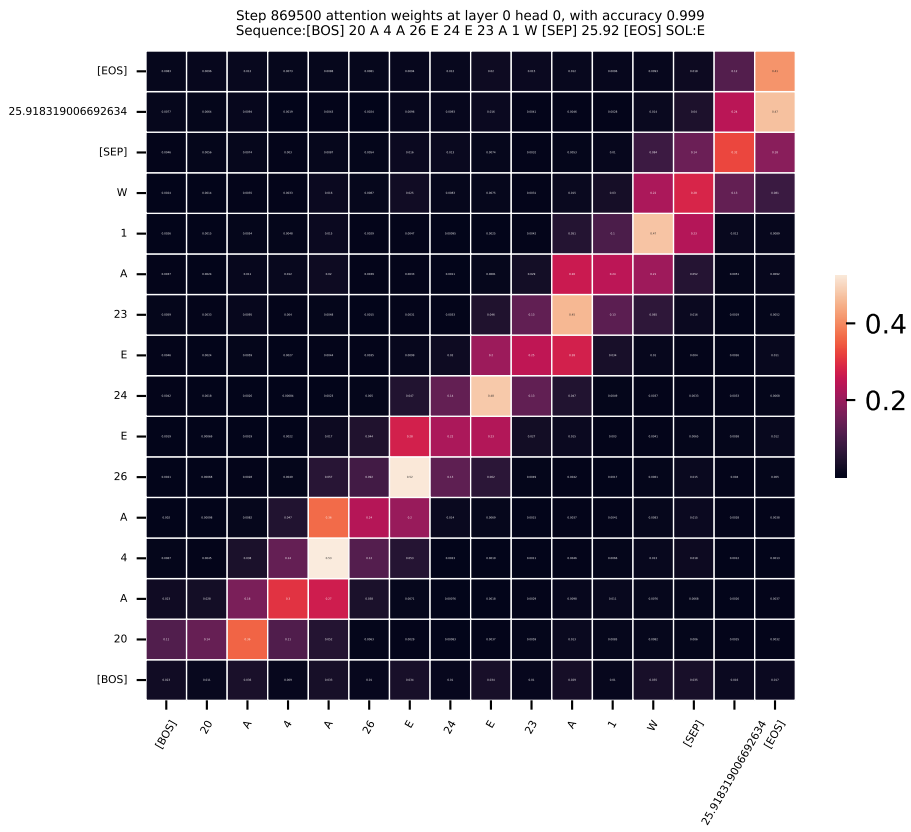

Step 860000 attention weights at layer 0 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

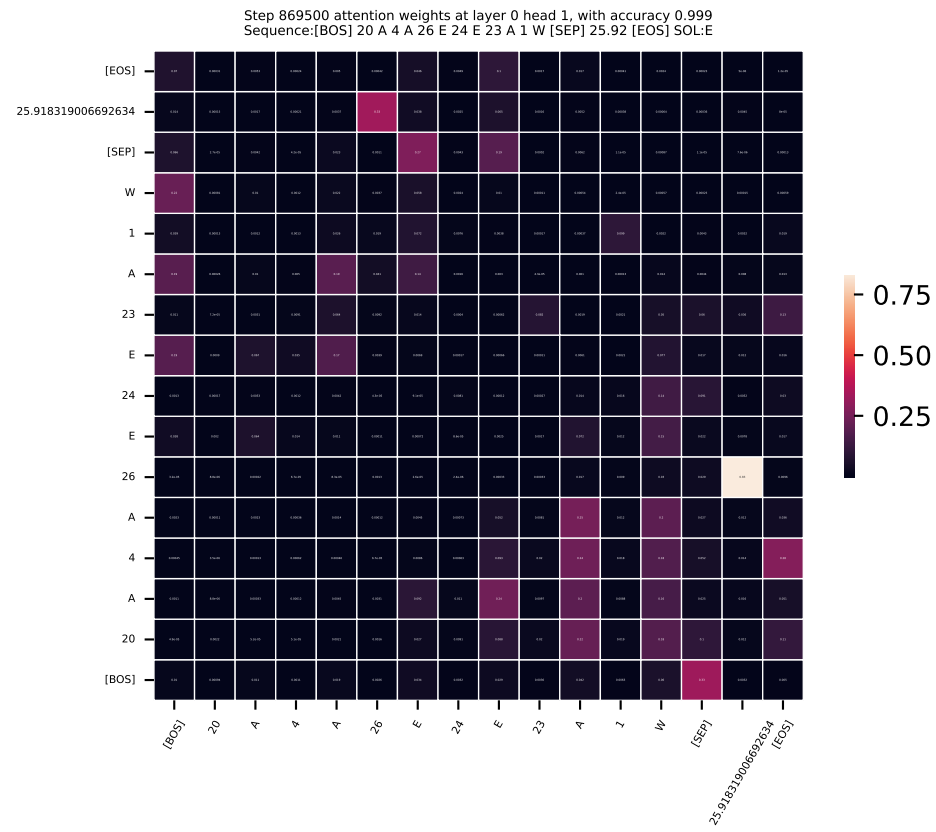

Step 860000 attention weights at layer 1 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

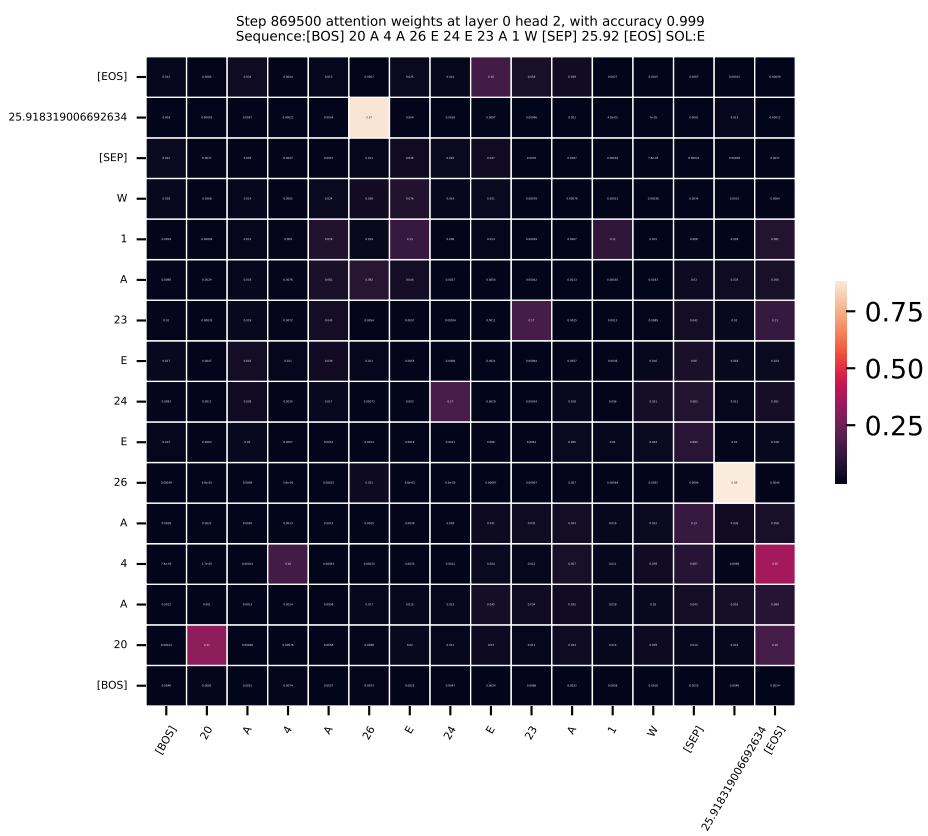

Step 860000 attention weights at layer 1 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

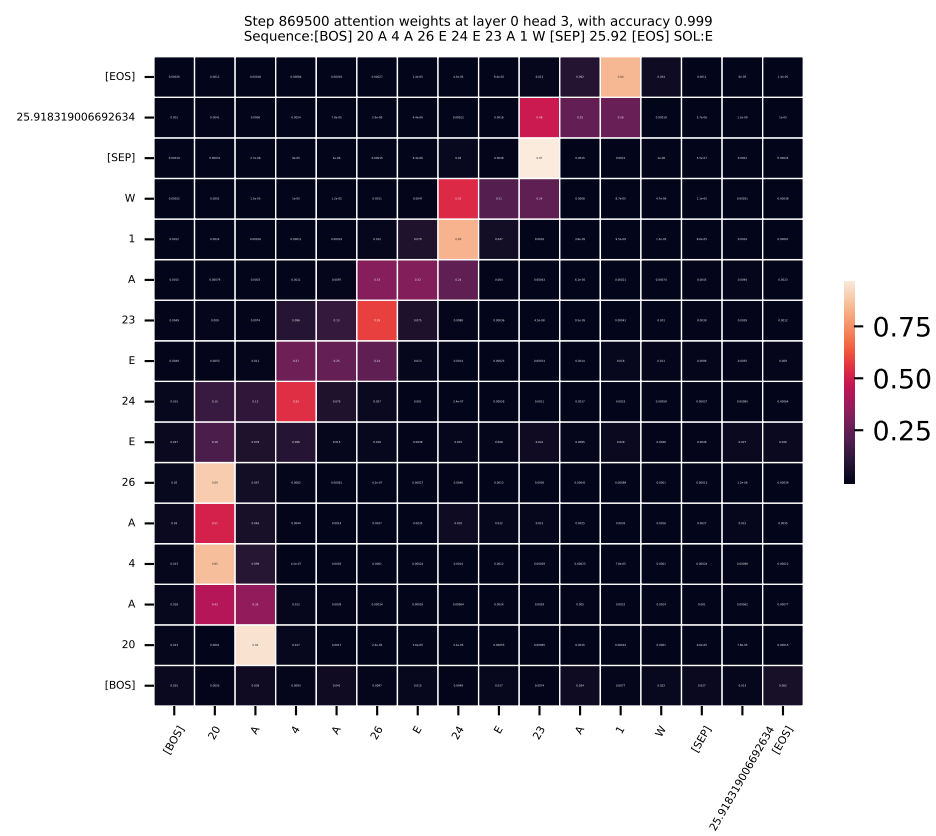

Step 860000 attention weights at layer 1 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

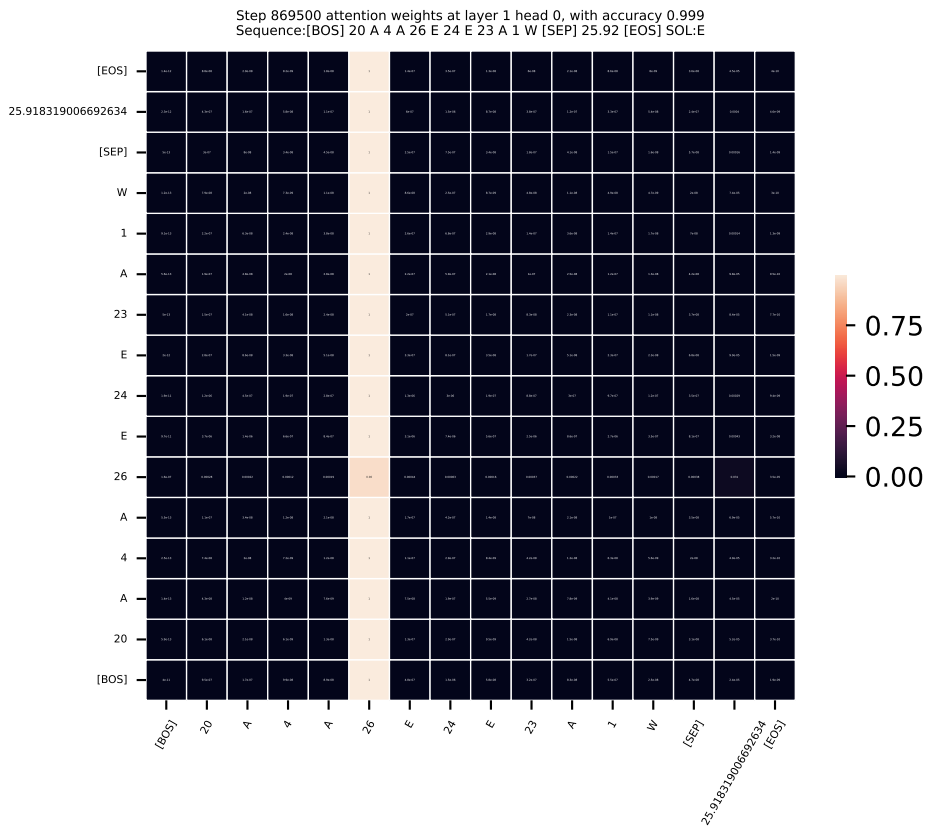

Step 860000 attention weights at layer 1 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

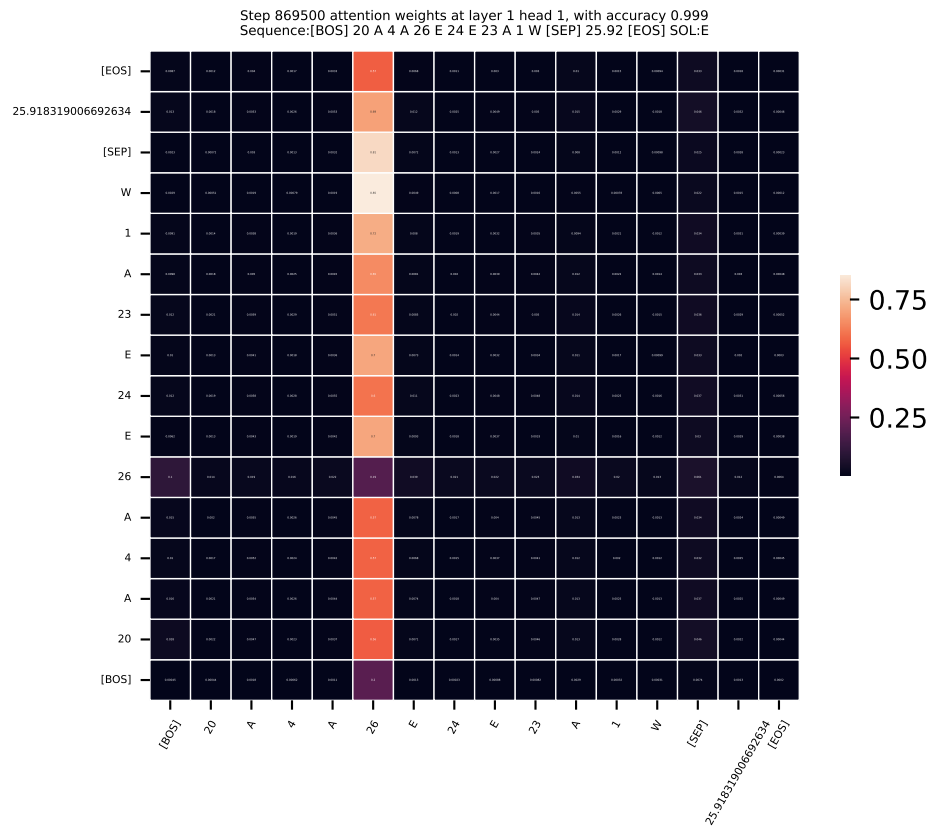

Step 869500 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

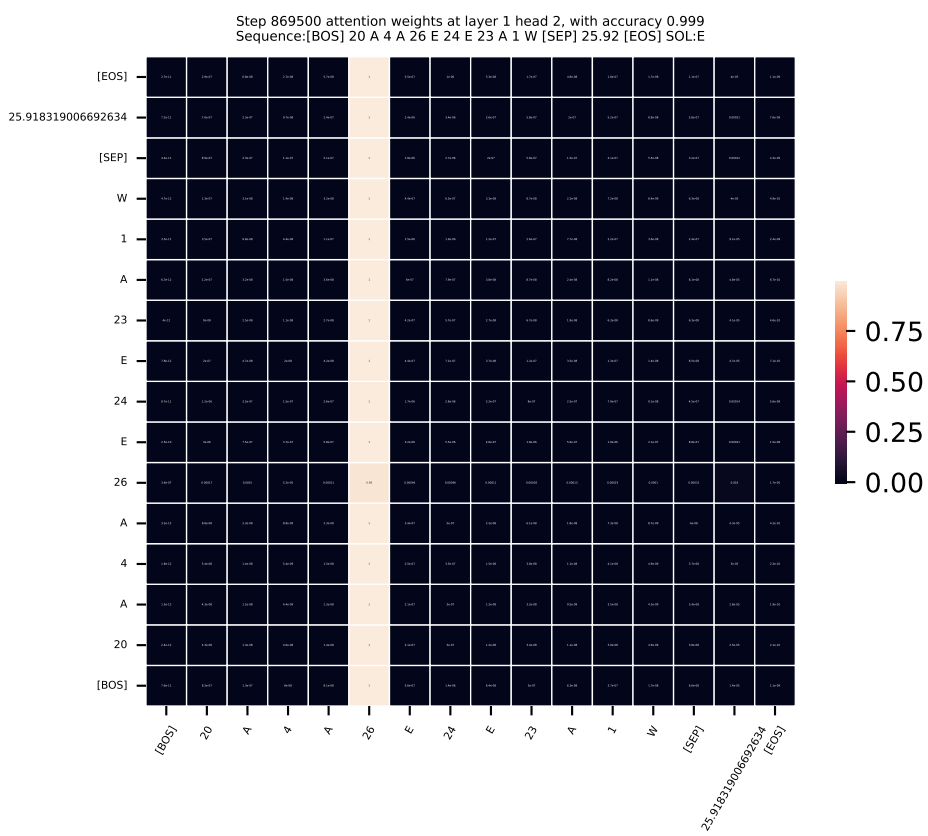

Step 869500 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

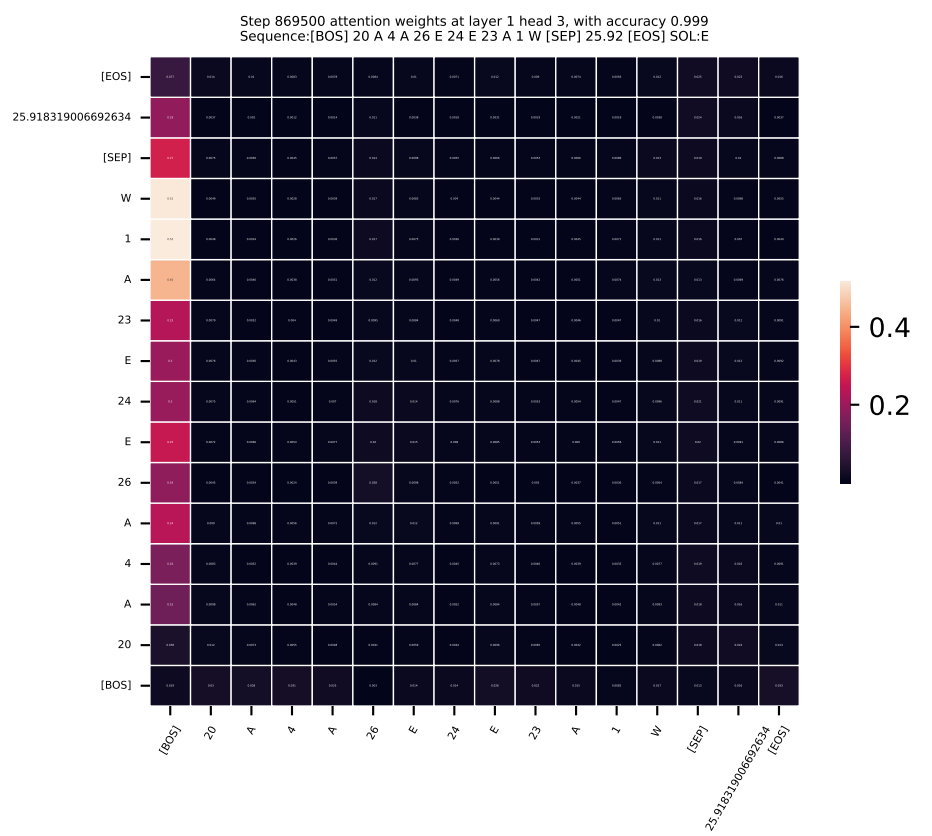

Step 869500 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

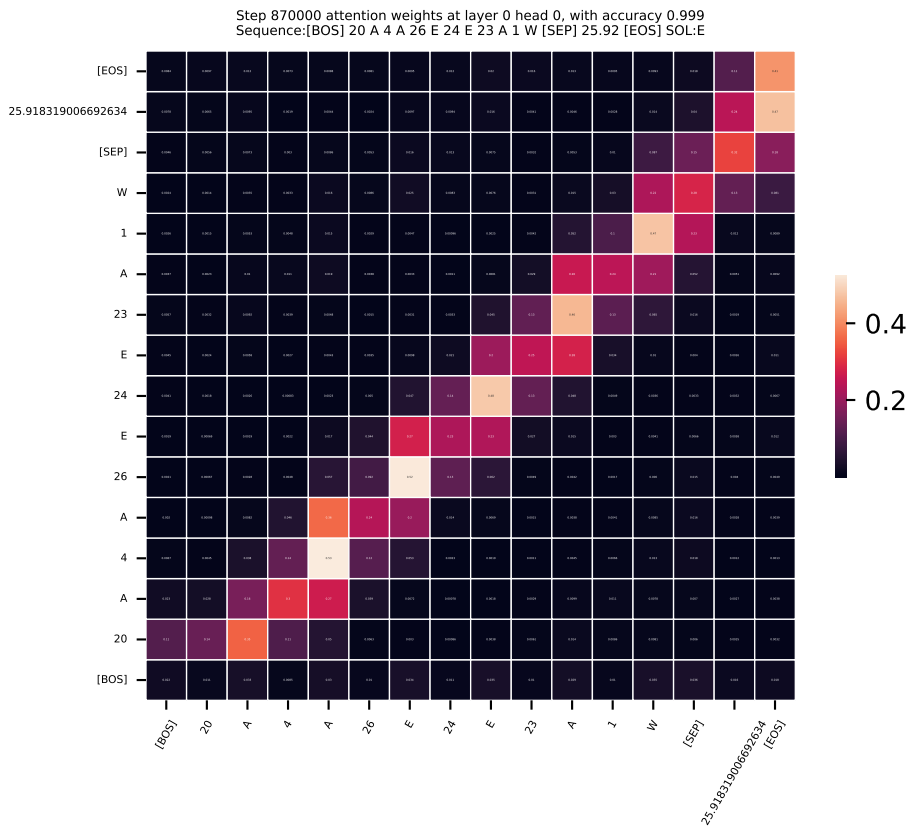

Step 869500 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

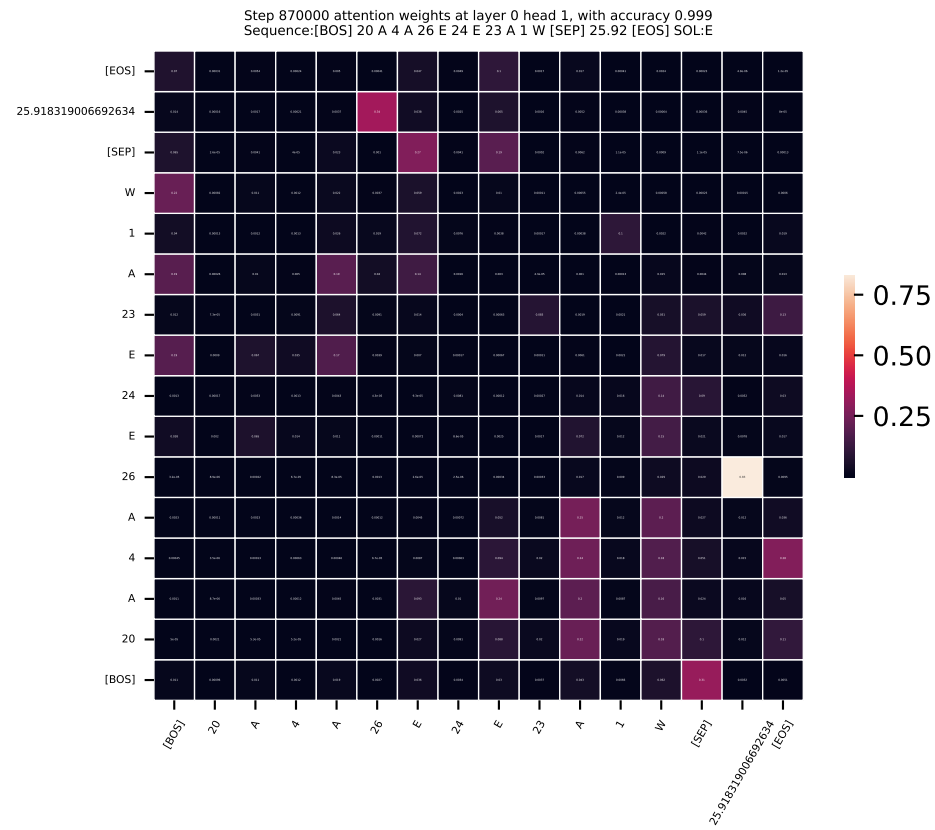

Step 869500 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

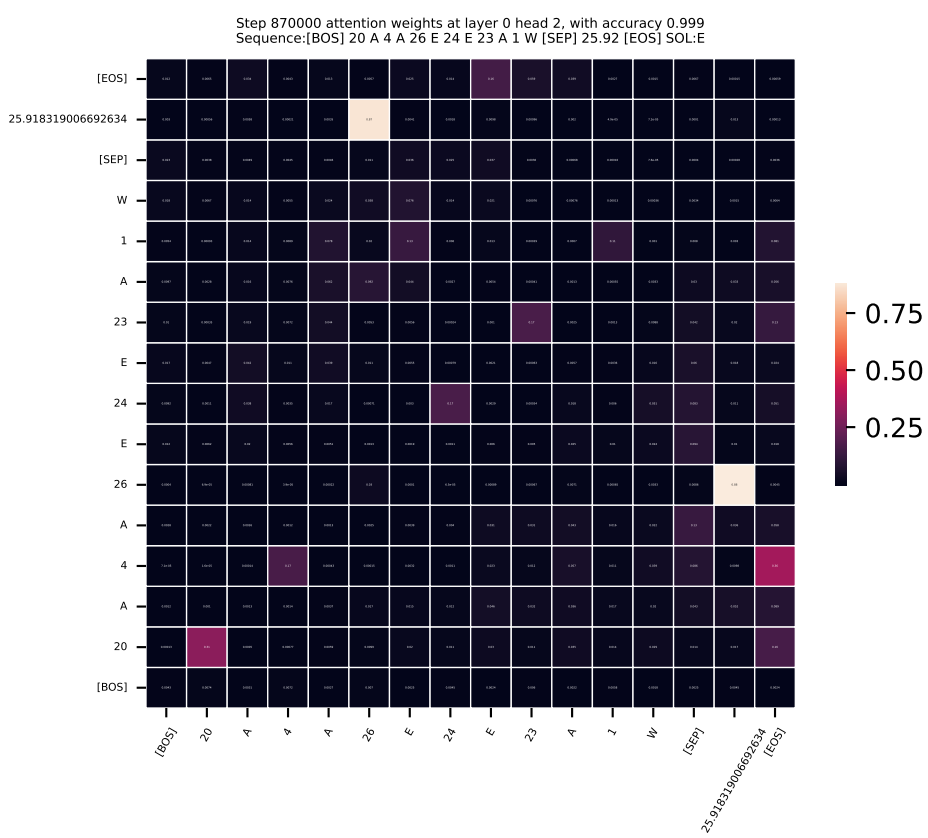

Step 869500 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

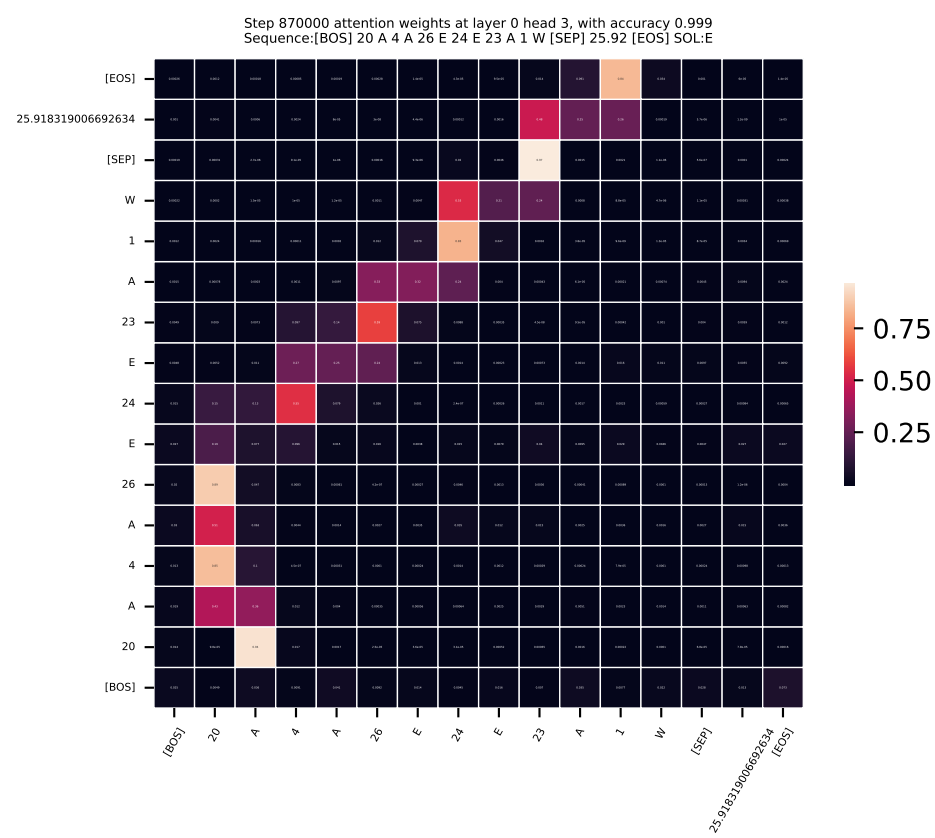

Step 869500 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

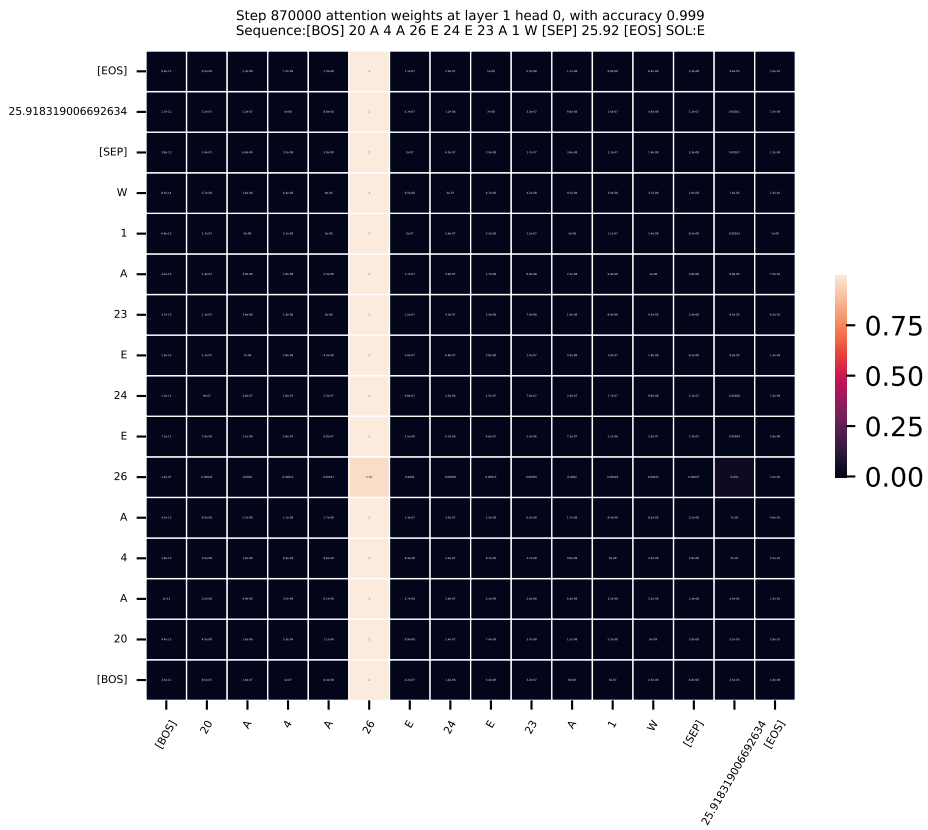

Step 869500 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

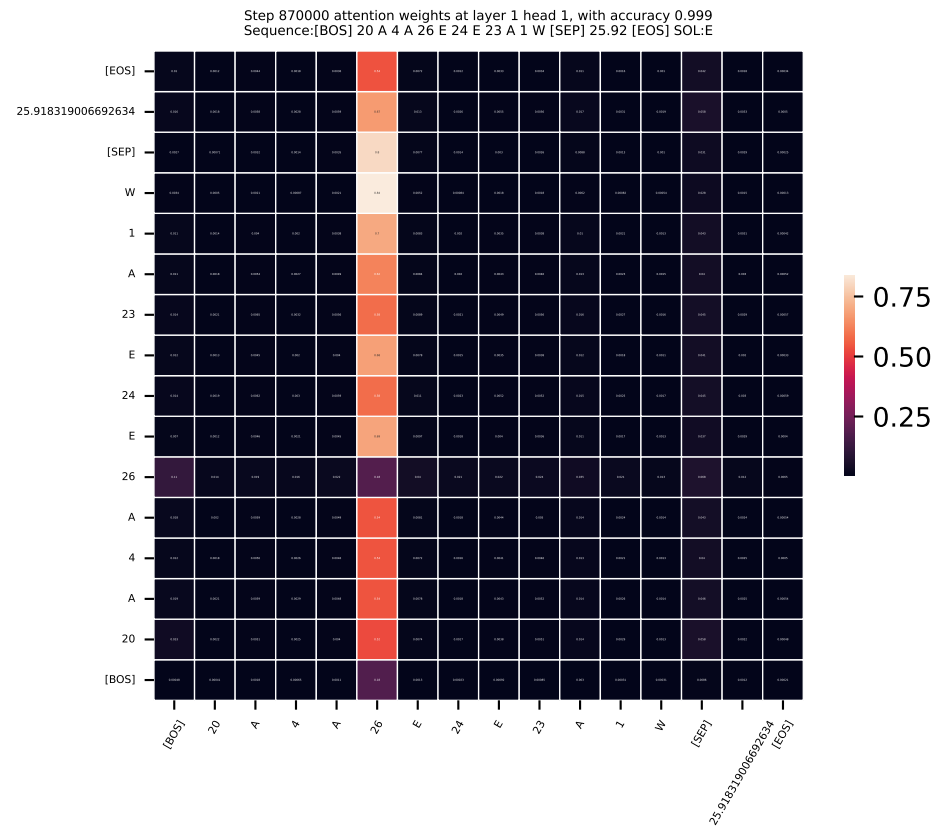

Step 870000 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

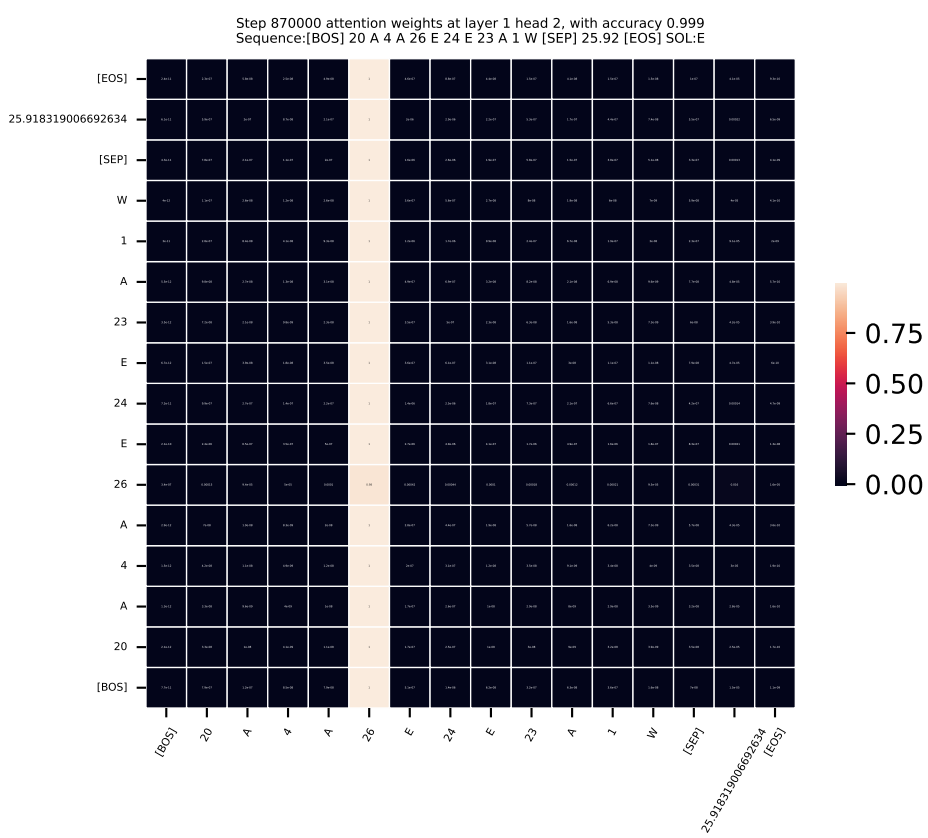

Step 870000 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

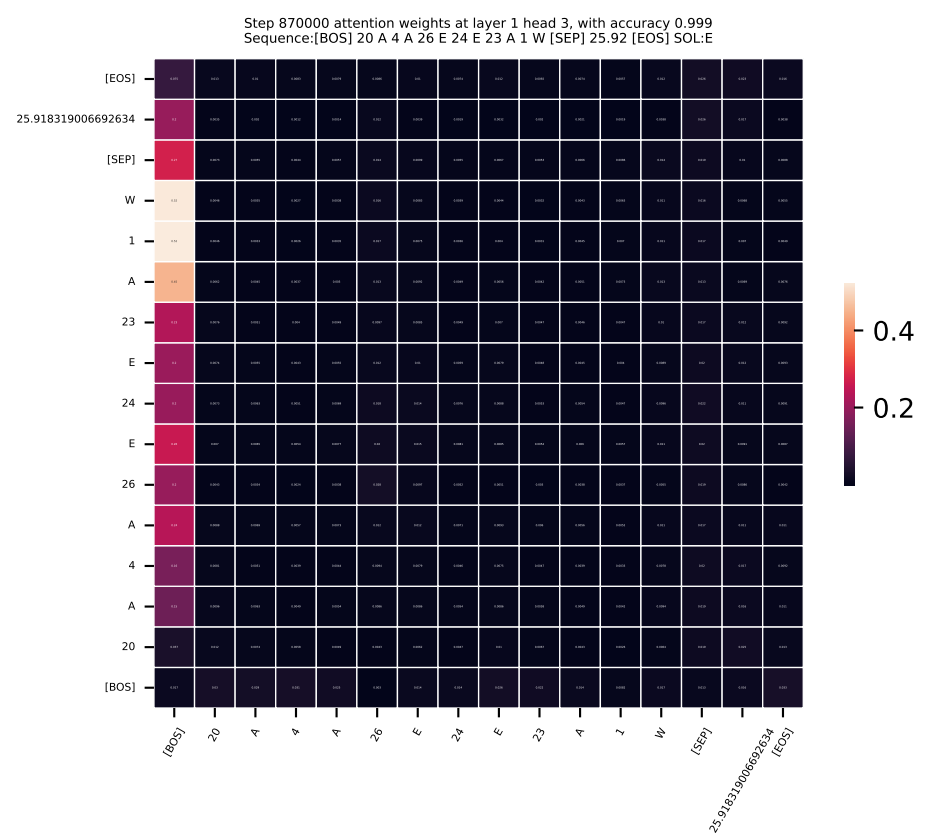

Step 870000 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

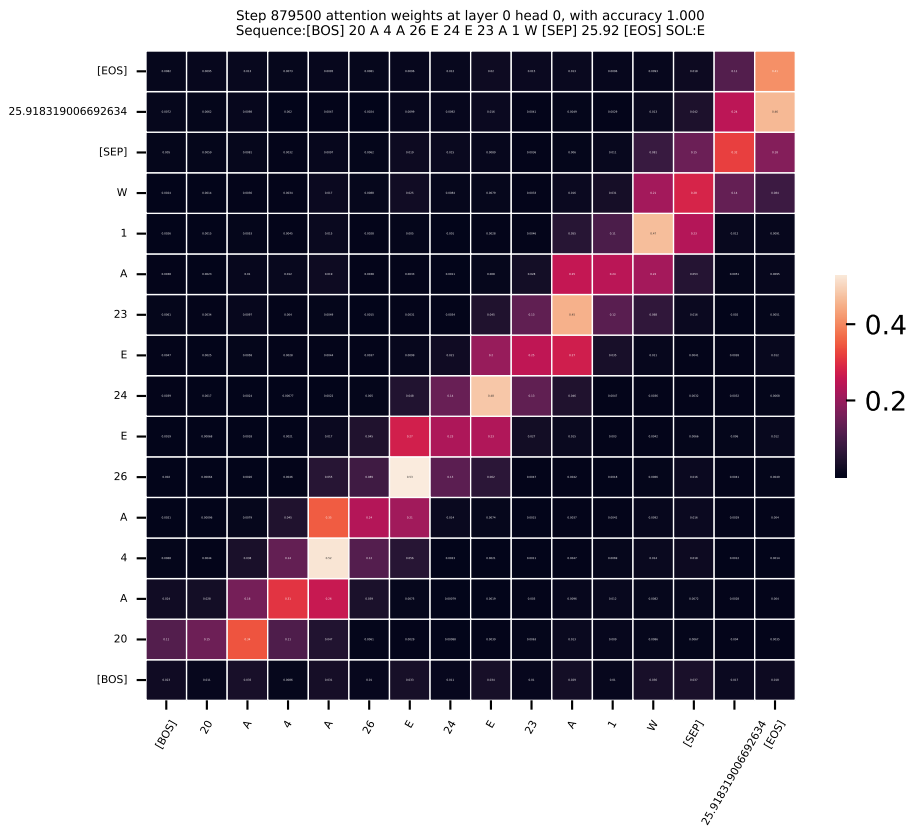

Step 870000 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

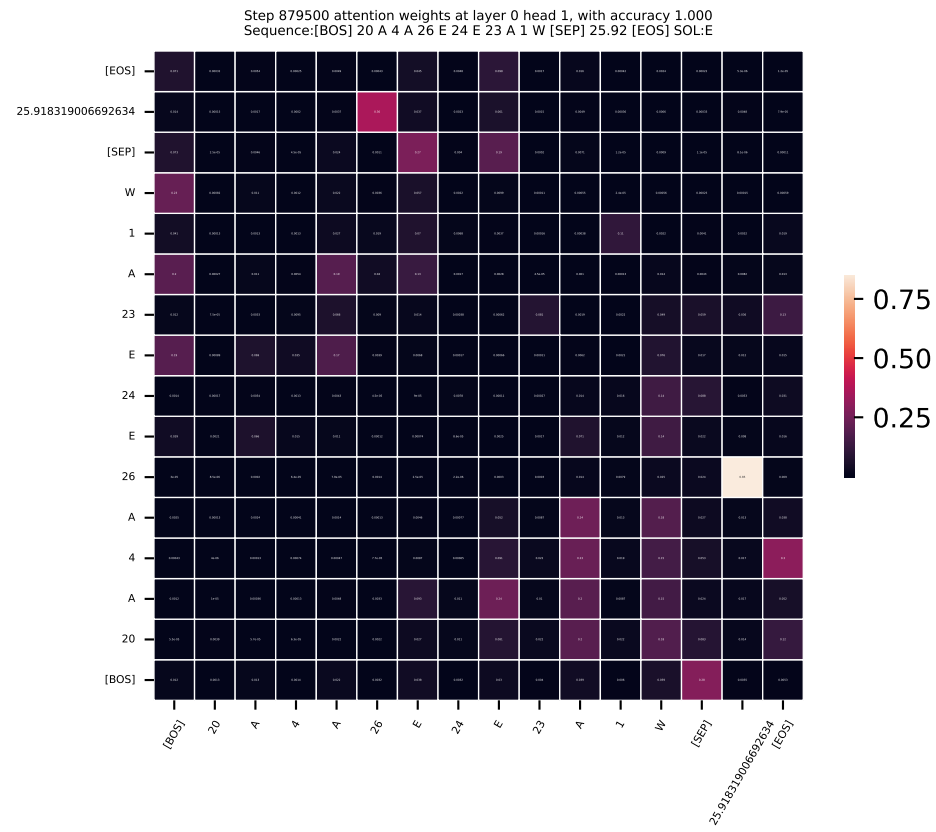

Step 870000 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

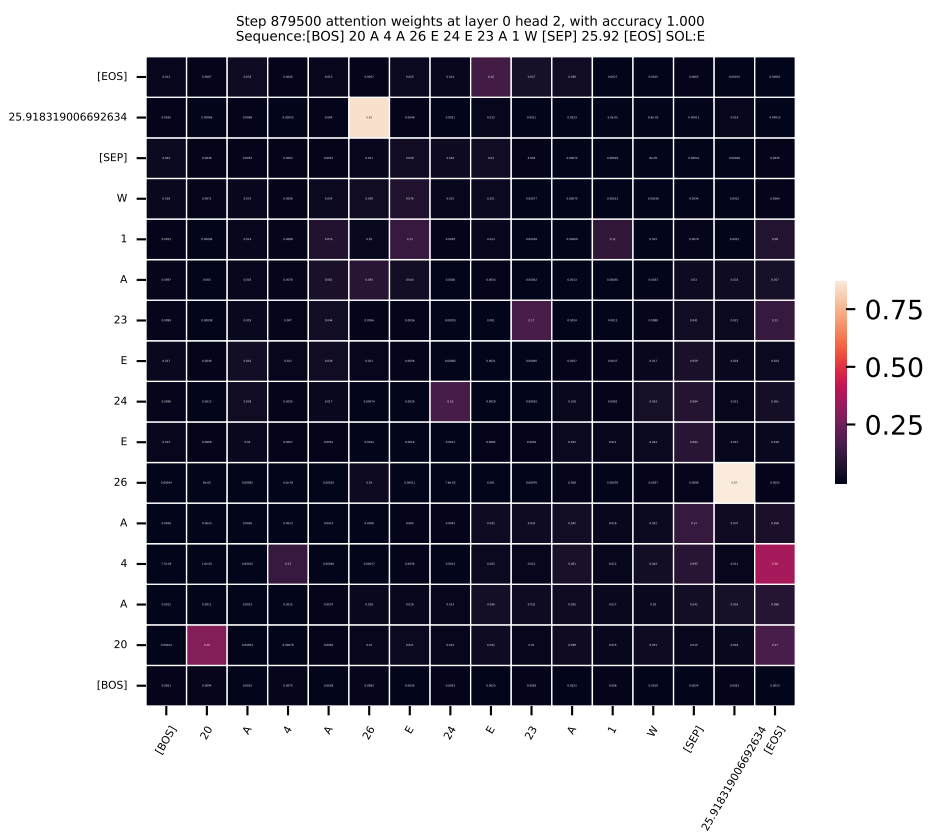

Step 870000 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

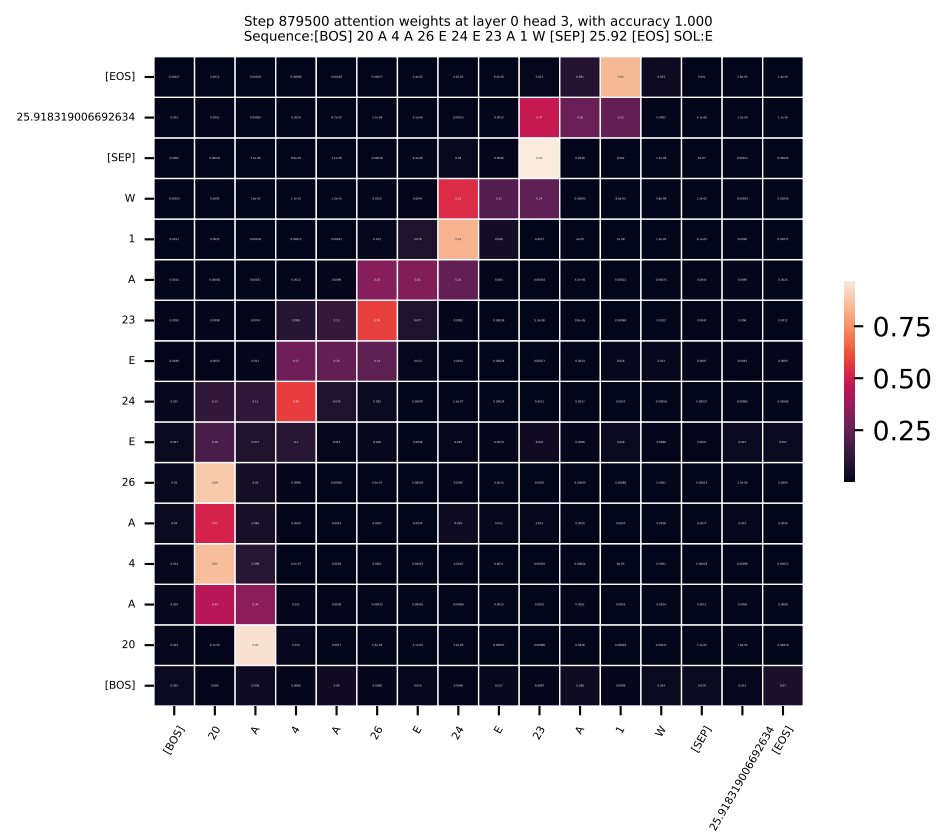

Step 870000 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

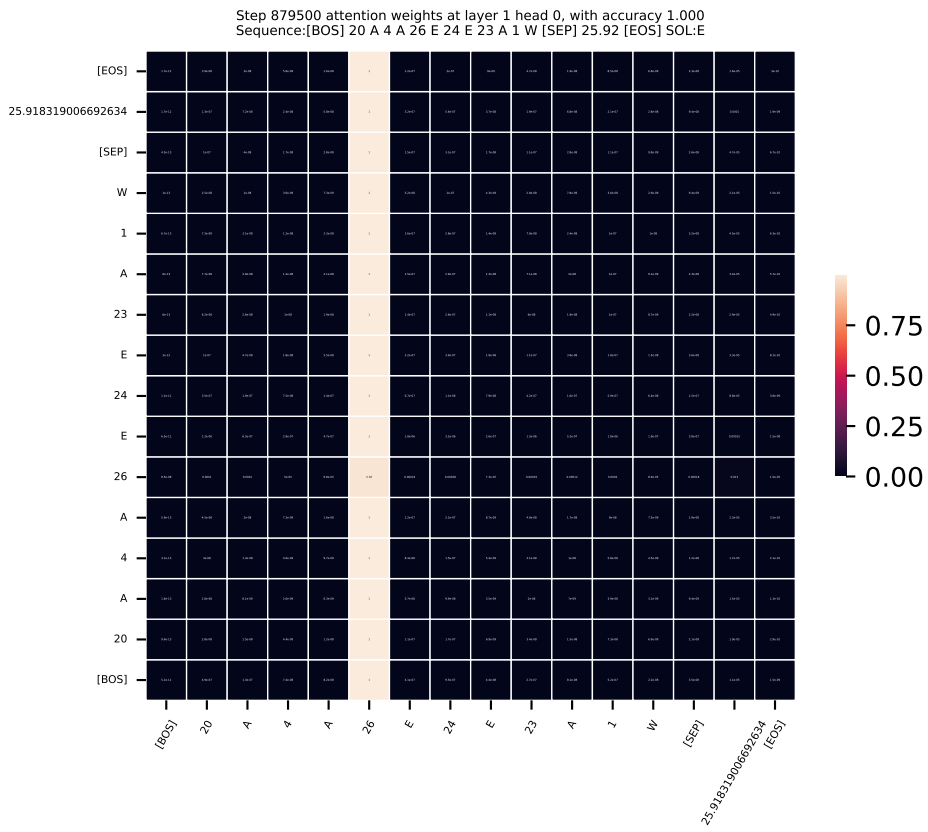

Step 870000 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

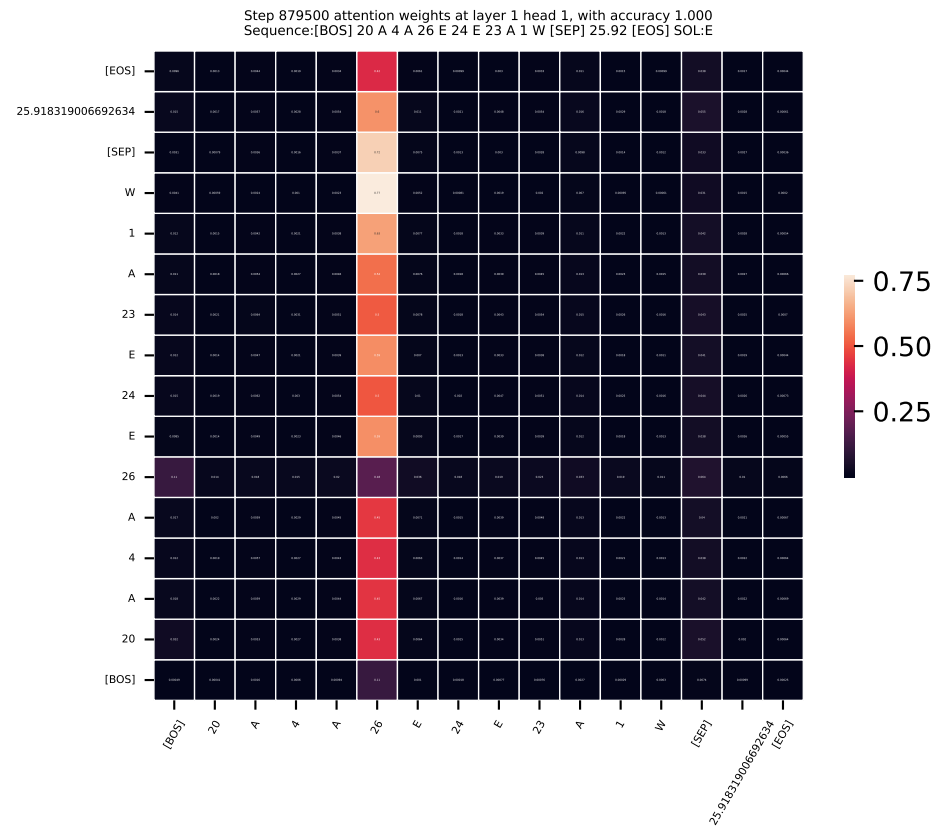

Step 879500 attention weights at layer 0 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

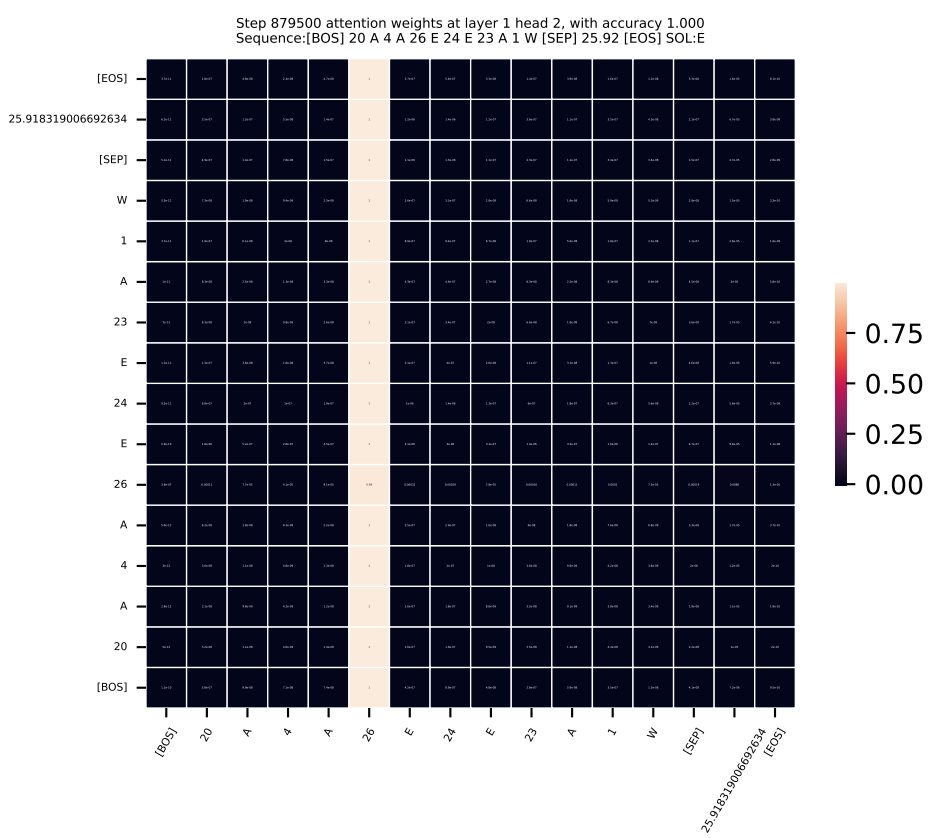

Step 879500 attention weights at layer 0 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

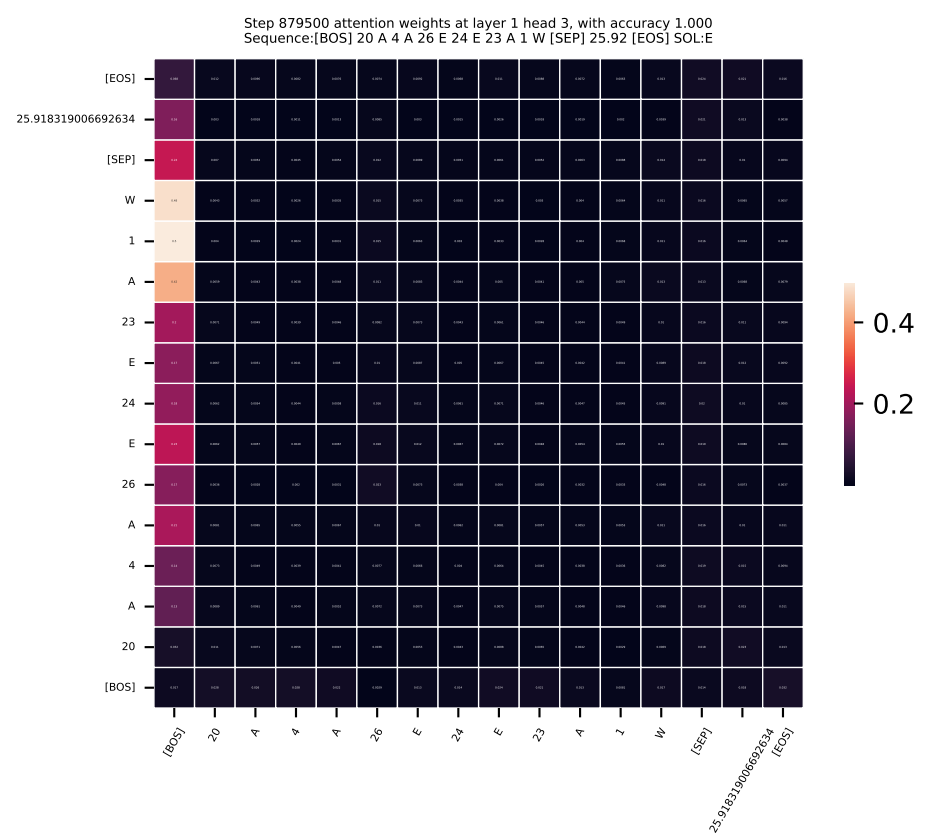

Step 879500 attention weights at layer 0 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

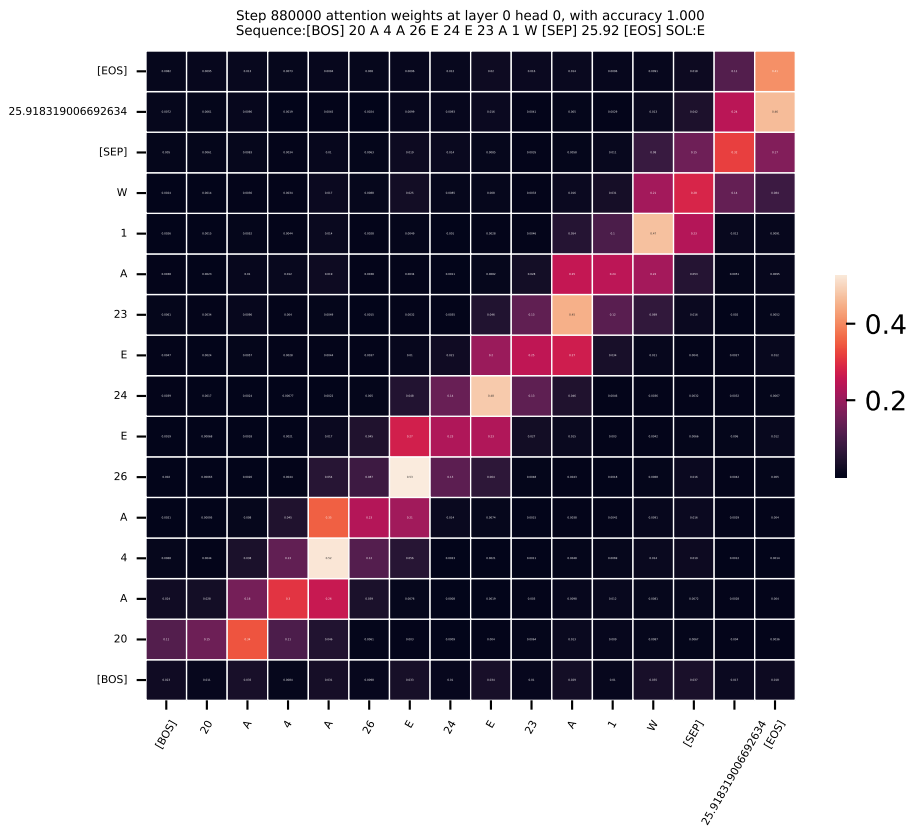

Step 879500 attention weights at layer 0 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

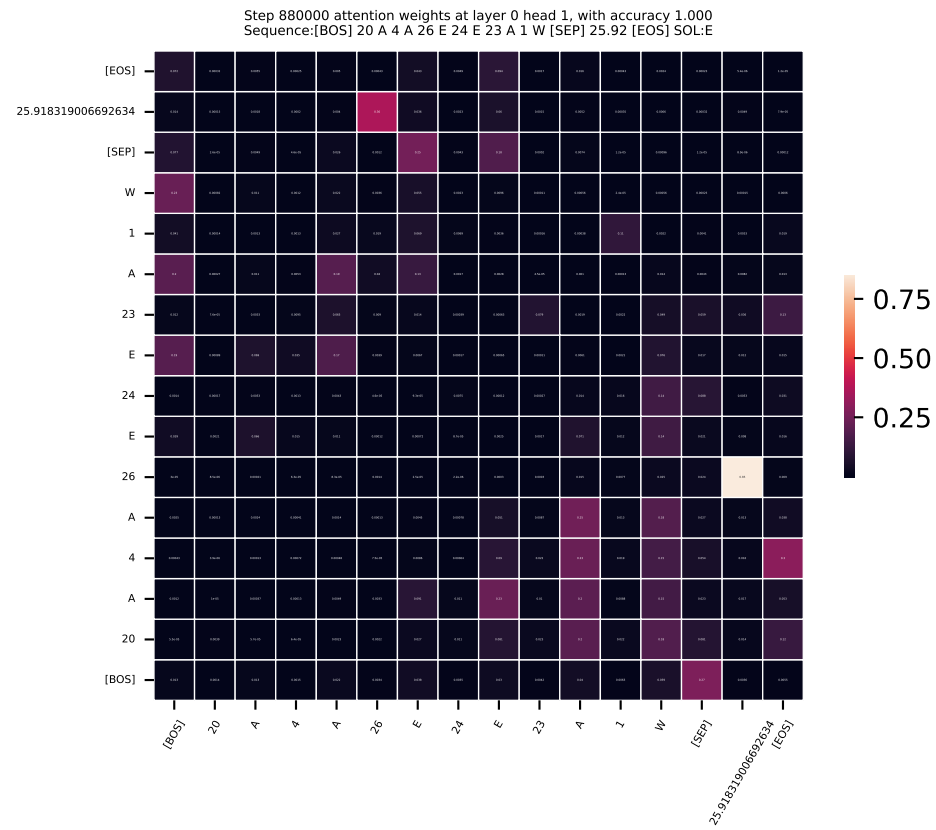

Step 879500 attention weights at layer 1 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

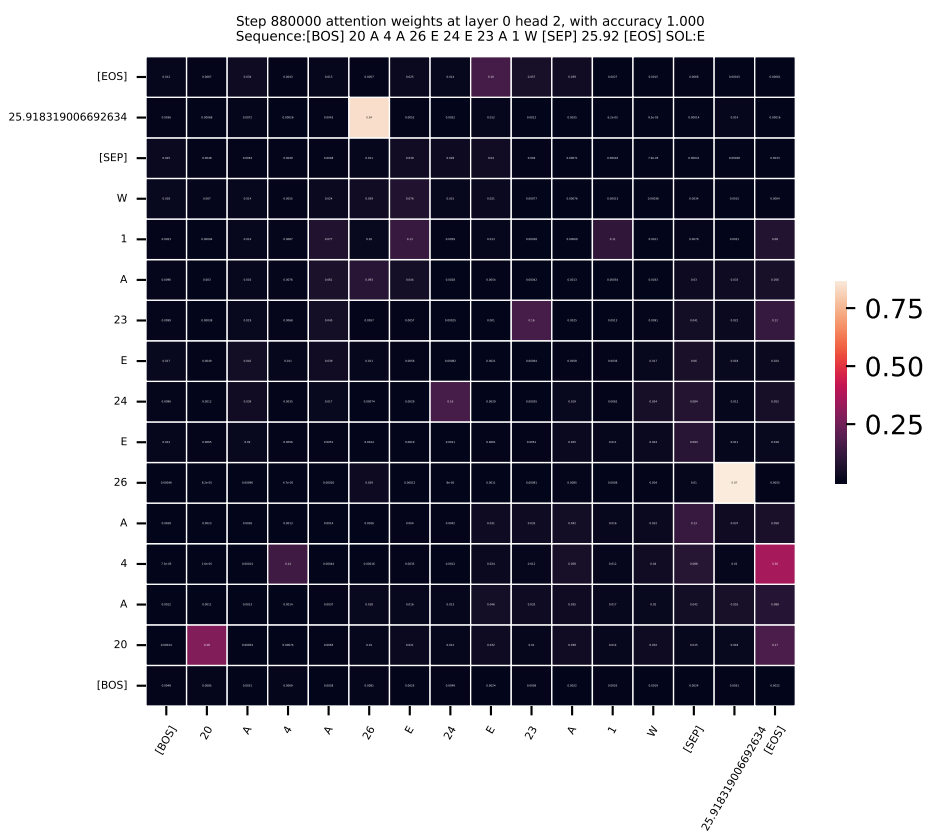

Step 879500 attention weights at layer 1 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

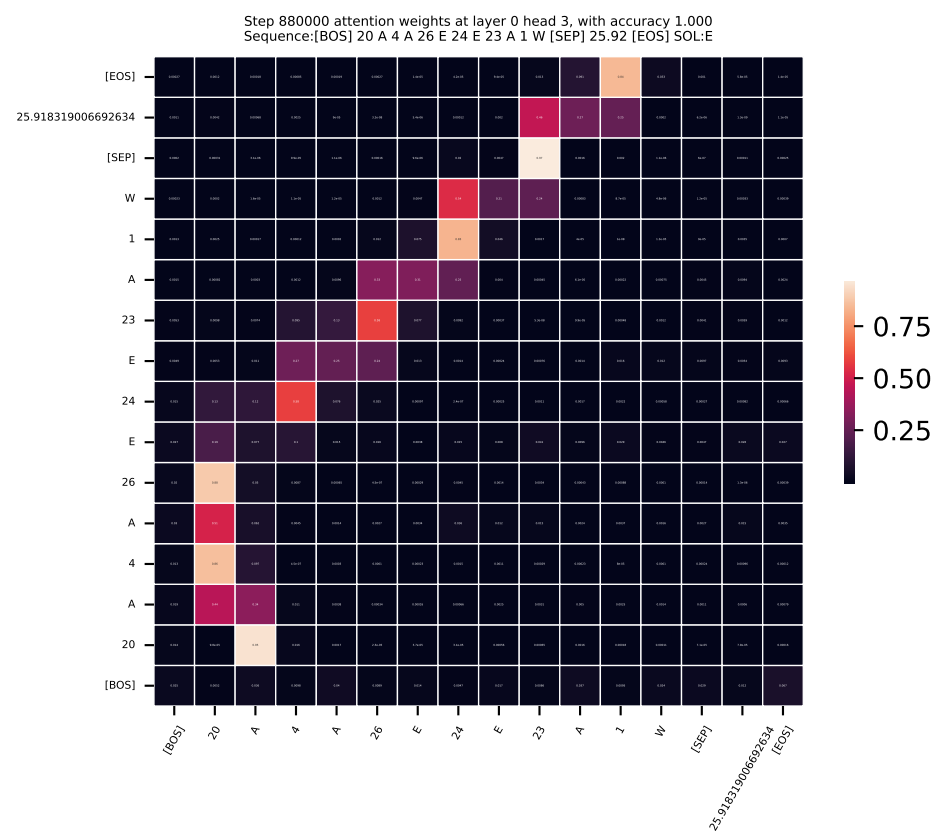

Step 879500 attention weights at layer 1 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

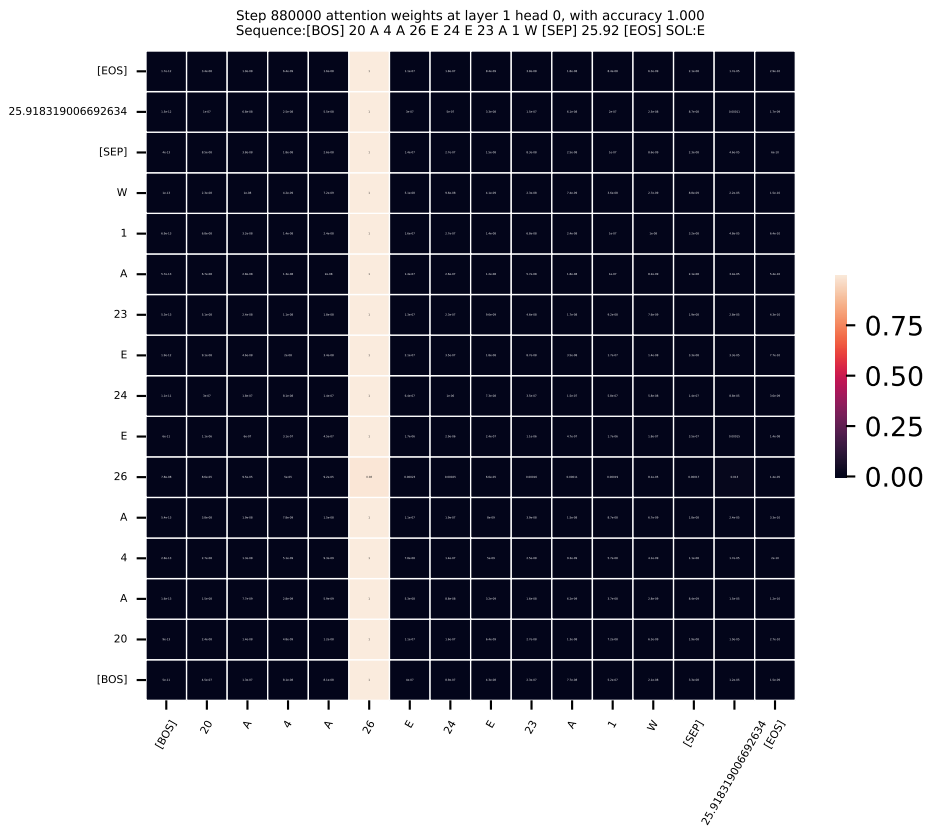

Step 879500 attention weights at layer 1 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

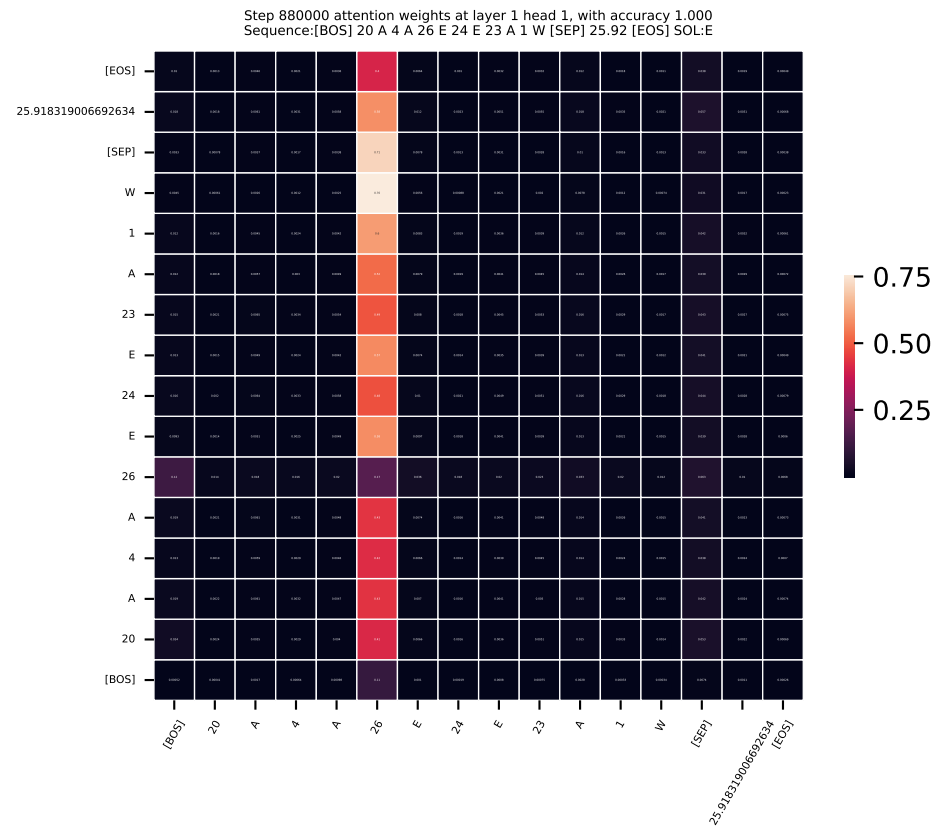

Step 880000 attention weights at layer 0 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

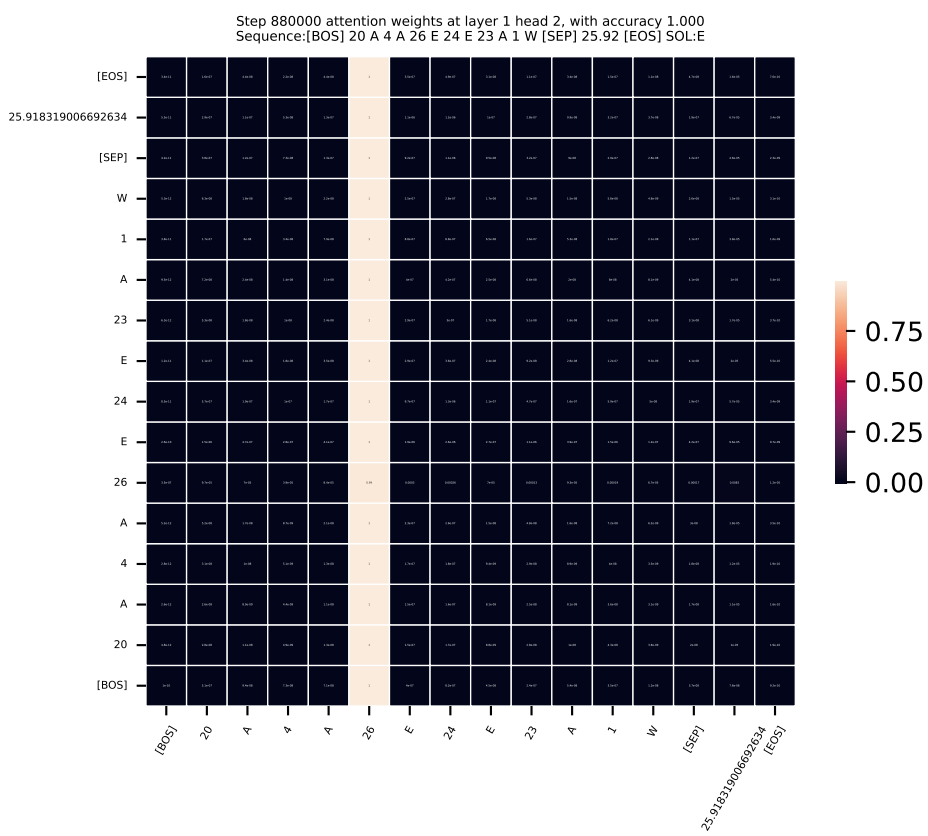

Step 880000 attention weights at layer 0 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

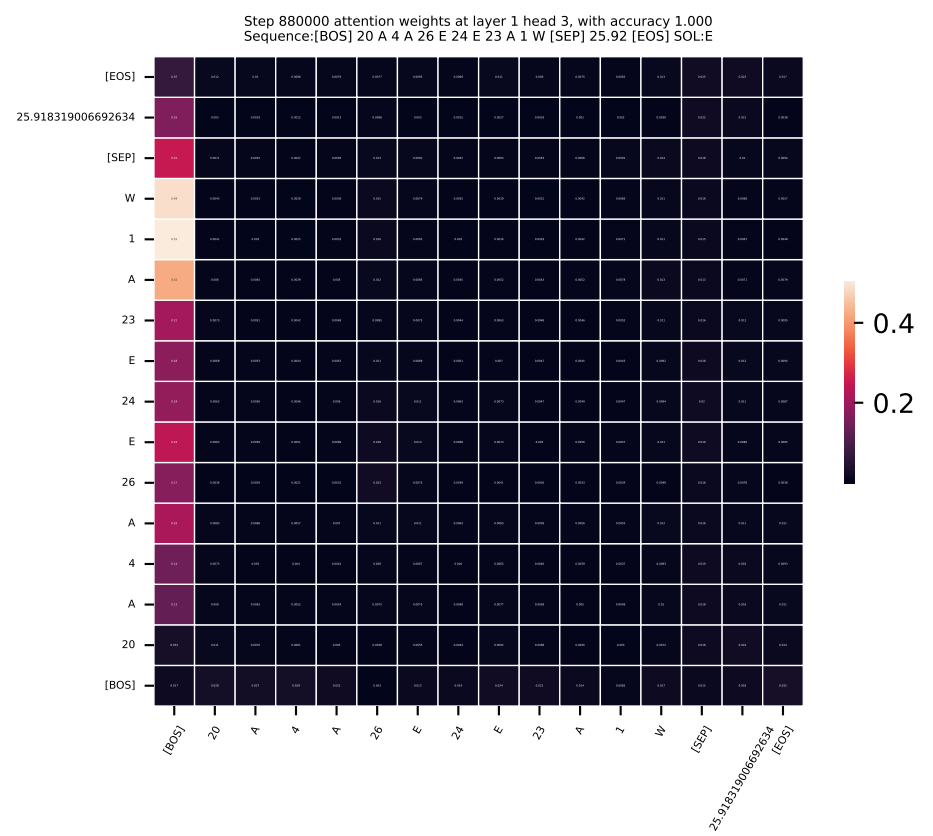

Step 880000 attention weights at layer 0 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

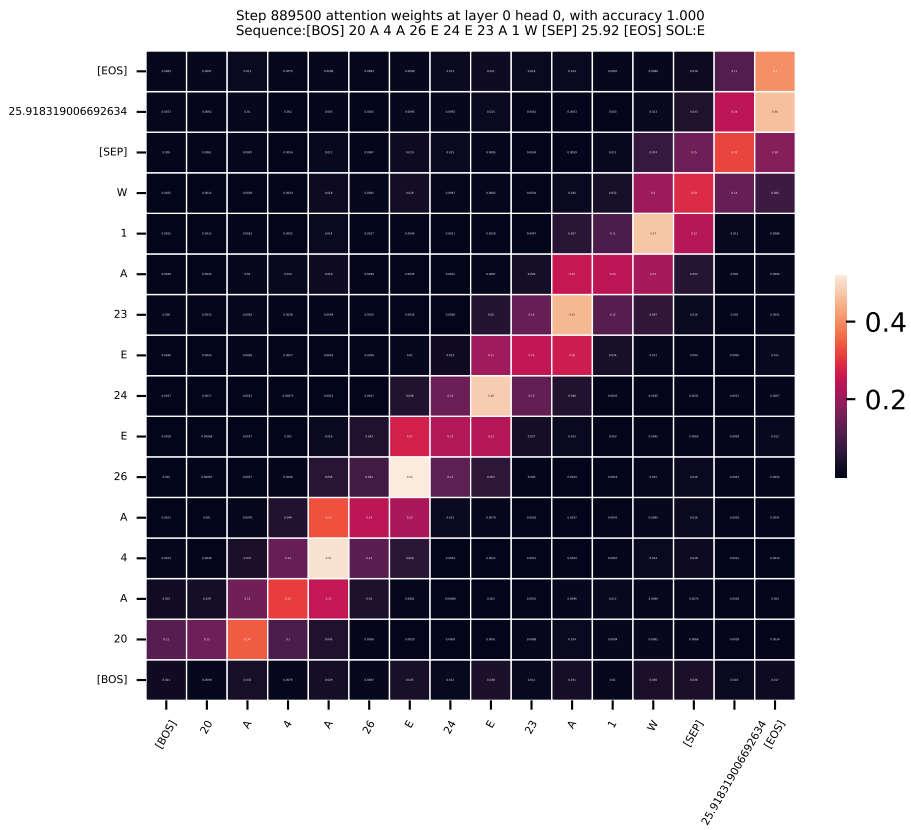

Step 880000 attention weights at layer 0 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

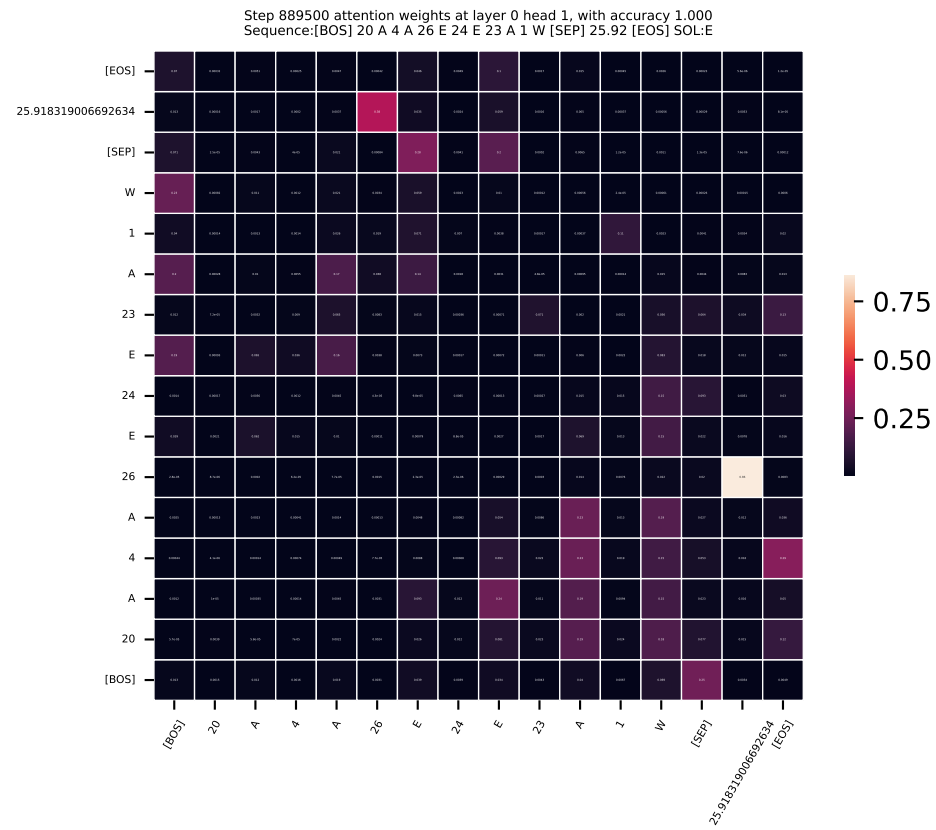

Step 880000 attention weights at layer 1 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

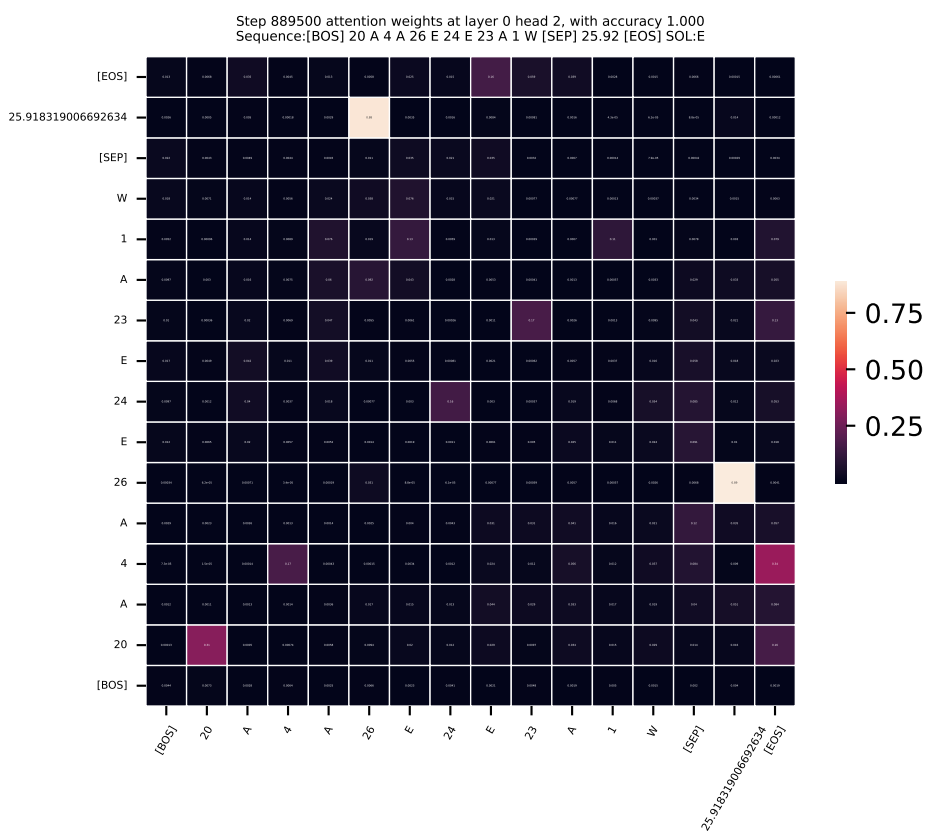

Step 880000 attention weights at layer 1 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

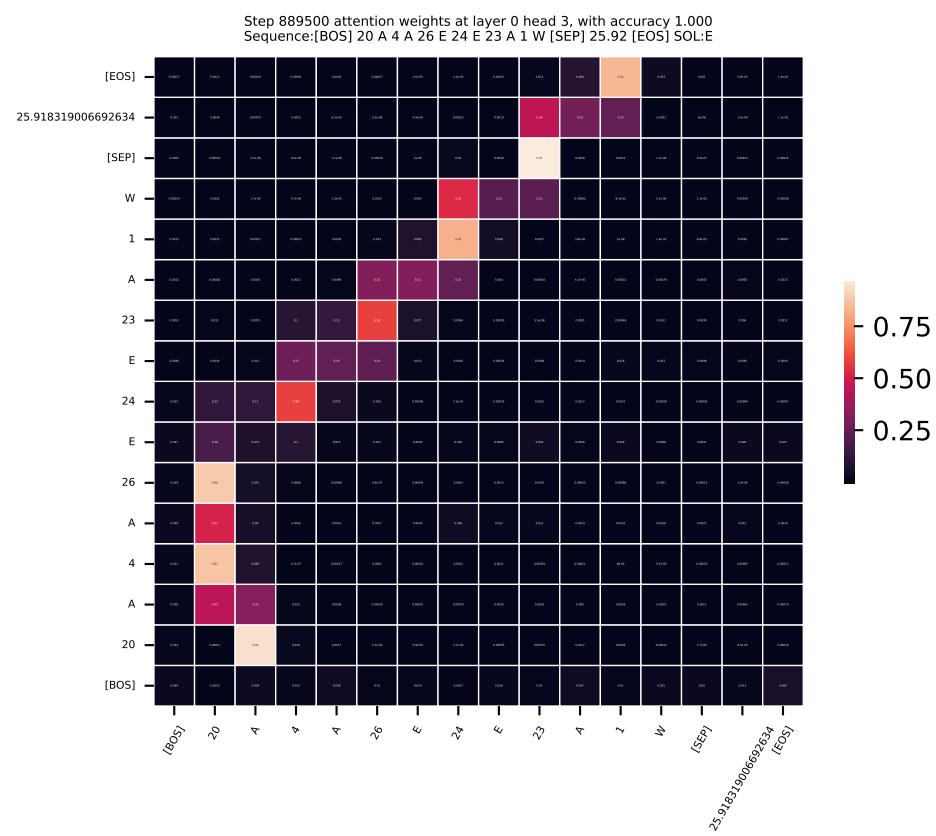

Step 880000 attention weights at layer 1 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

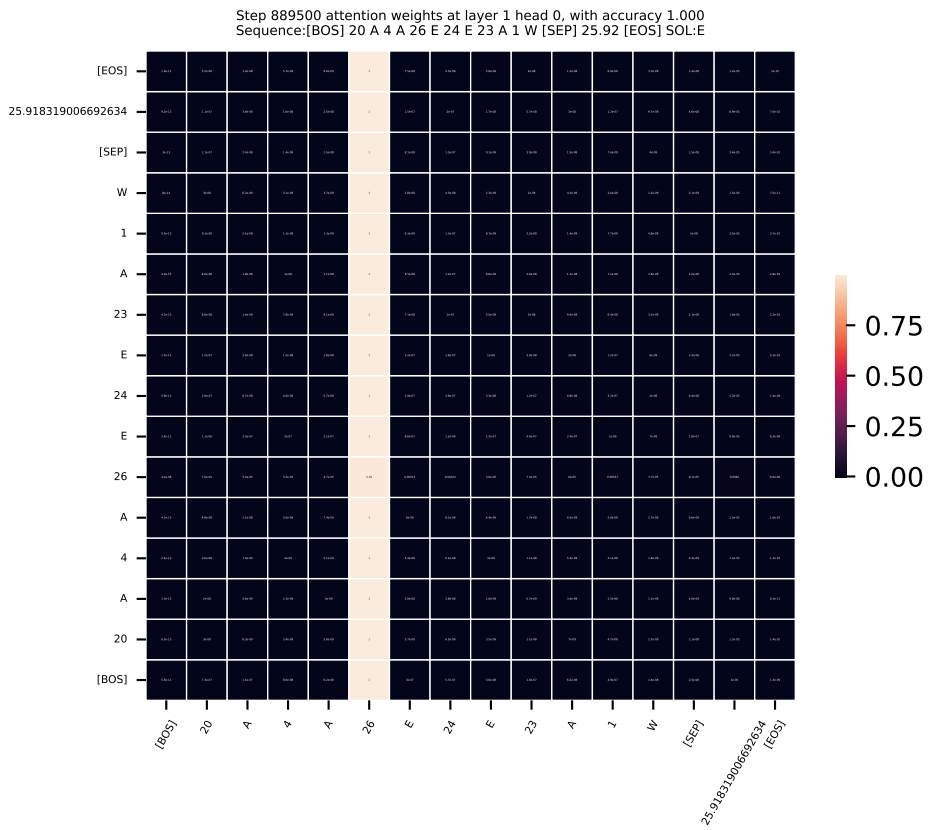

Step 880000 attention weights at layer 1 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

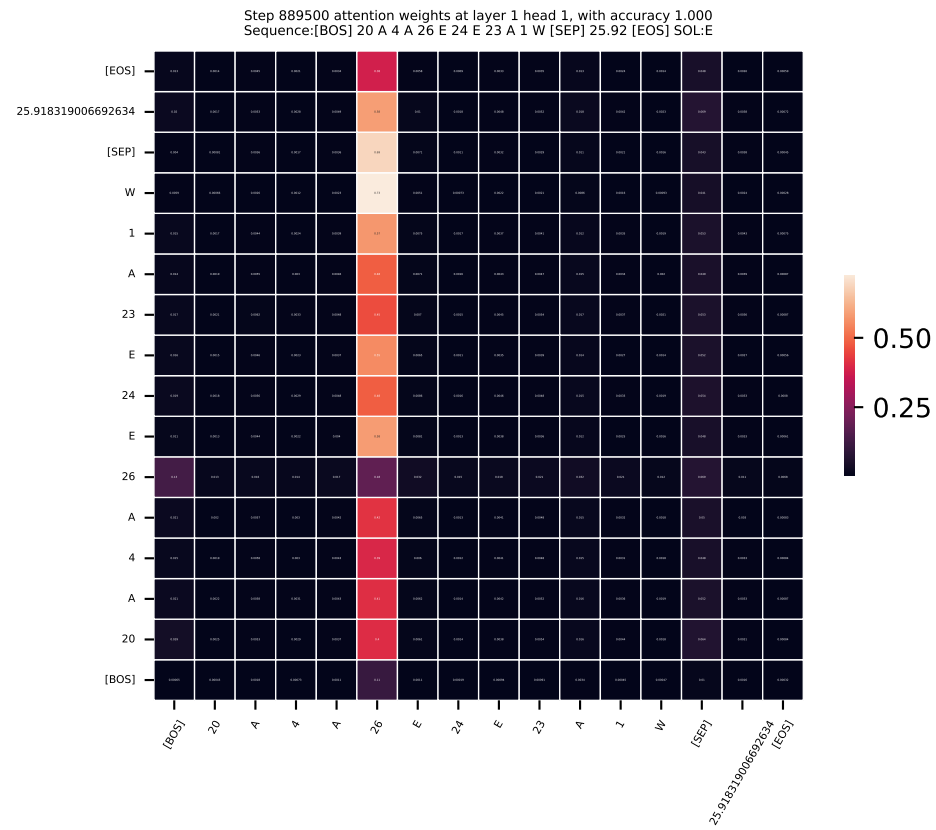

Step 889500 attention weights at layer 0 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

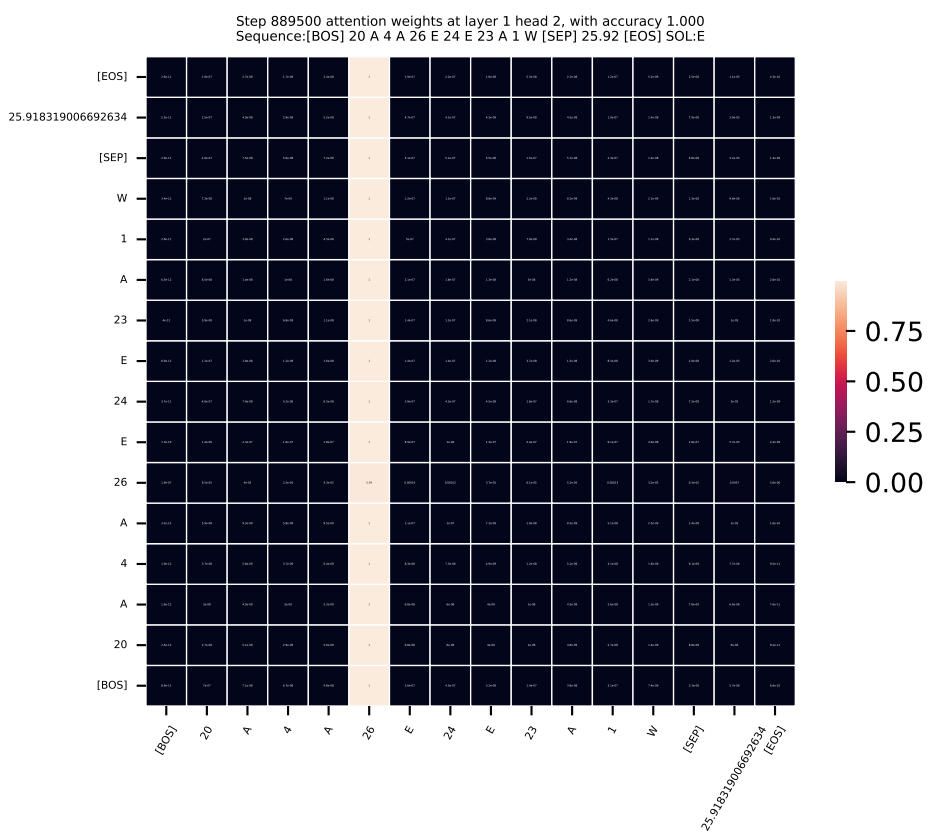

Step 889500 attention weights at layer 0 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

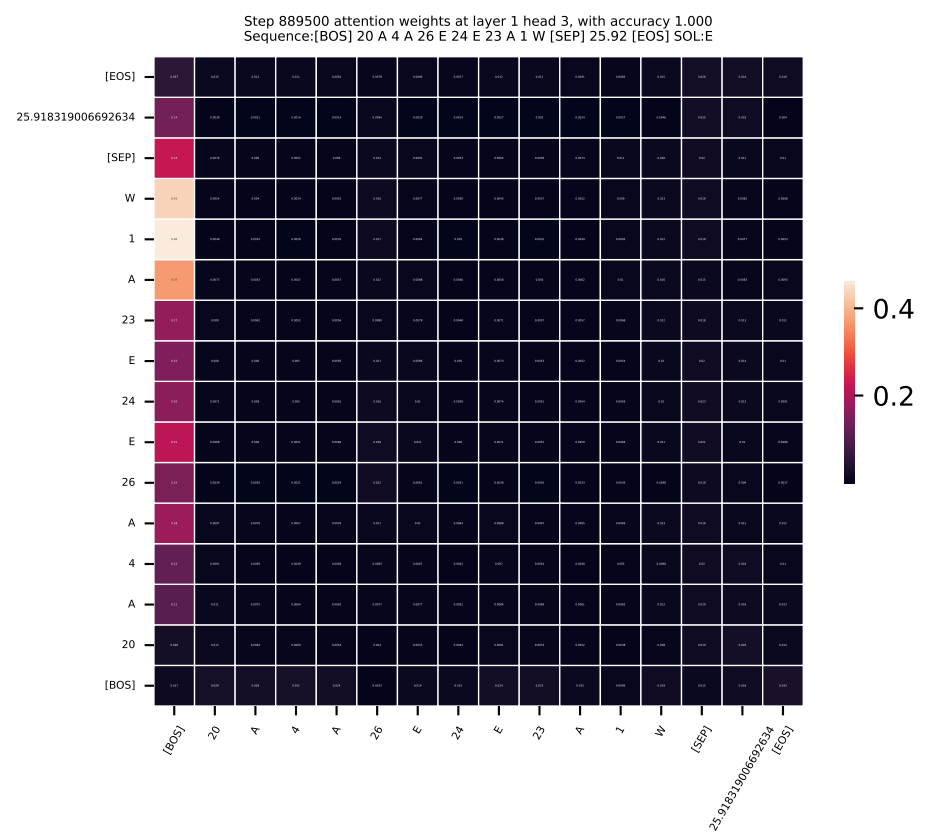

Step 889500 attention weights at layer 0 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

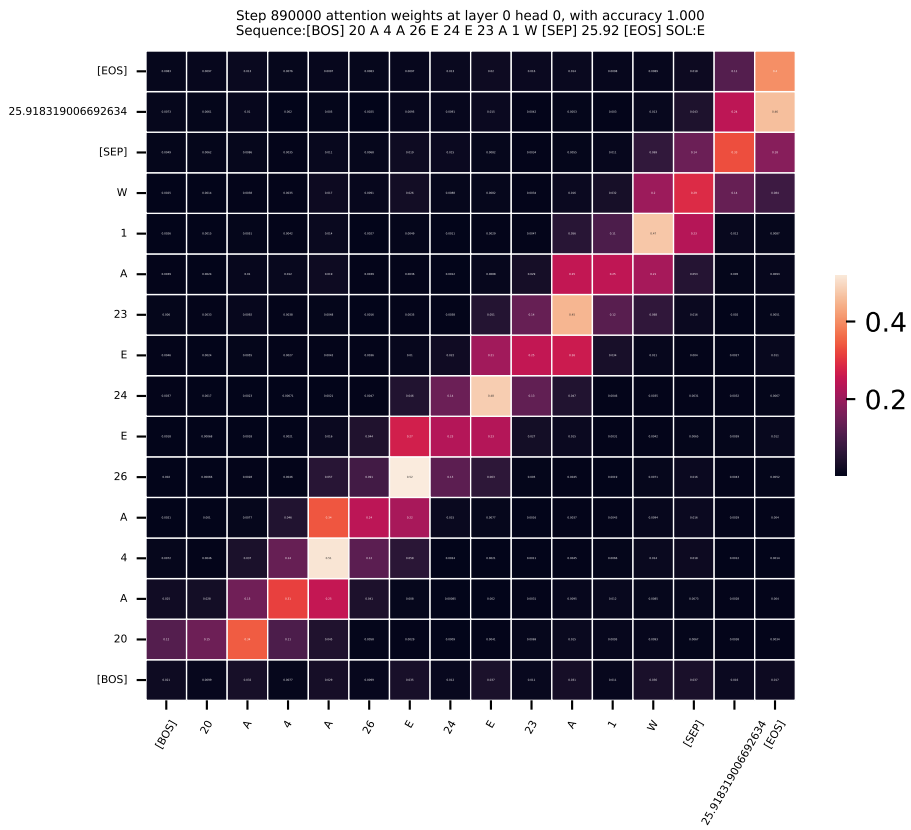

Step 889500 attention weights at layer 0 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

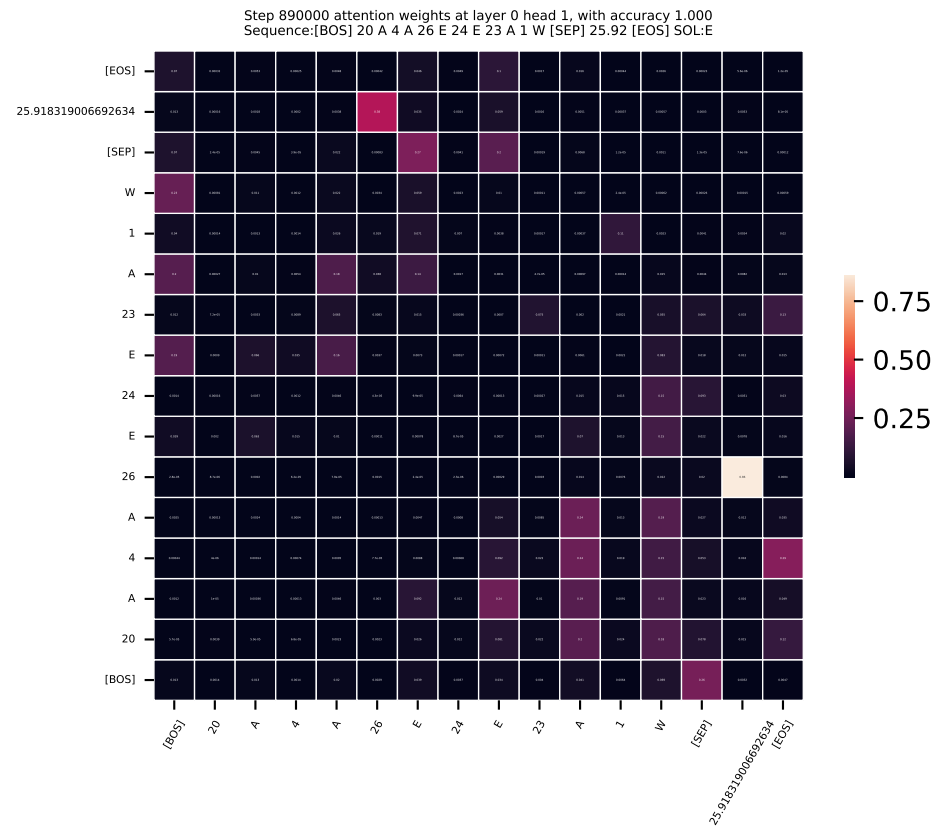

Step 889500 attention weights at layer 1 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

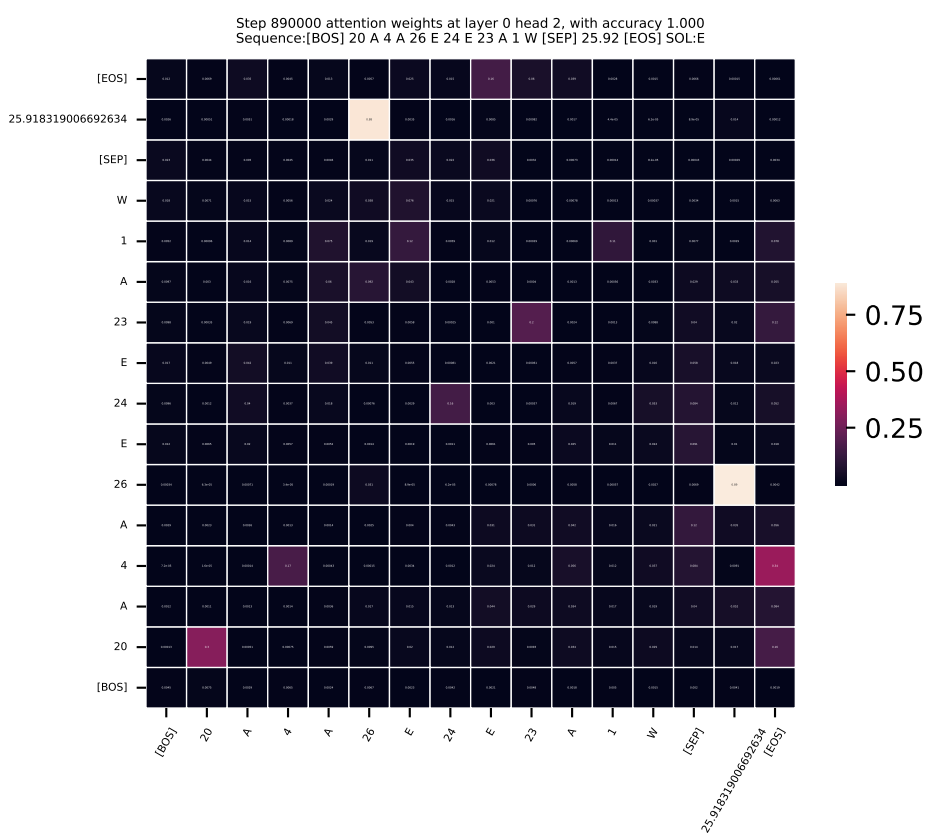

Step 889500 attention weights at layer 1 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

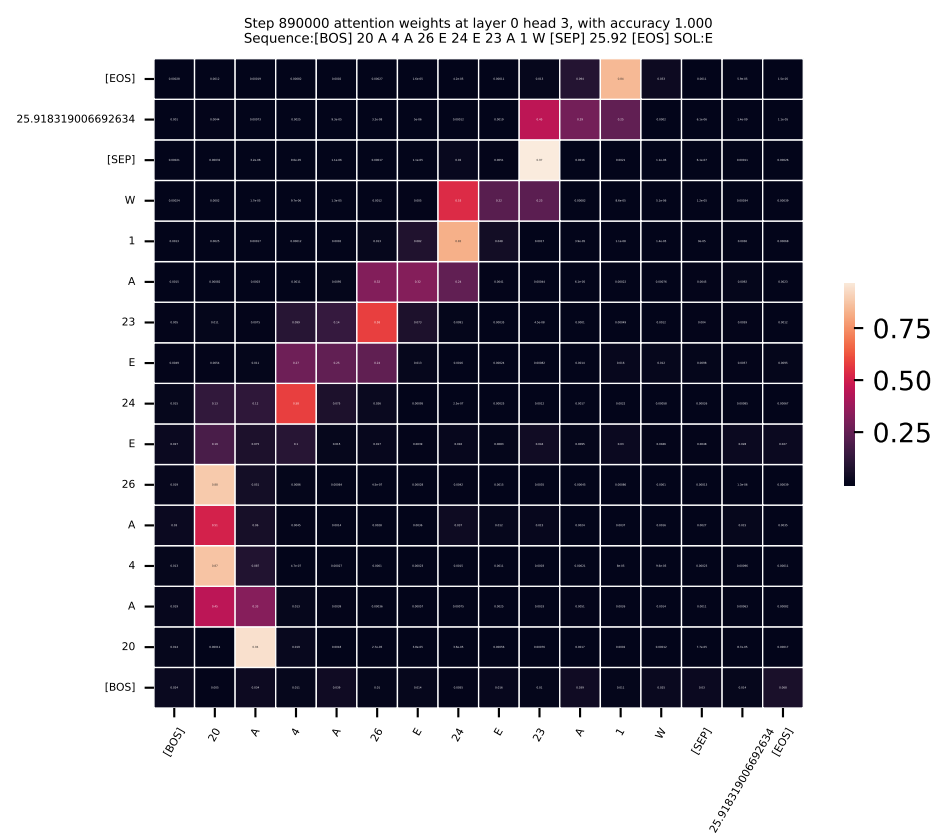

Step 889500 attention weights at layer 1 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

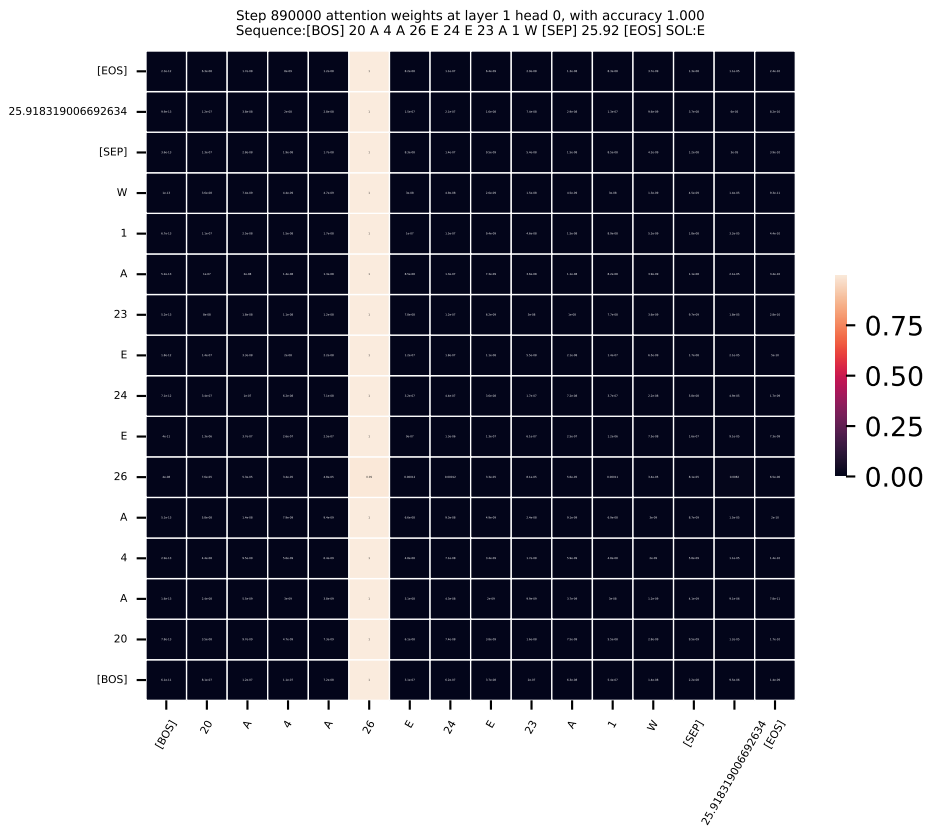

Step 889500 attention weights at layer 1 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

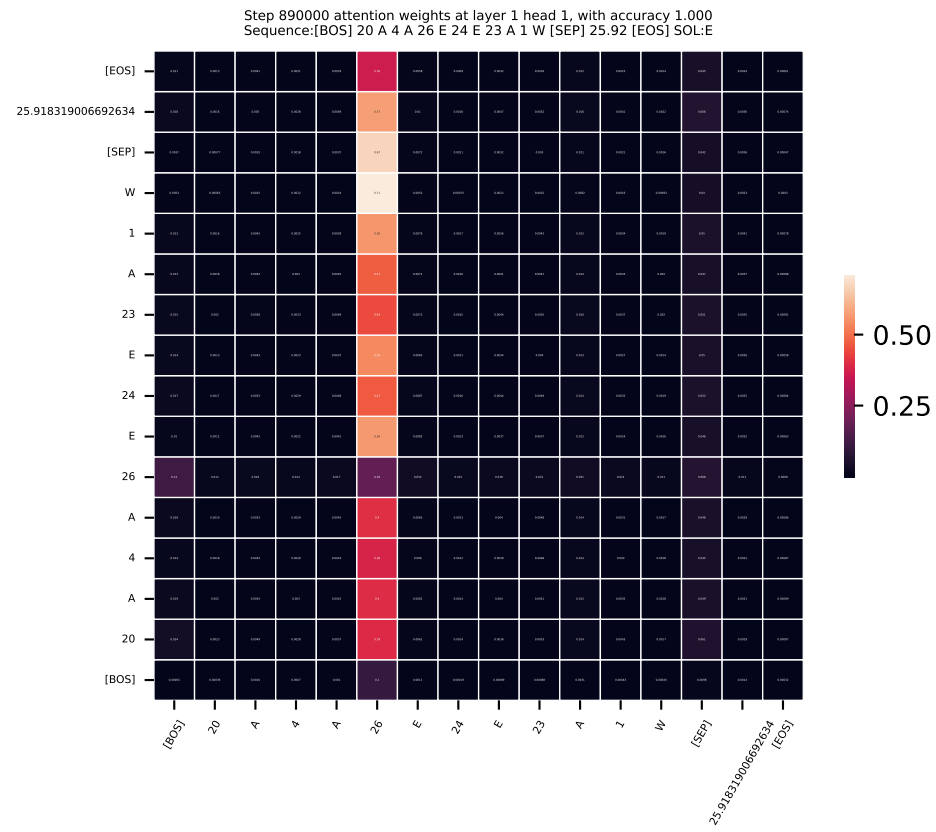

Step 890000 attention weights at layer 0 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

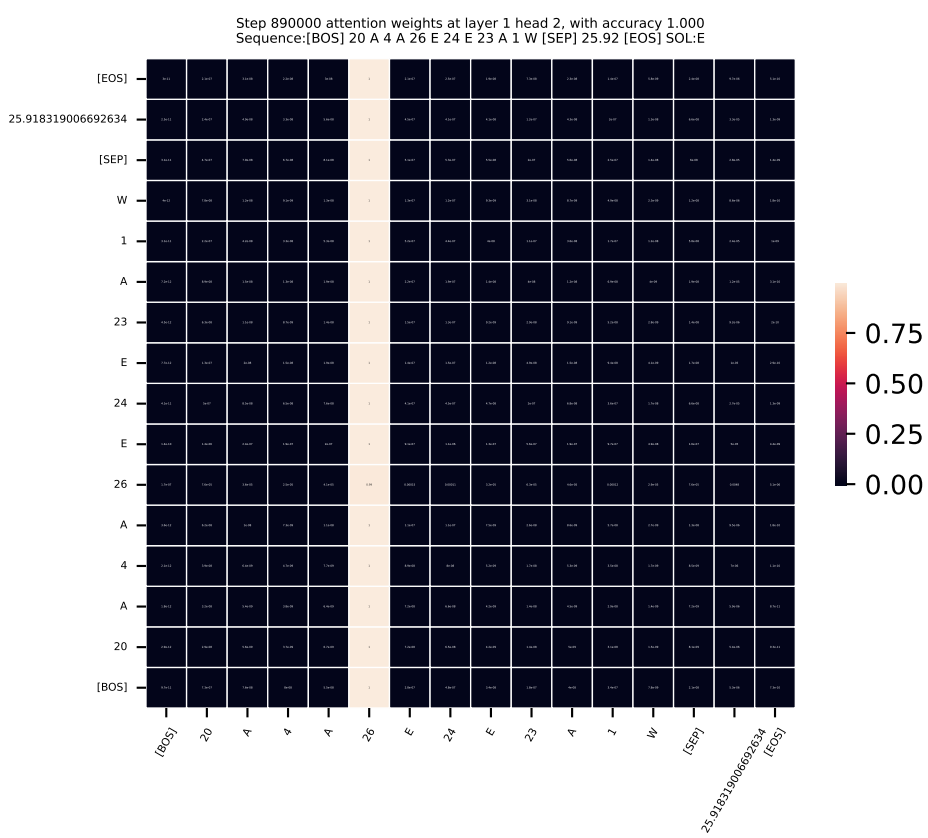

Step 890000 attention weights at layer 0 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

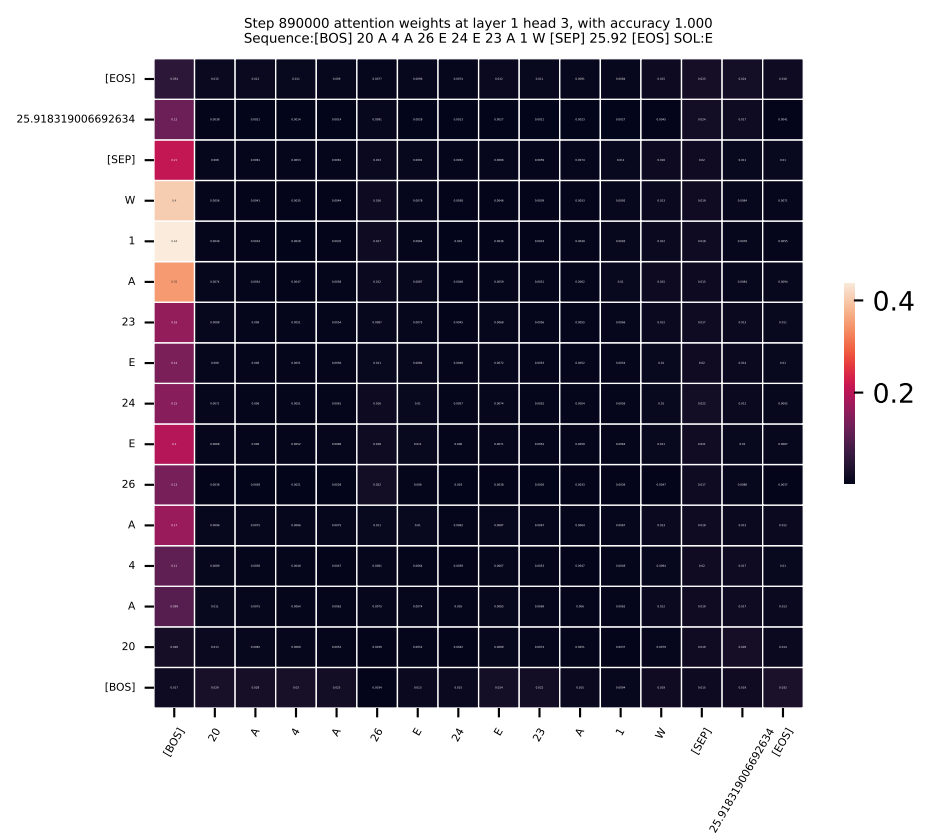

Step 890000 attention weights at layer 0 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

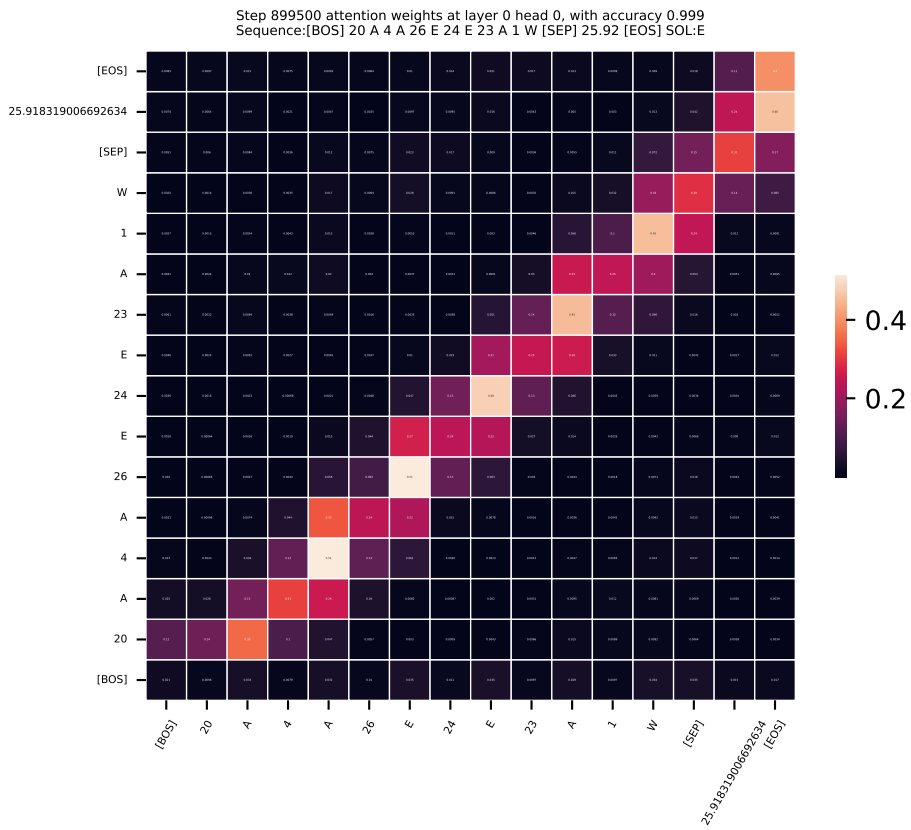

Step 890000 attention weights at layer 0 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25

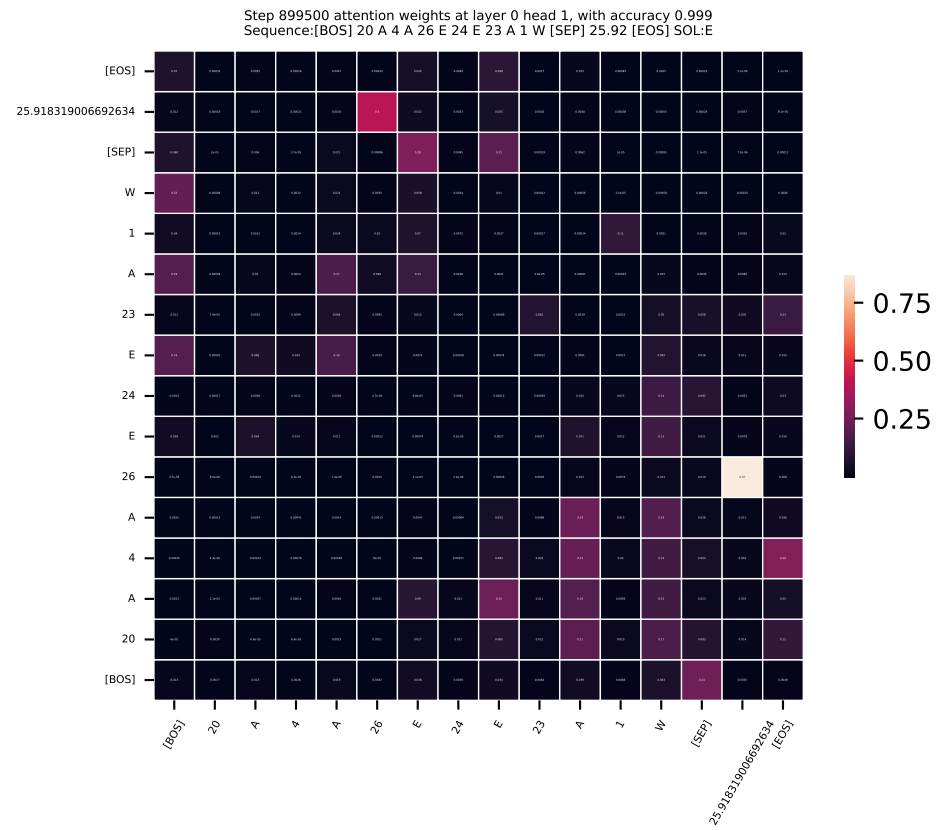

Step 890000 attention weights at layer 1 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

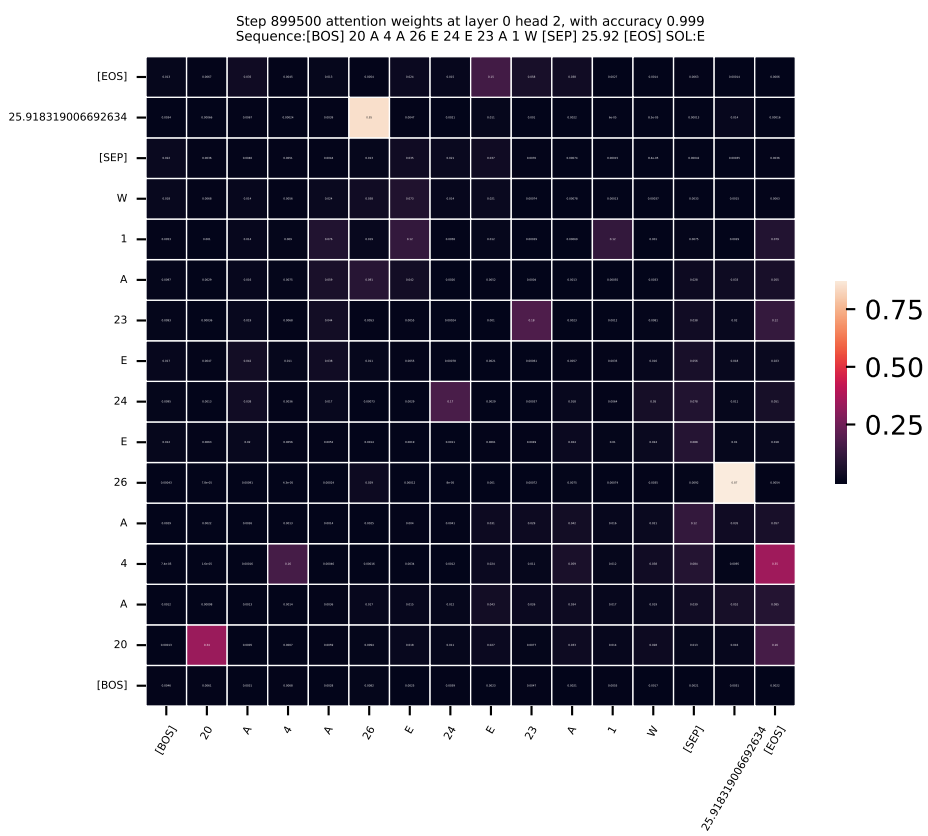

Step 890000 attention weights at layer 1 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

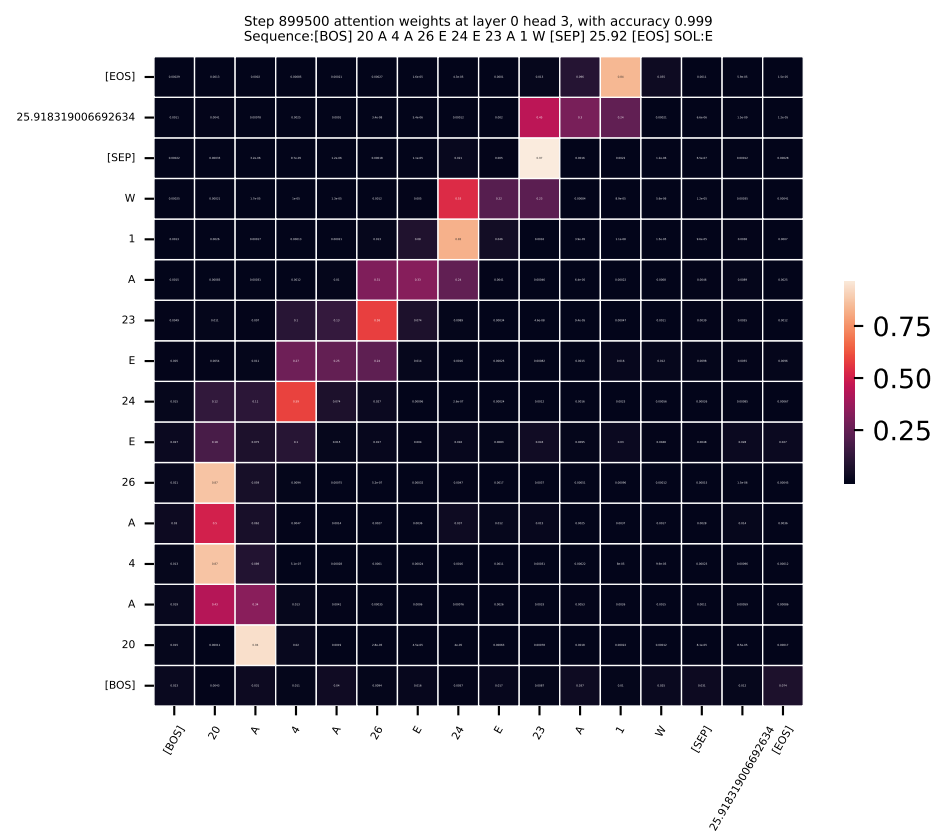

Step 890000 attention weights at layer 1 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

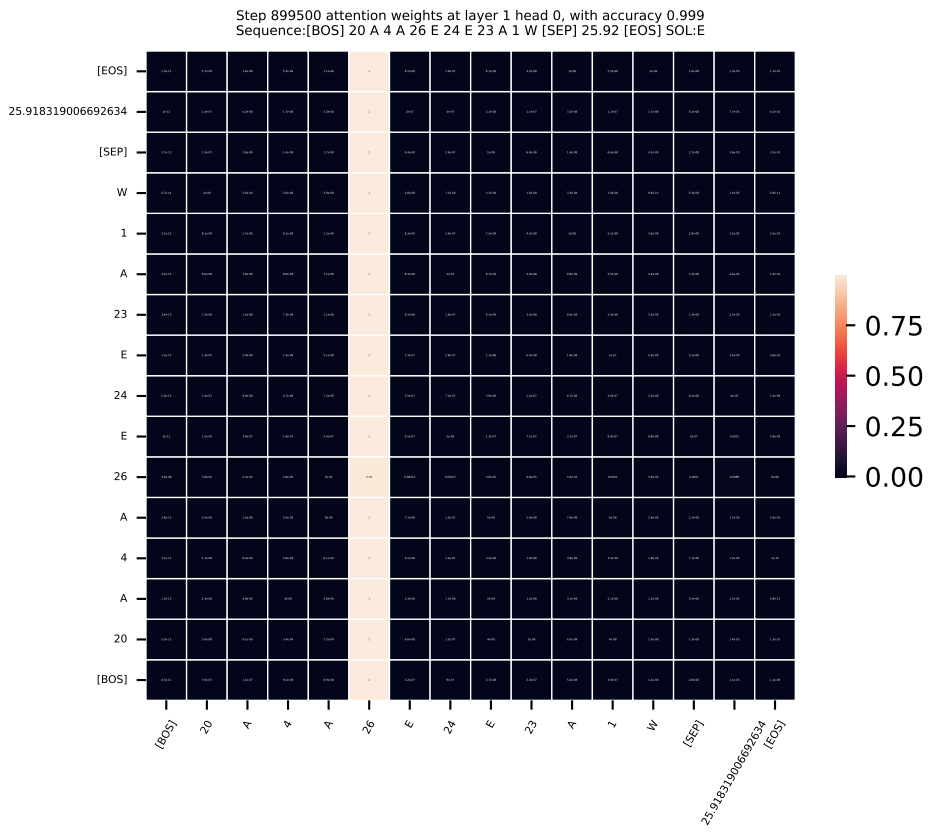

Step 890000 attention weights at layer 1 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

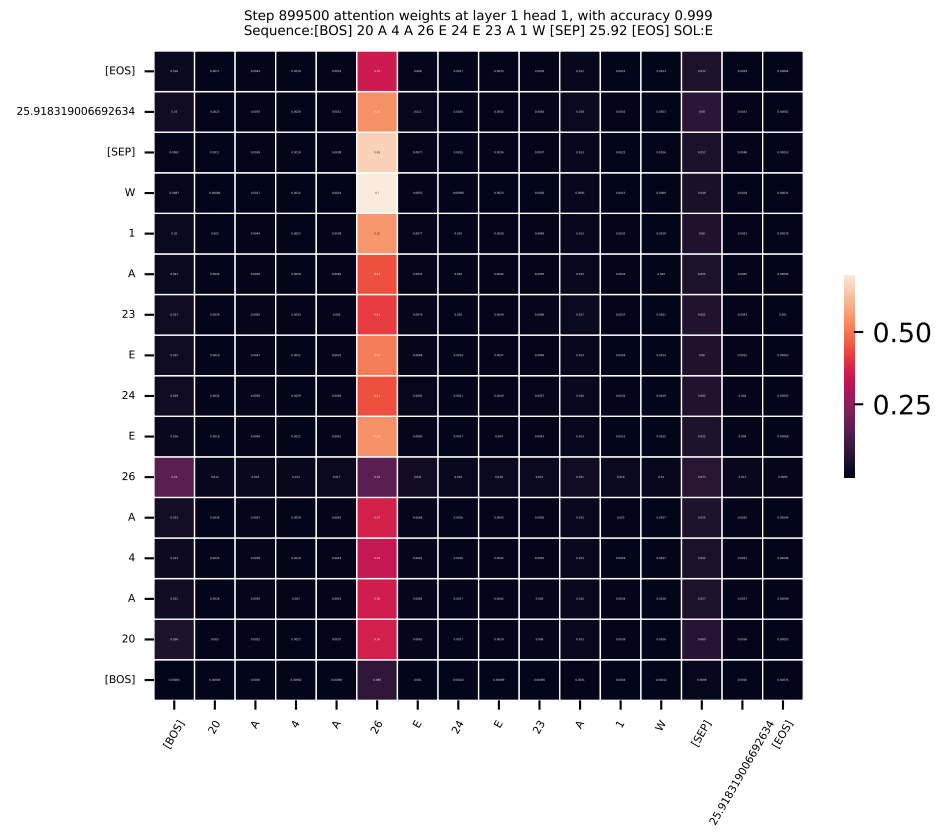

Step 899500 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

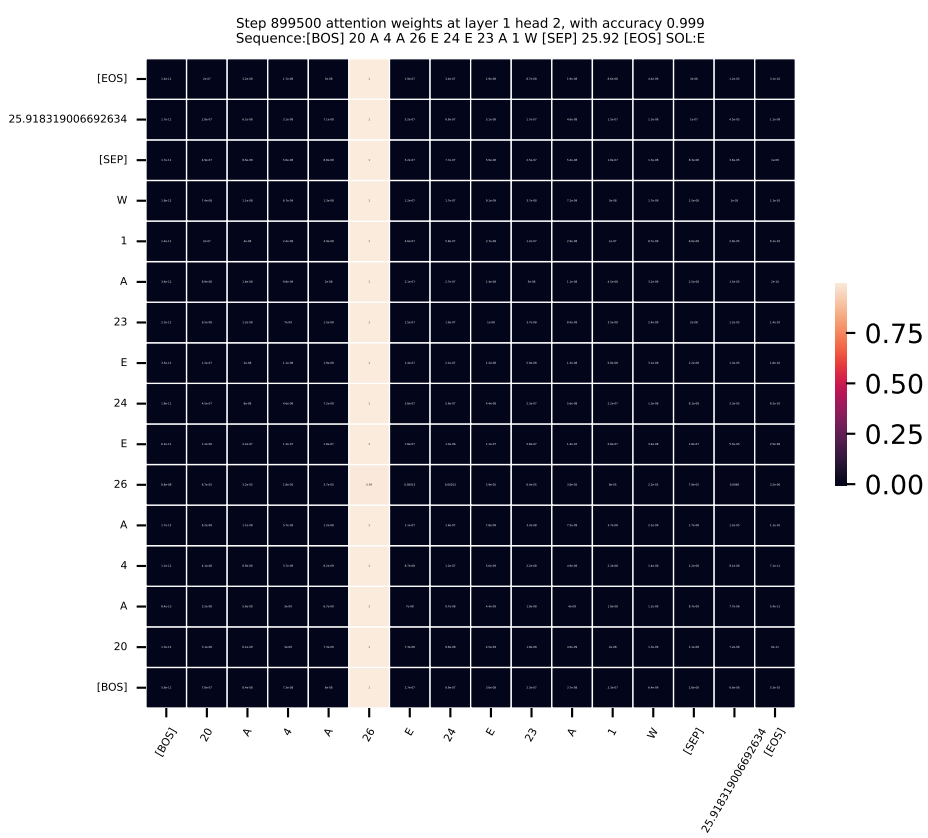

Step 899500 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

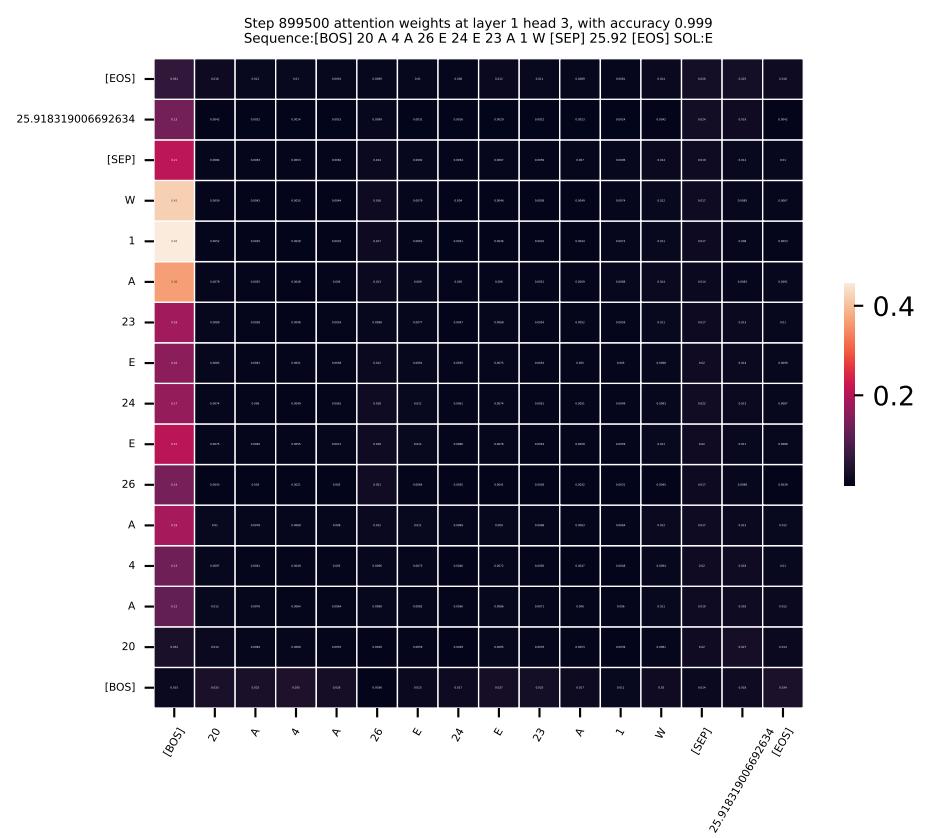

Step 899500 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

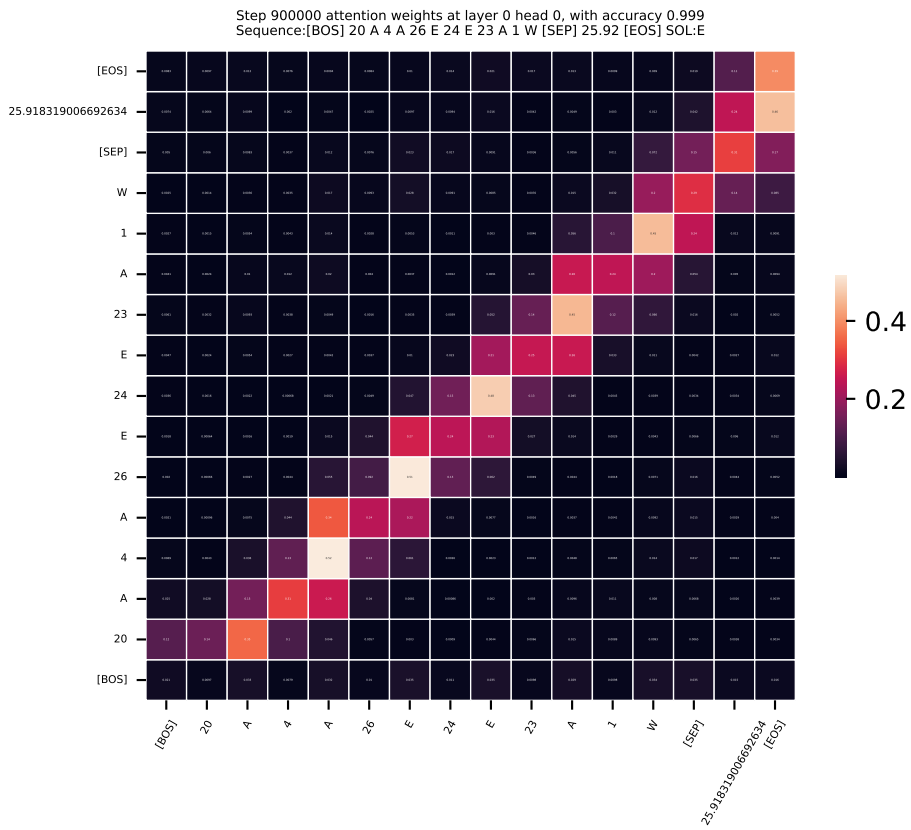

Step 899500 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

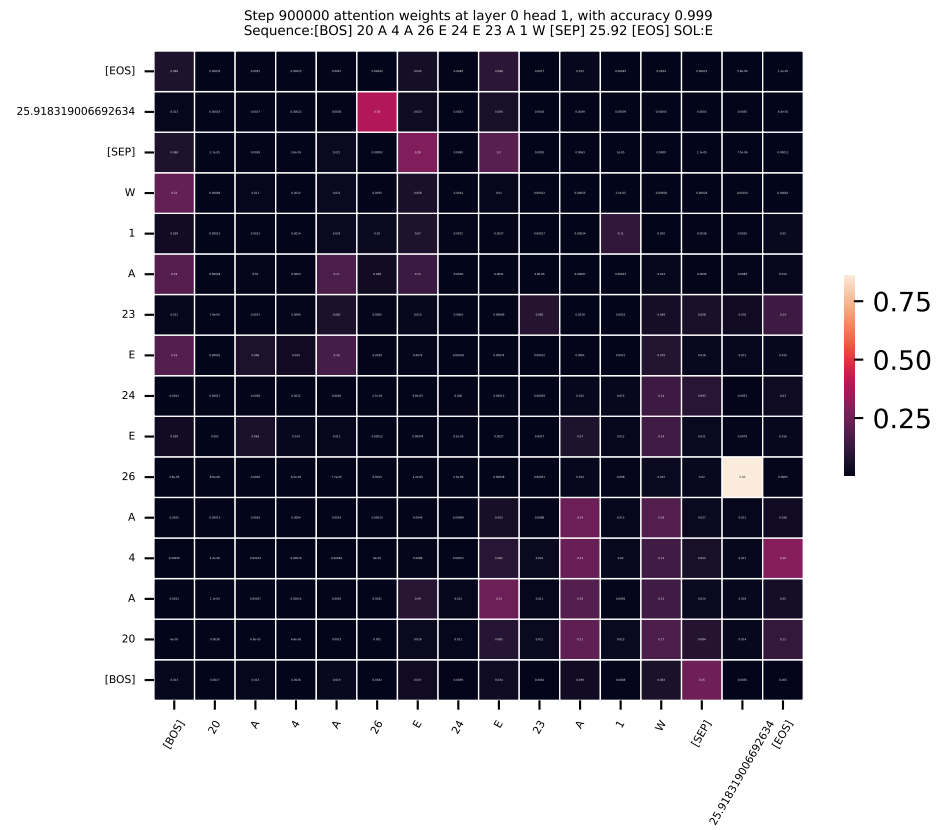

Step 899500 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

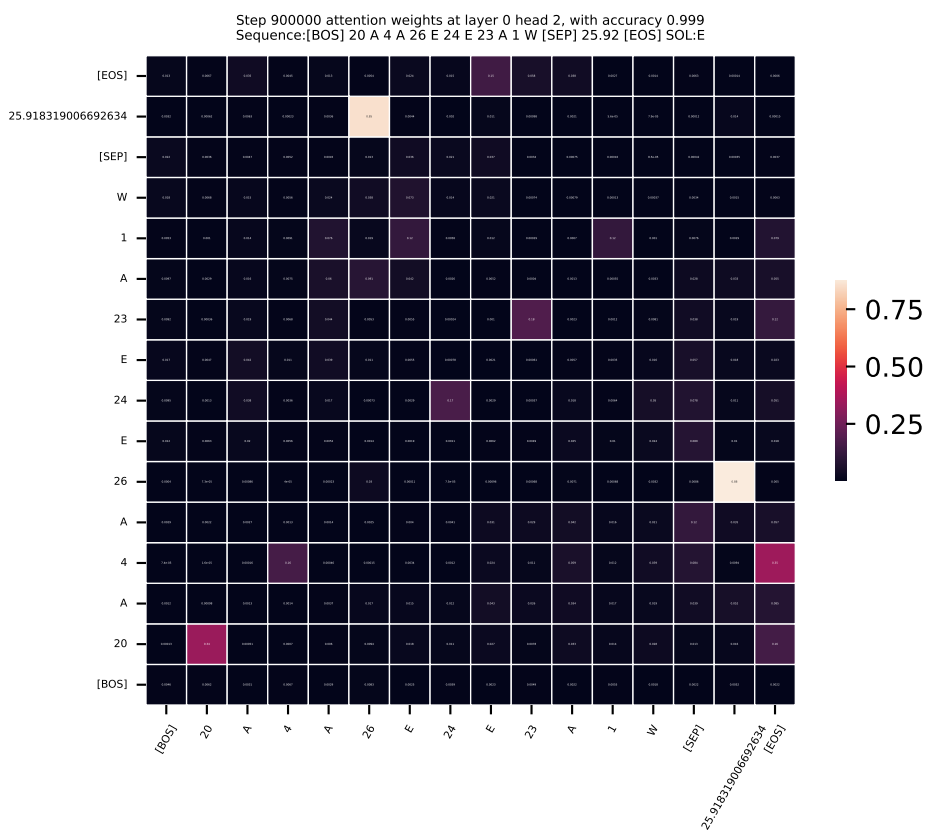

Step 899500 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

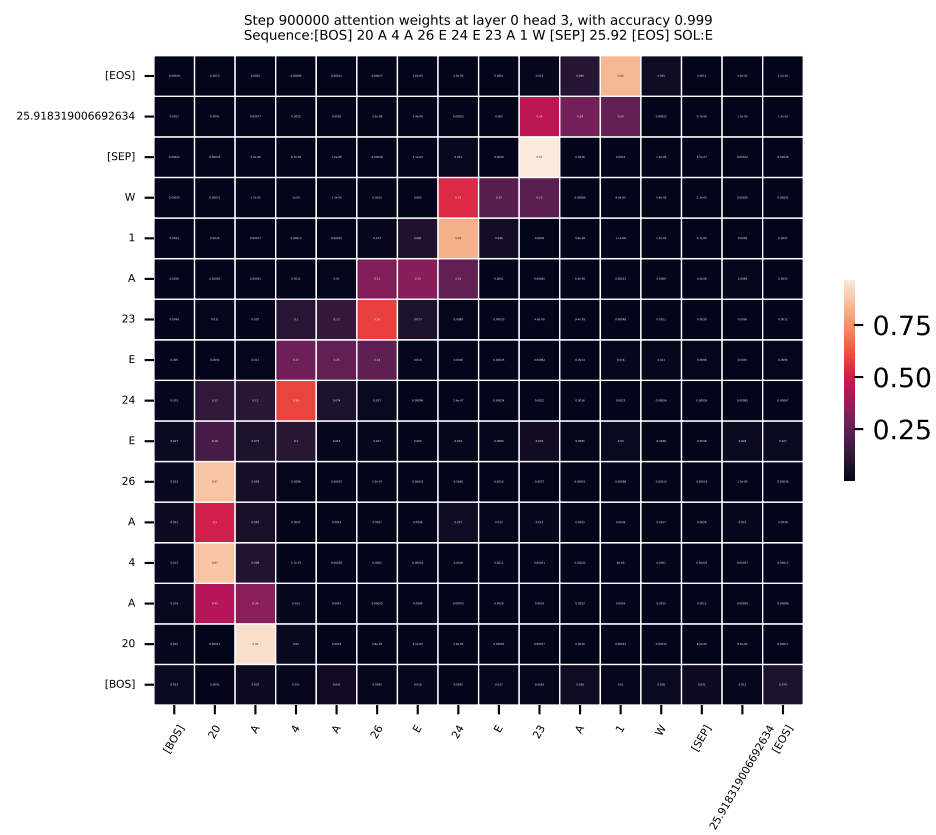

Step 899500 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

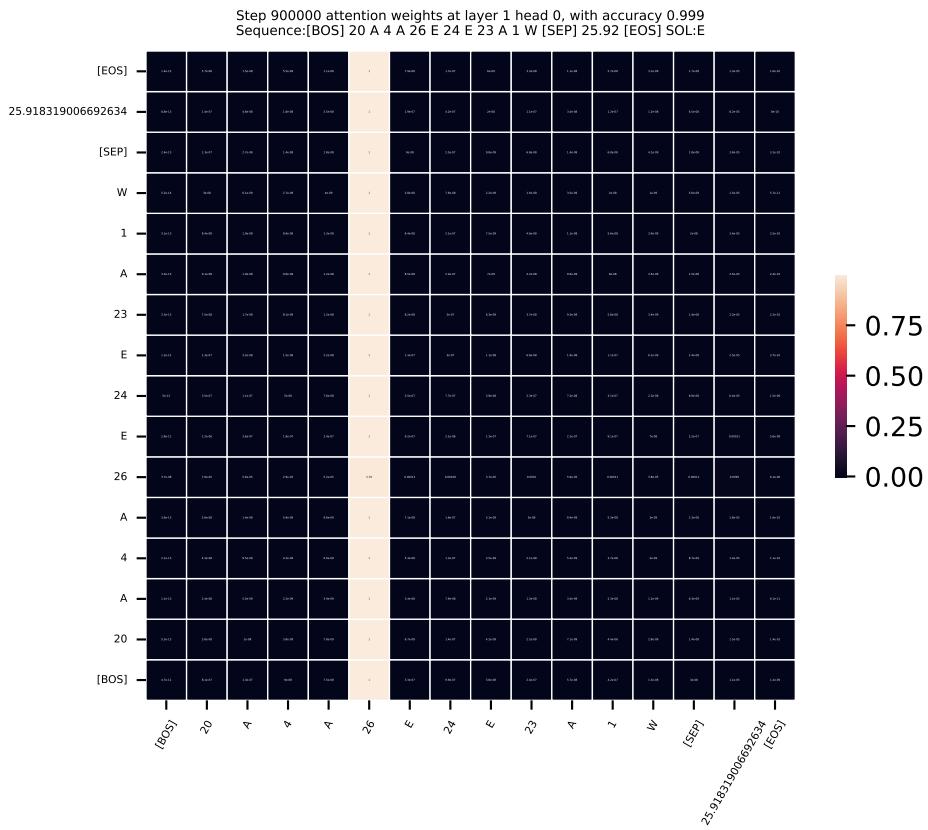

Step 899500 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

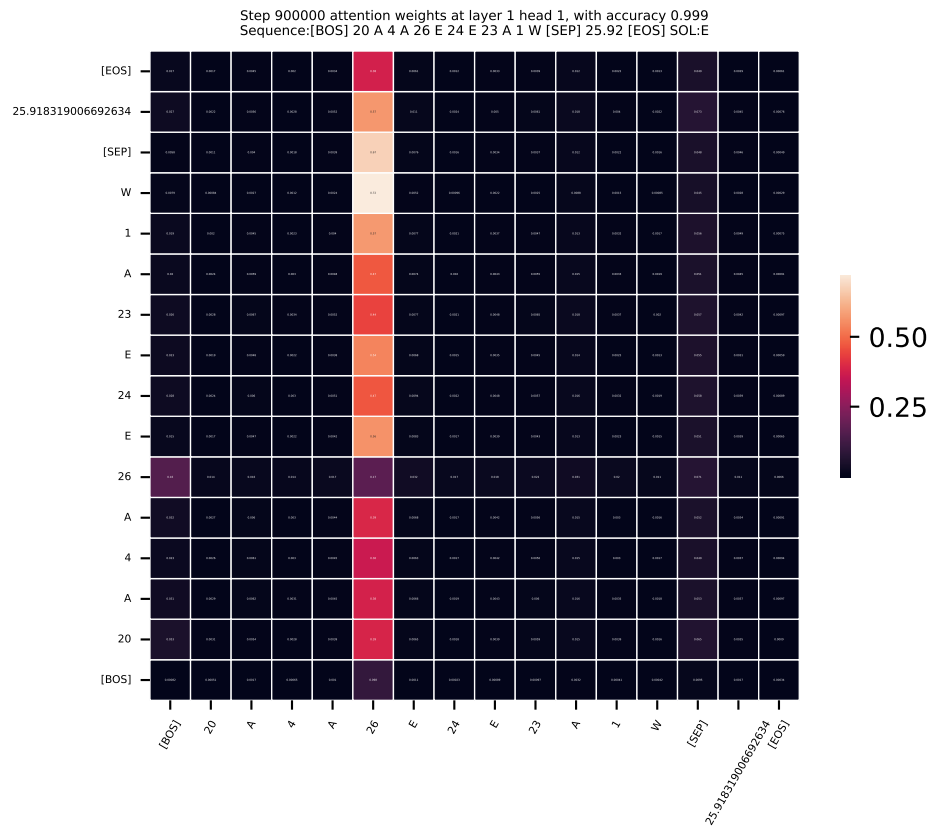

Step 900000 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

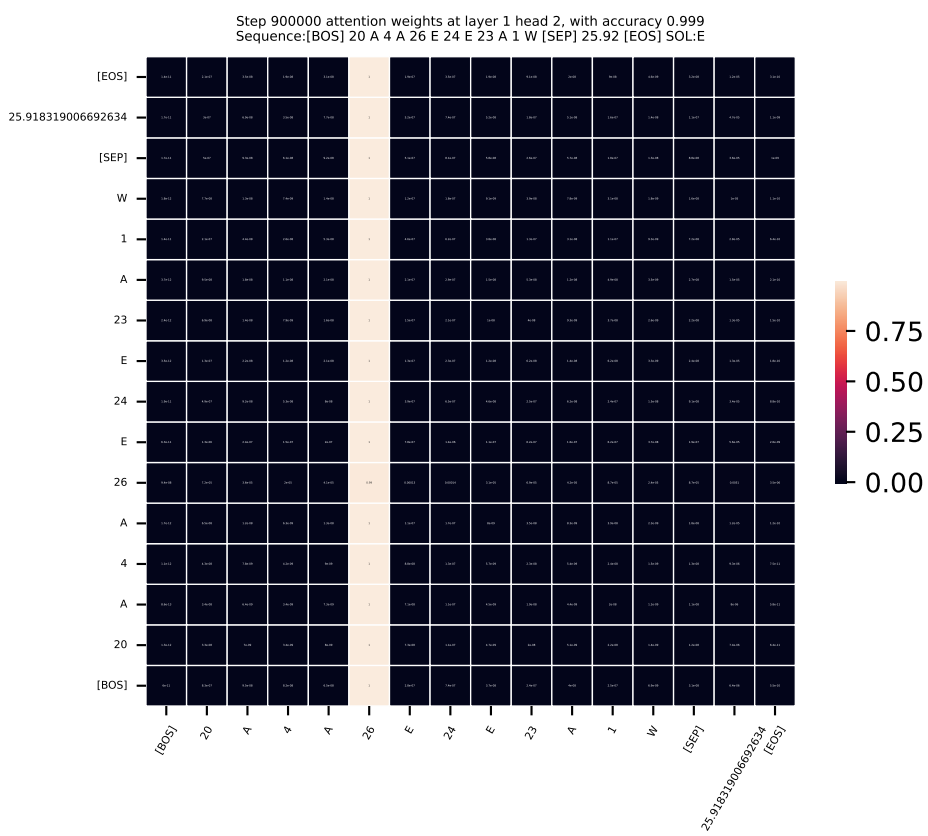

Step 900000 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

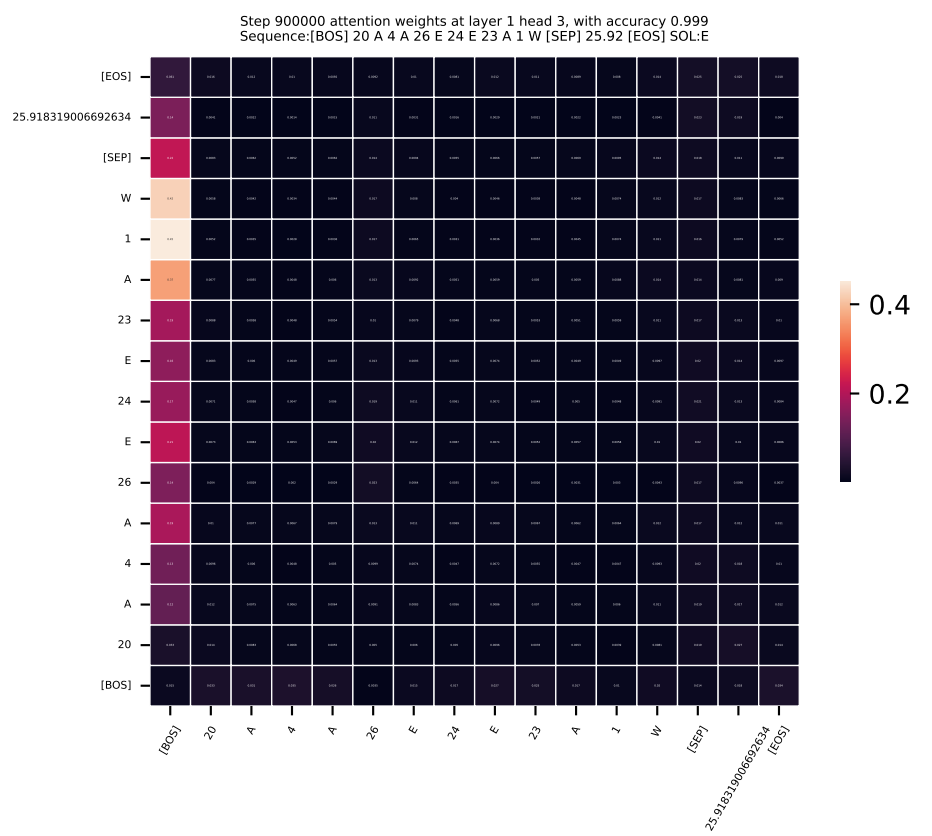

Step 900000 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

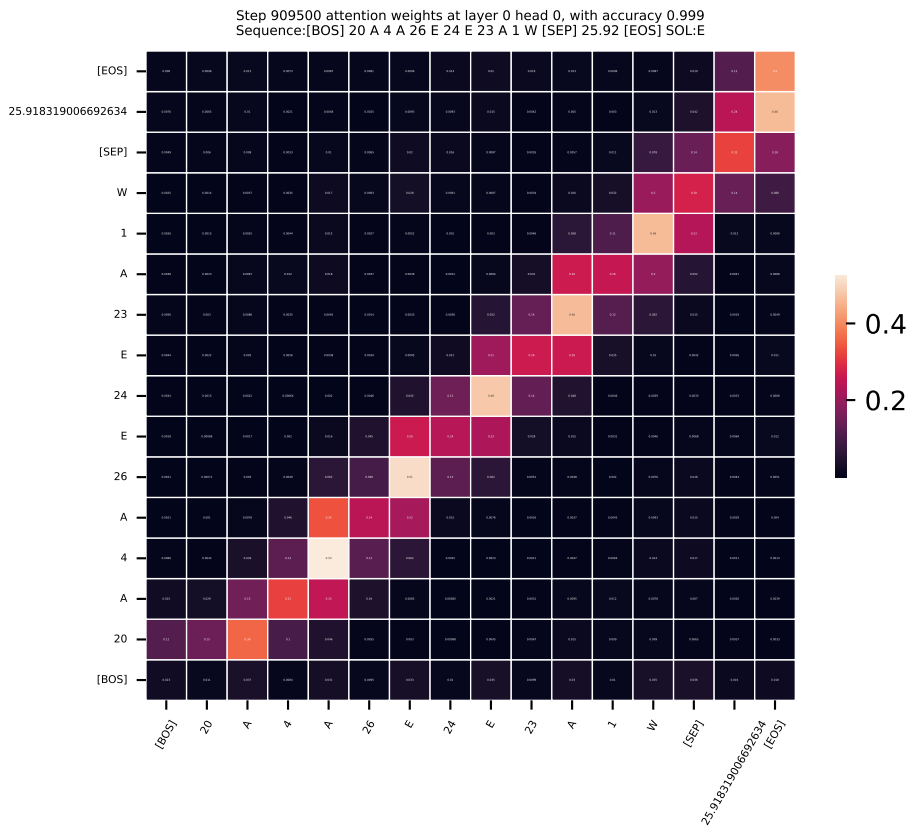

Step 900000 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

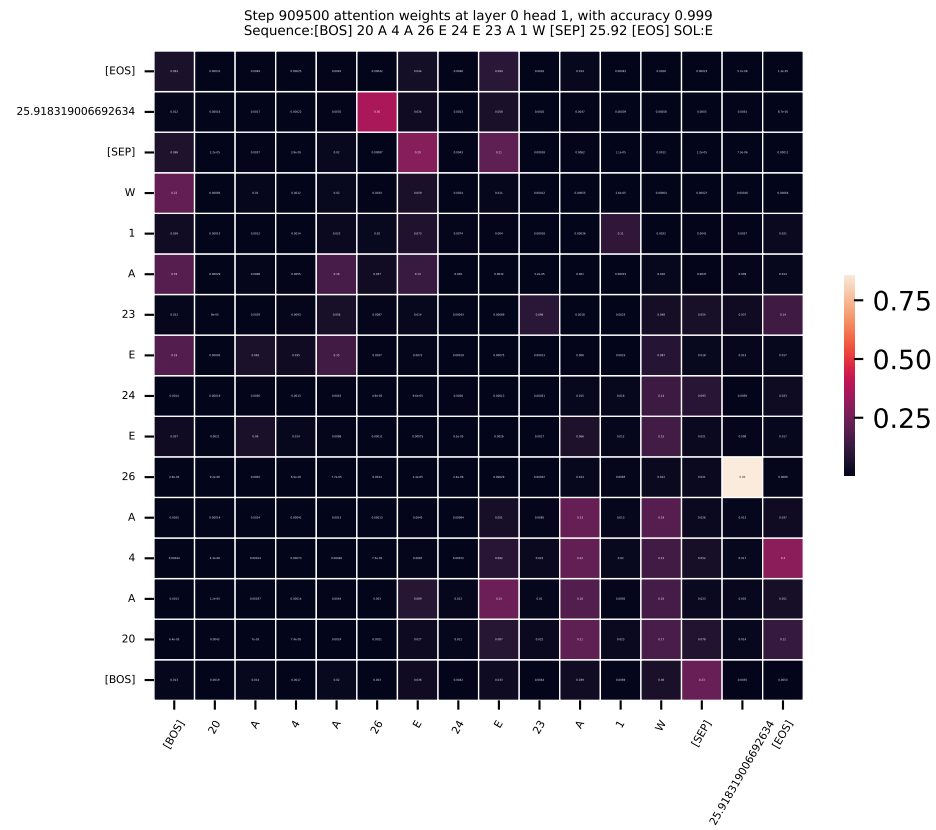

Step 900000 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

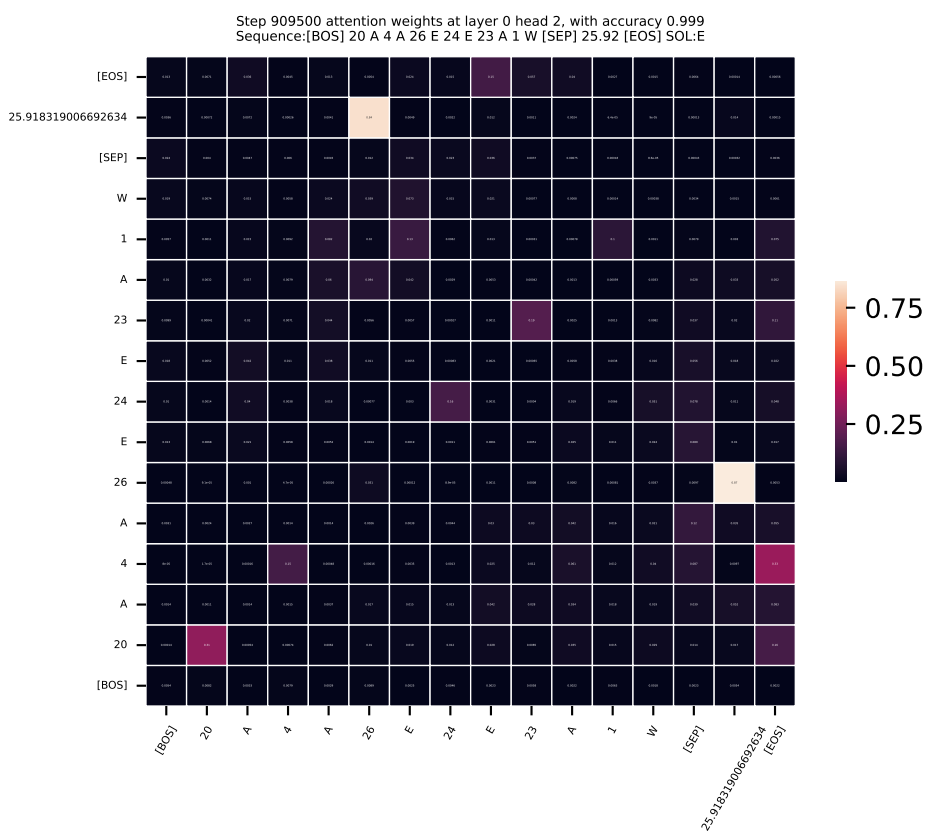

Step 900000 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

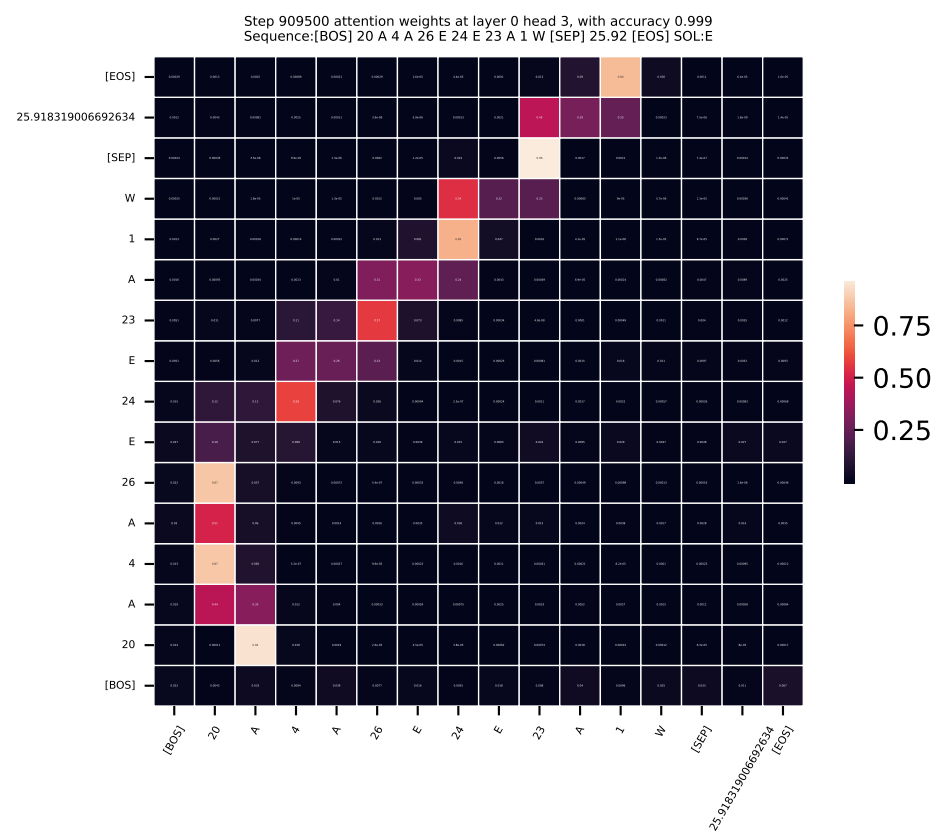

Step 900000 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

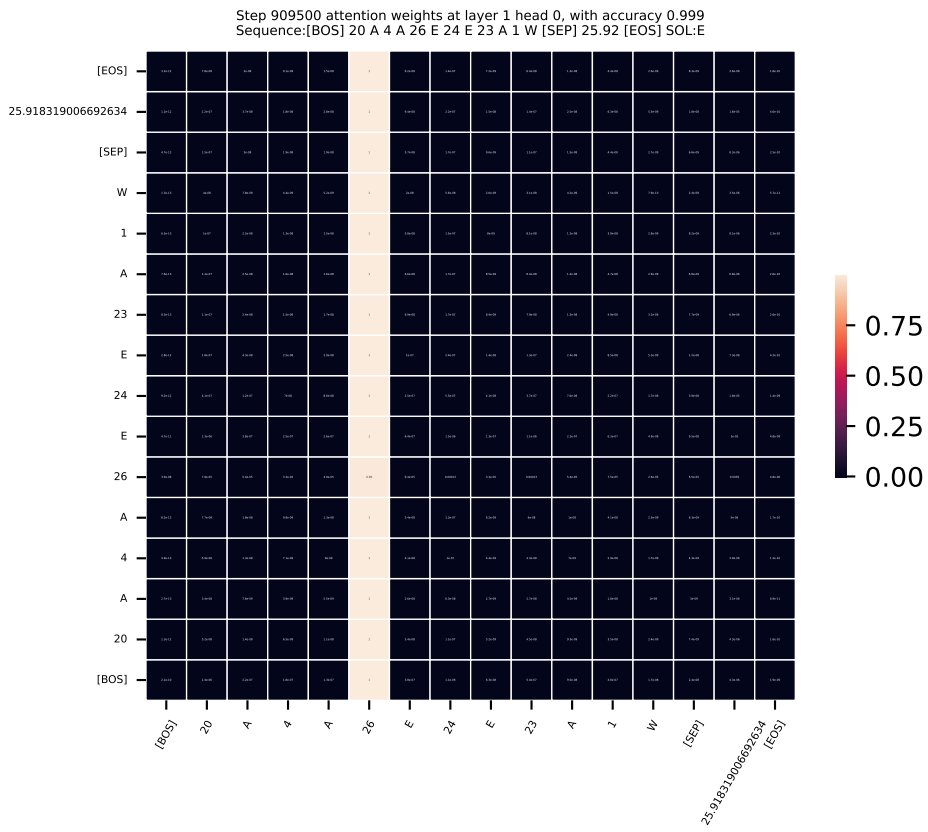

Step 900000 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.4
0.2

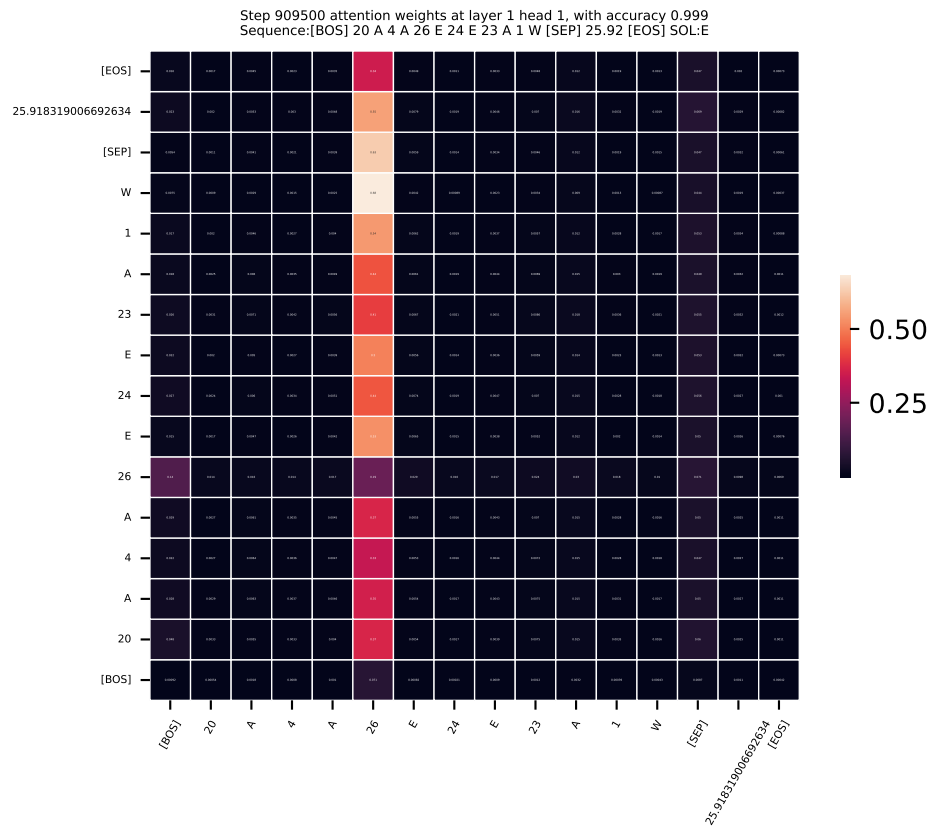

Step 909500 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

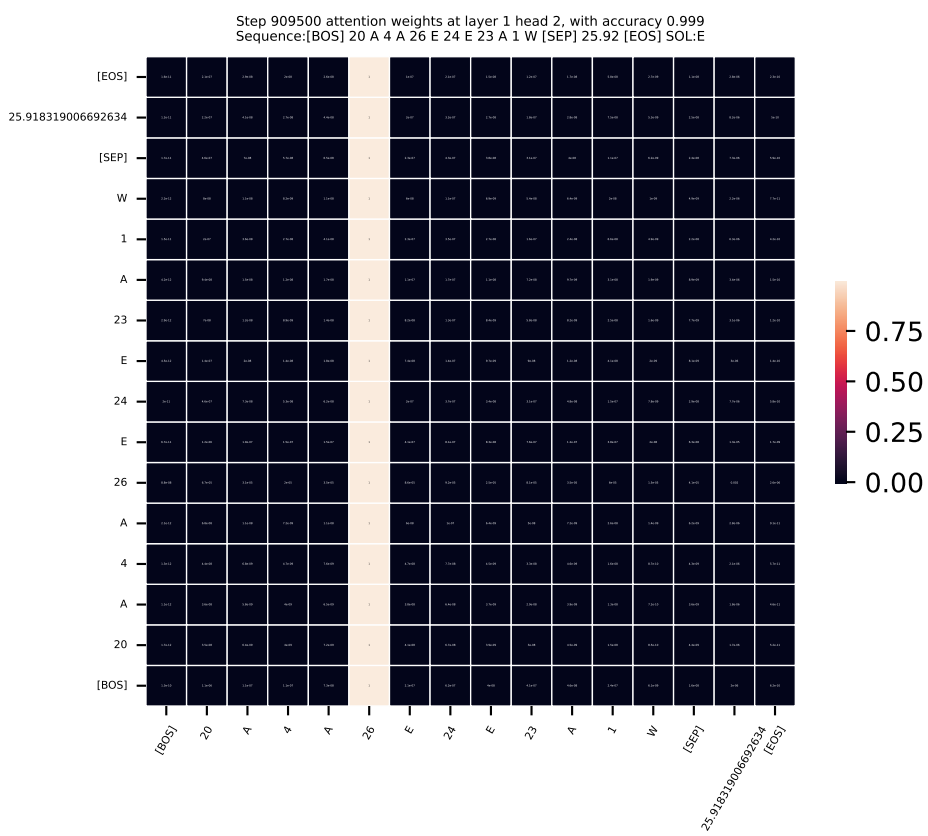

Step 909500 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

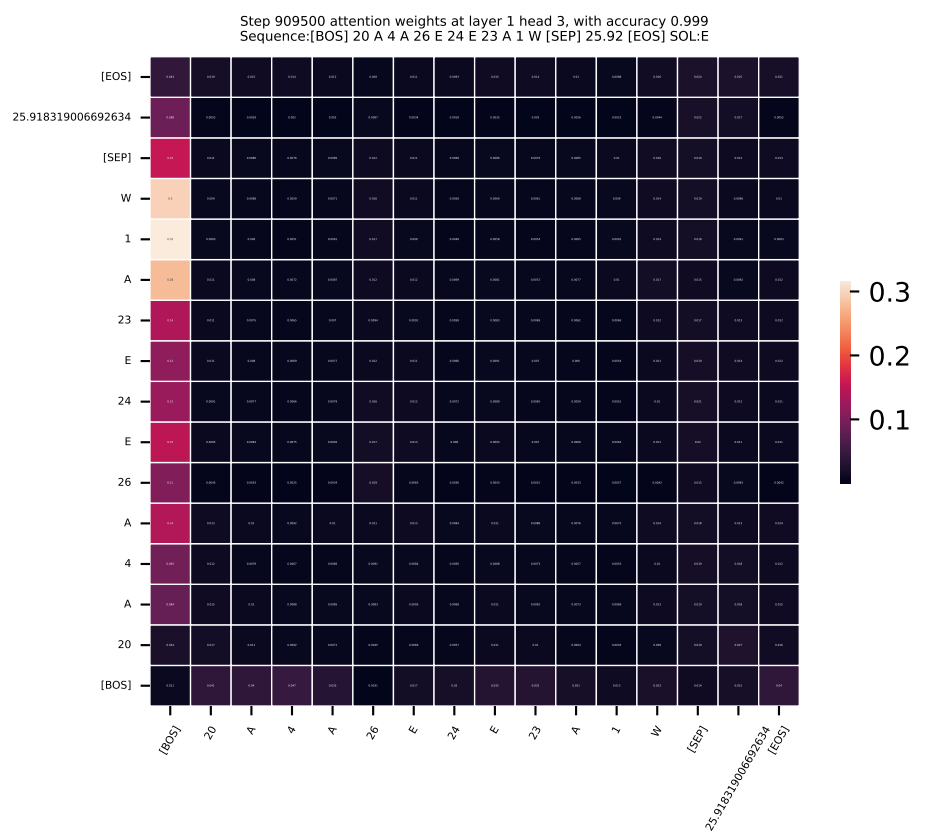

Step 909500 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

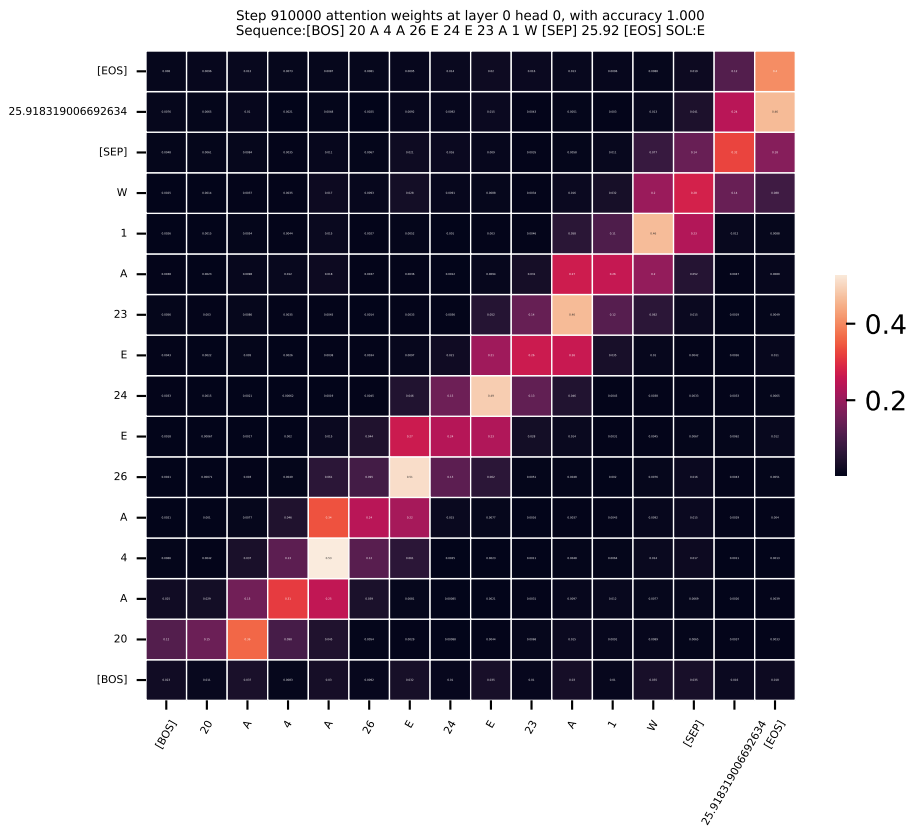

Step 909500 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

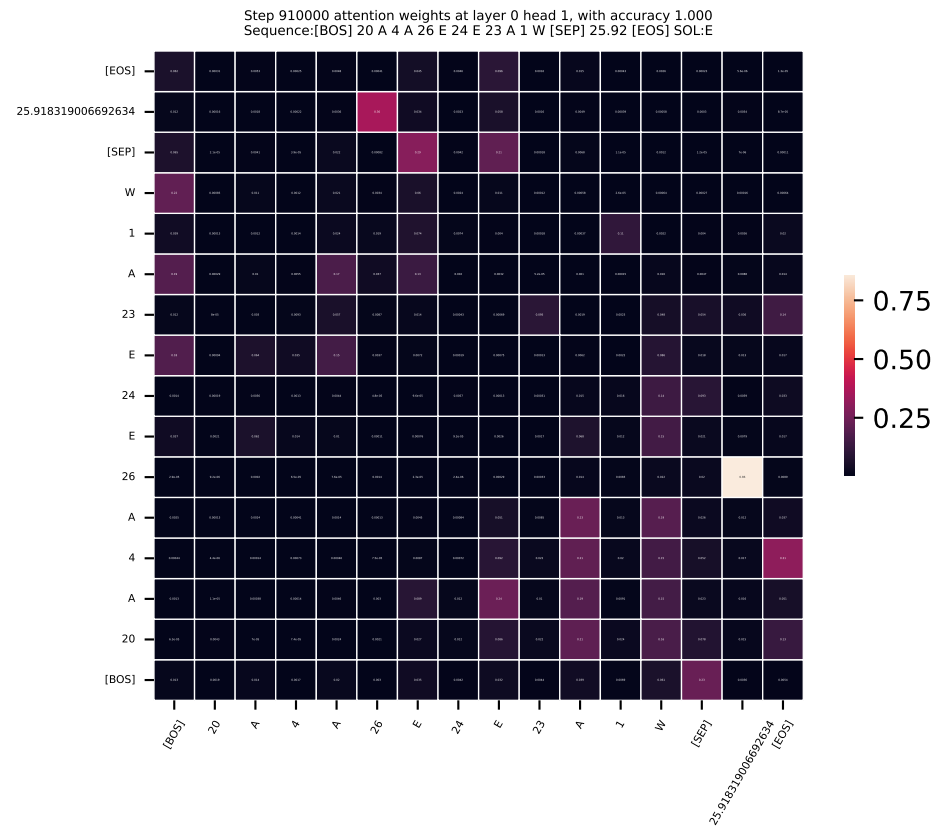

Step 909500 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

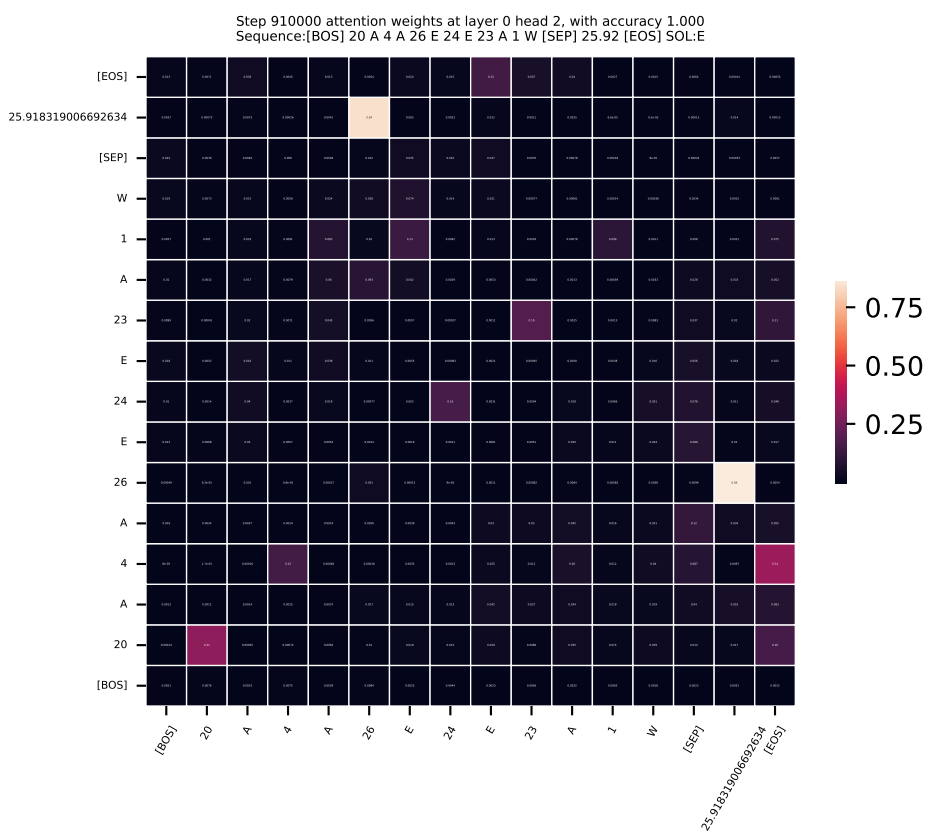

Step 909500 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

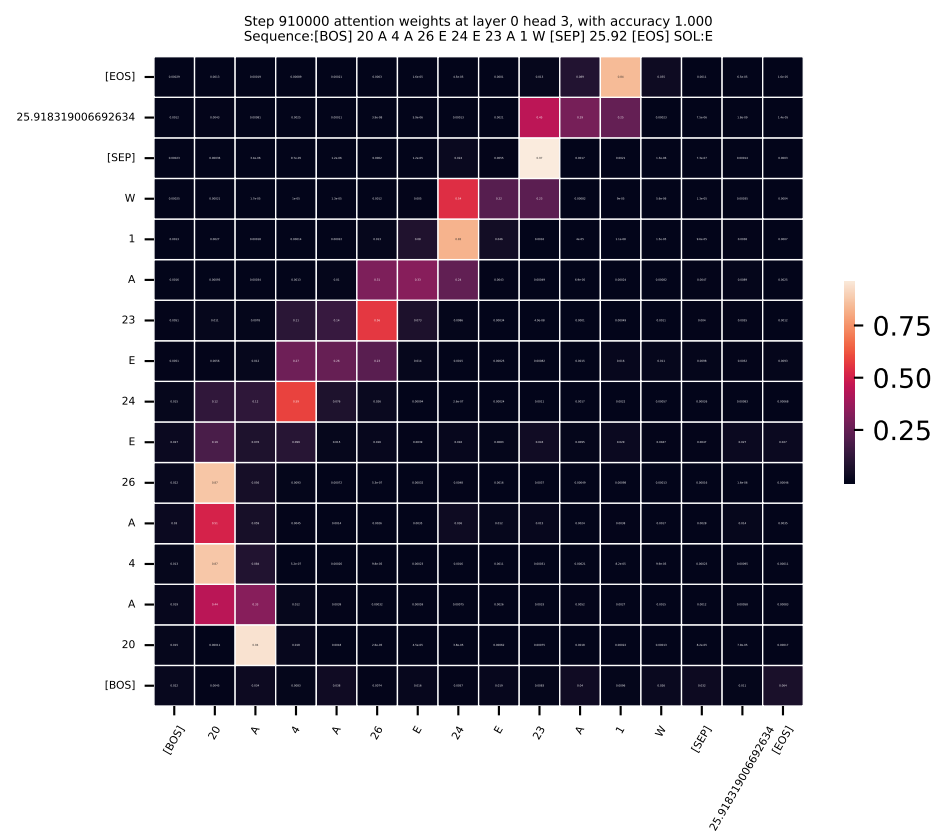

Step 909500 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

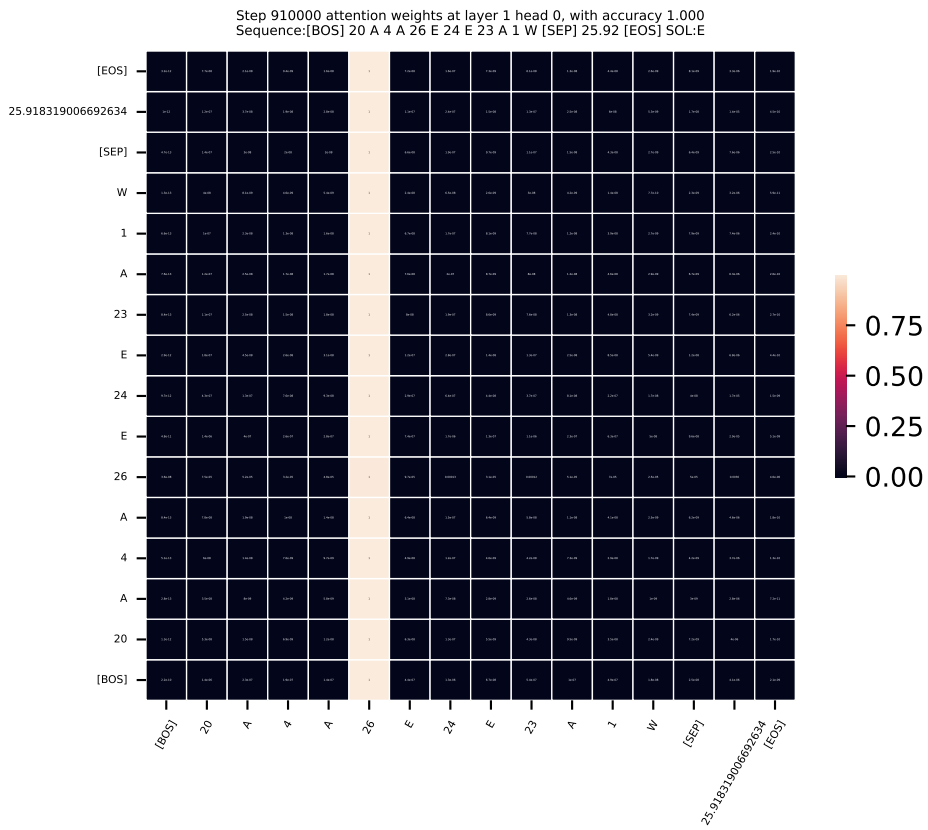

Step 909500 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

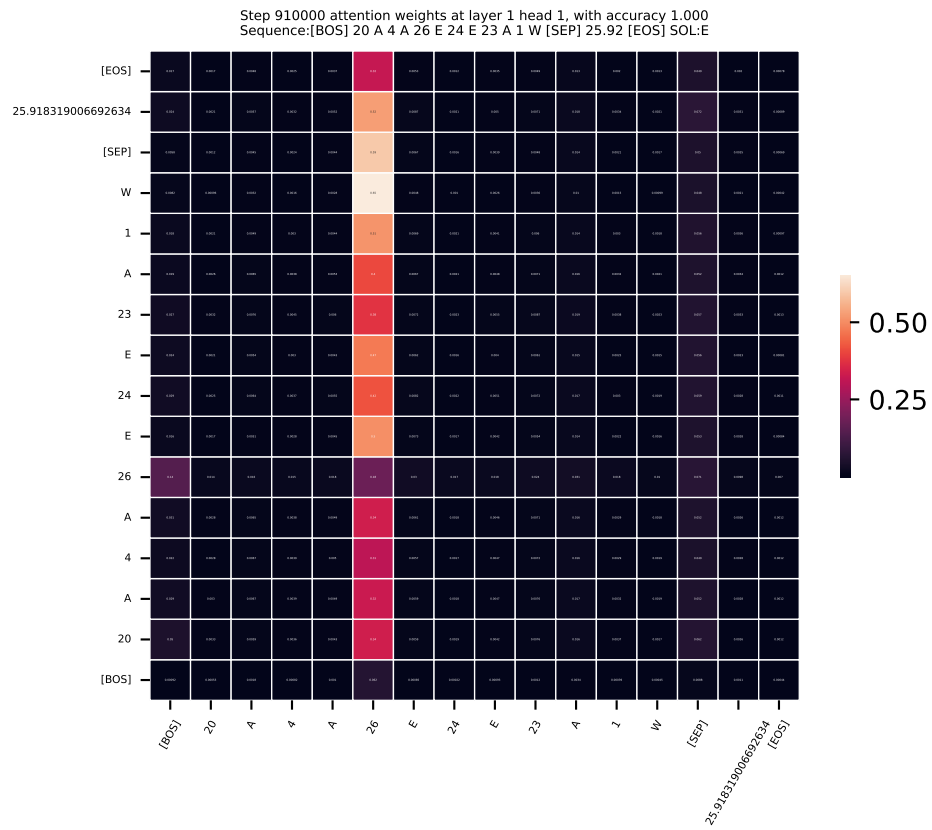

Step 910000 attention weights at layer 0 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

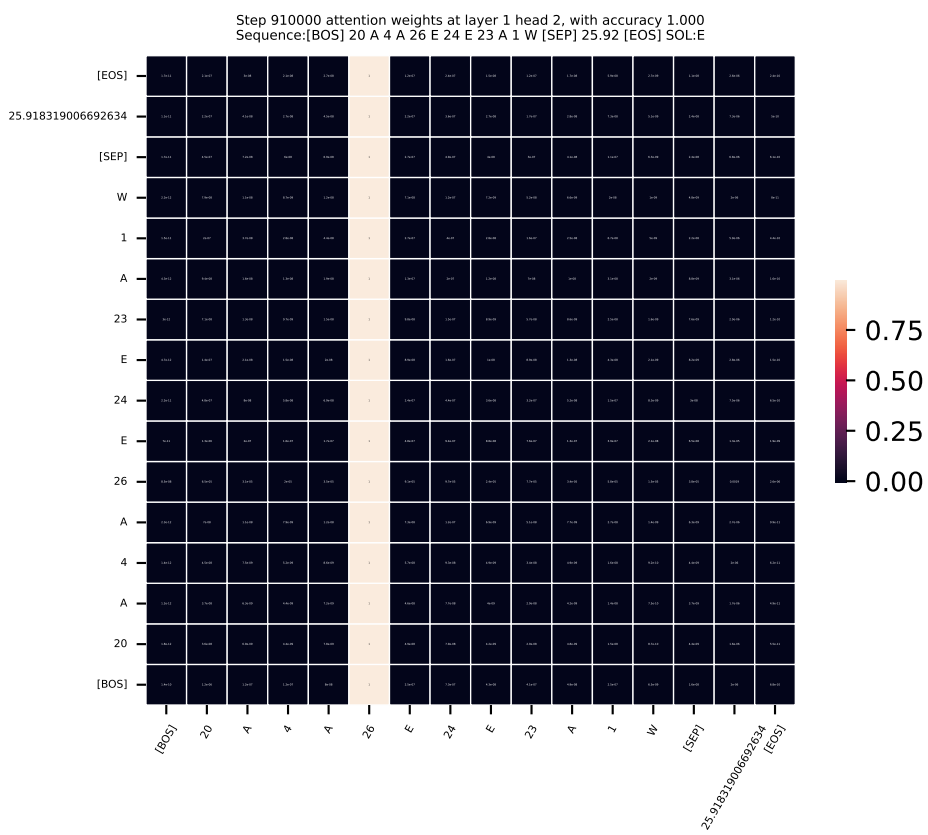

Step 910000 attention weights at layer 0 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

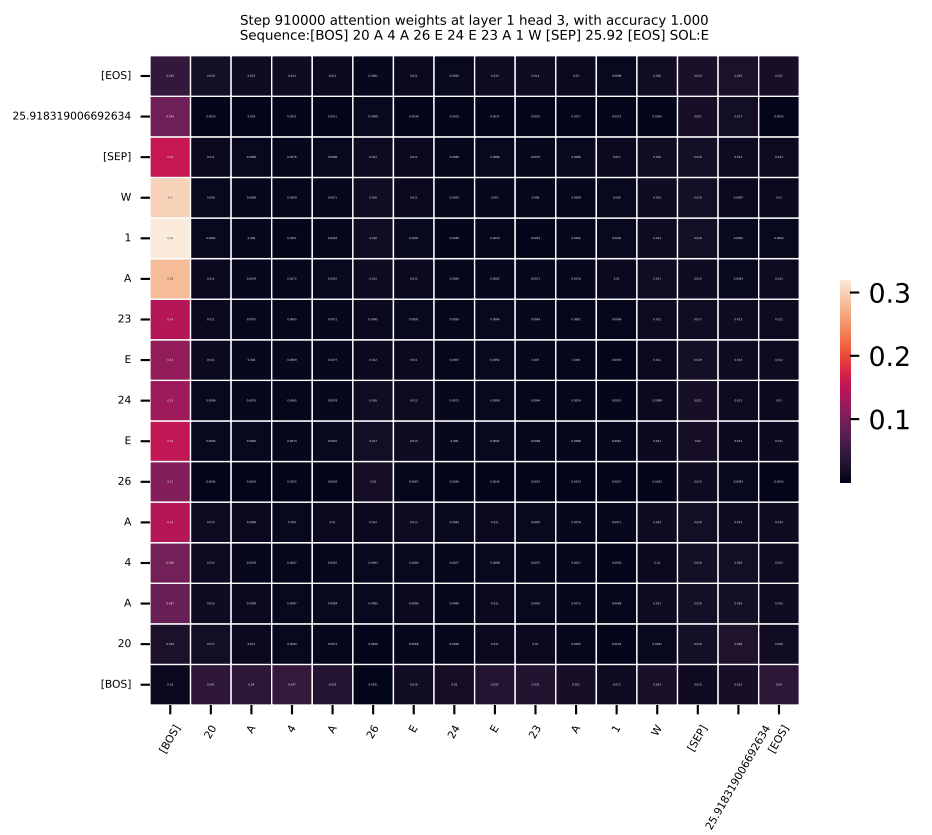

Step 910000 attention weights at layer 0 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

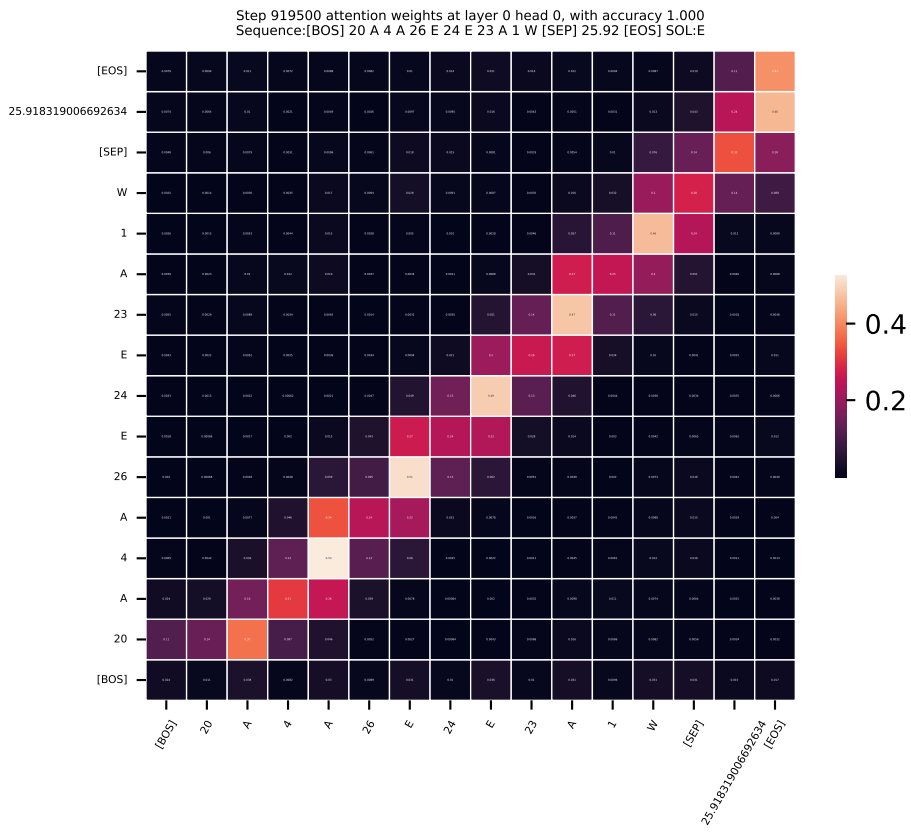

Step 910000 attention weights at layer 0 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

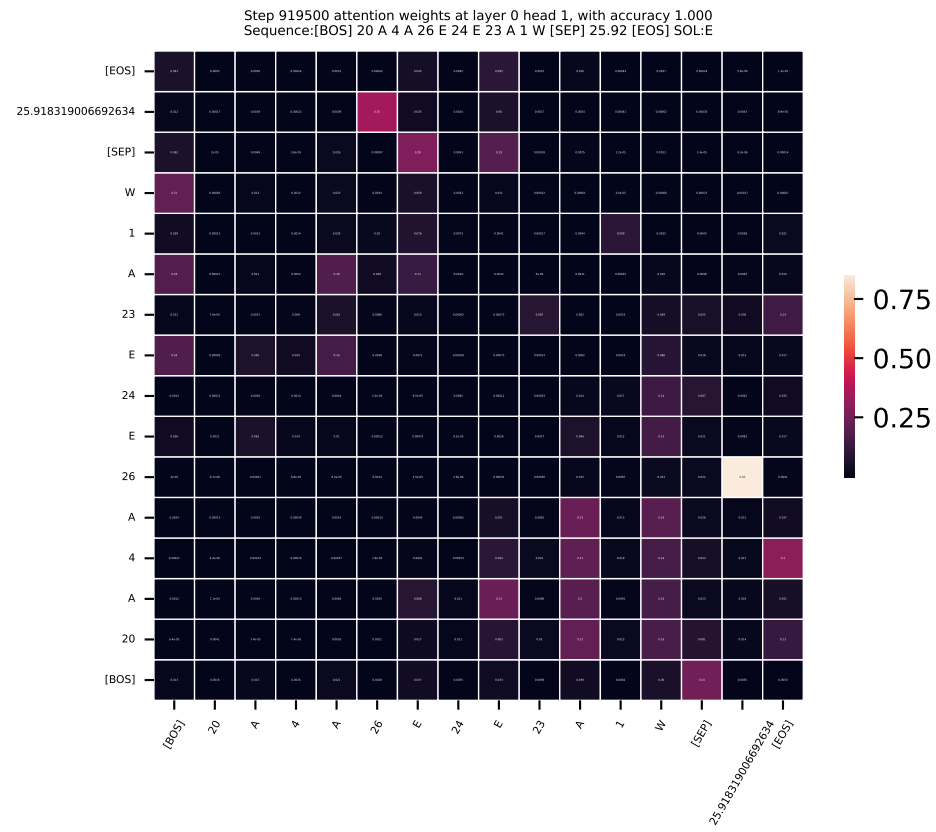

Step 910000 attention weights at layer 1 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

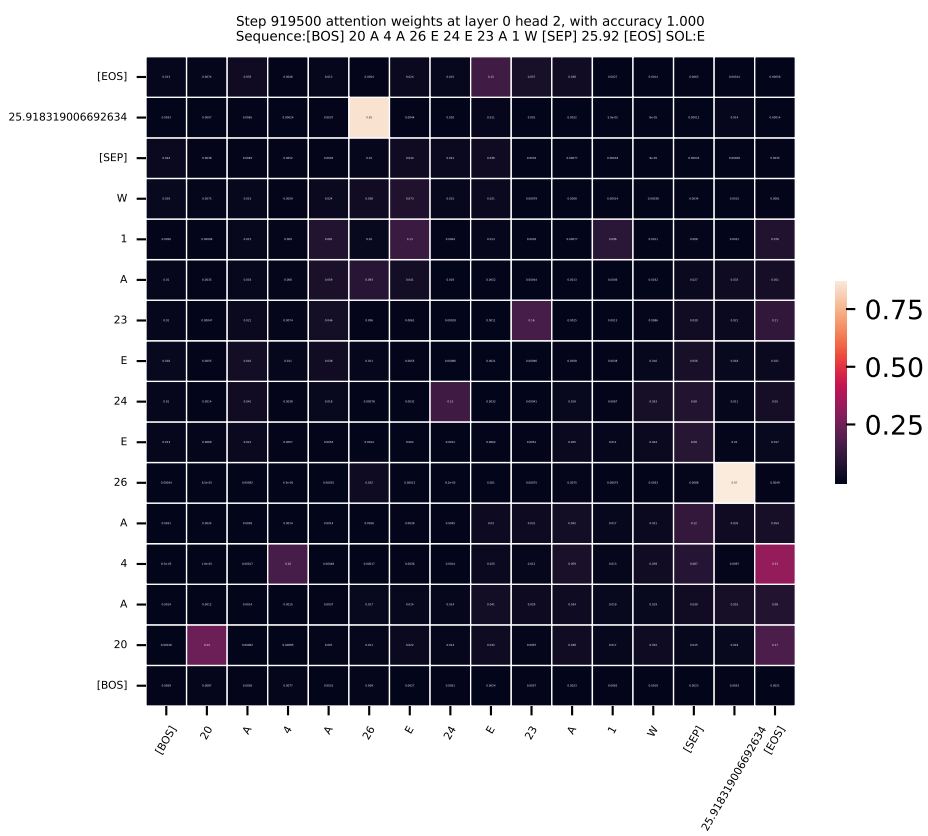

Step 910000 attention weights at layer 1 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

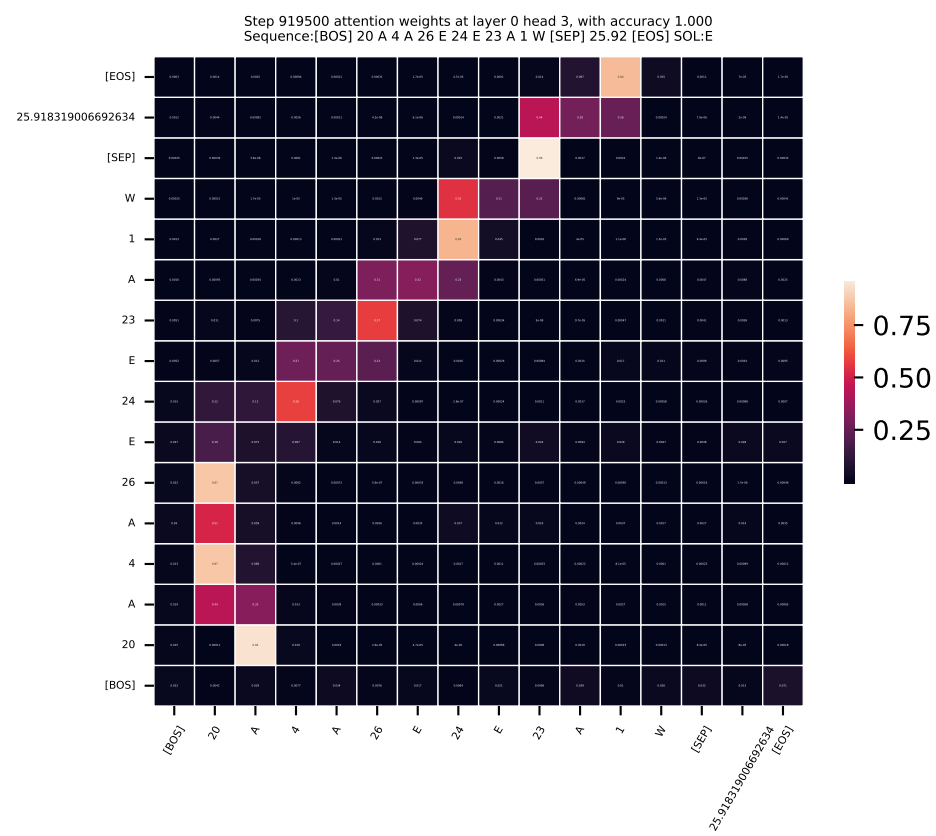

Step 910000 attention weights at layer 1 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

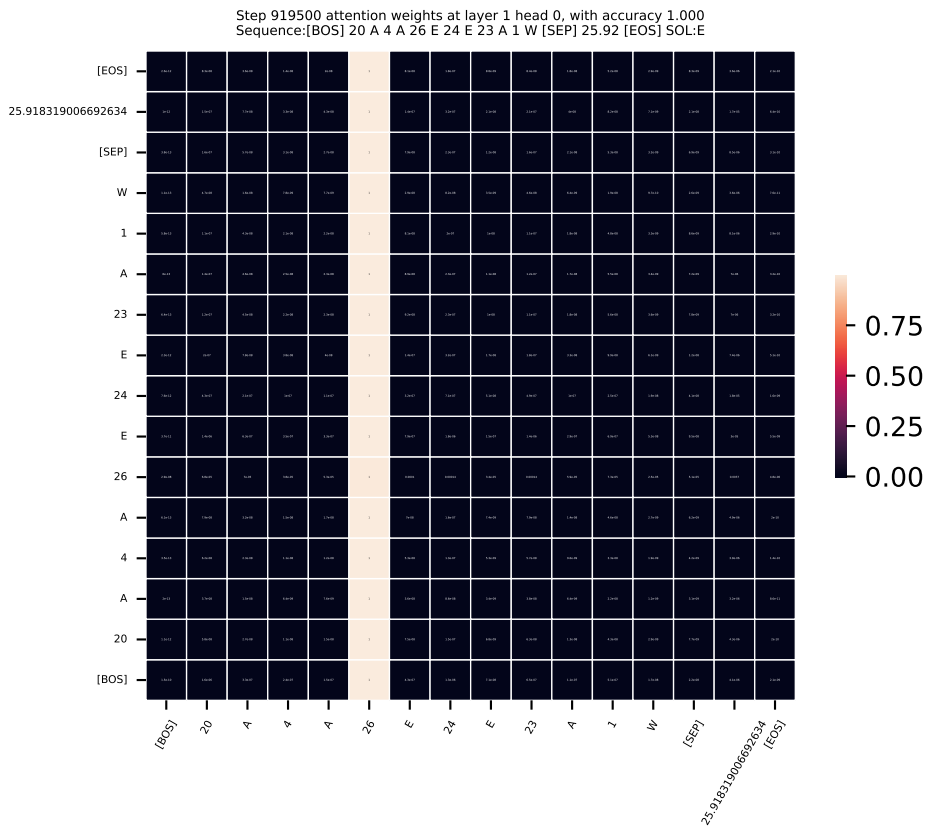

Step 910000 attention weights at layer 1 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

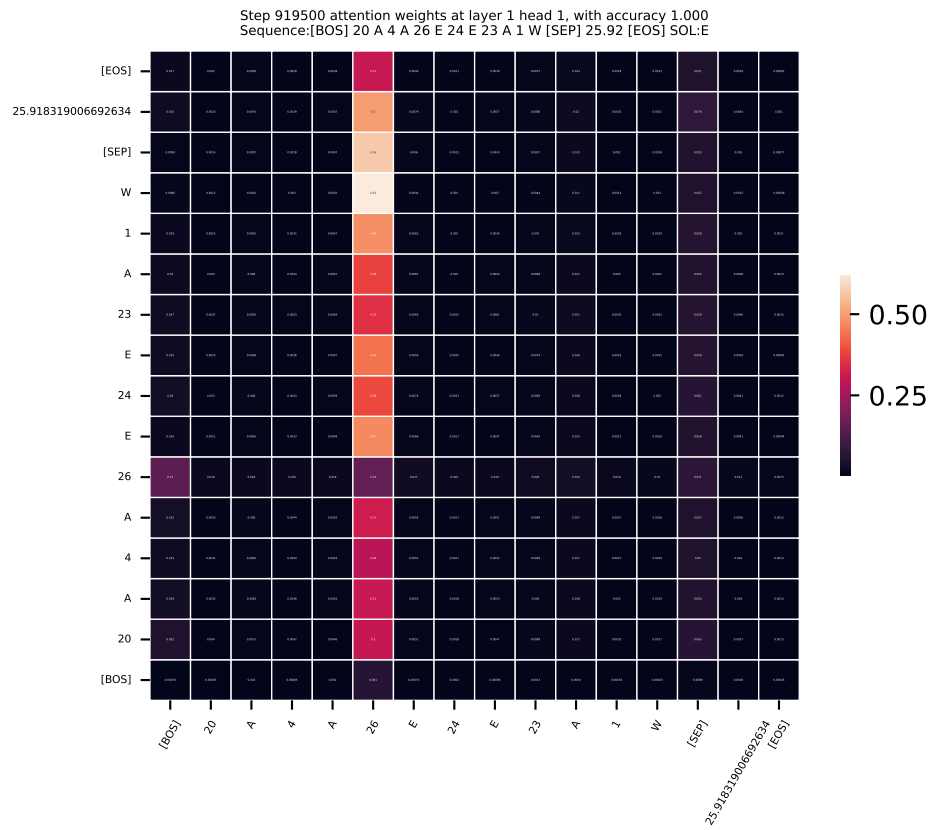

Step 919500 attention weights at layer 0 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

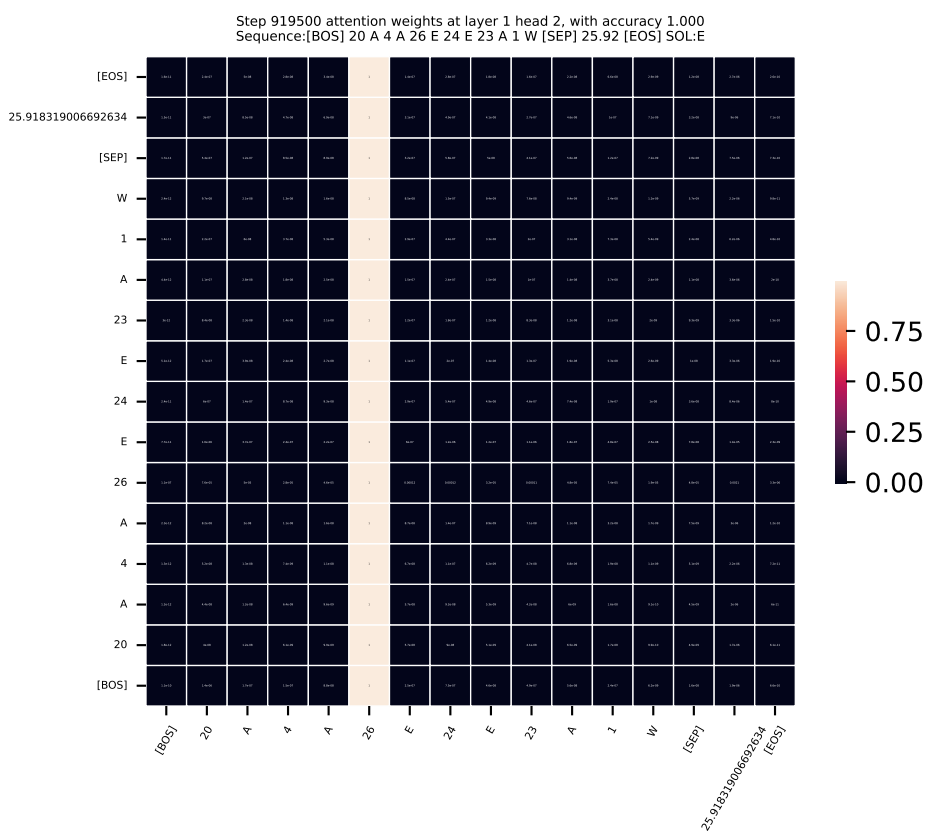

Step 919500 attention weights at layer 0 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

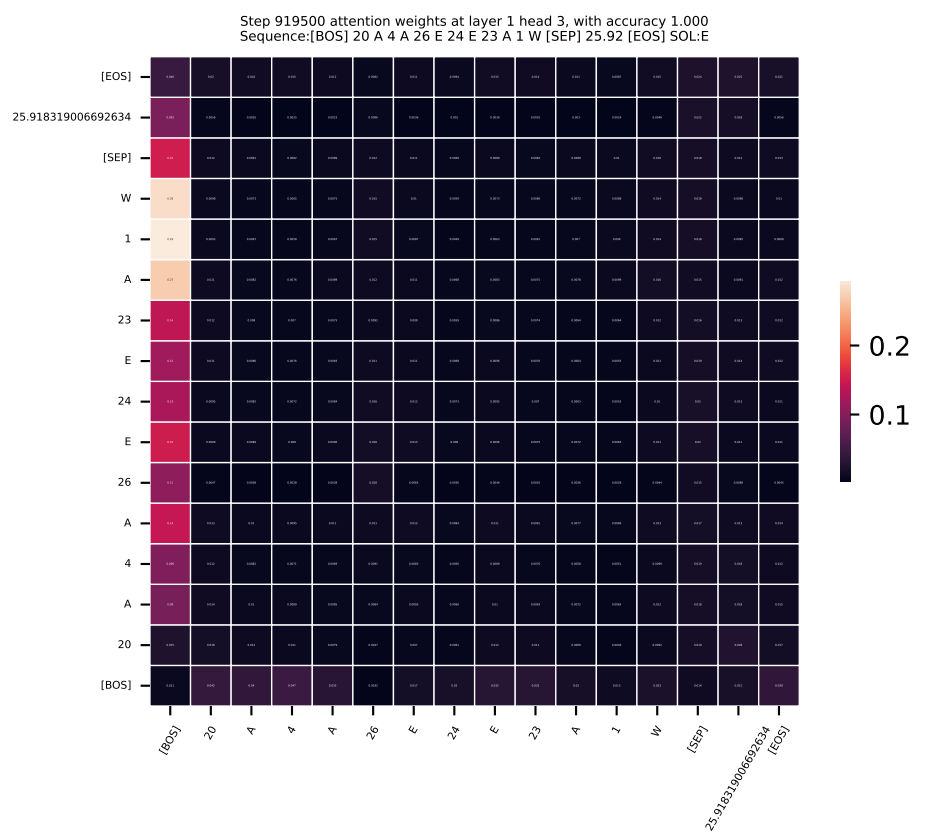

Step 919500 attention weights at layer 0 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

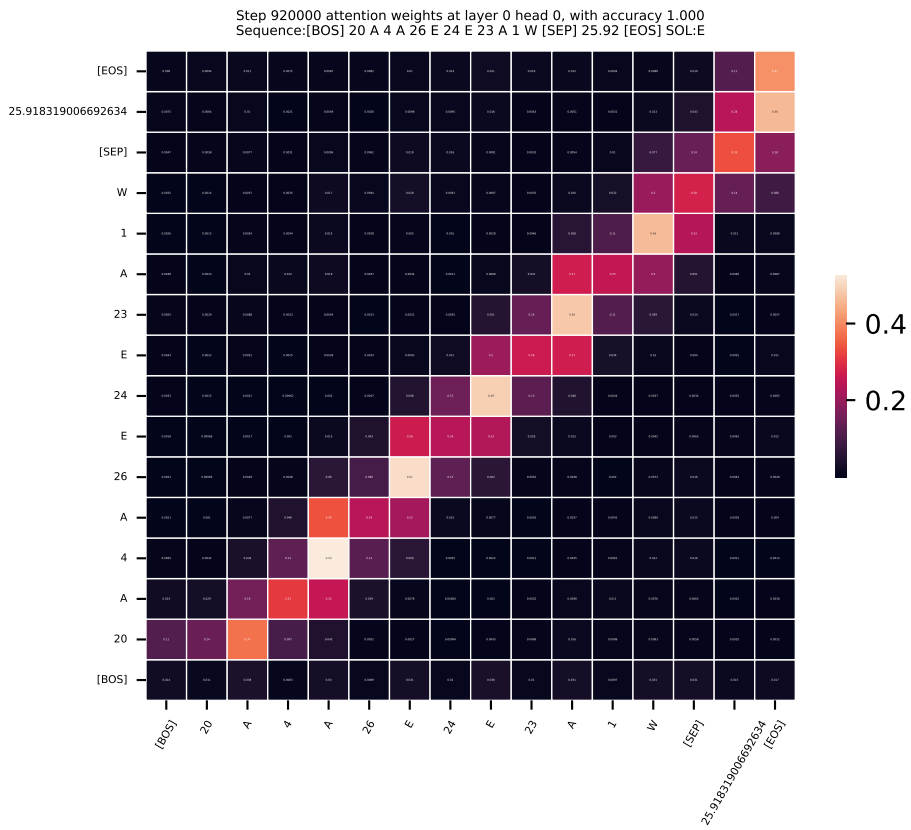

Step 919500 attention weights at layer 0 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

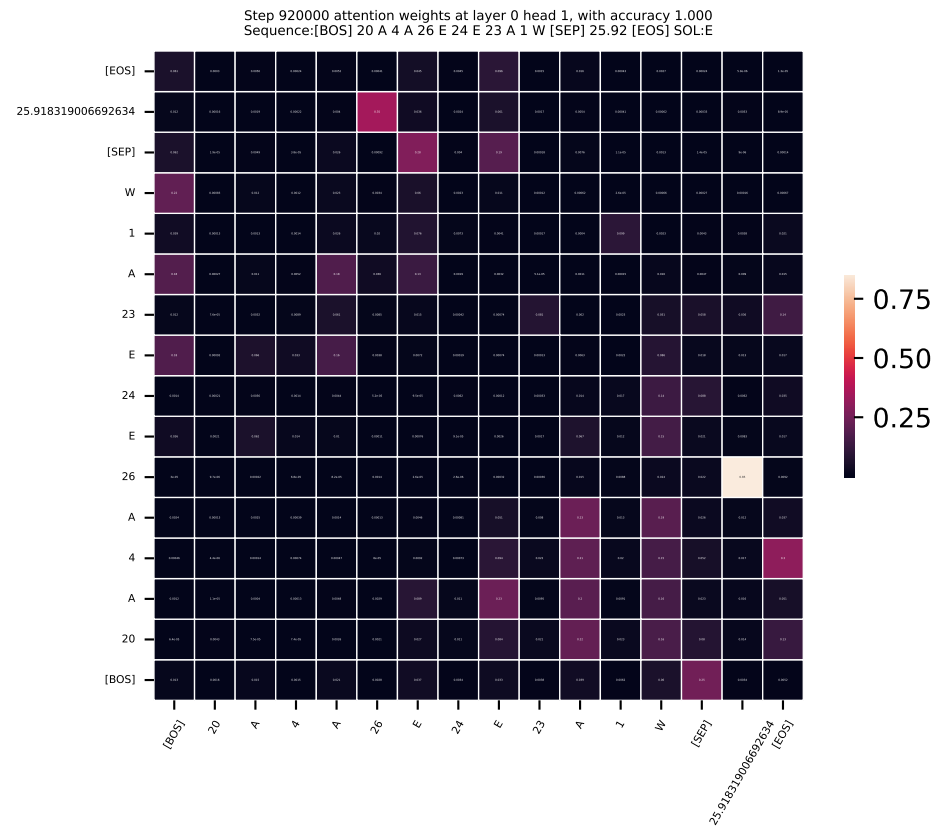

Step 919500 attention weights at layer 1 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

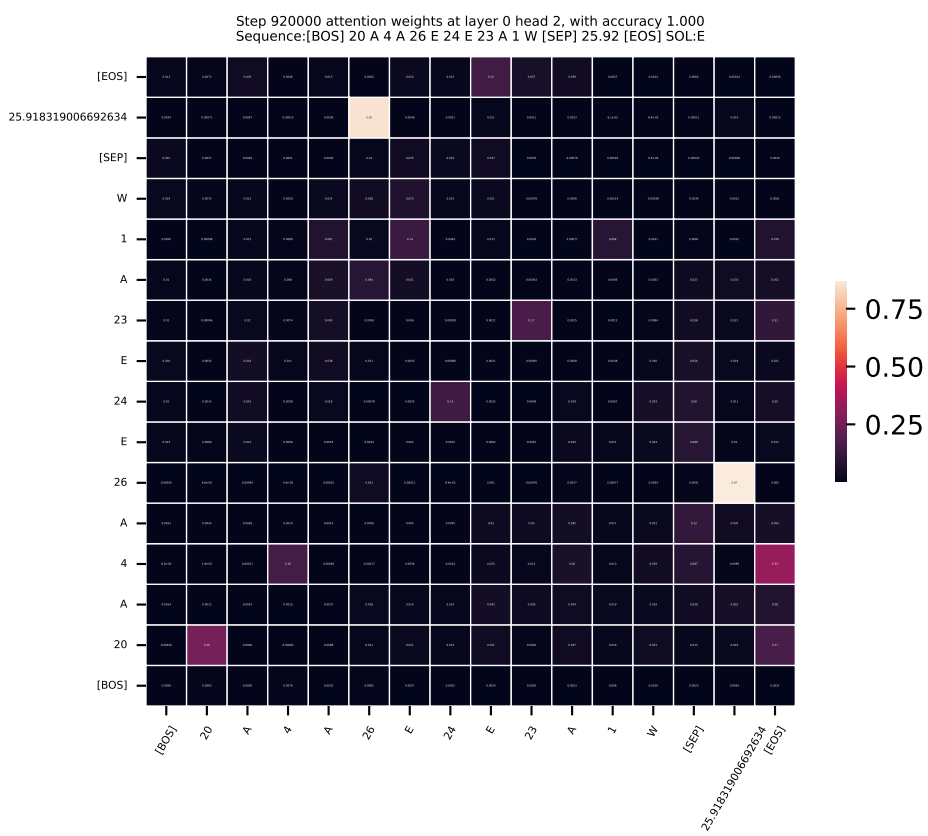

Step 919500 attention weights at layer 1 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

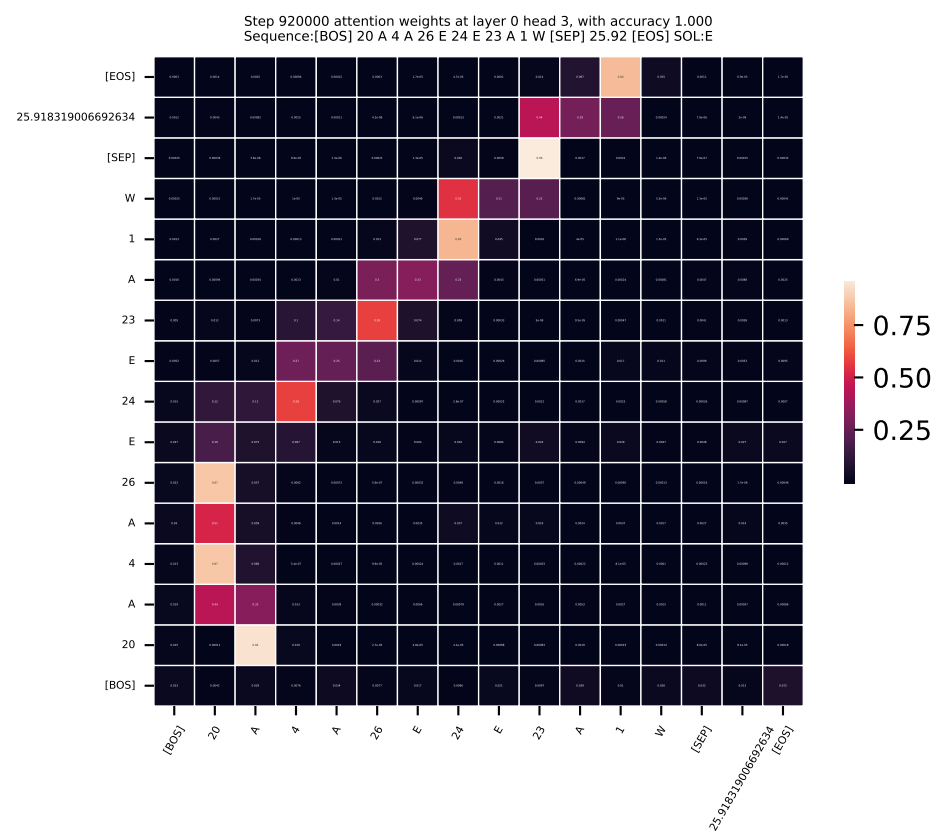

Step 919500 attention weights at layer 1 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

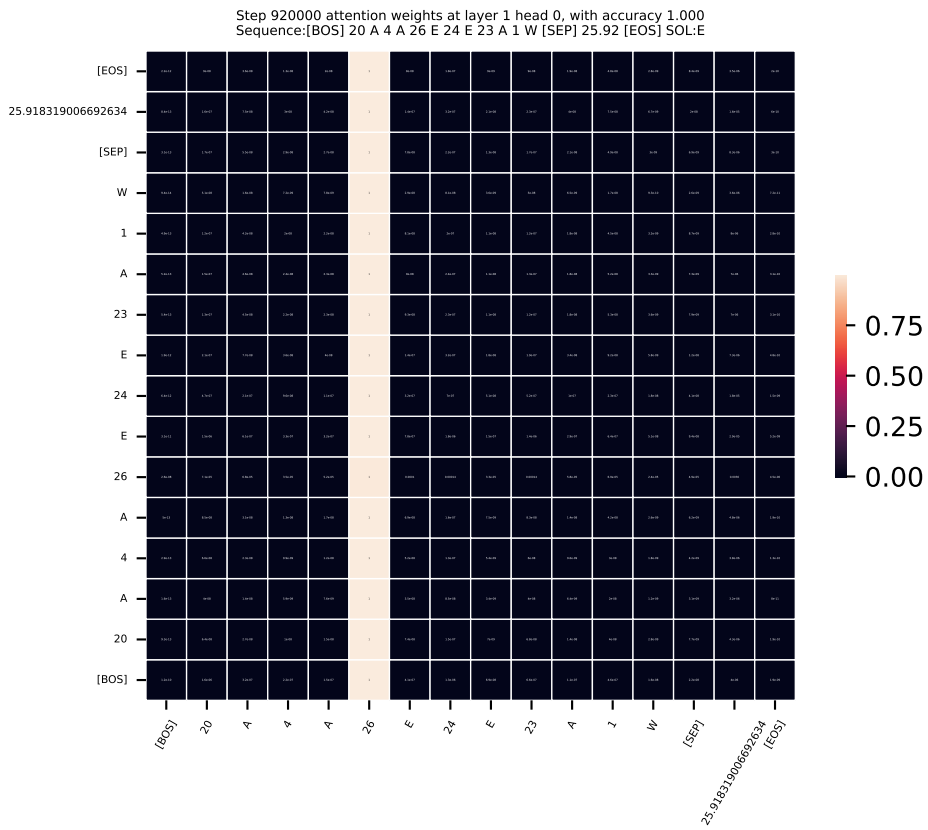

Step 919500 attention weights at layer 1 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.2
0.1

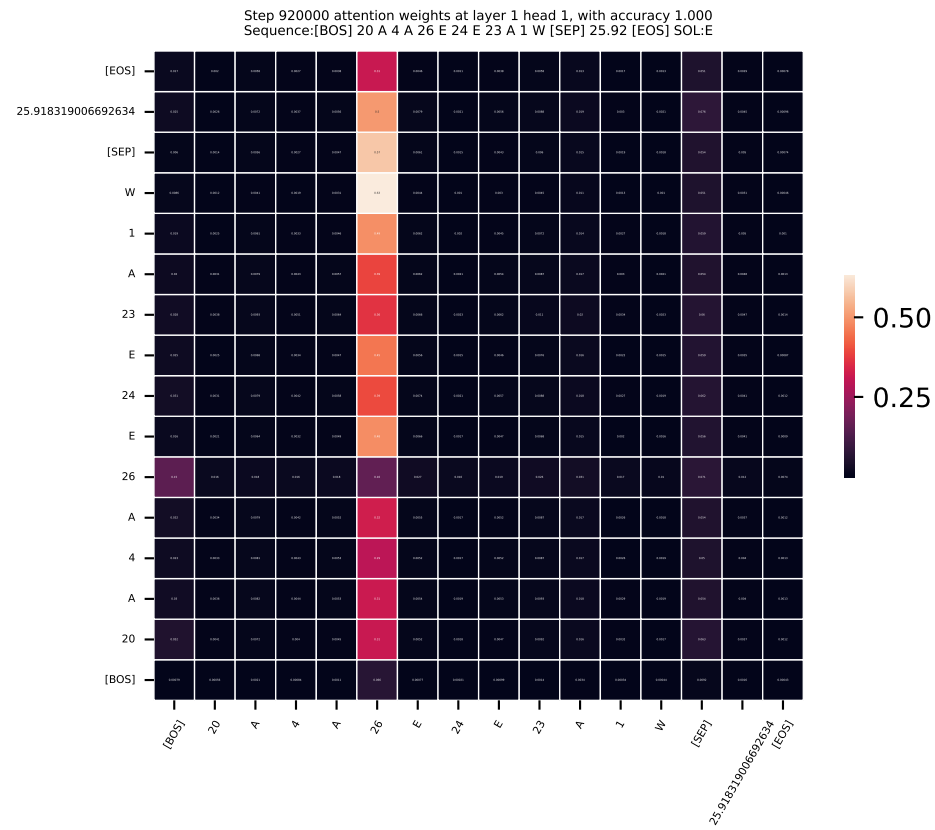

Step 920000 attention weights at layer 0 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

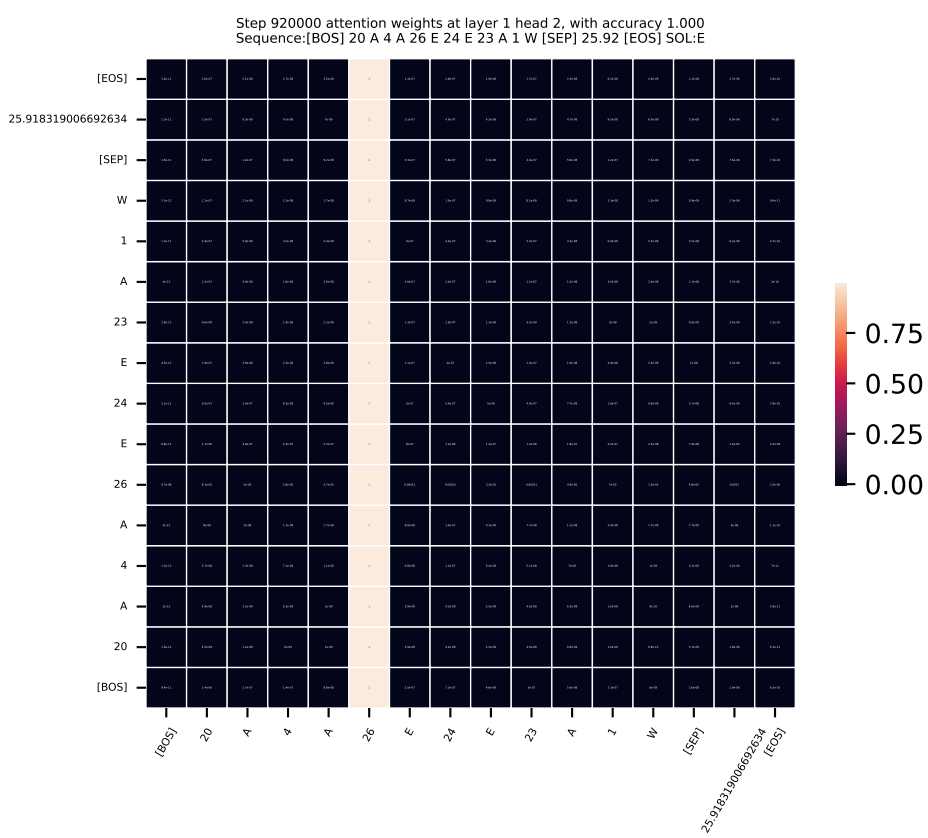

Step 920000 attention weights at layer 0 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

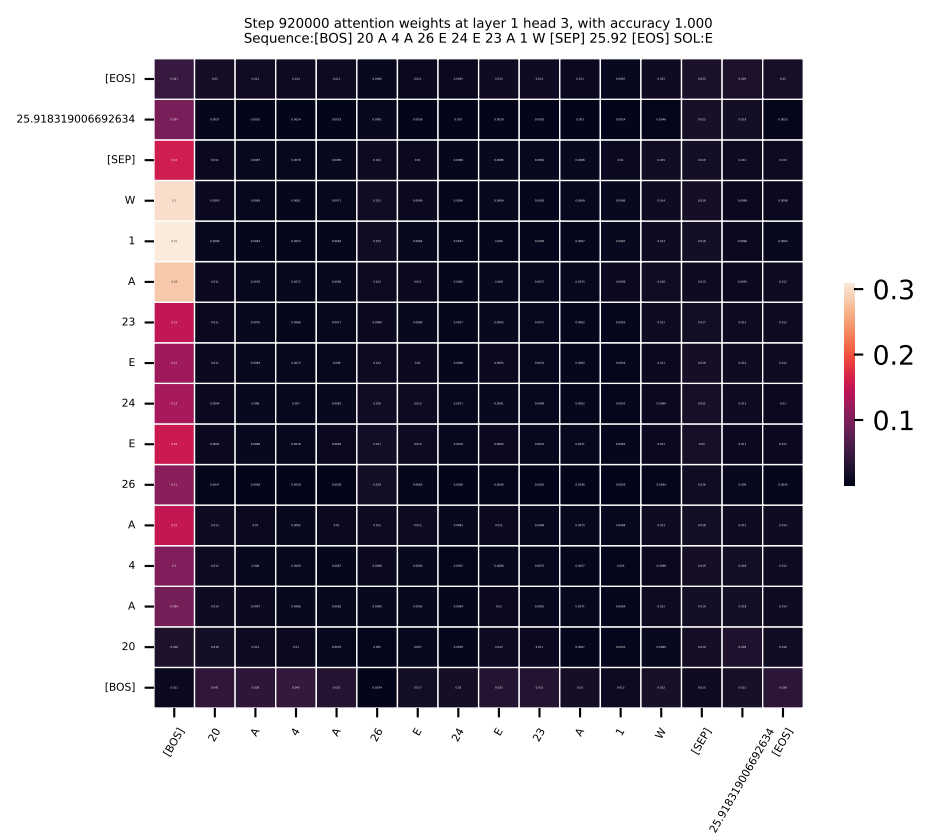

Step 920000 attention weights at layer 0 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

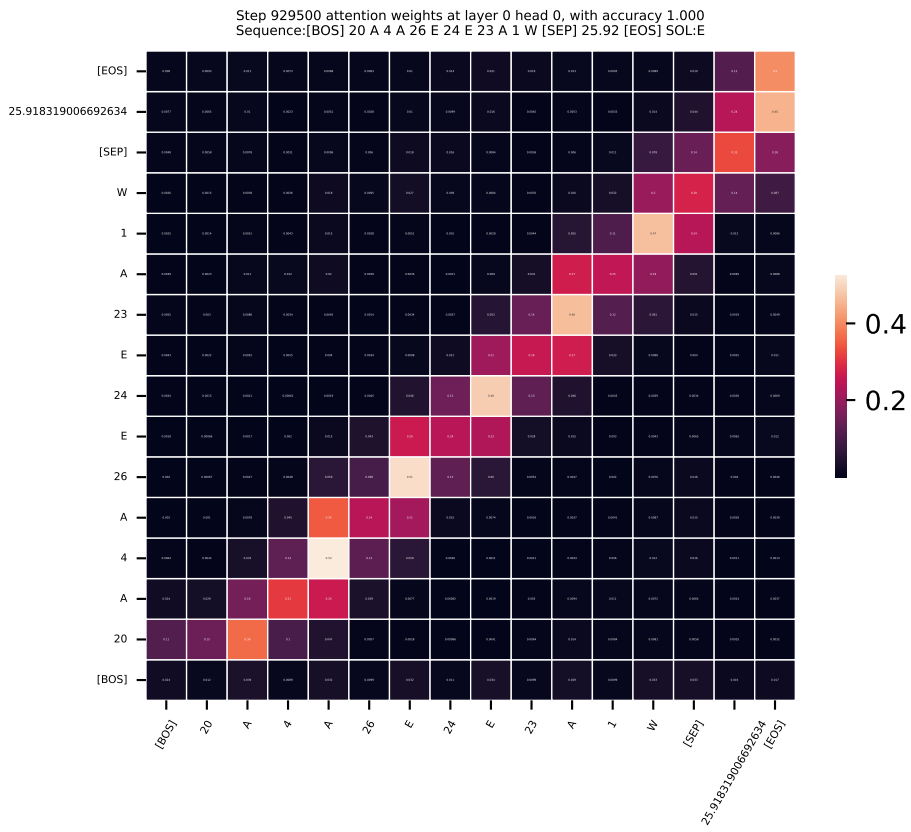

Step 920000 attention weights at layer 0 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

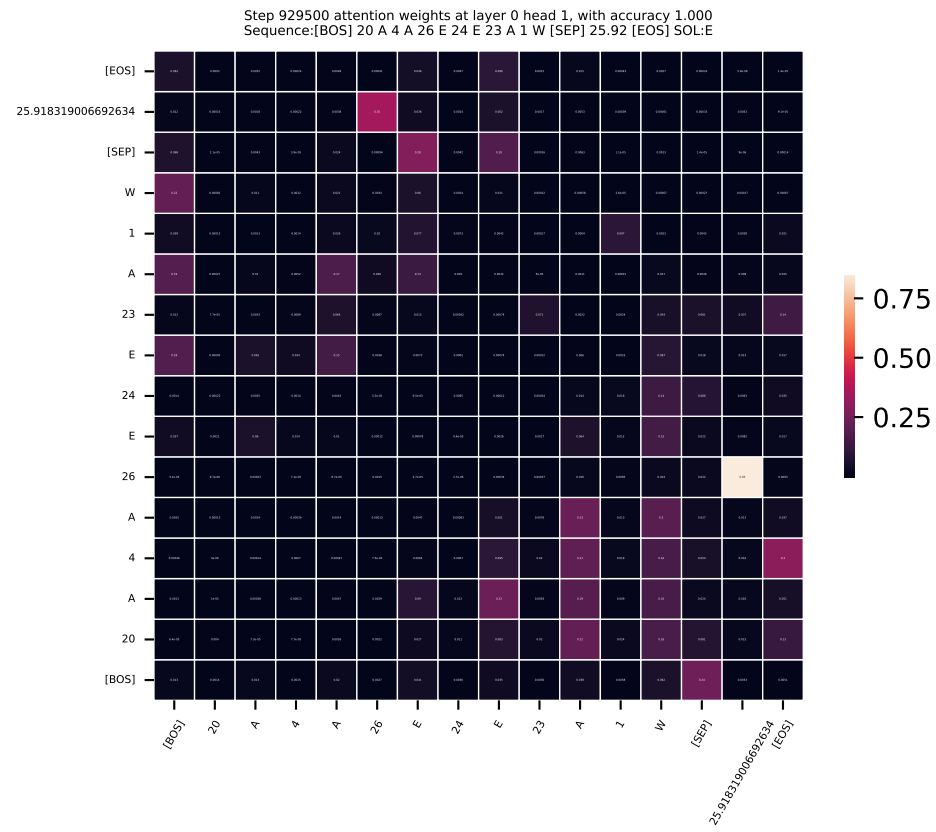

Step 920000 attention weights at layer 1 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

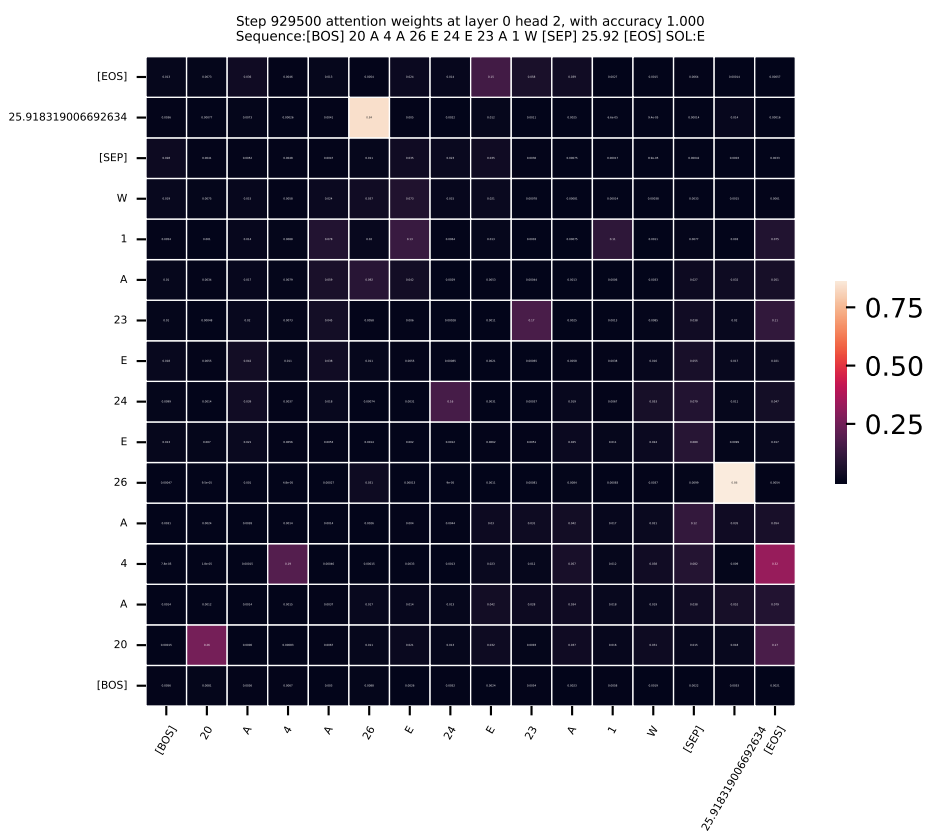

Step 920000 attention weights at layer 1 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

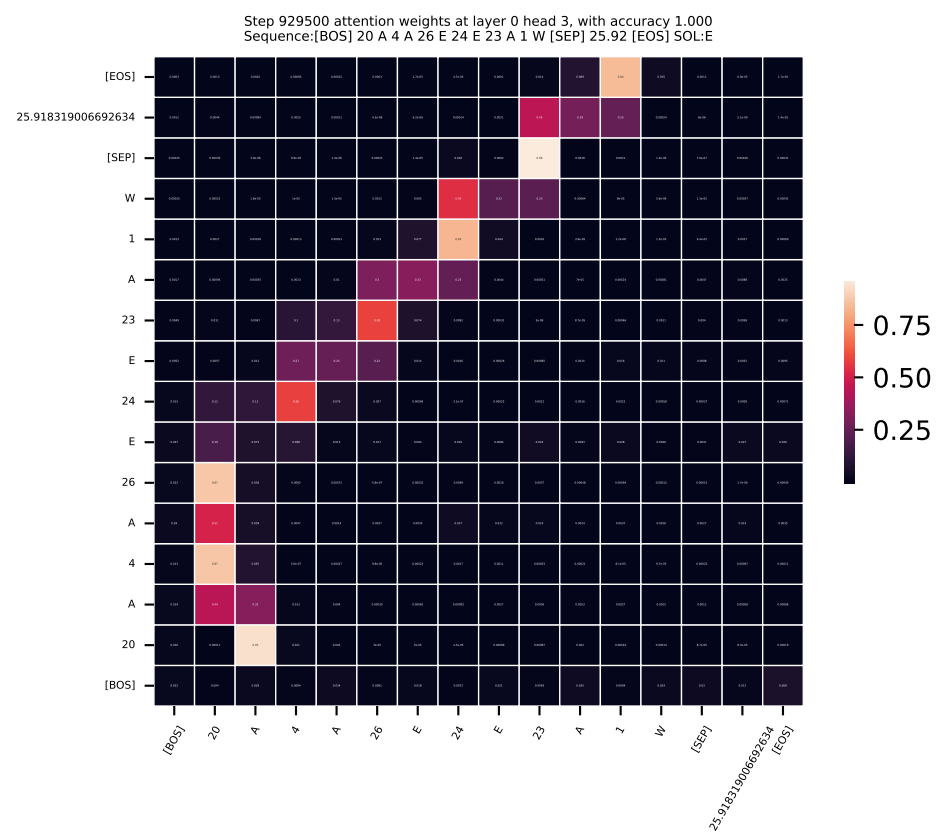

Step 920000 attention weights at layer 1 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

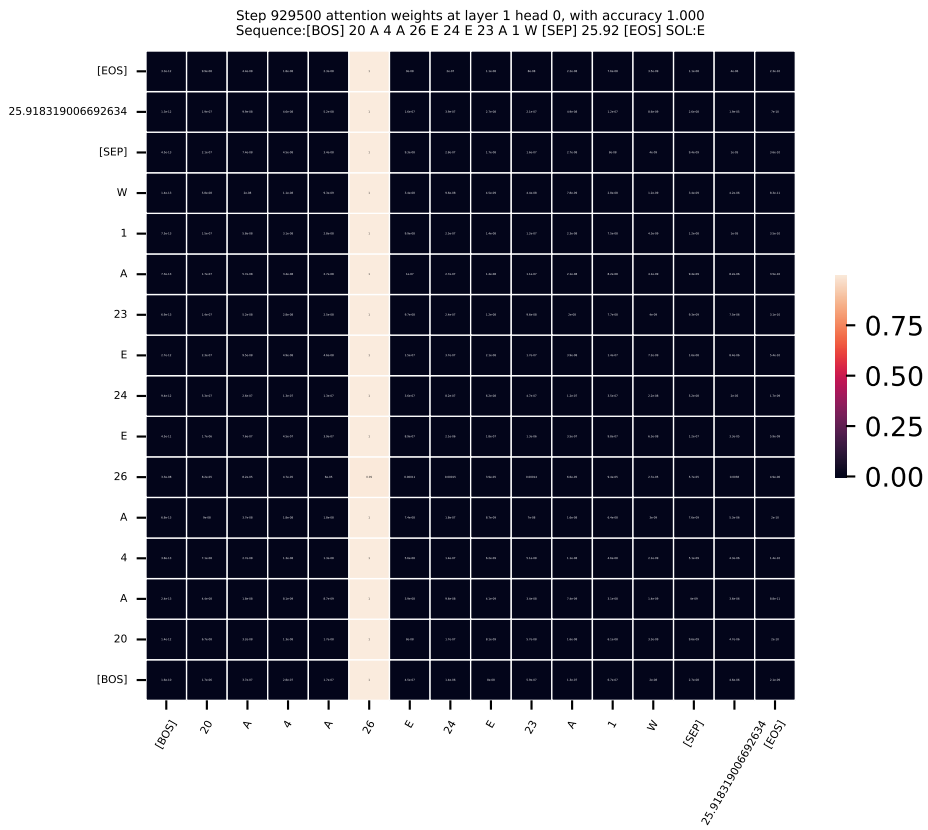

Step 920000 attention weights at layer 1 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

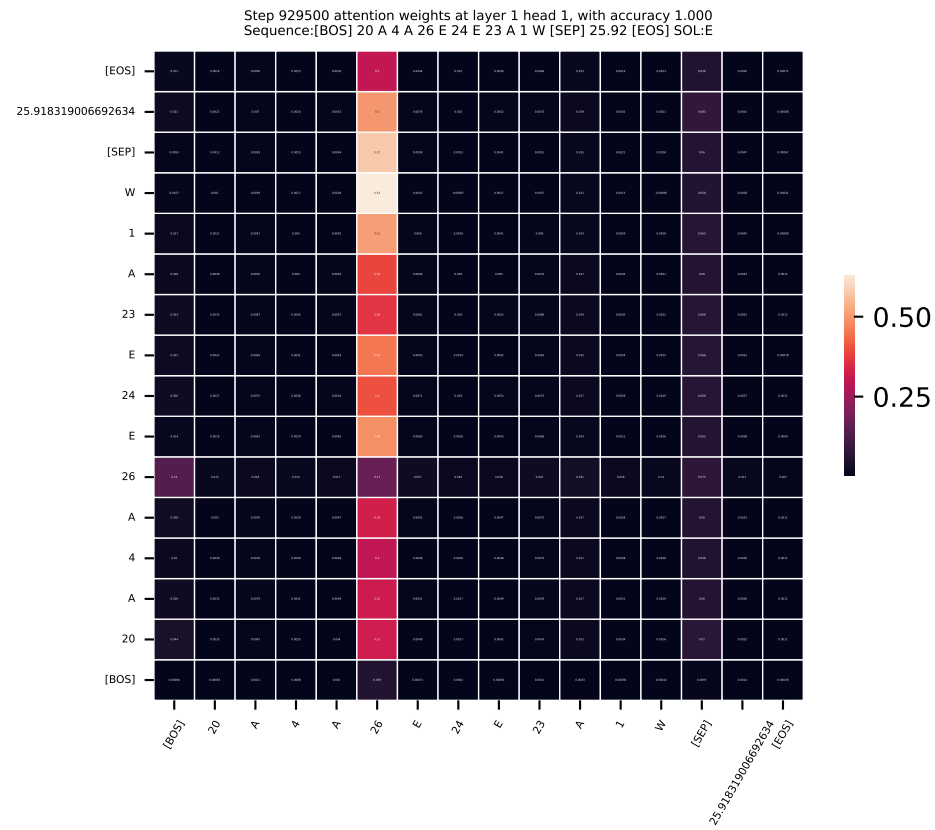

Step 929500 attention weights at layer 0 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

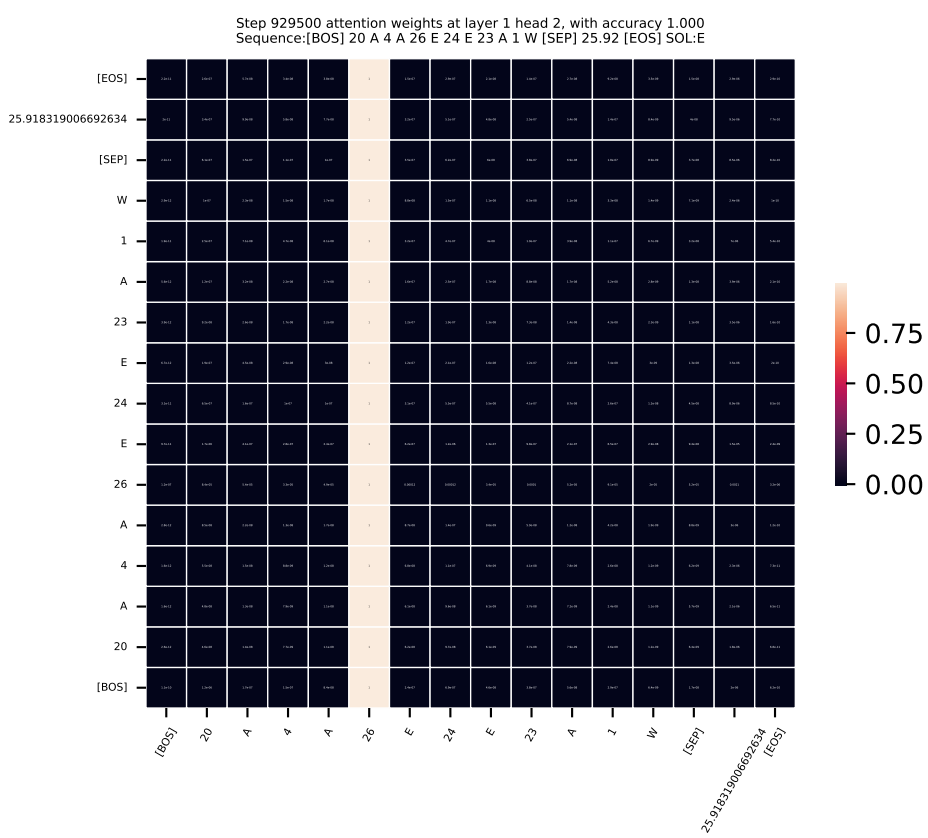

Step 929500 attention weights at layer 0 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

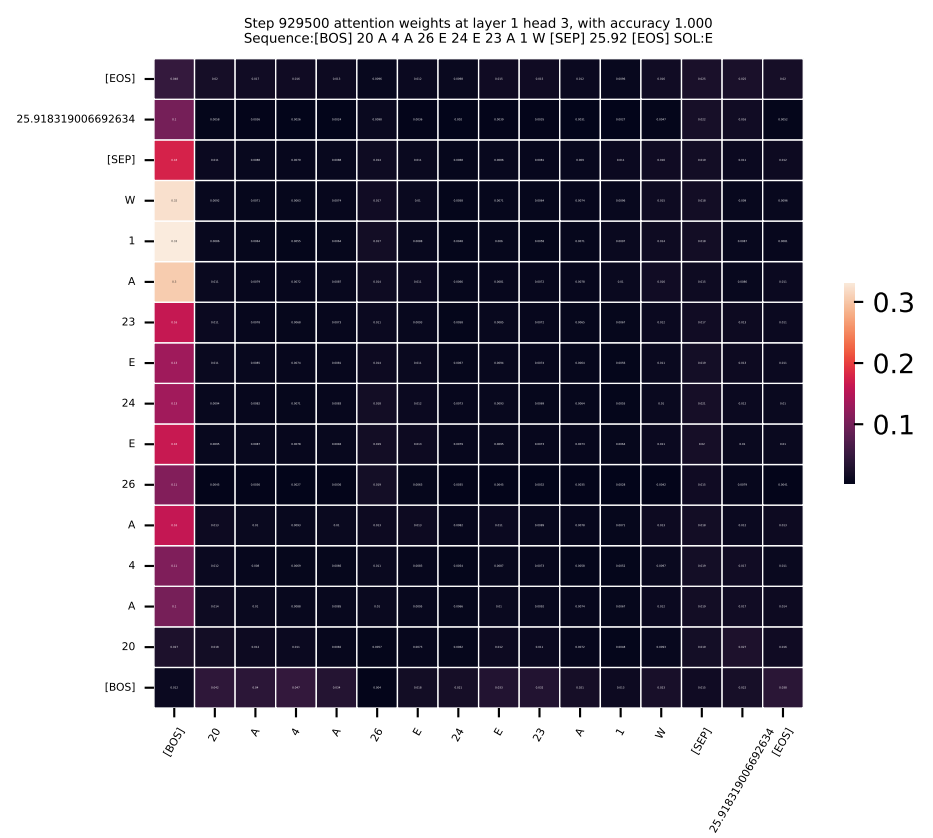

Step 929500 attention weights at layer 0 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

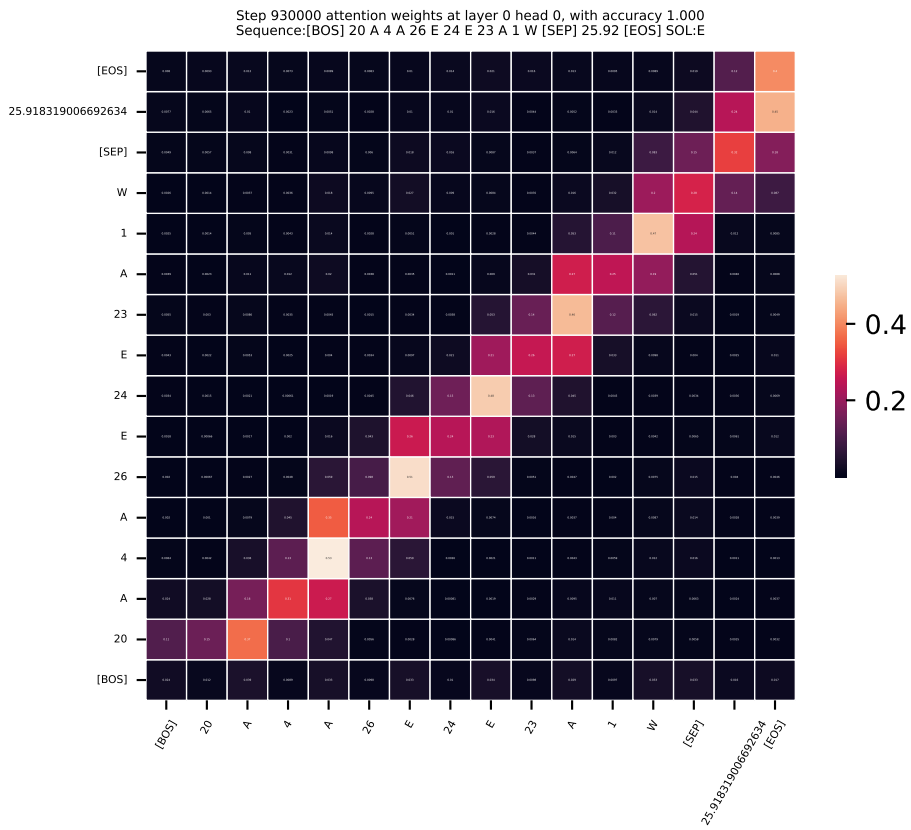

Step 929500 attention weights at layer 0 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

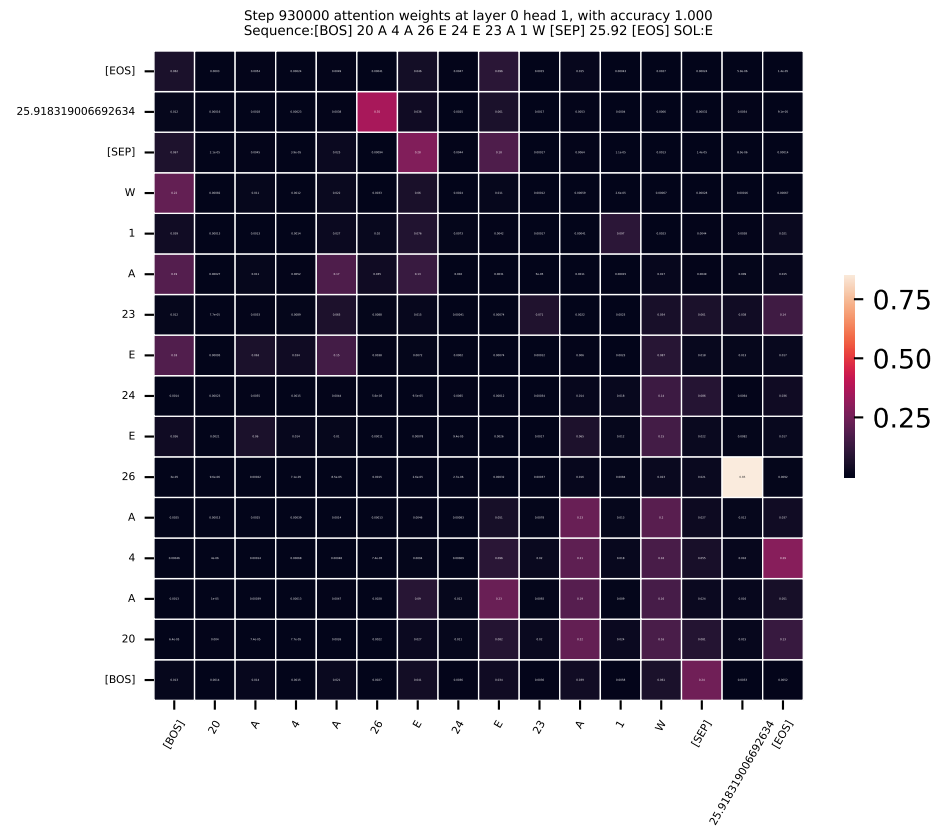

Step 929500 attention weights at layer 1 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

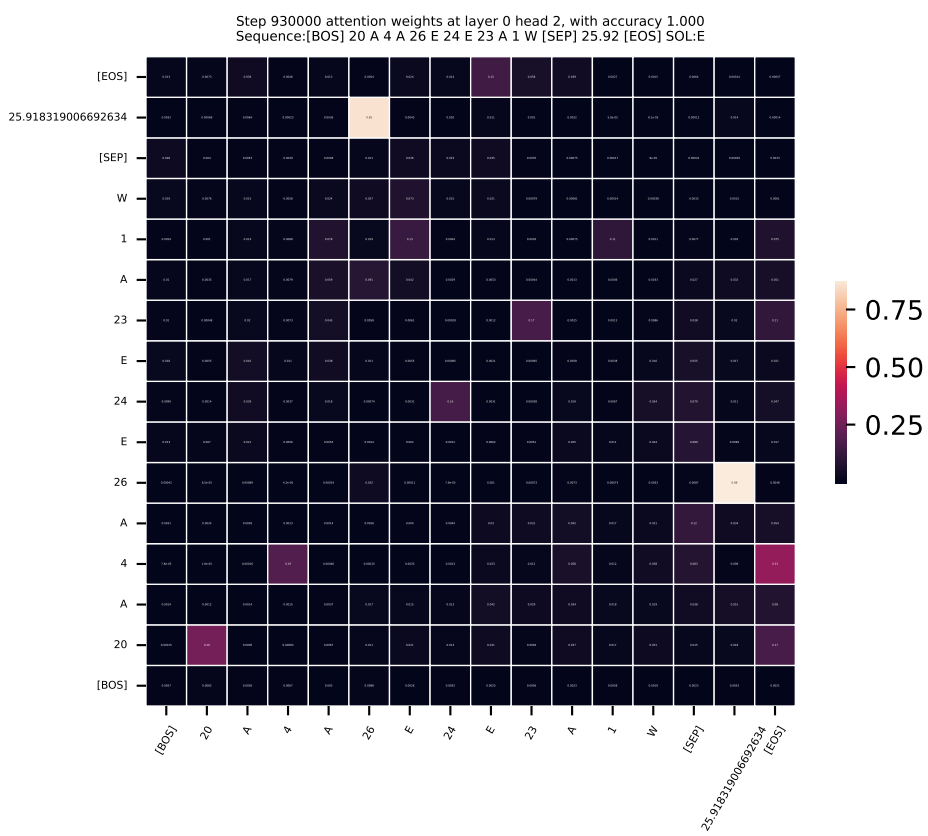

Step 929500 attention weights at layer 1 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

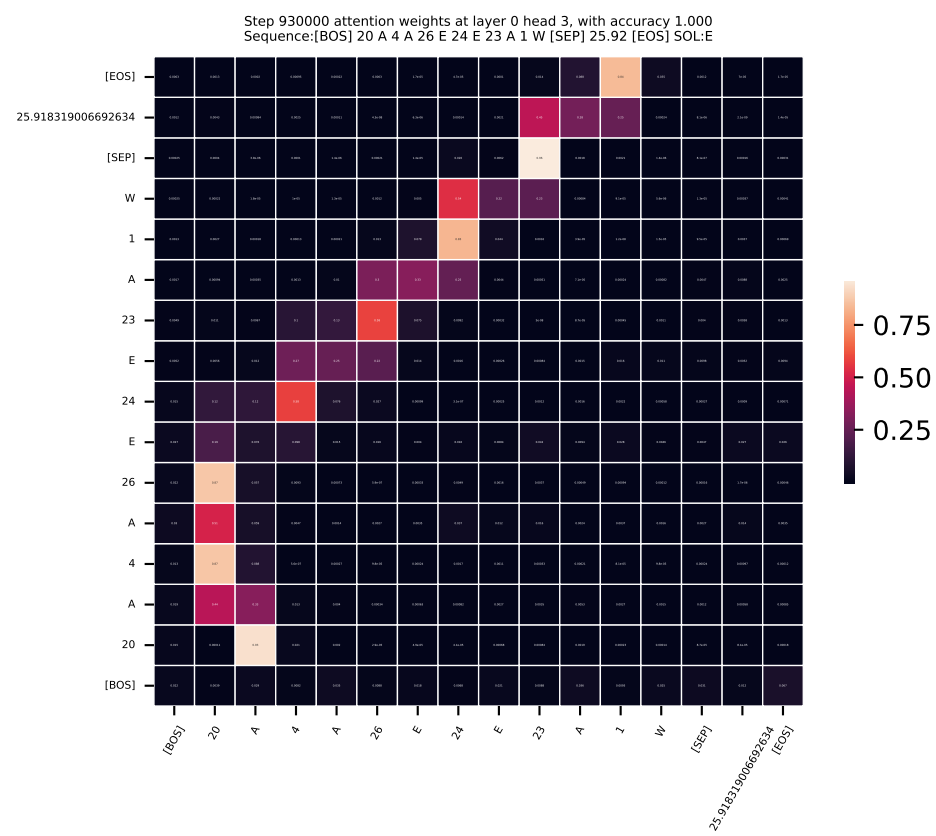

Step 929500 attention weights at layer 1 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

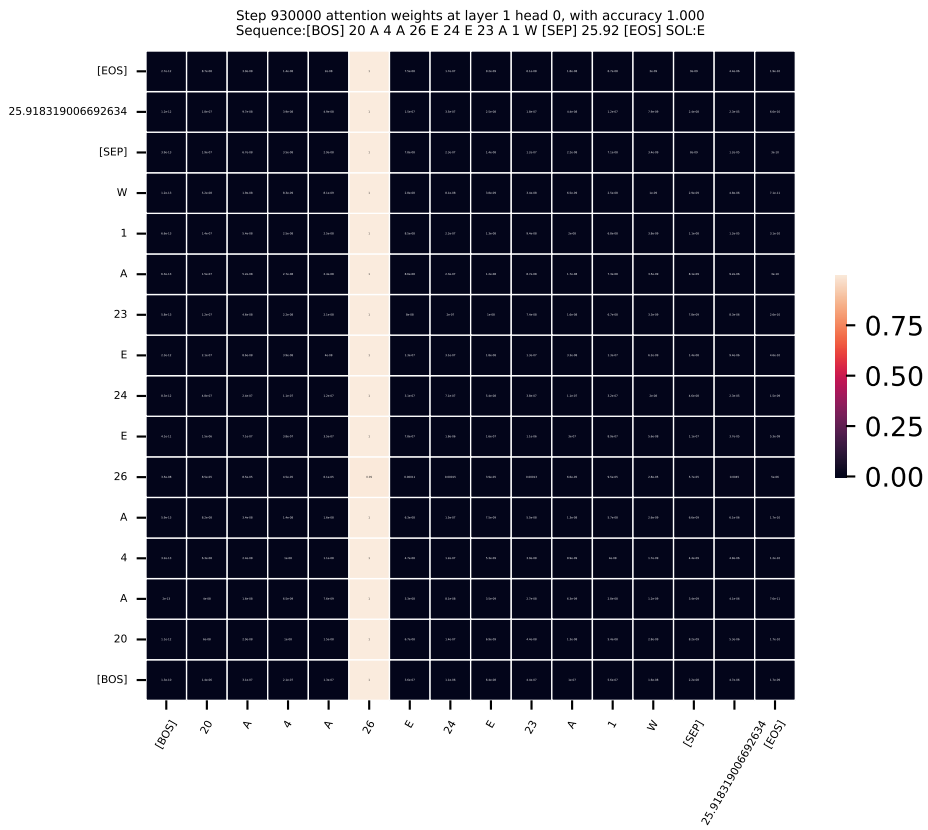

Step 929500 attention weights at layer 1 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

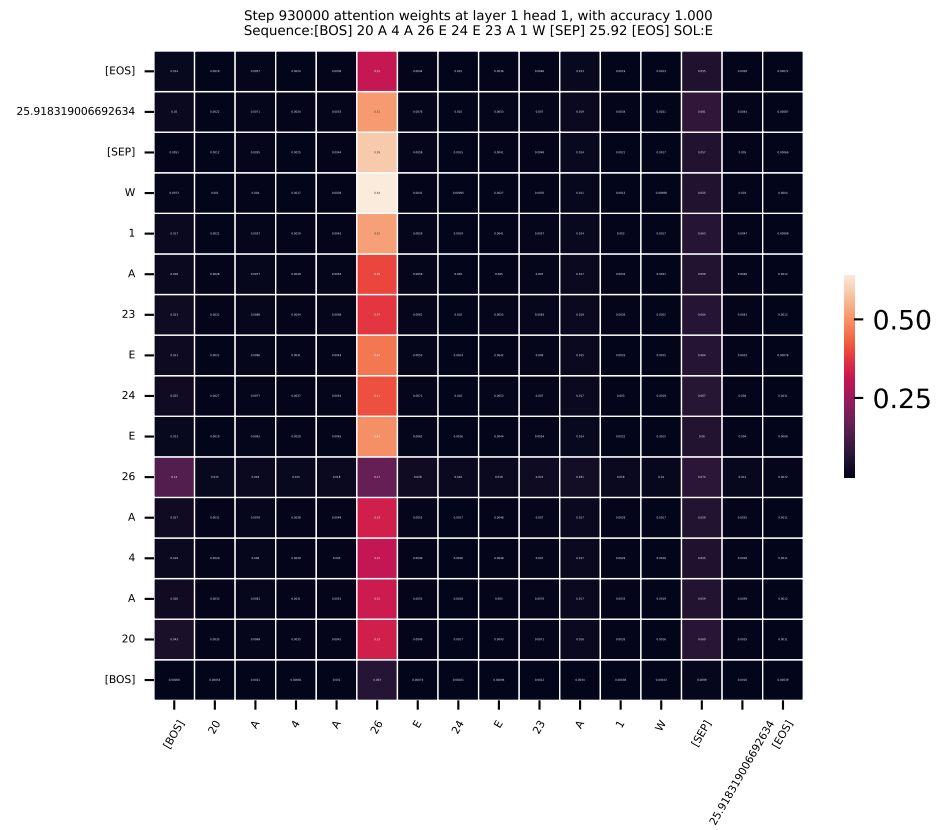

Step 930000 attention weights at layer 0 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

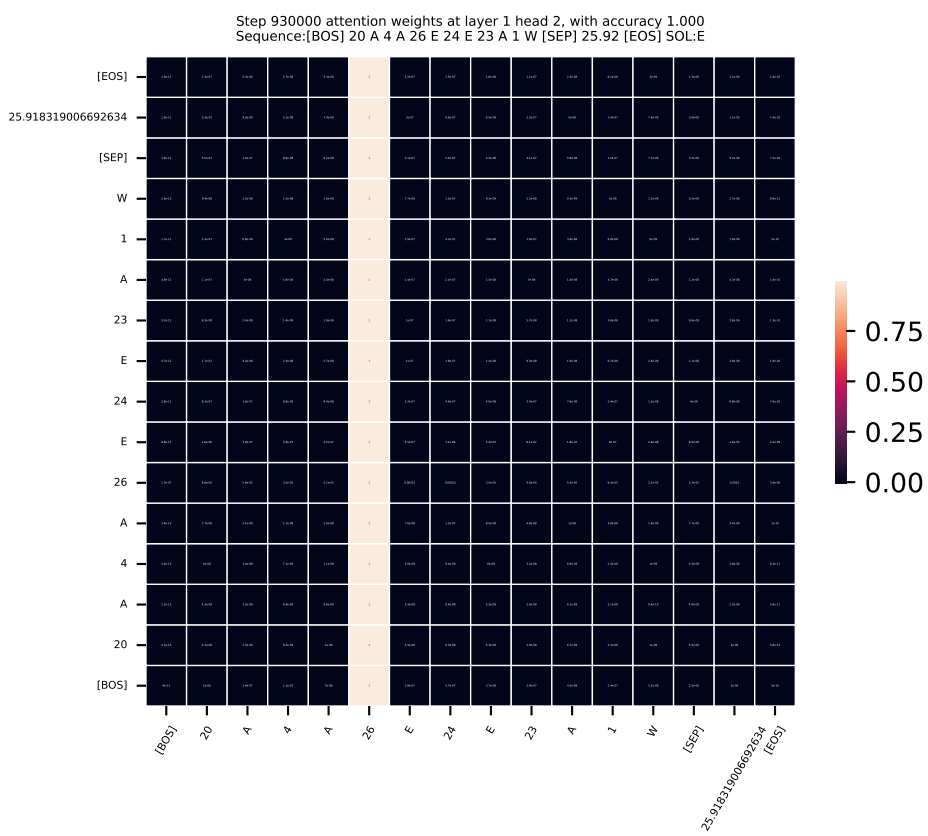

Step 930000 attention weights at layer 0 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

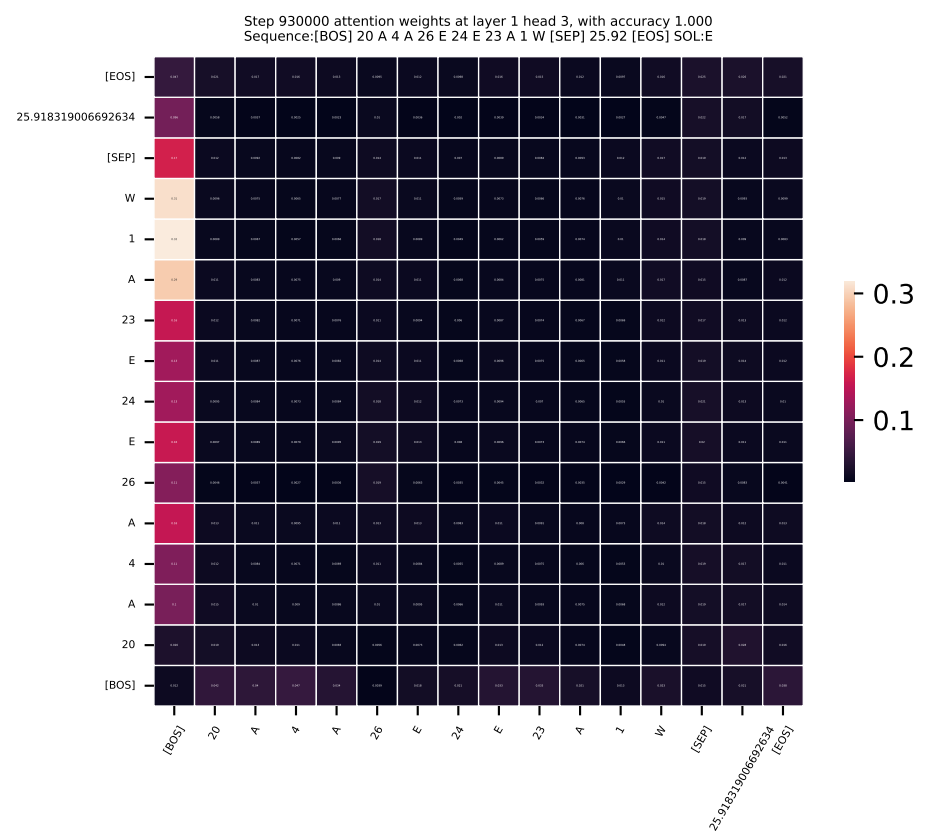

Step 930000 attention weights at layer 0 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

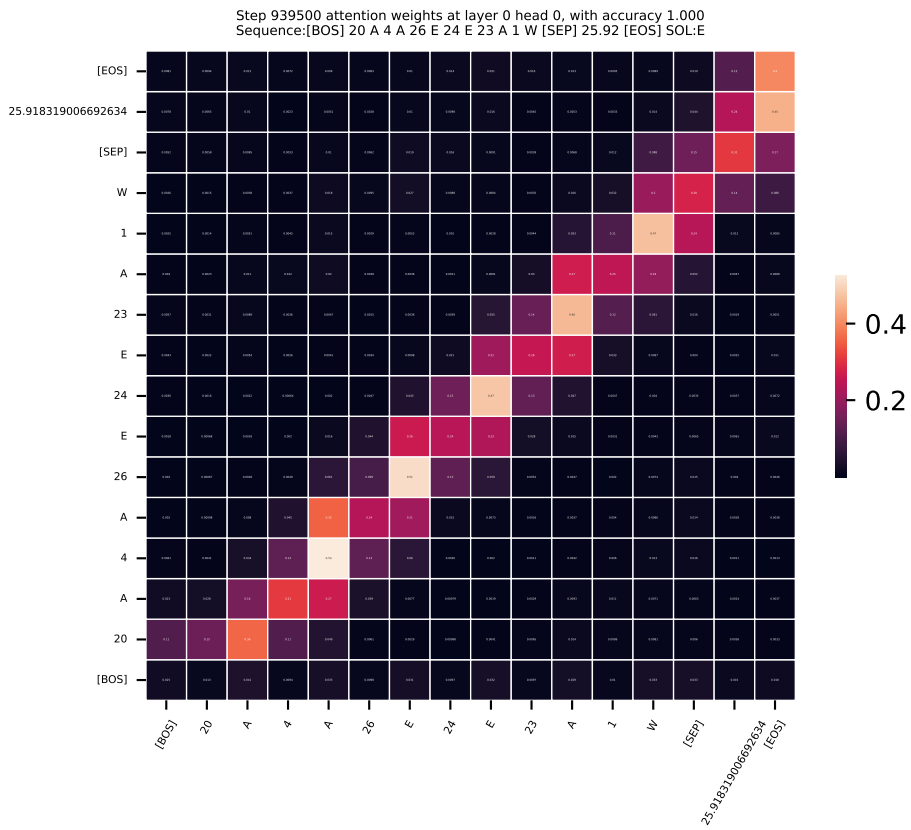

Step 930000 attention weights at layer 0 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

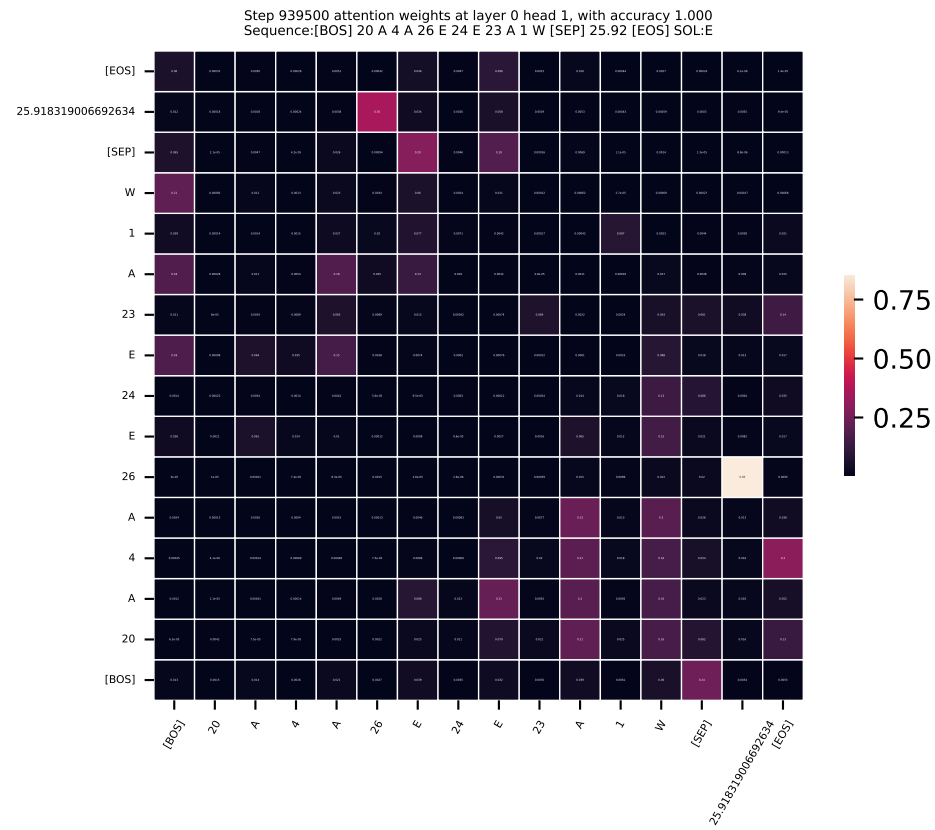

Step 930000 attention weights at layer 1 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

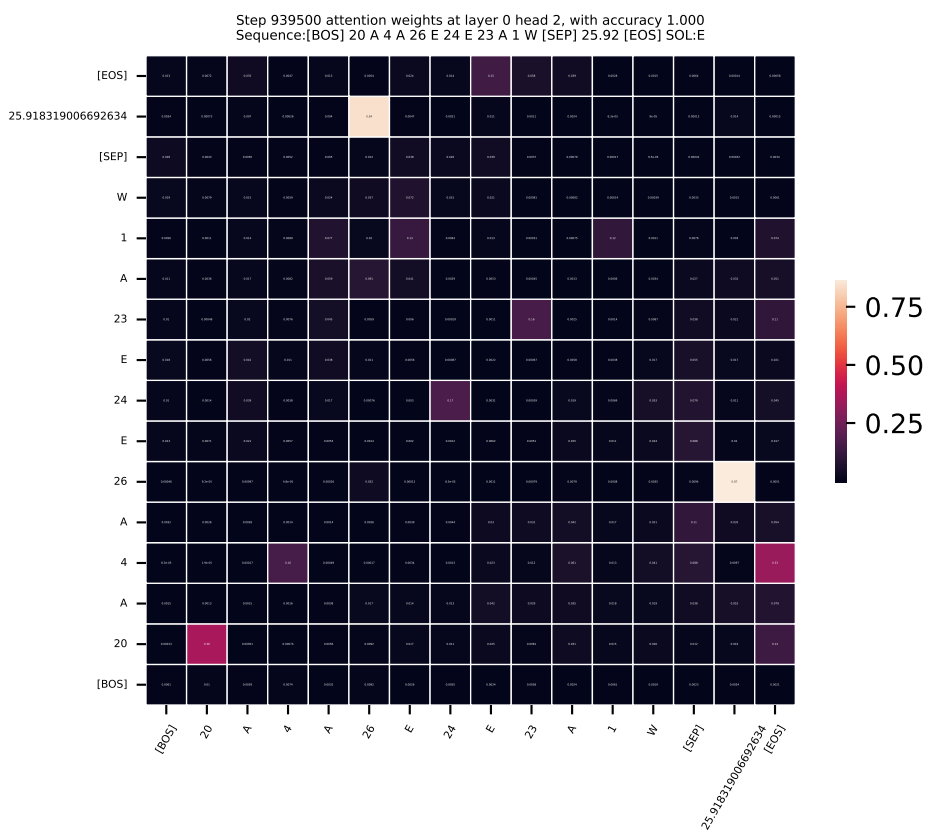

Step 930000 attention weights at layer 1 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

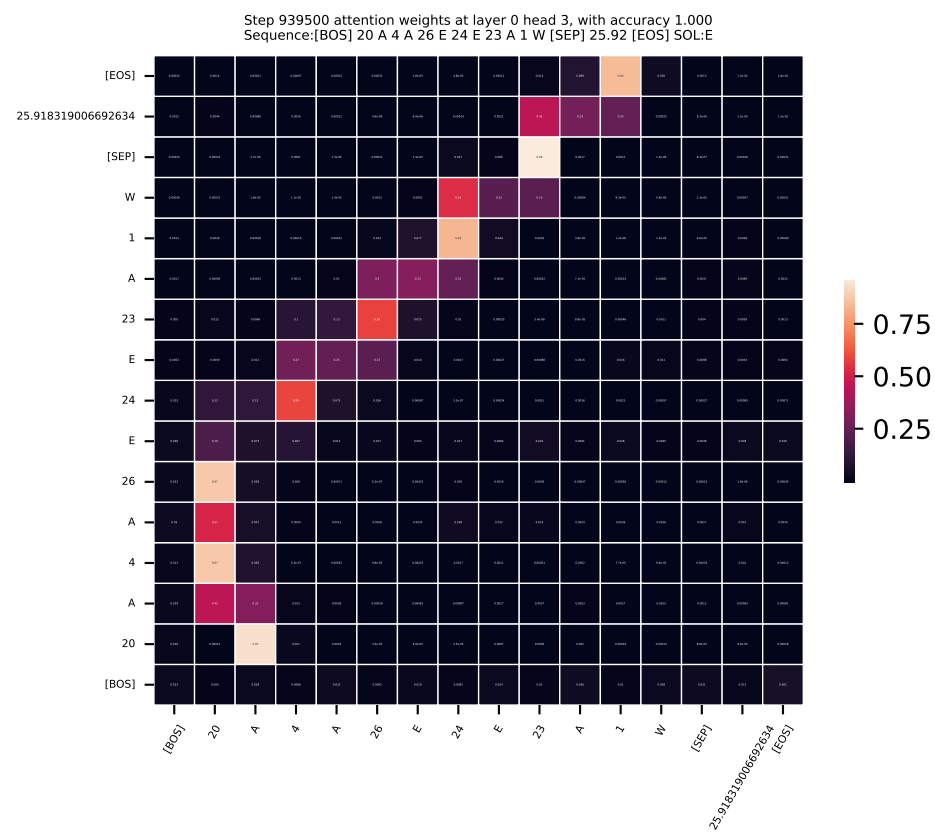

Step 930000 attention weights at layer 1 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

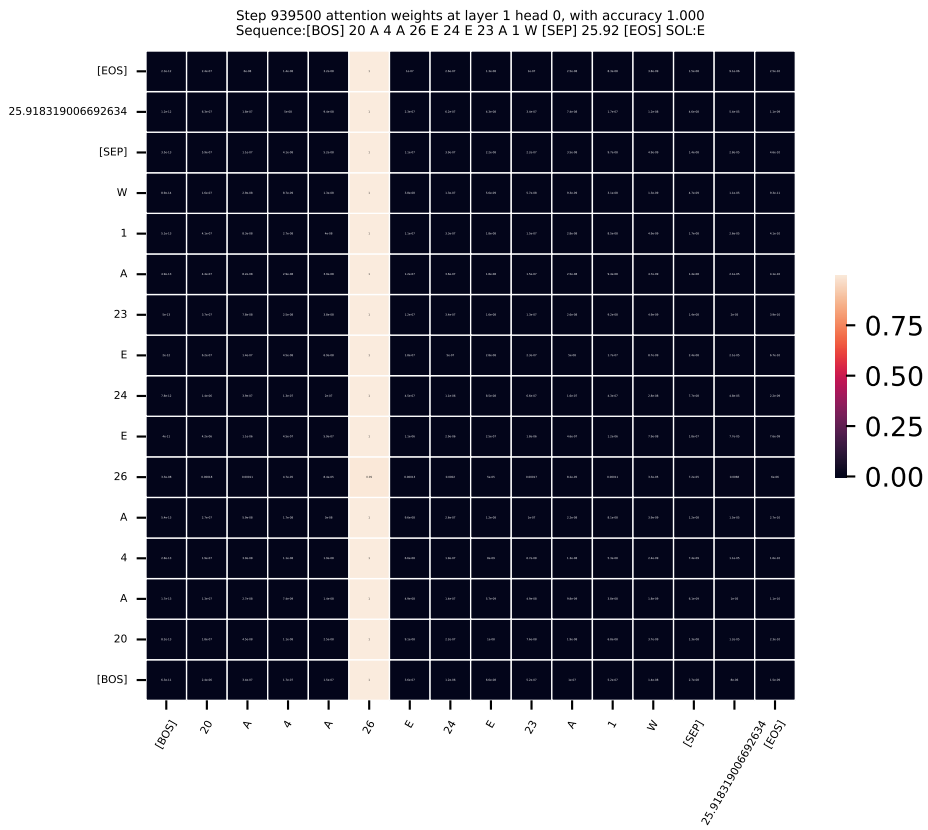

Step 930000 attention weights at layer 1 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

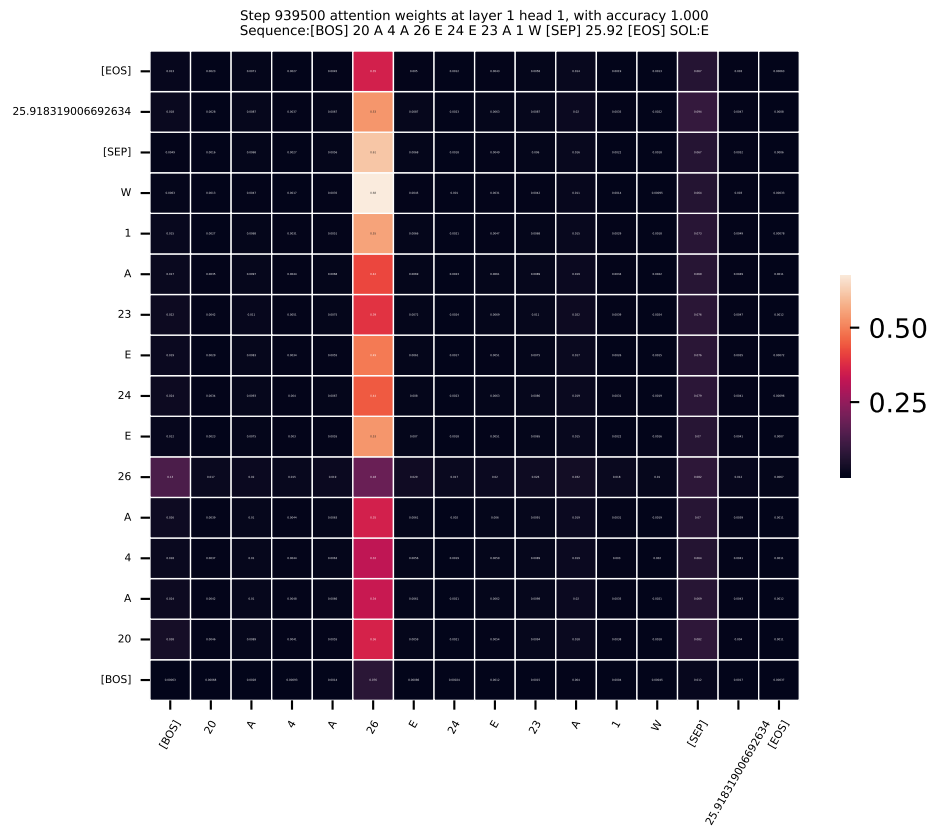

Step 939500 attention weights at layer 0 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

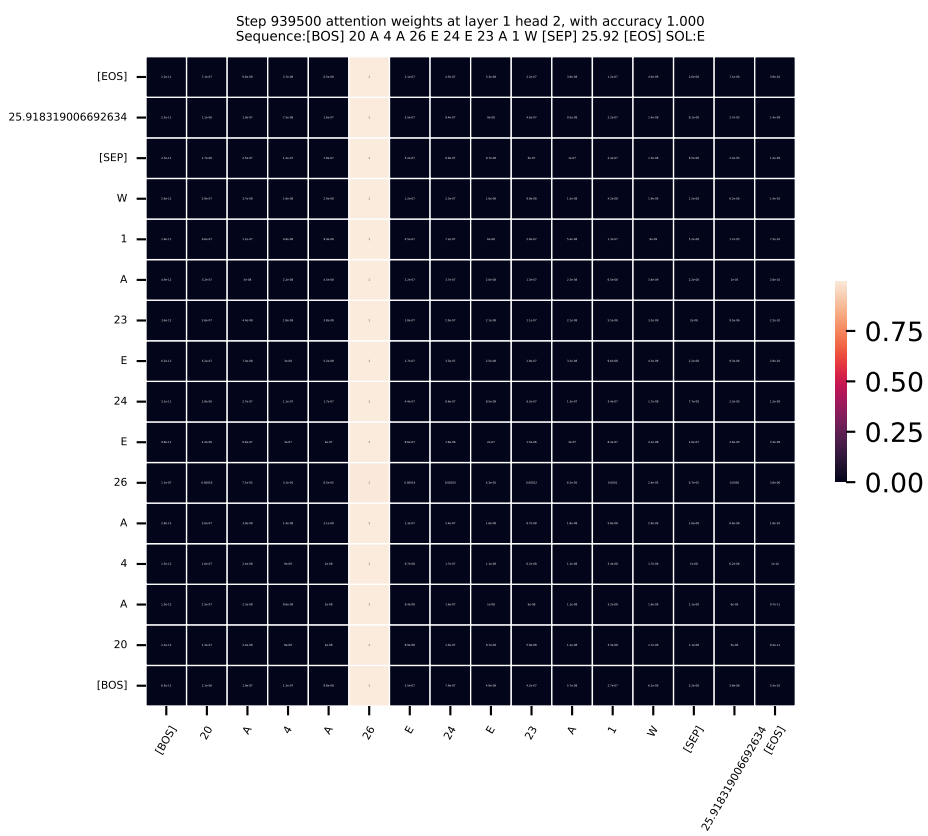

Step 939500 attention weights at layer 0 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

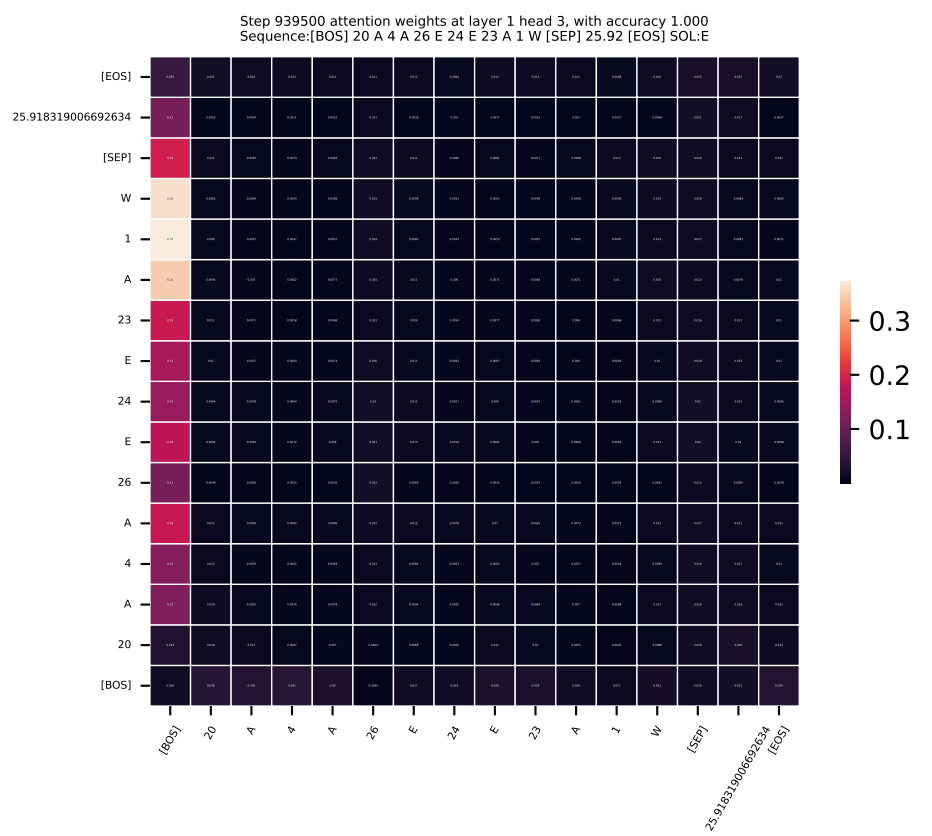

Step 939500 attention weights at layer 0 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

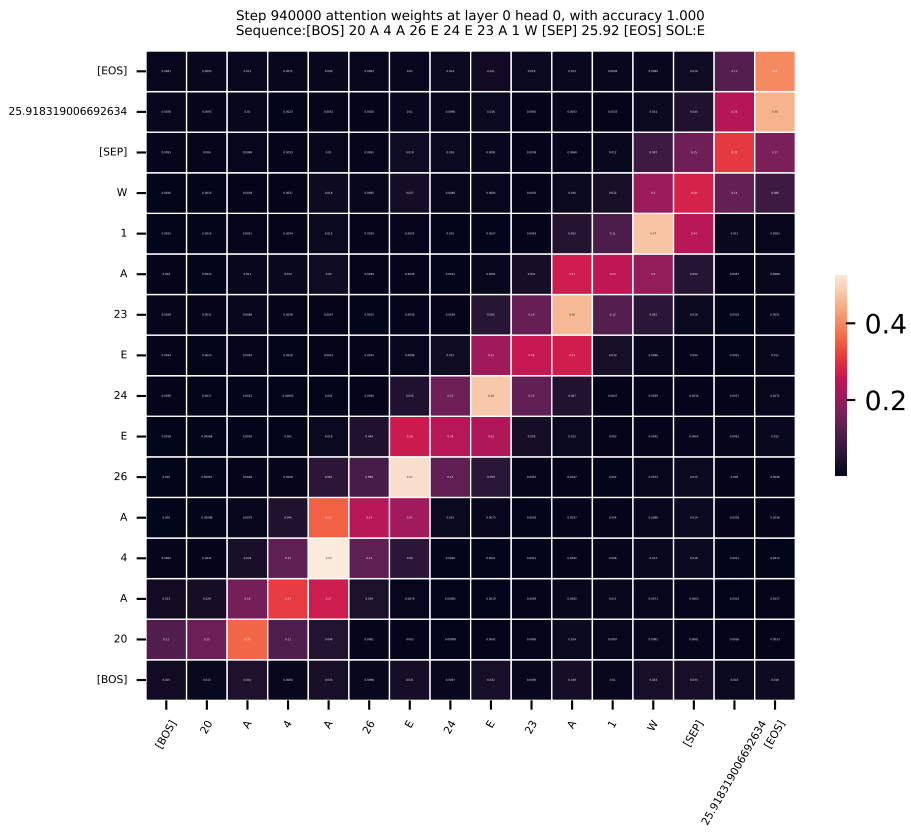

Step 939500 attention weights at layer 0 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

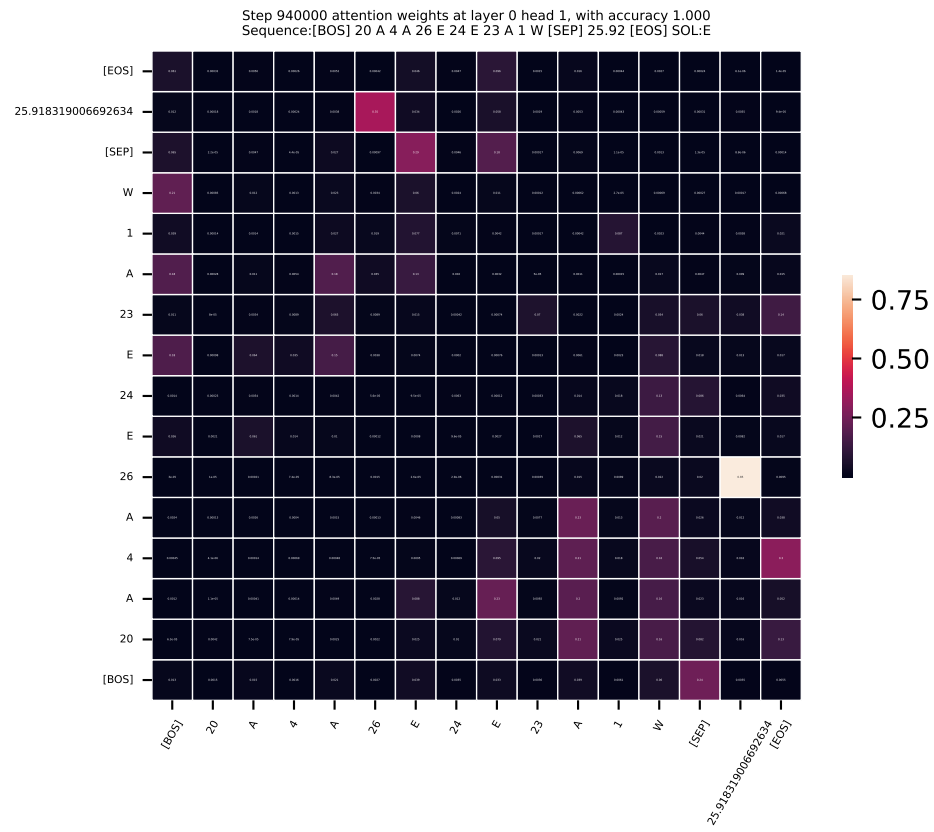

Step 939500 attention weights at layer 1 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

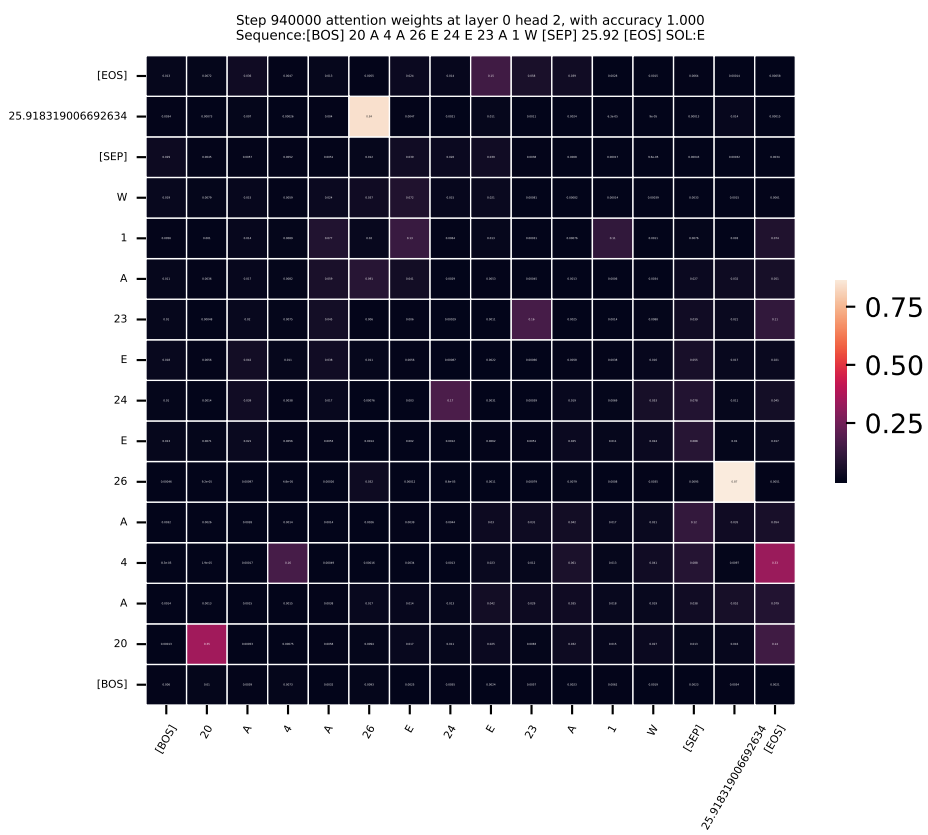

Step 939500 attention weights at layer 1 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

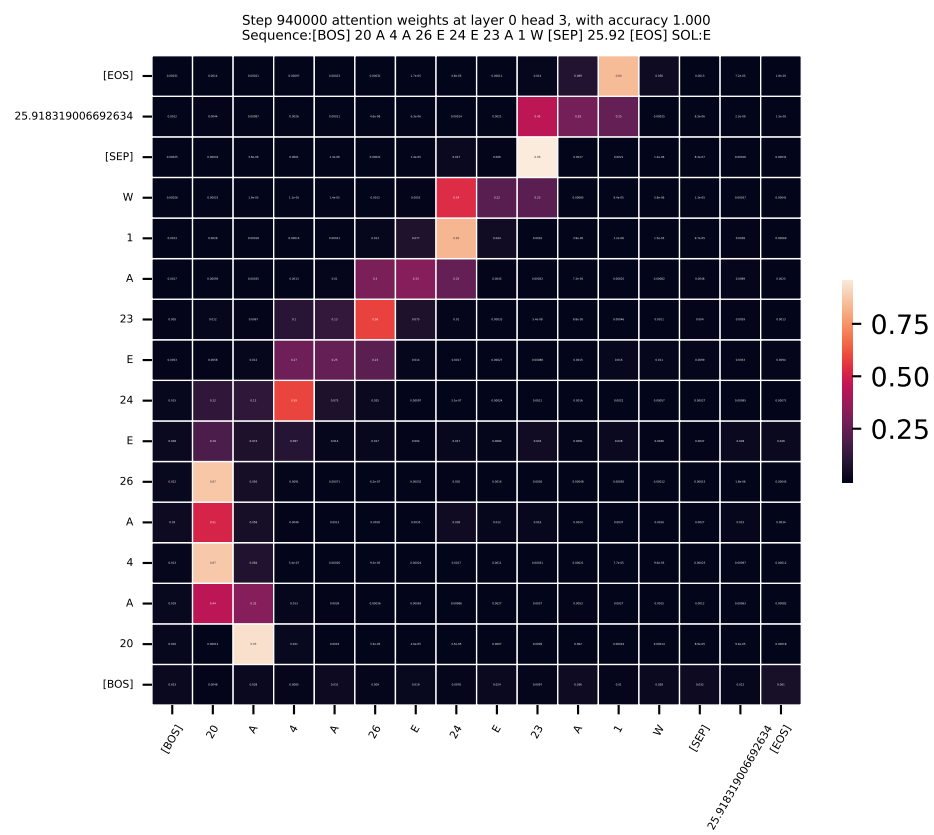

Step 939500 attention weights at layer 1 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

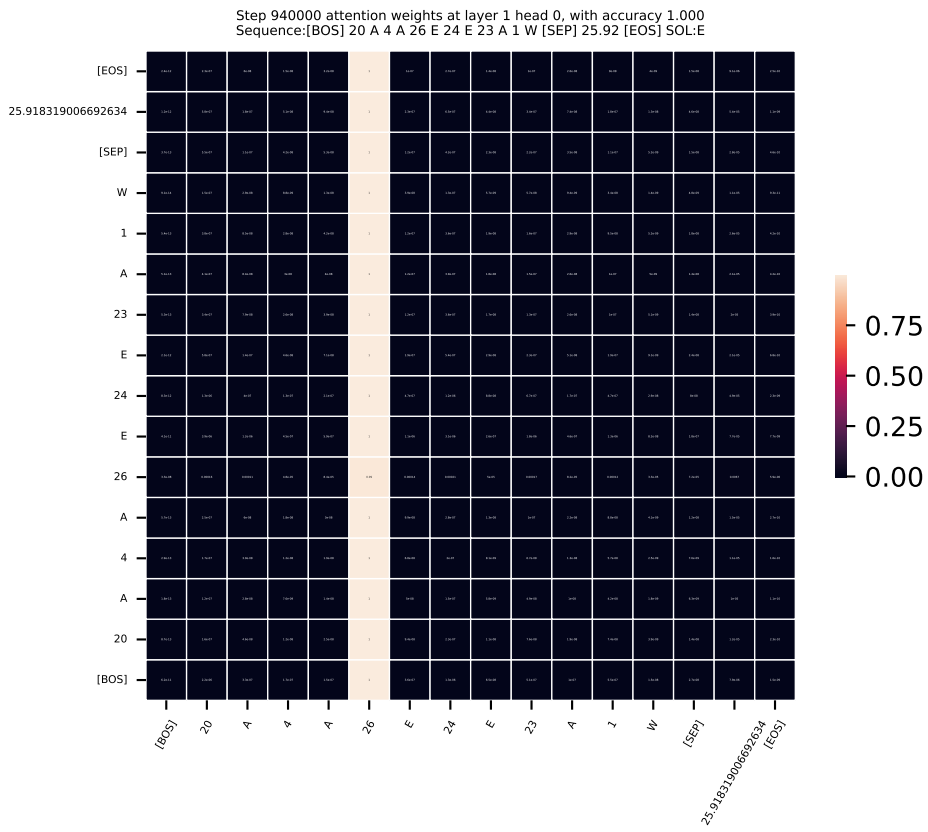

Step 939500 attention weights at layer 1 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

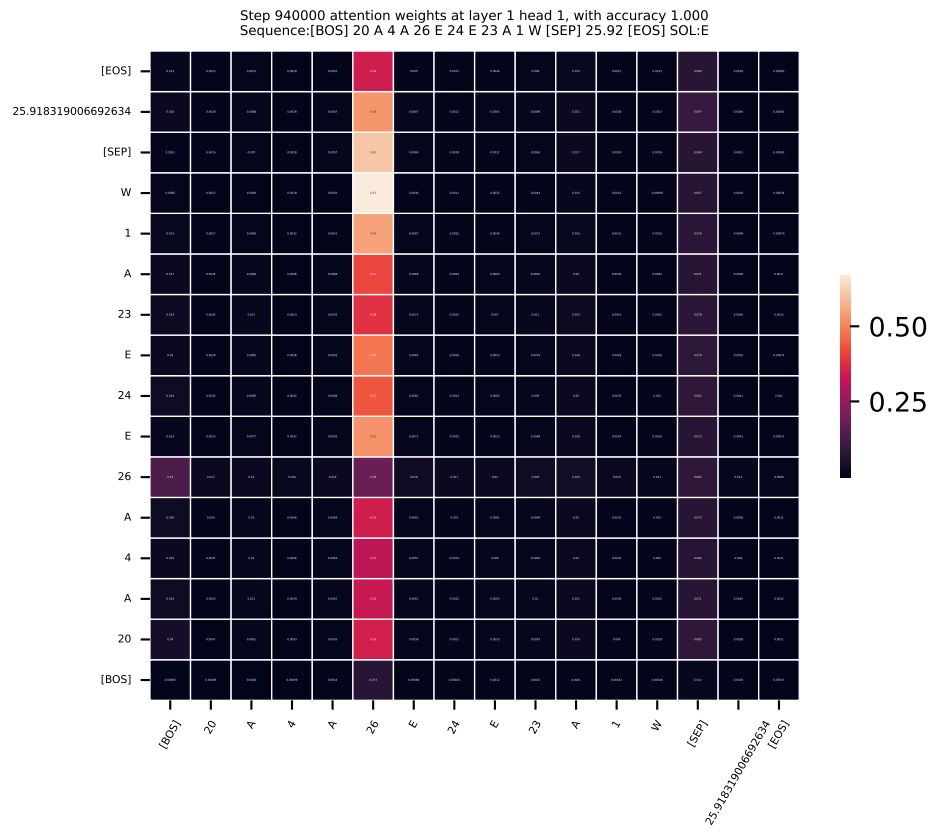

Step 940000 attention weights at layer 0 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

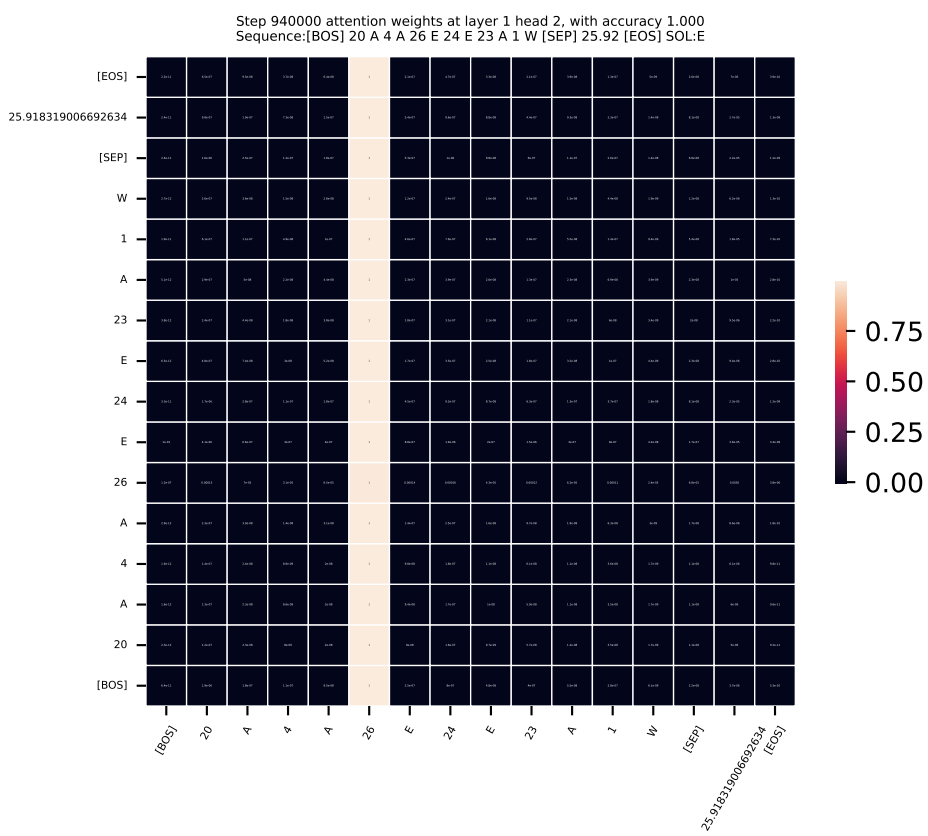

Step 940000 attention weights at layer 0 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

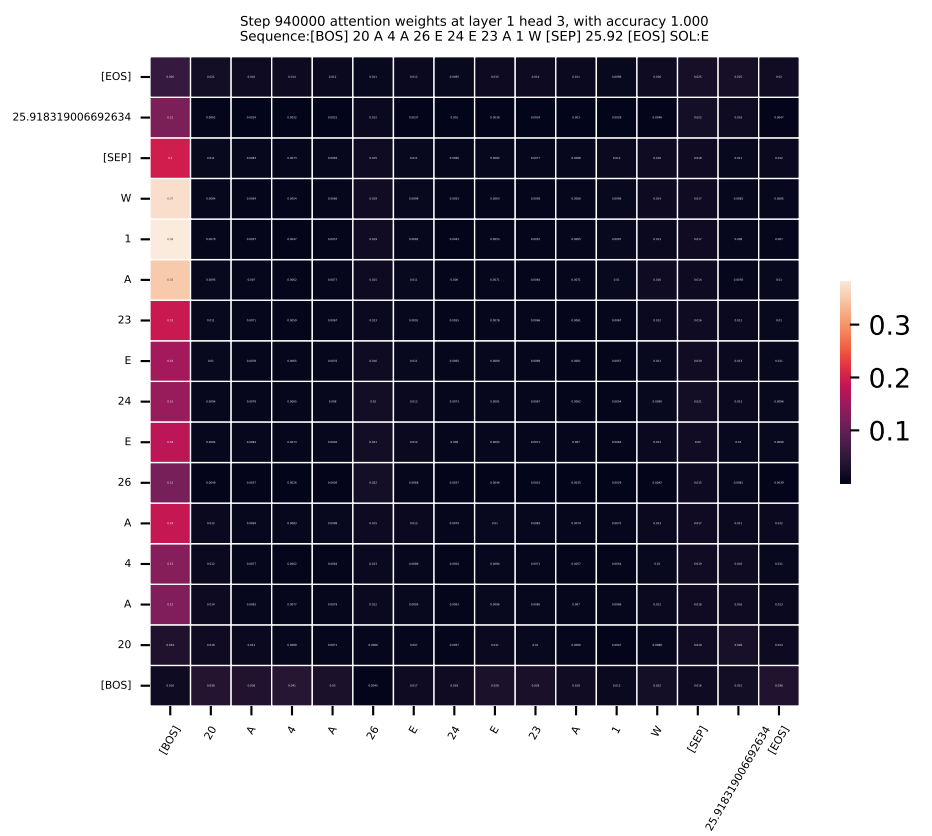

Step 940000 attention weights at layer 0 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

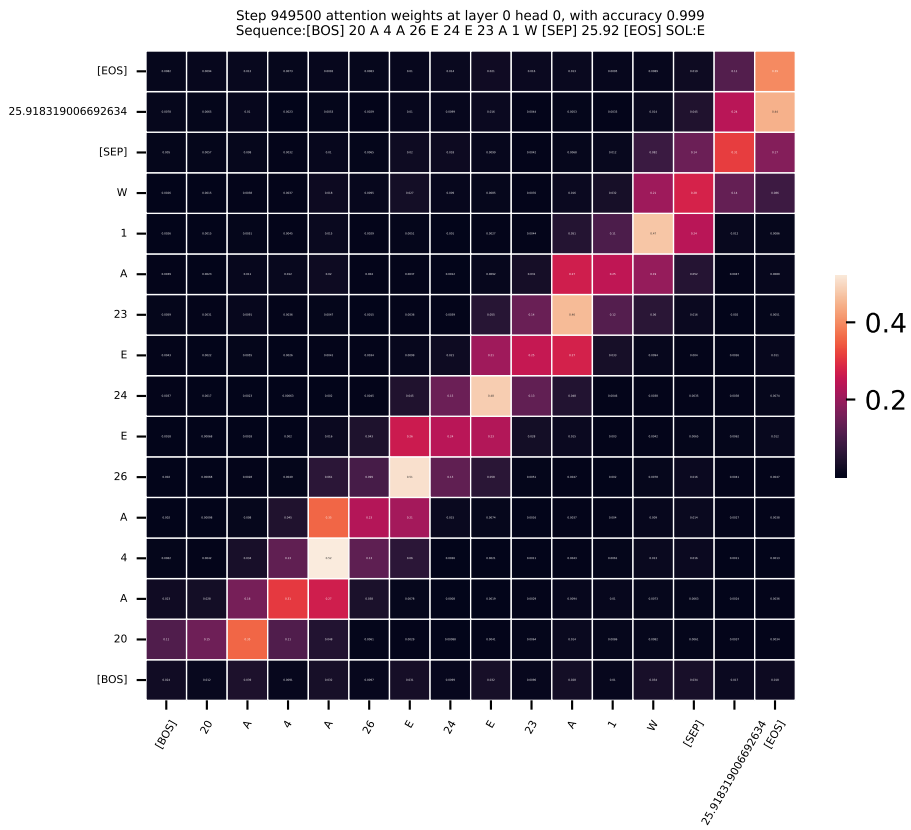

Step 940000 attention weights at layer 0 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

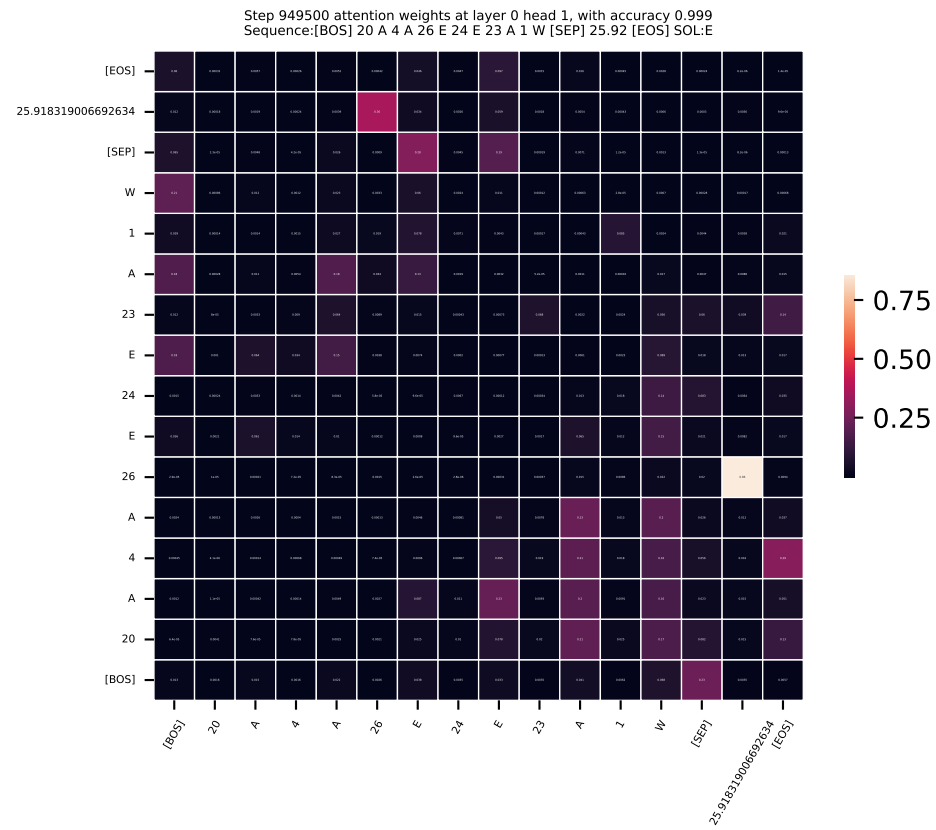

Step 940000 attention weights at layer 1 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

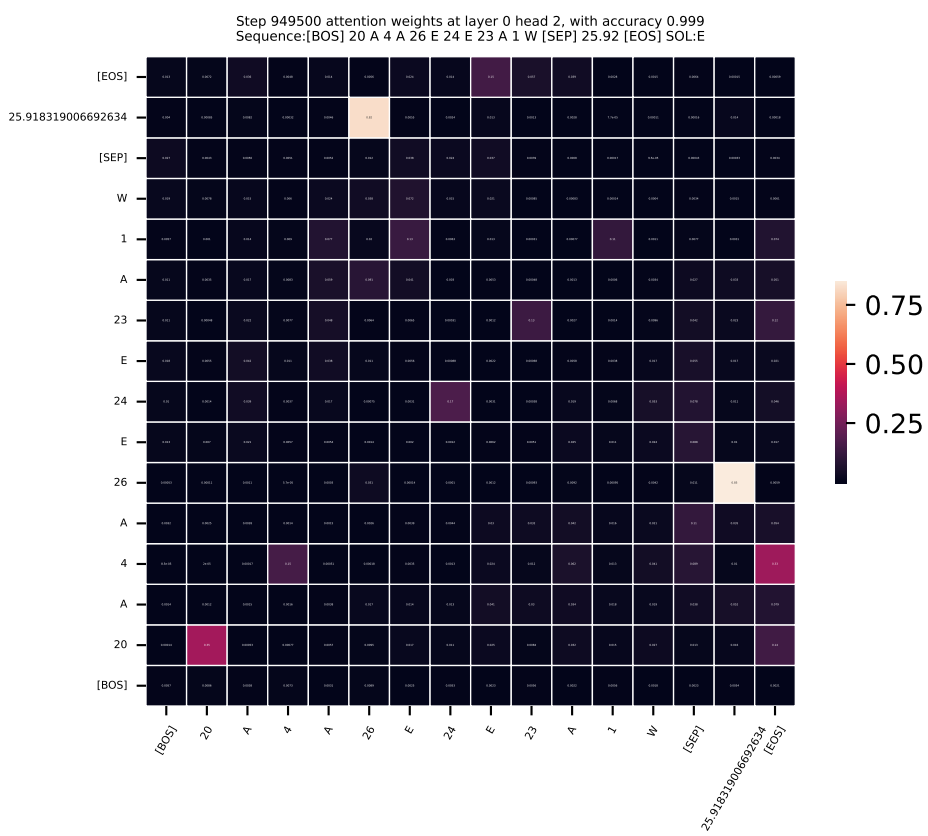

Step 940000 attention weights at layer 1 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

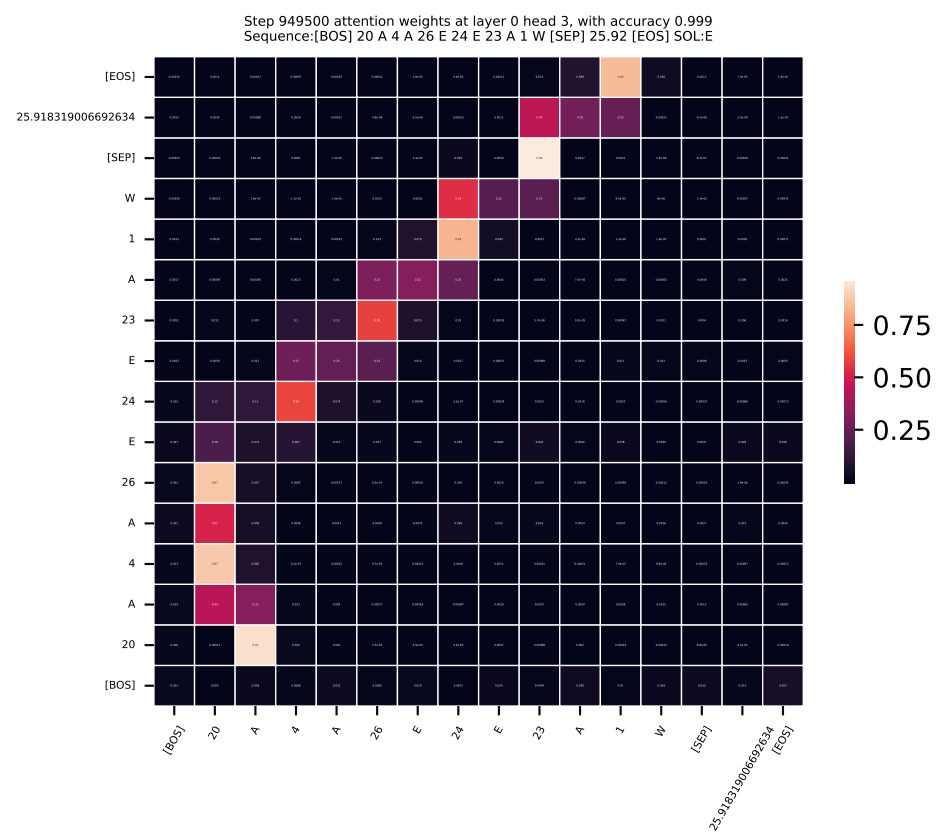

Step 940000 attention weights at layer 1 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

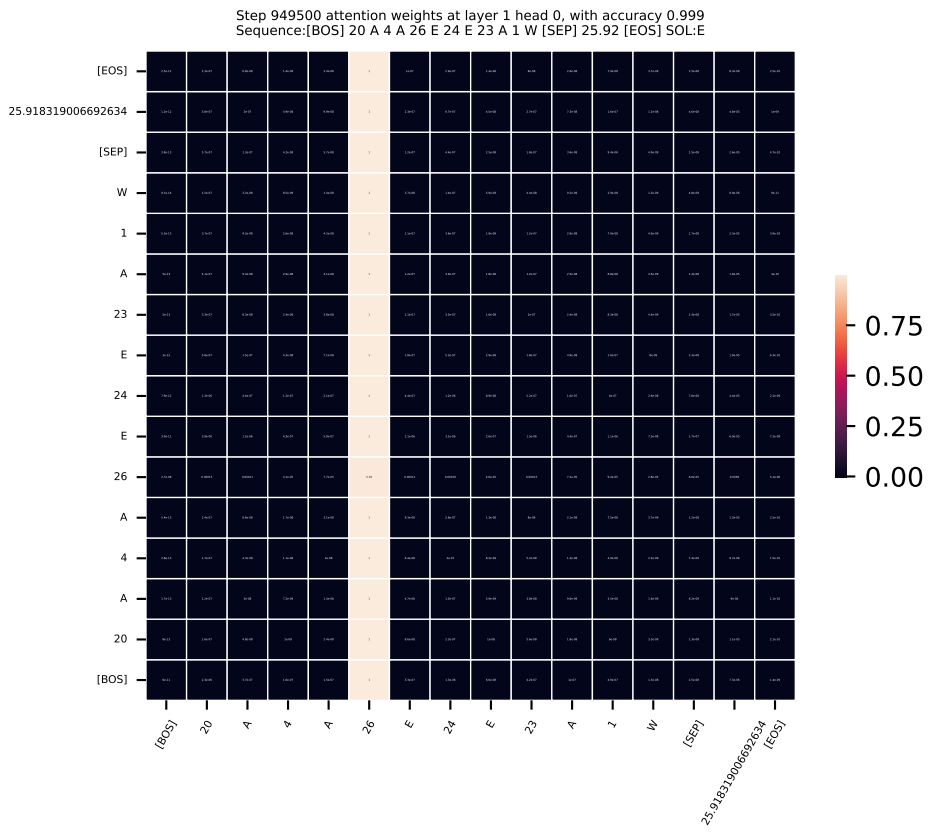

Step 940000 attention weights at layer 1 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

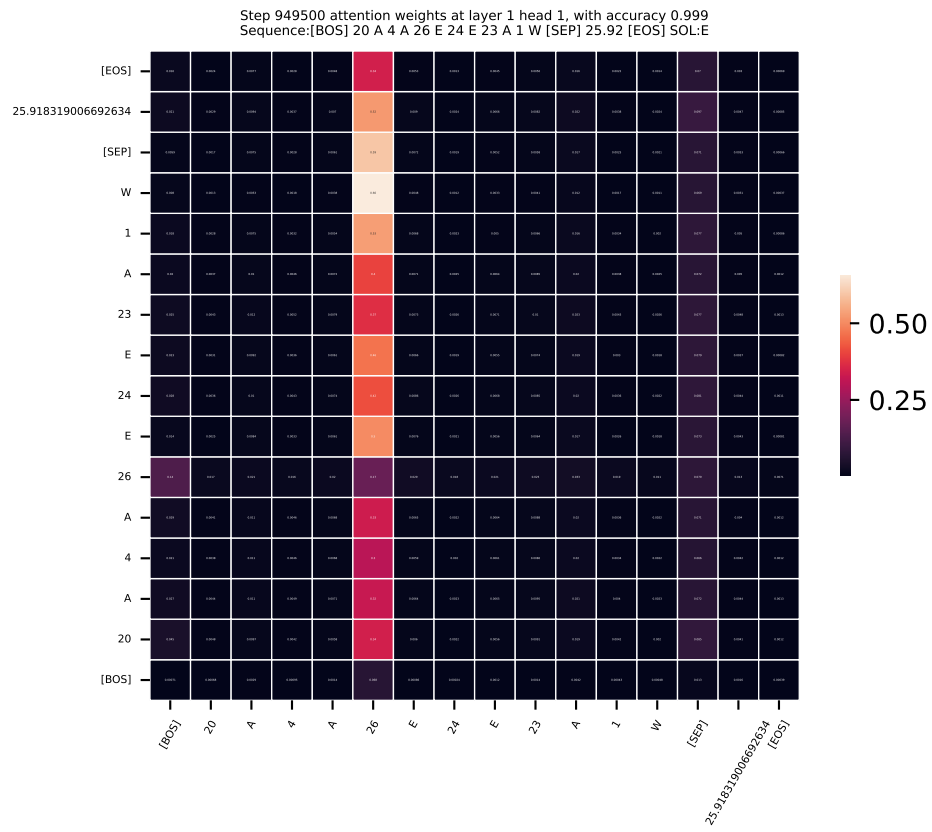

Step 949500 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

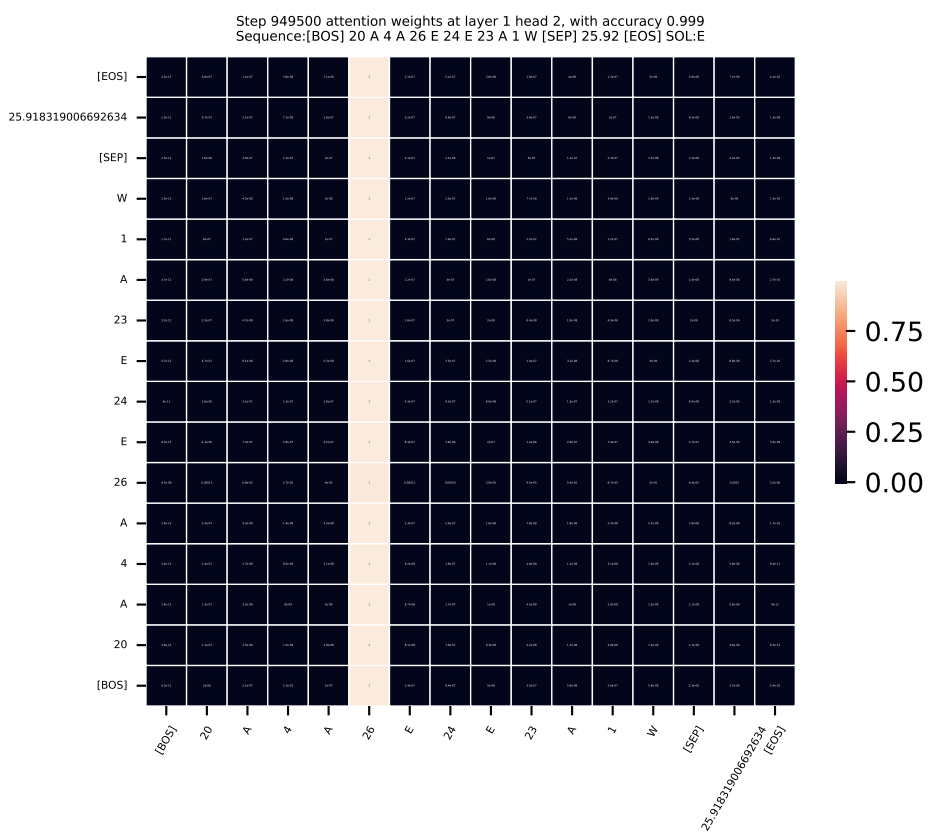

Step 949500 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

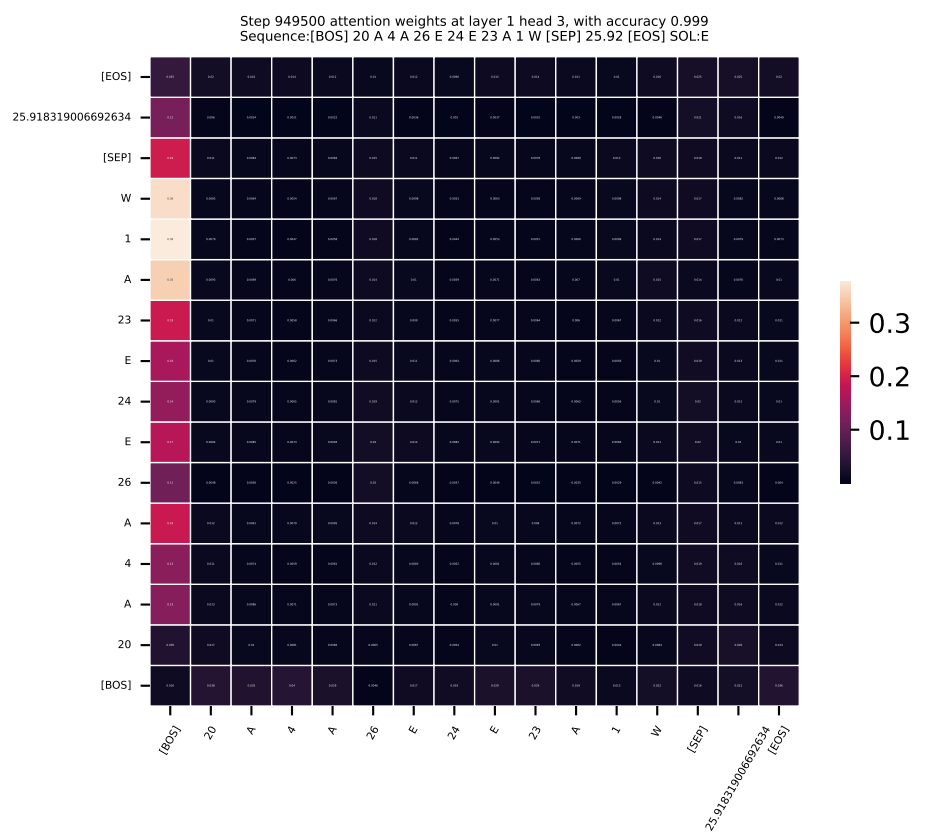

Step 949500 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

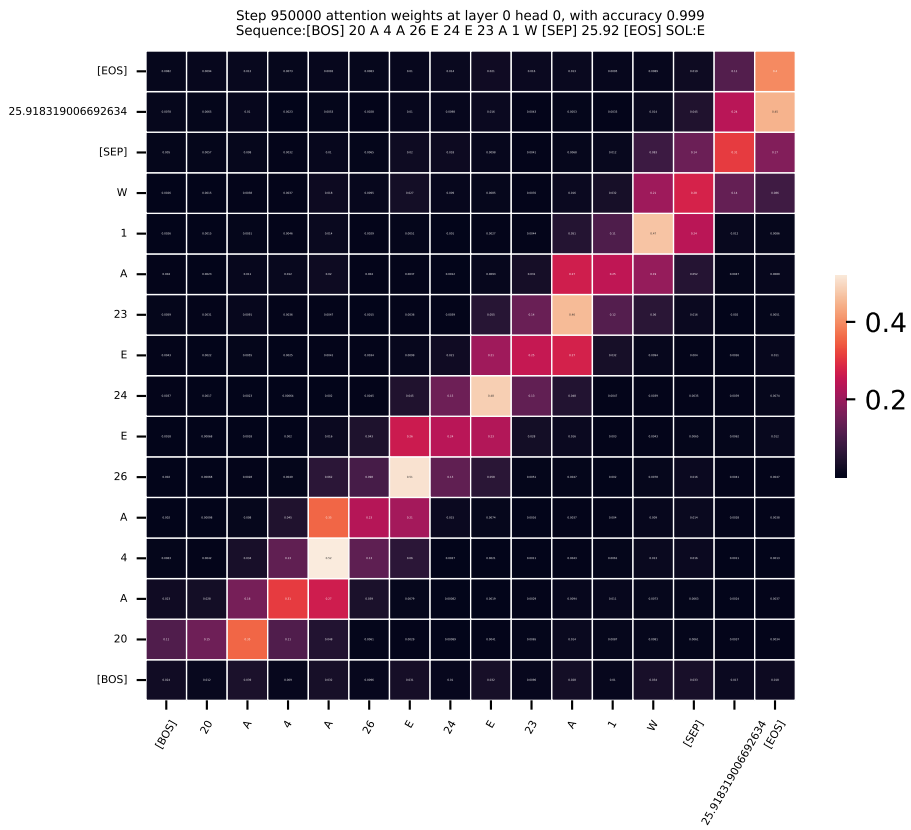

Step 949500 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

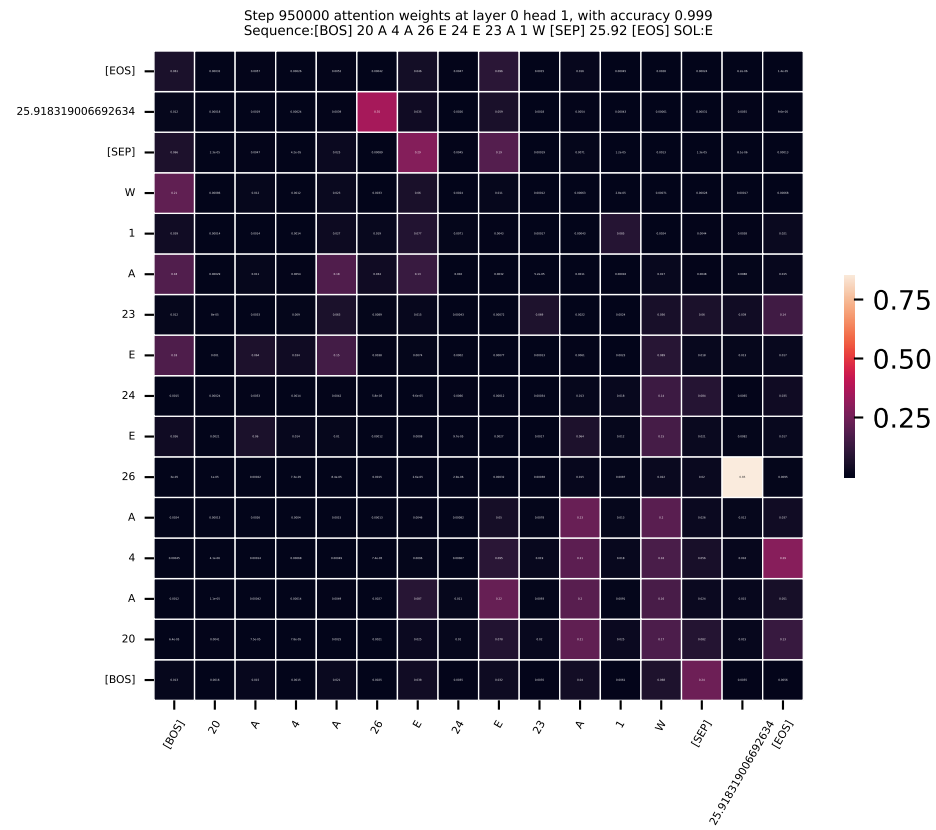

Step 949500 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

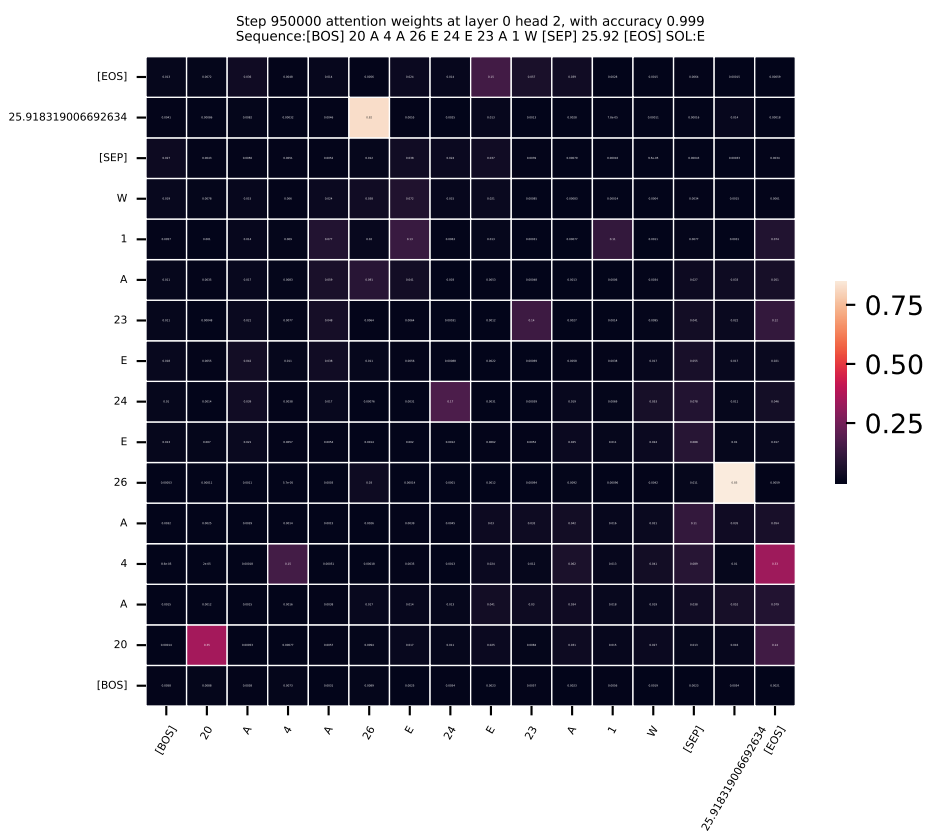

Step 949500 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

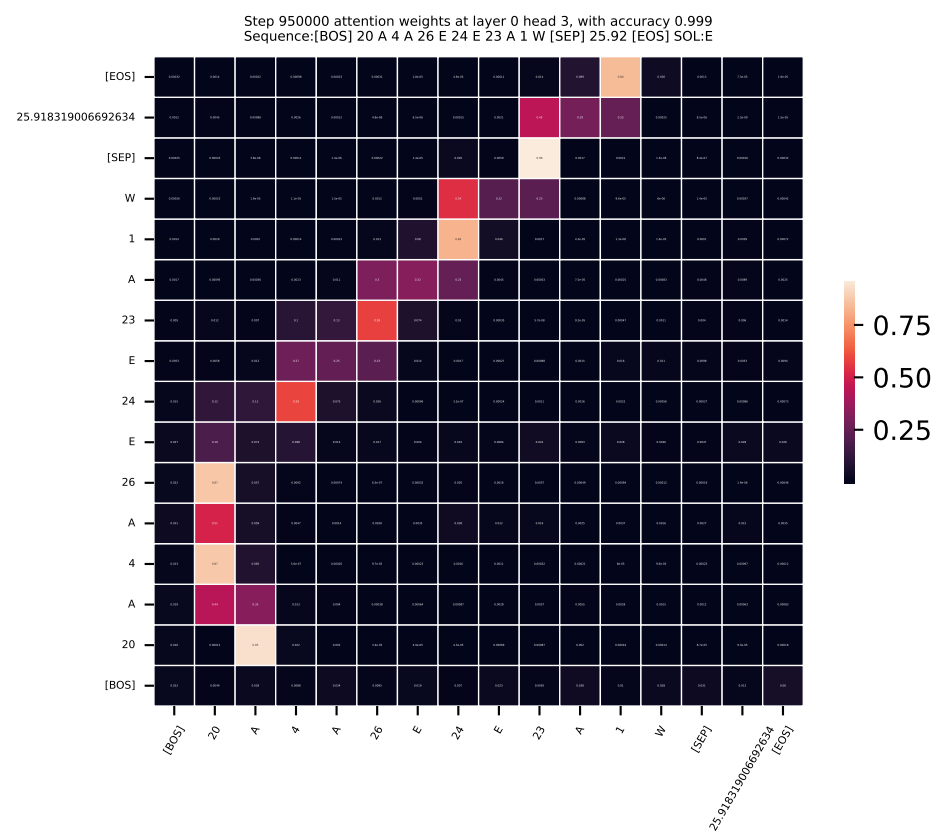

Step 949500 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

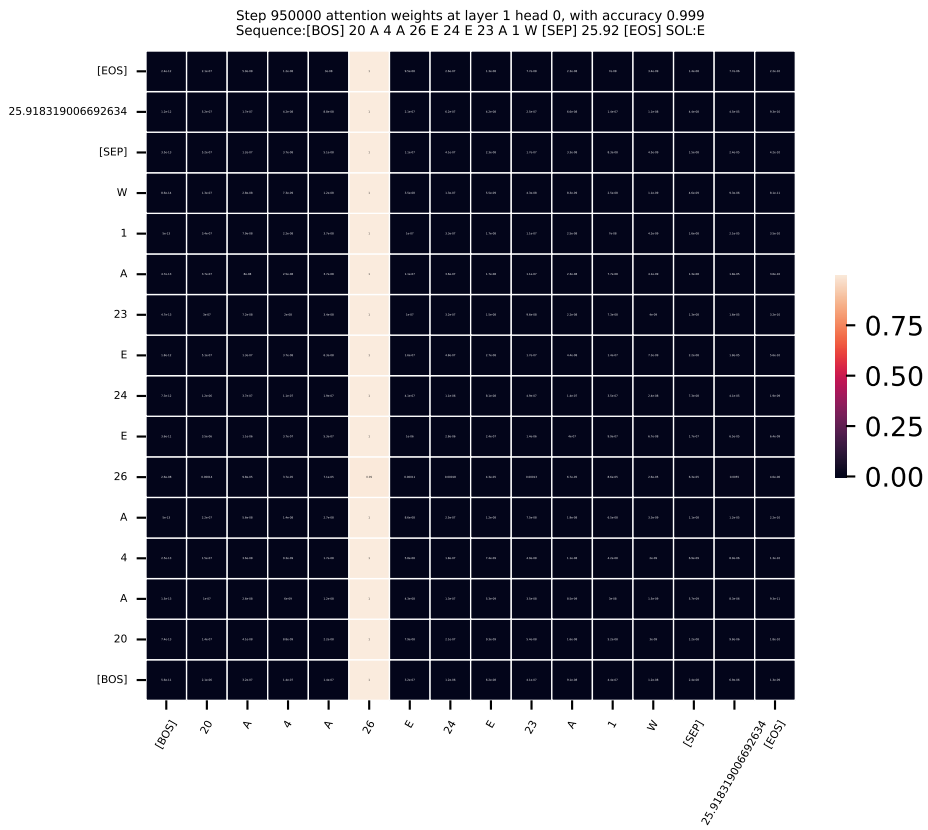

Step 949500 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

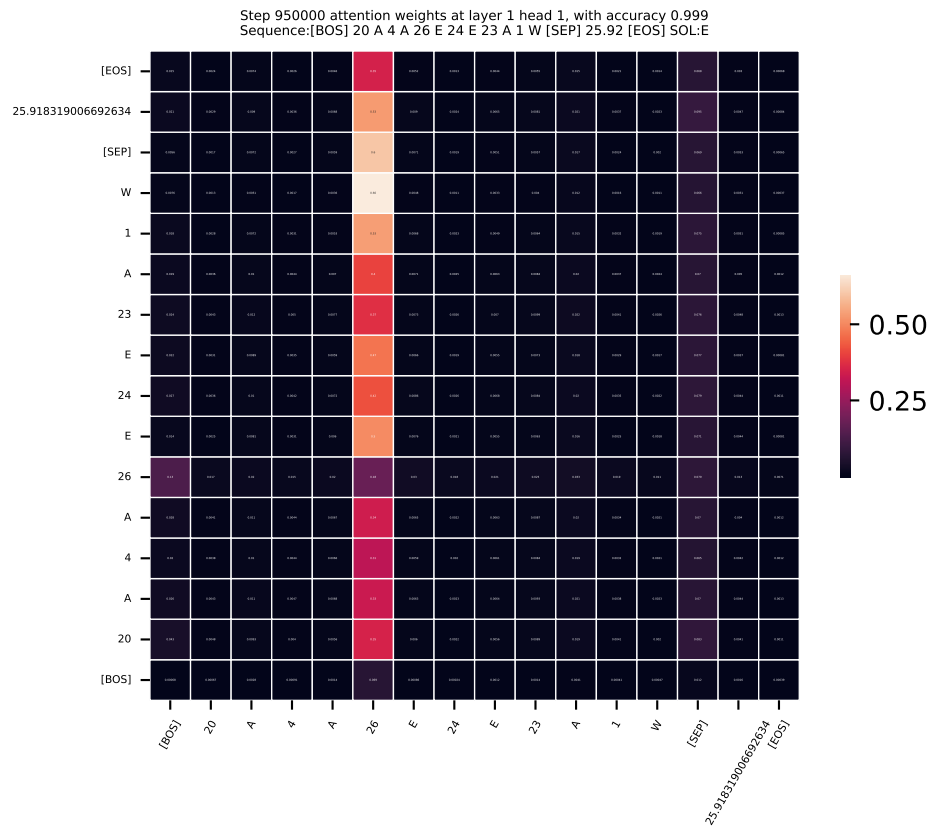

Step 950000 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

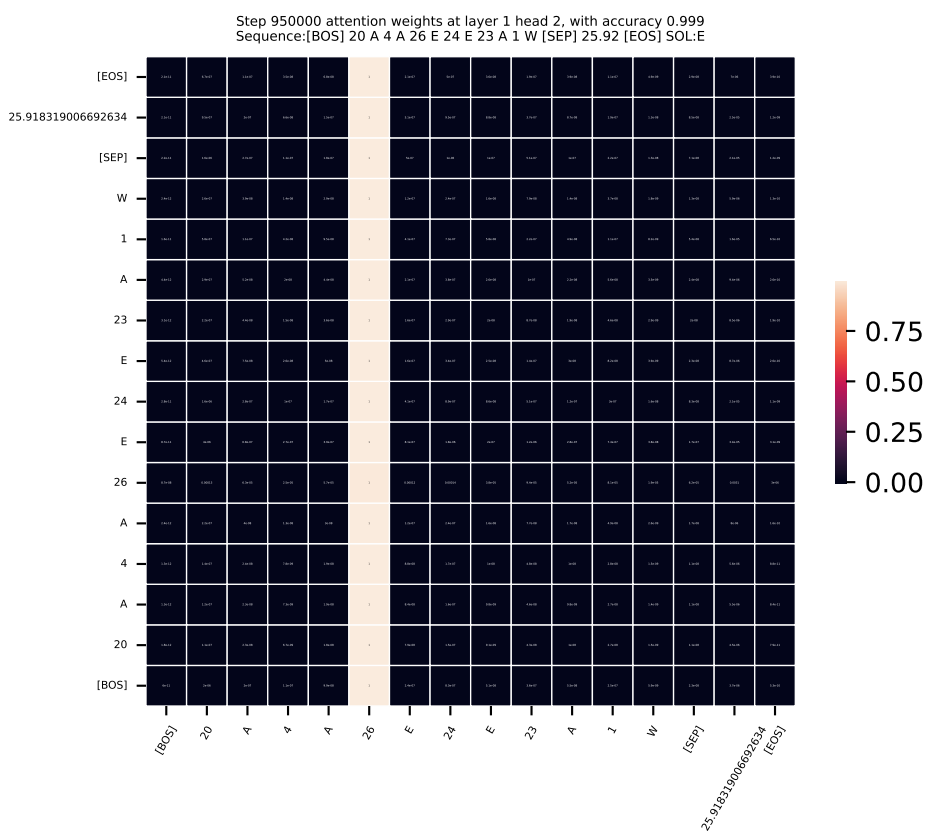

Step 950000 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

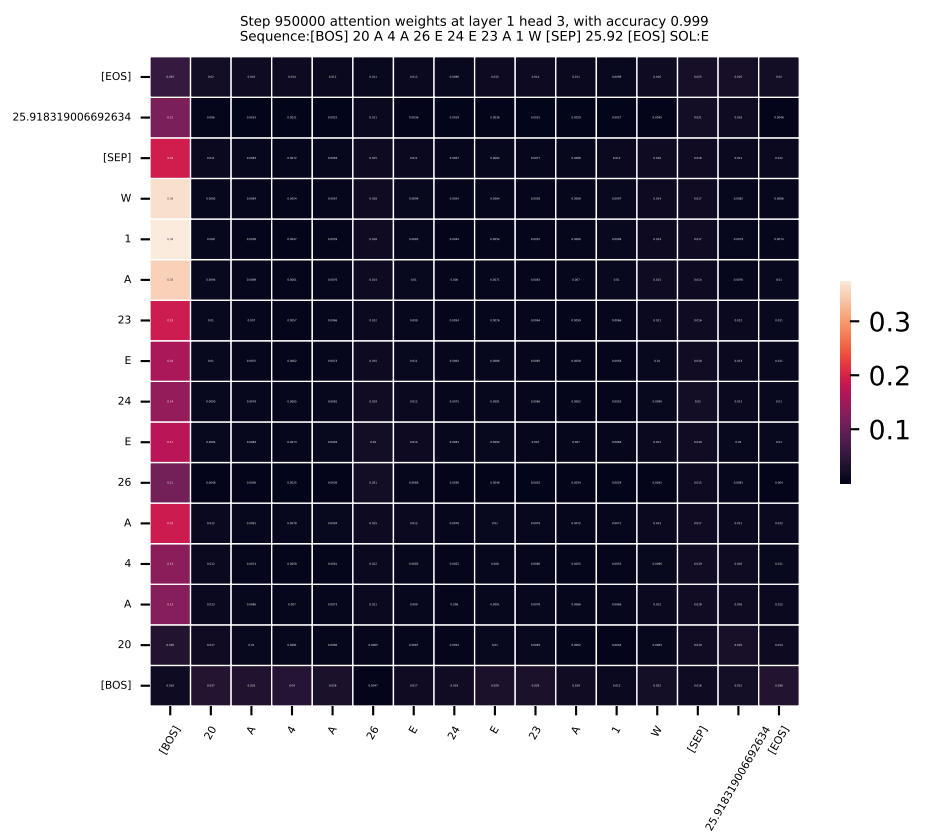

Step 950000 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

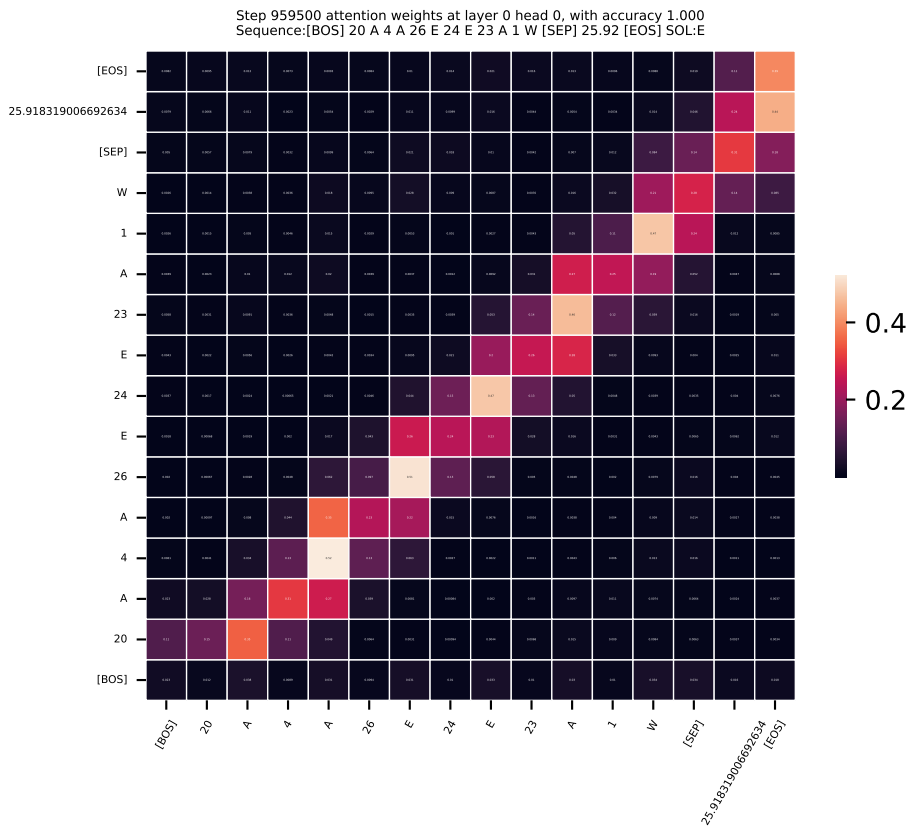

Step 950000 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

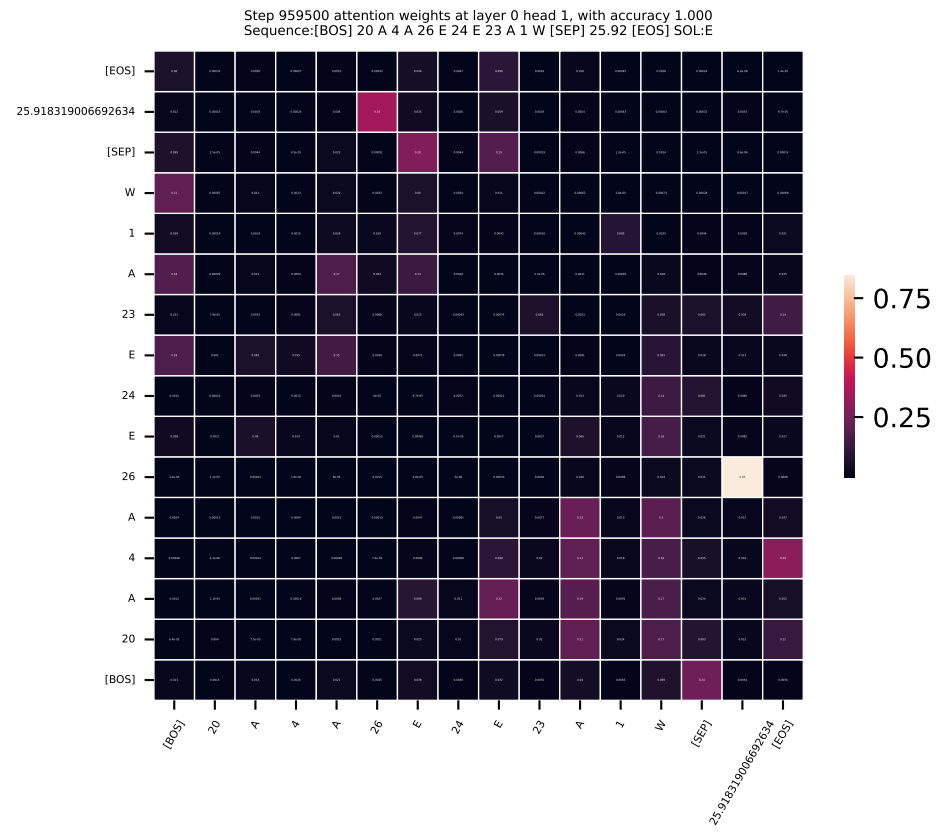

Step 950000 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

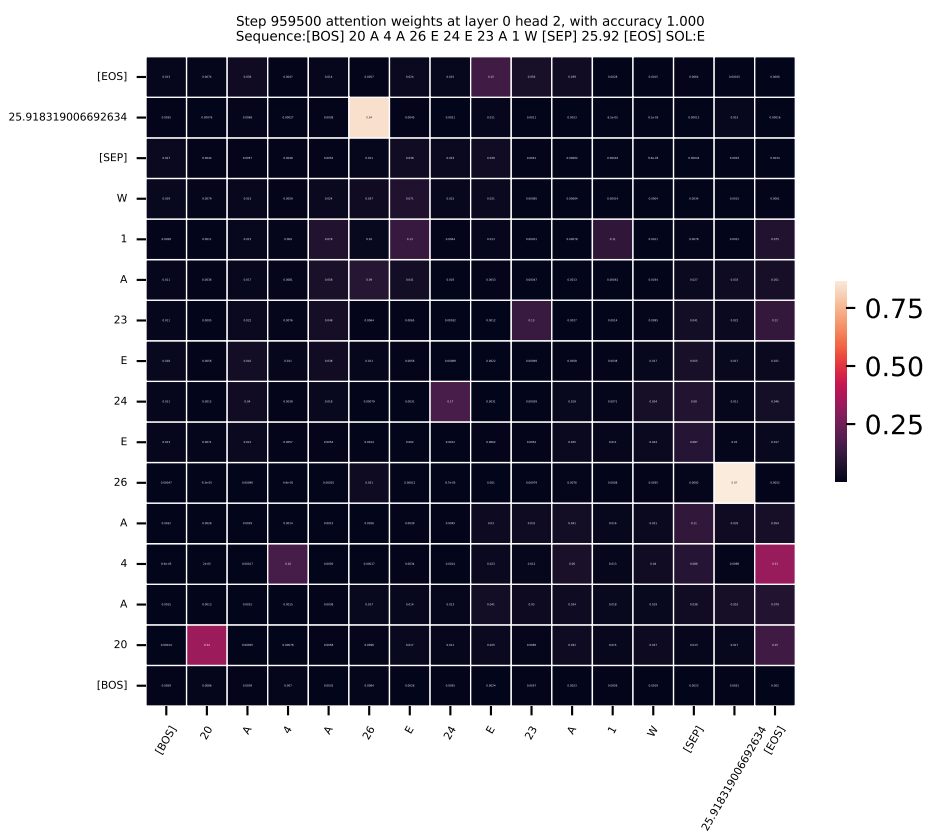

Step 950000 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

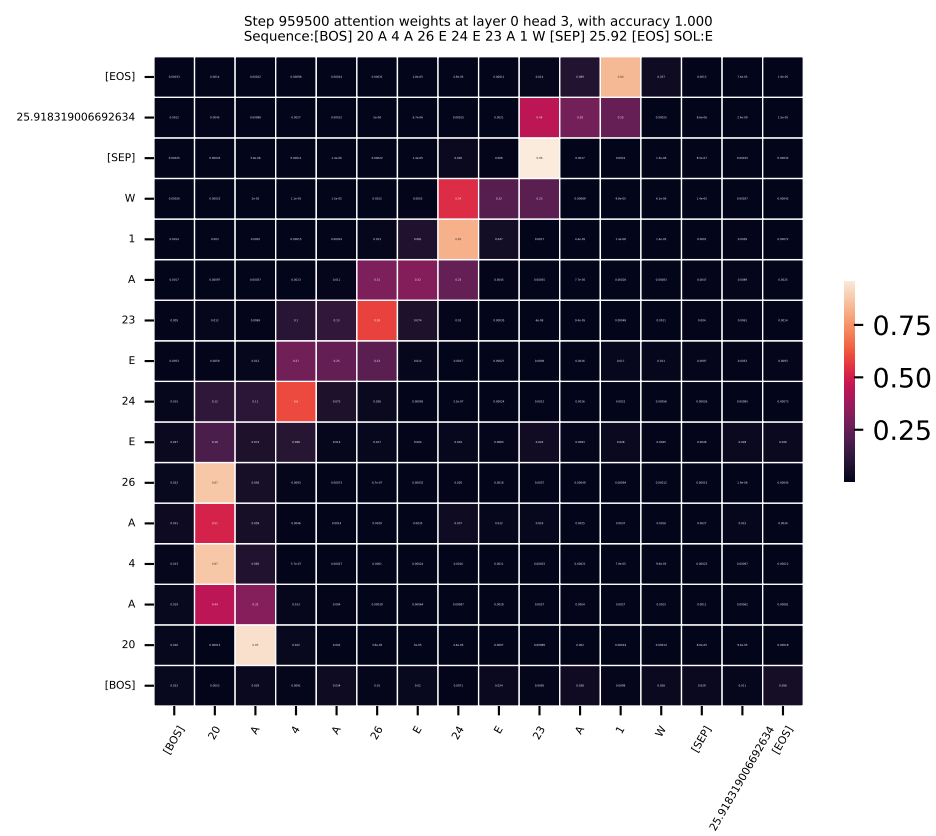

Step 950000 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

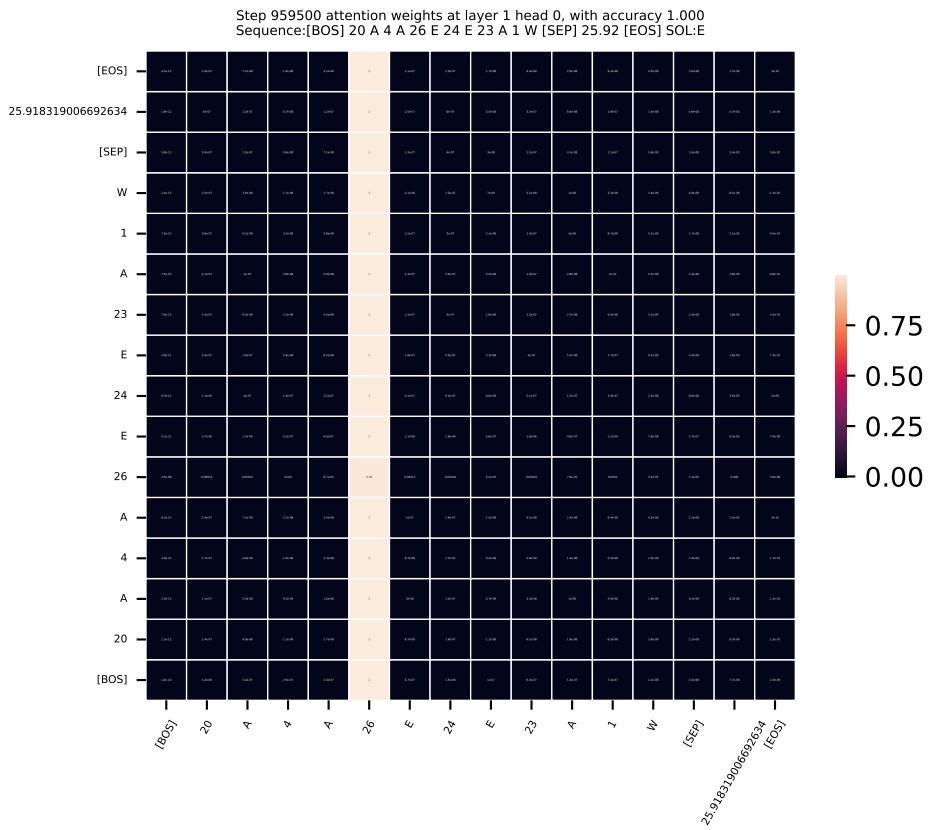

Step 950000 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

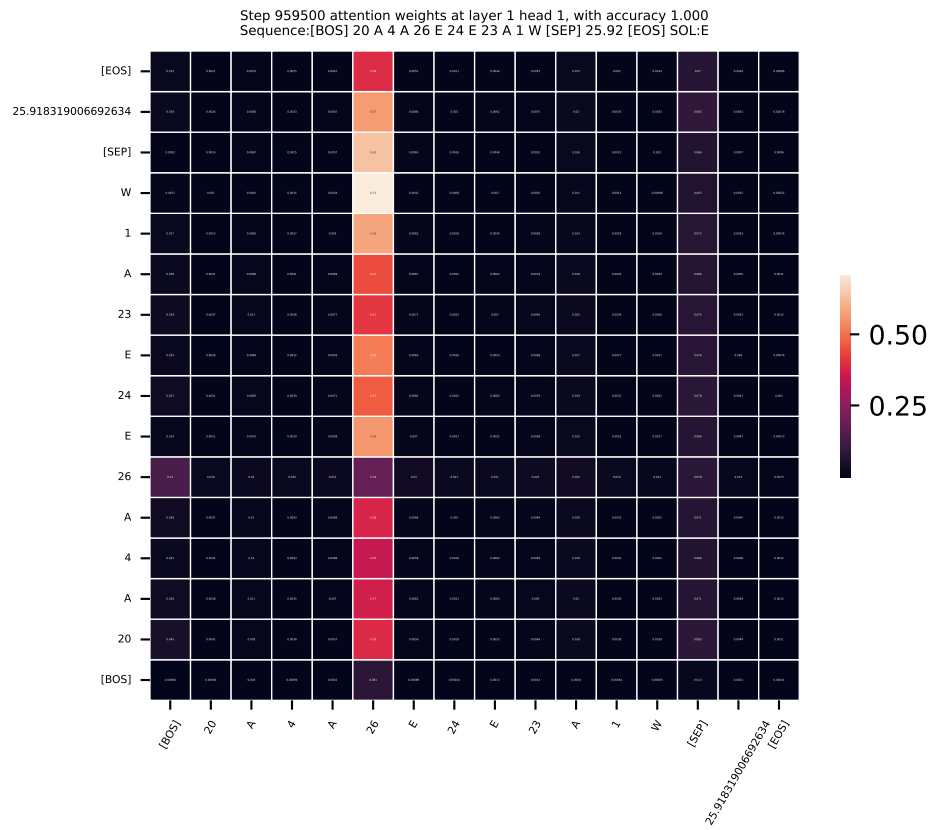

Step 959500 attention weights at layer 0 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

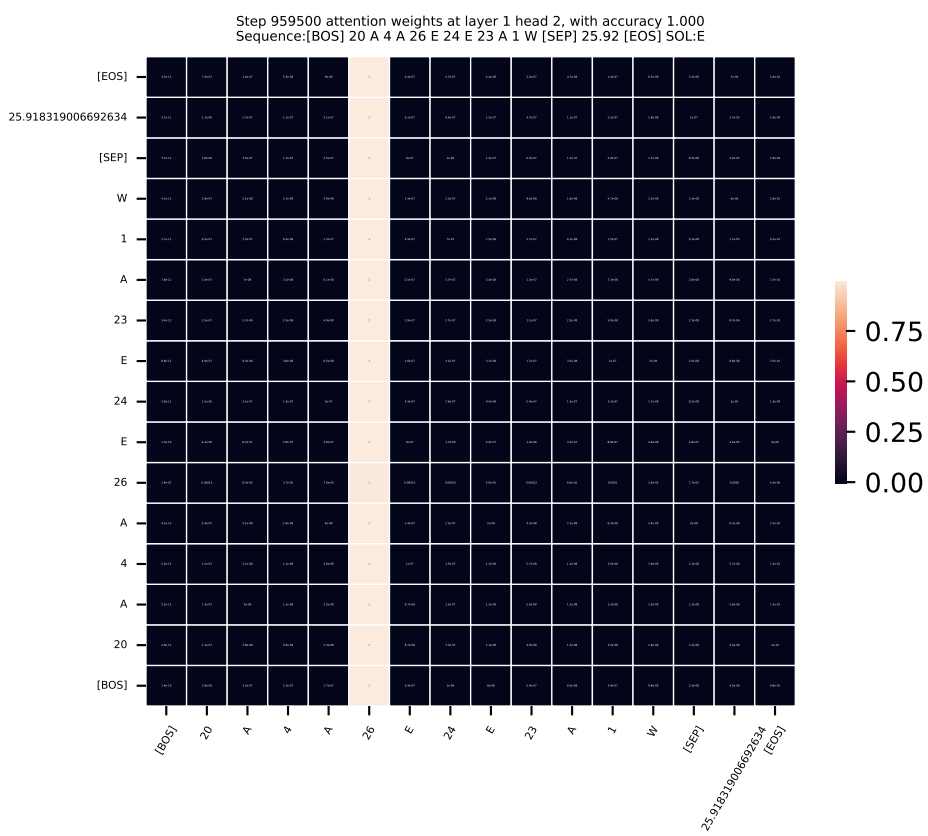

Step 959500 attention weights at layer 0 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

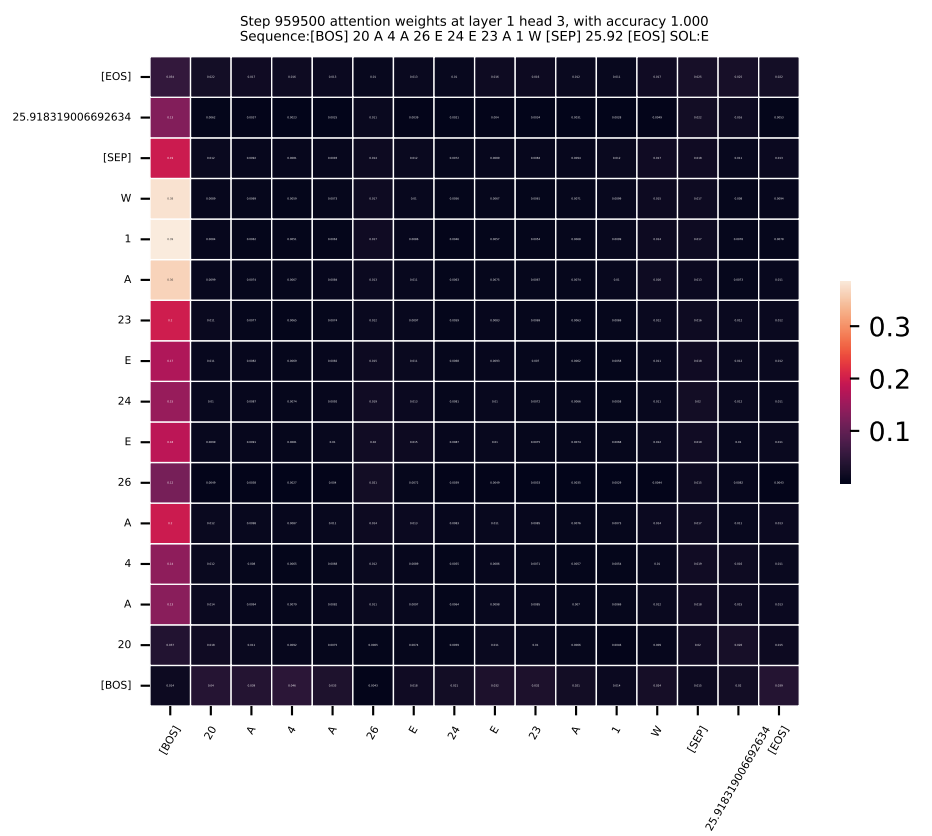

Step 959500 attention weights at layer 0 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

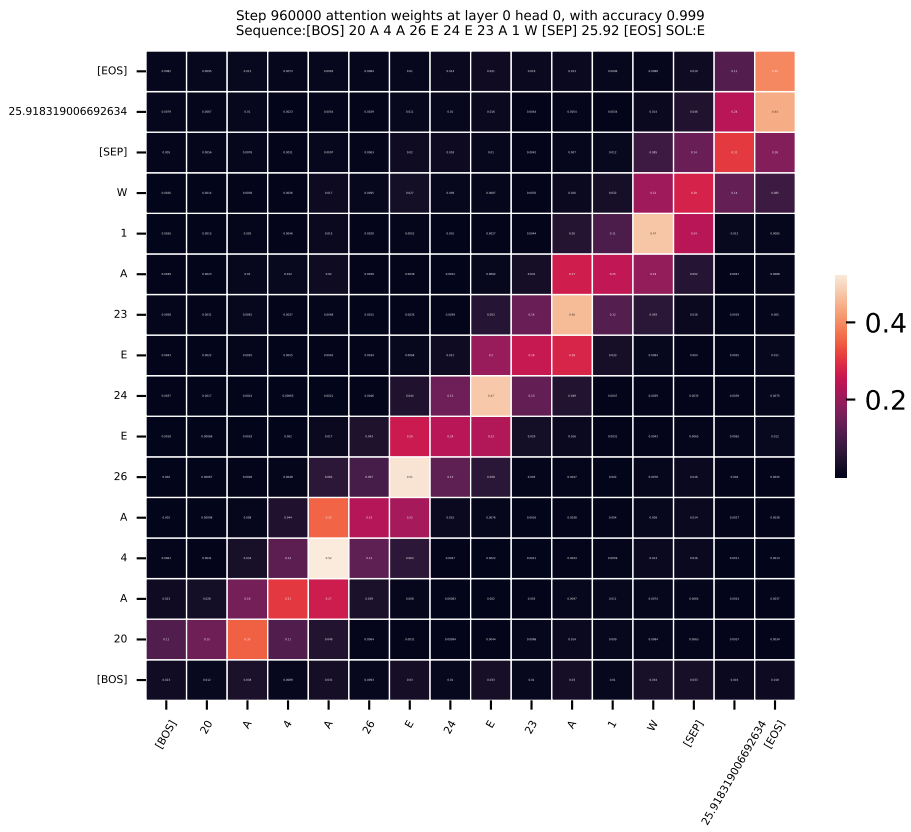

Step 959500 attention weights at layer 0 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

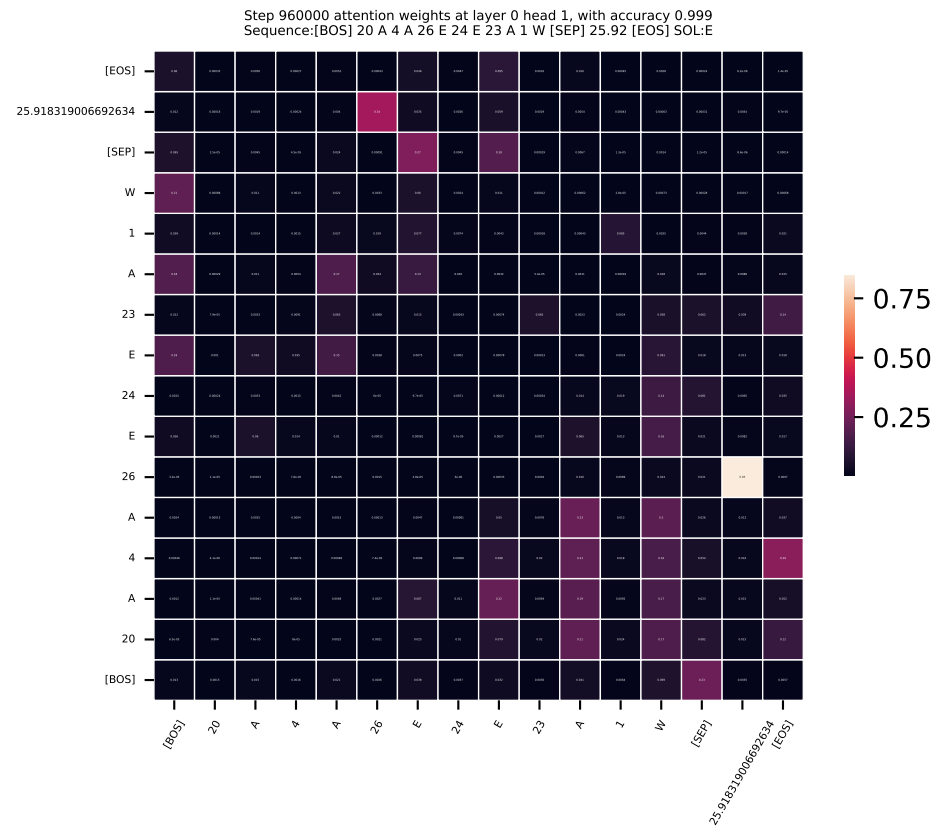

Step 959500 attention weights at layer 1 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

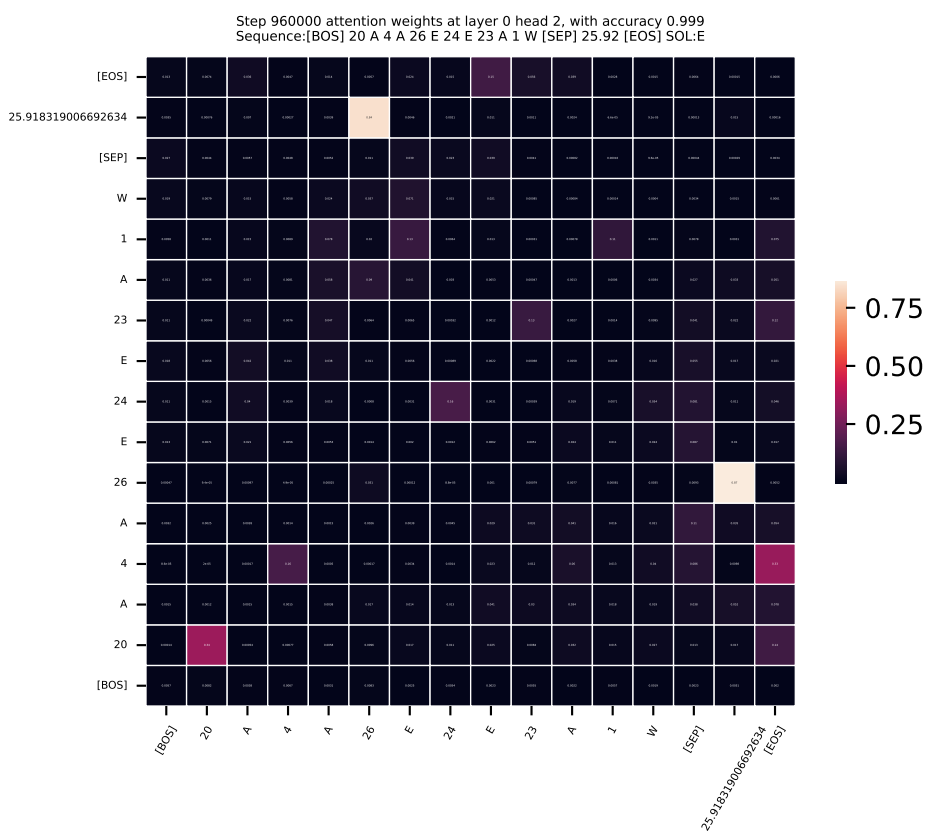

Step 959500 attention weights at layer 1 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

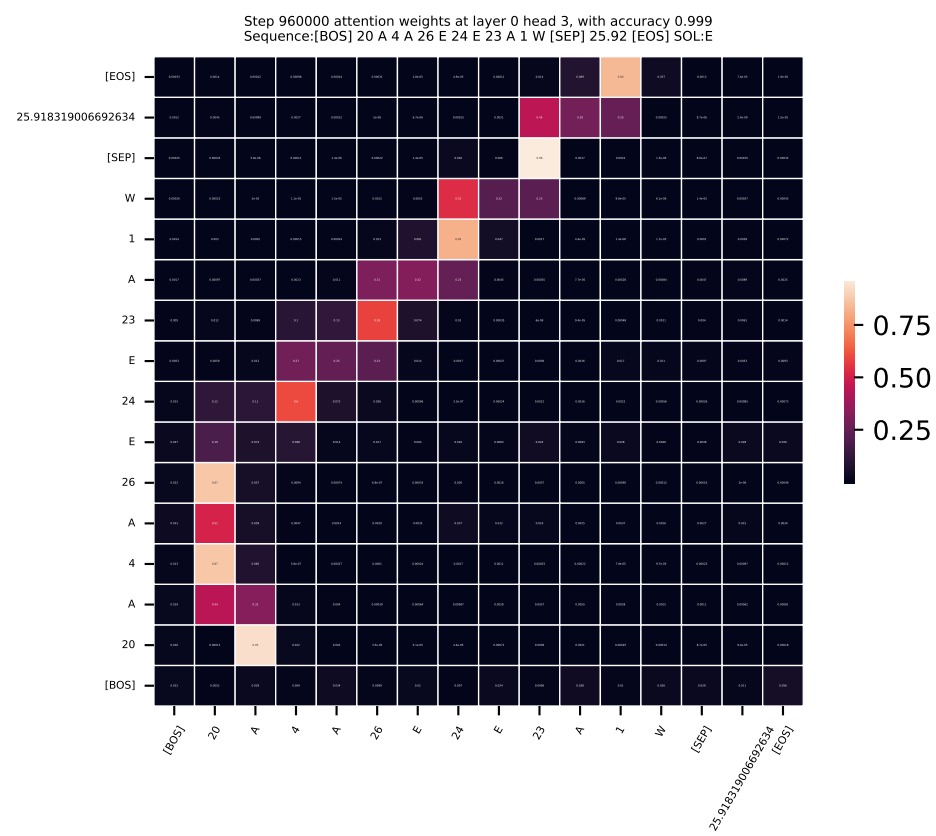

Step 959500 attention weights at layer 1 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

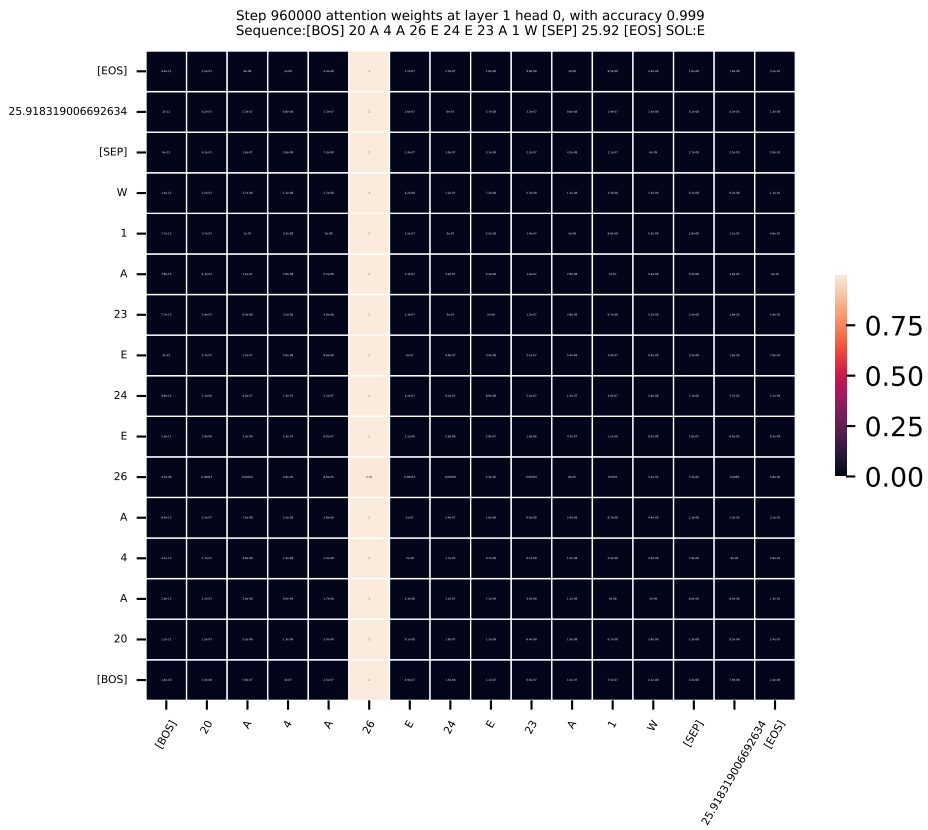

Step 959500 attention weights at layer 1 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

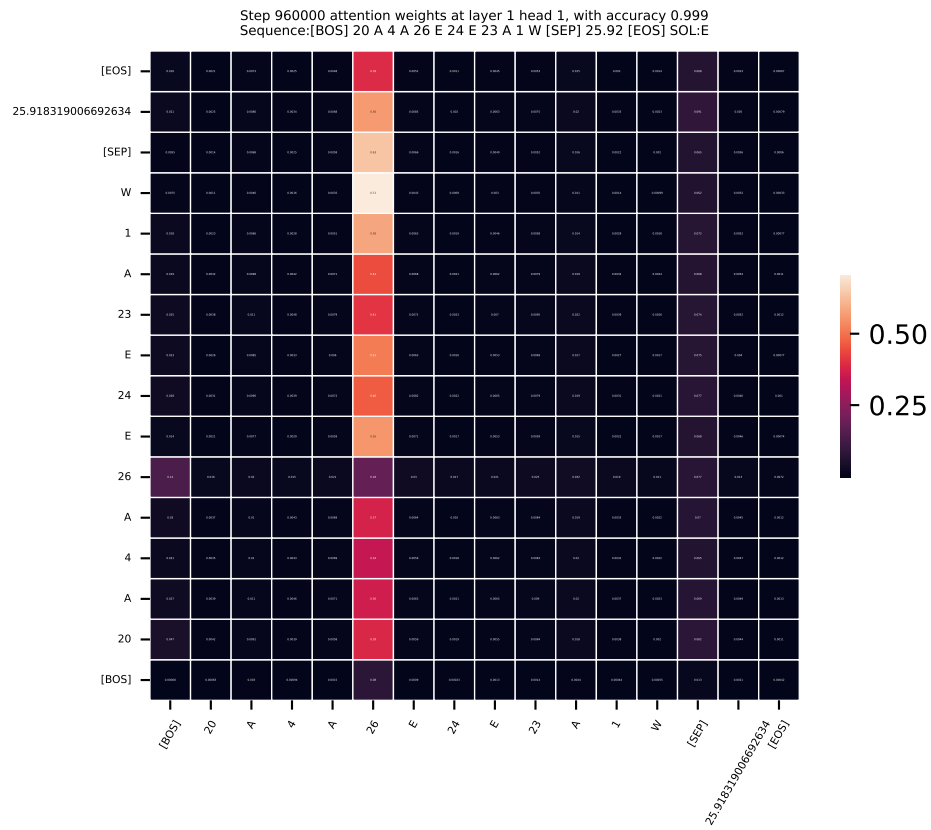

Step 960000 attention weights at layer 0 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

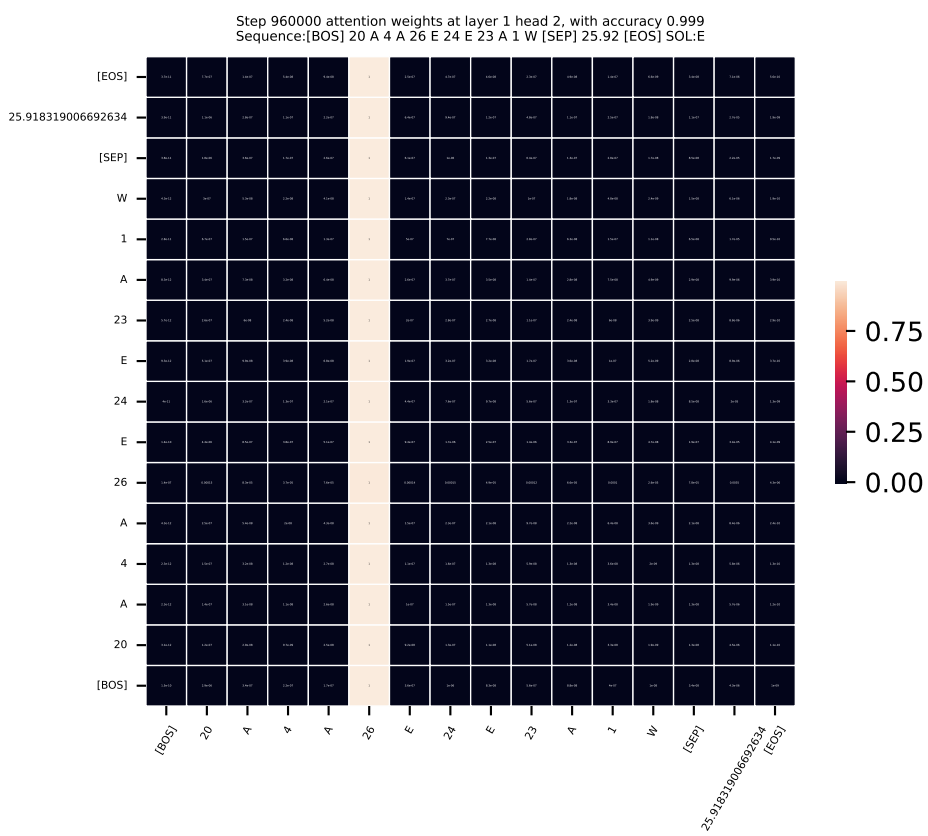

Step 960000 attention weights at layer 0 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

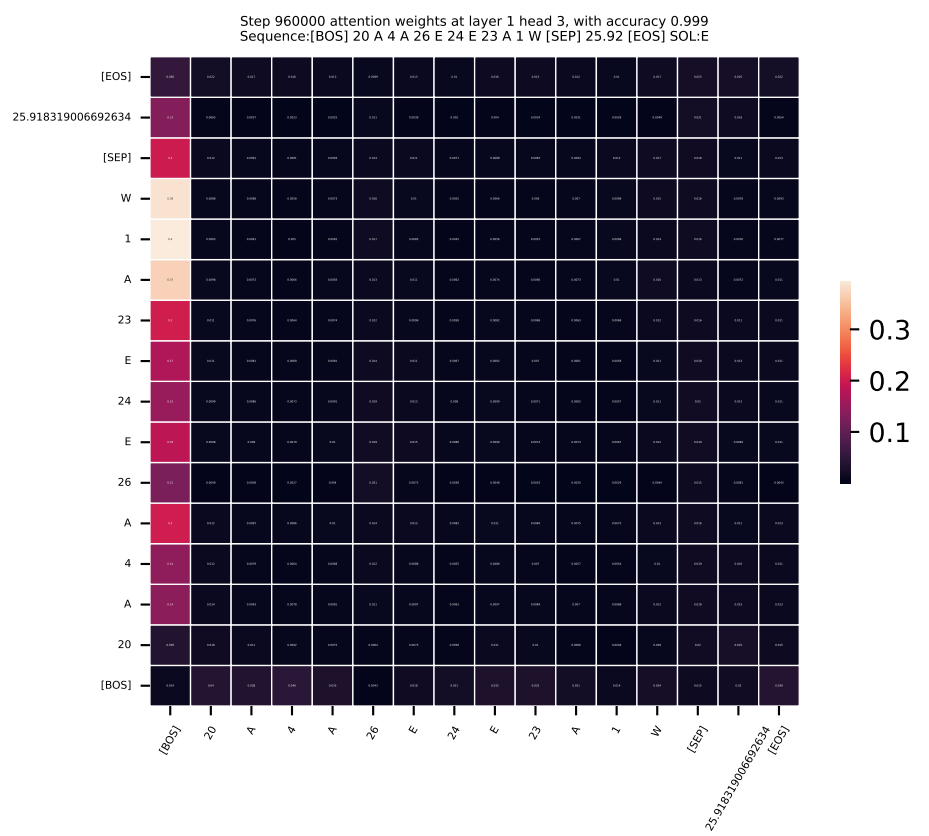

Step 960000 attention weights at layer 0 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

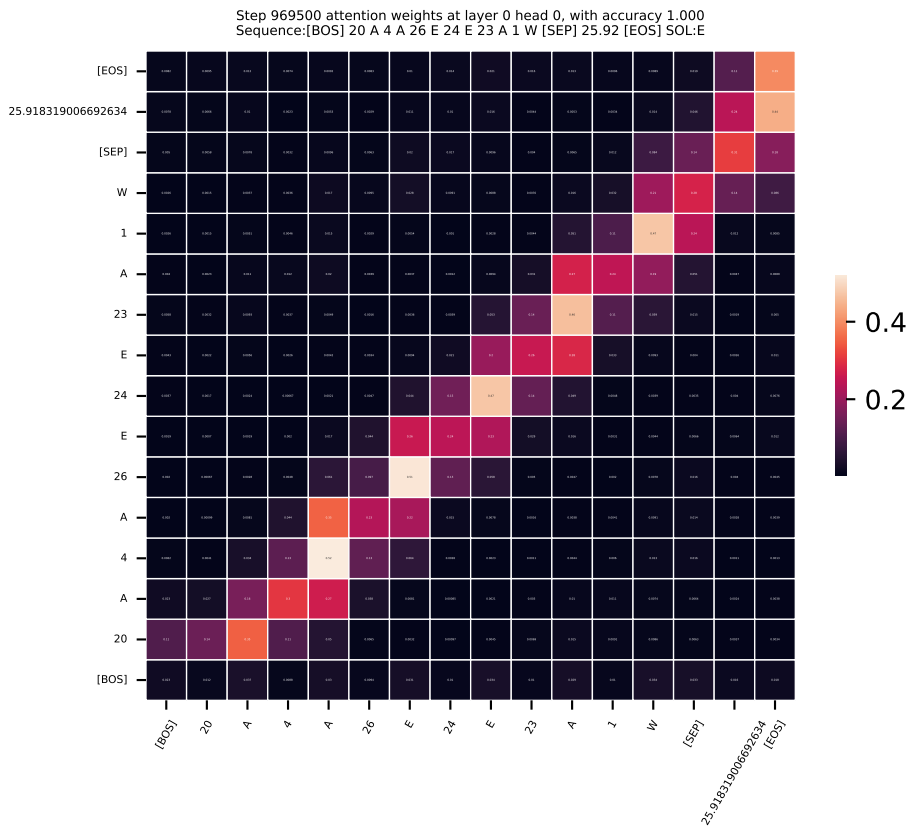

Step 960000 attention weights at layer 0 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

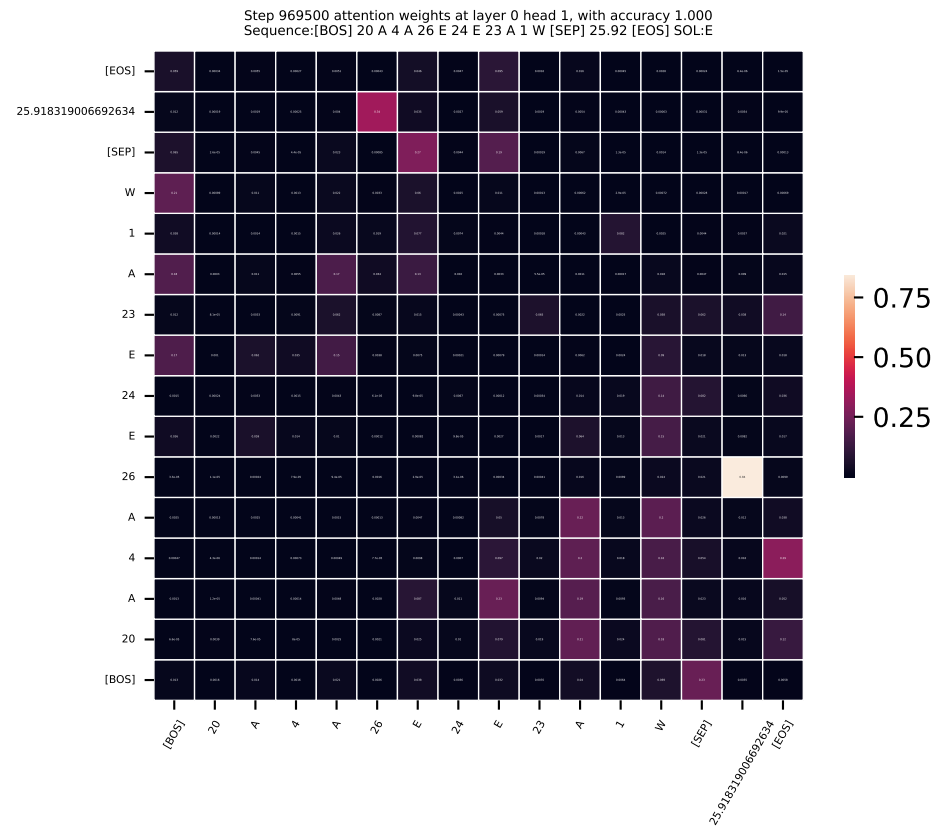

Step 960000 attention weights at layer 1 head 0, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

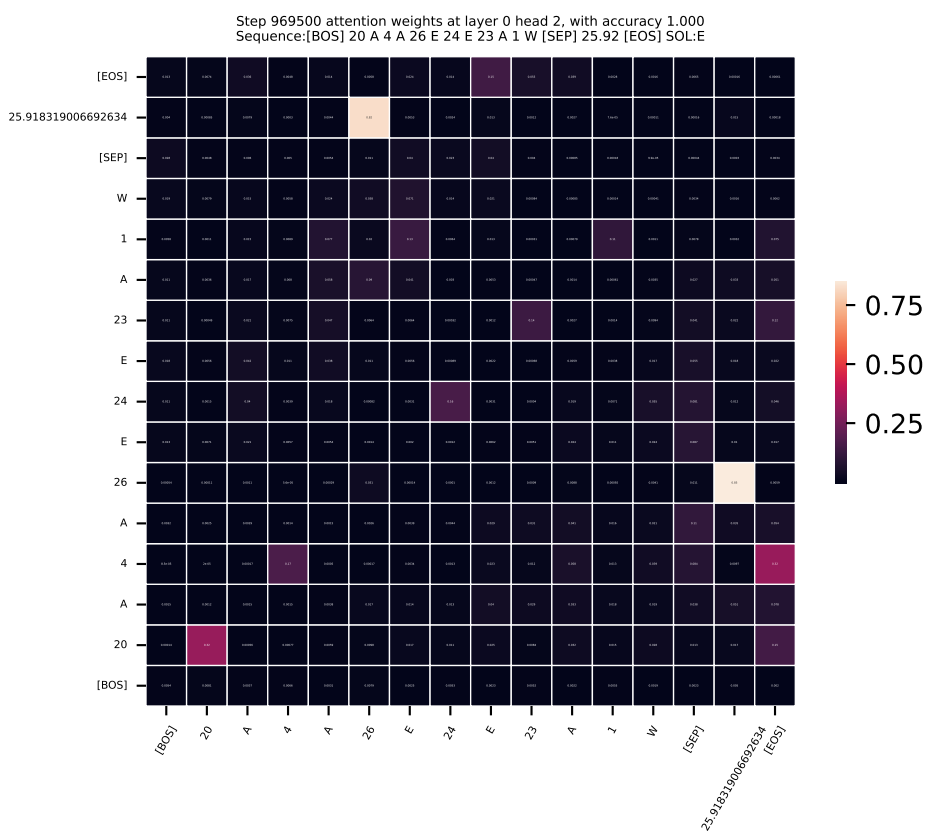

Step 960000 attention weights at layer 1 head 1, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

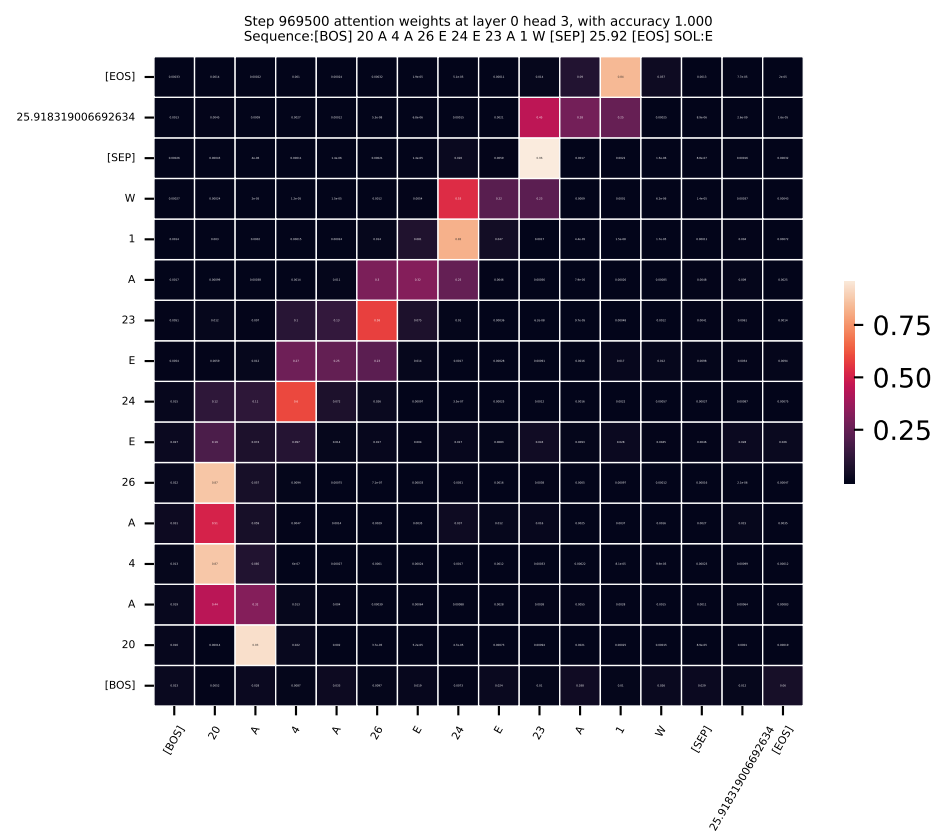

Step 960000 attention weights at layer 1 head 2, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

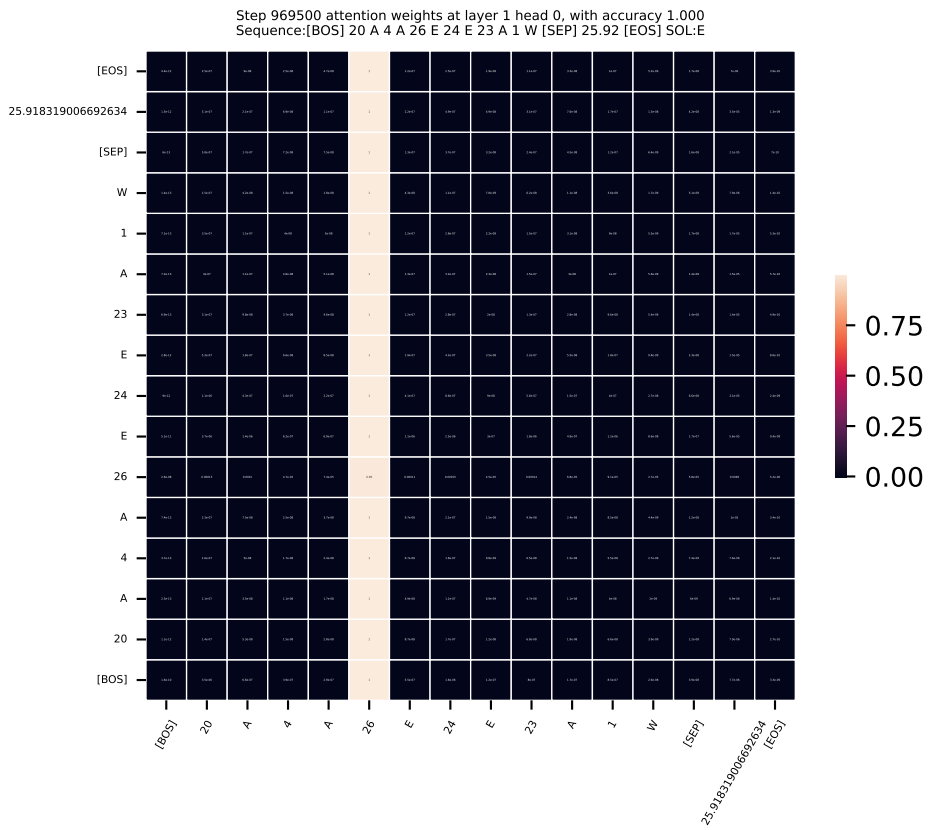

Step 960000 attention weights at layer 1 head 3, with accuracy 0.999
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

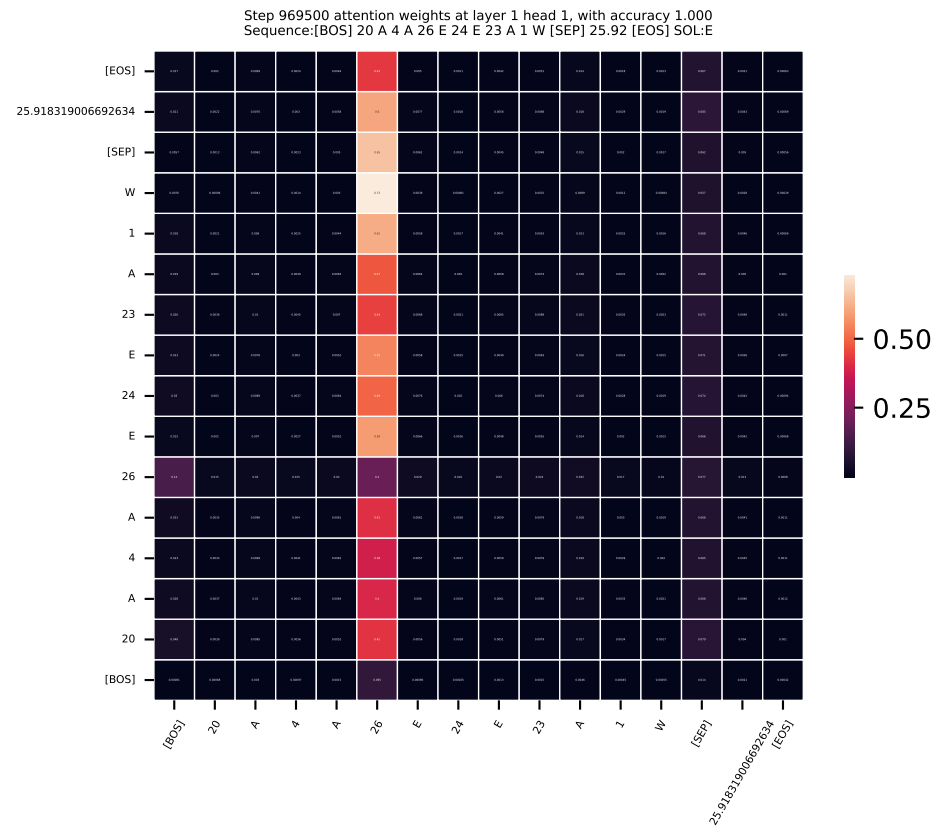

Step 969500 attention weights at layer 0 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

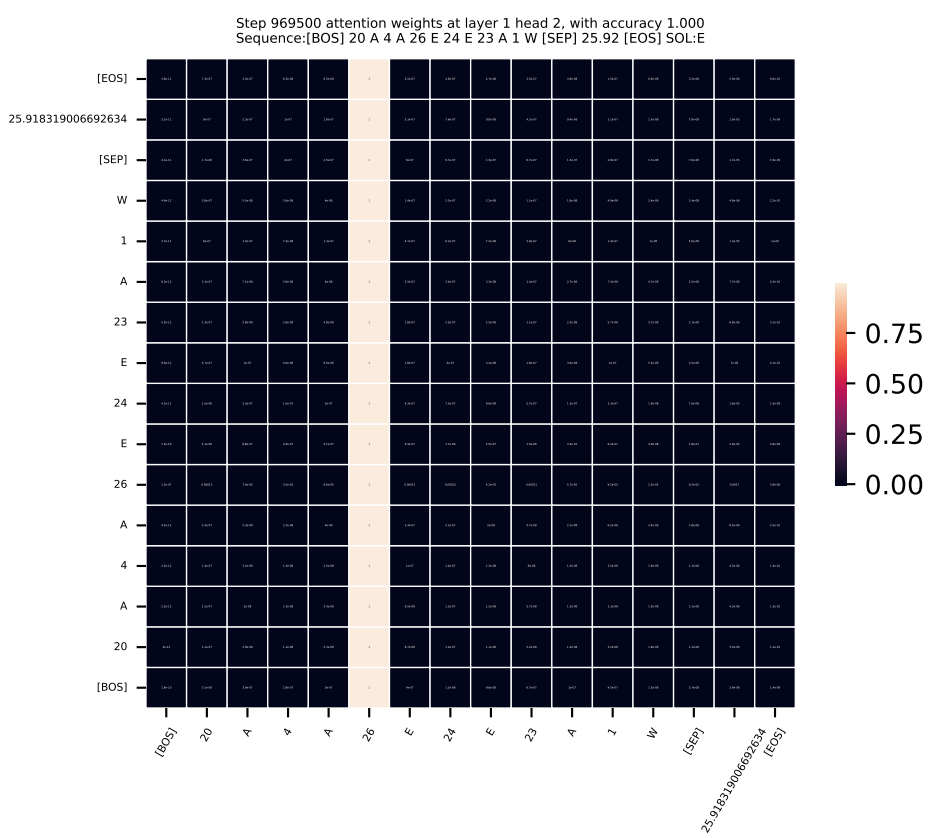

Step 969500 attention weights at layer 0 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

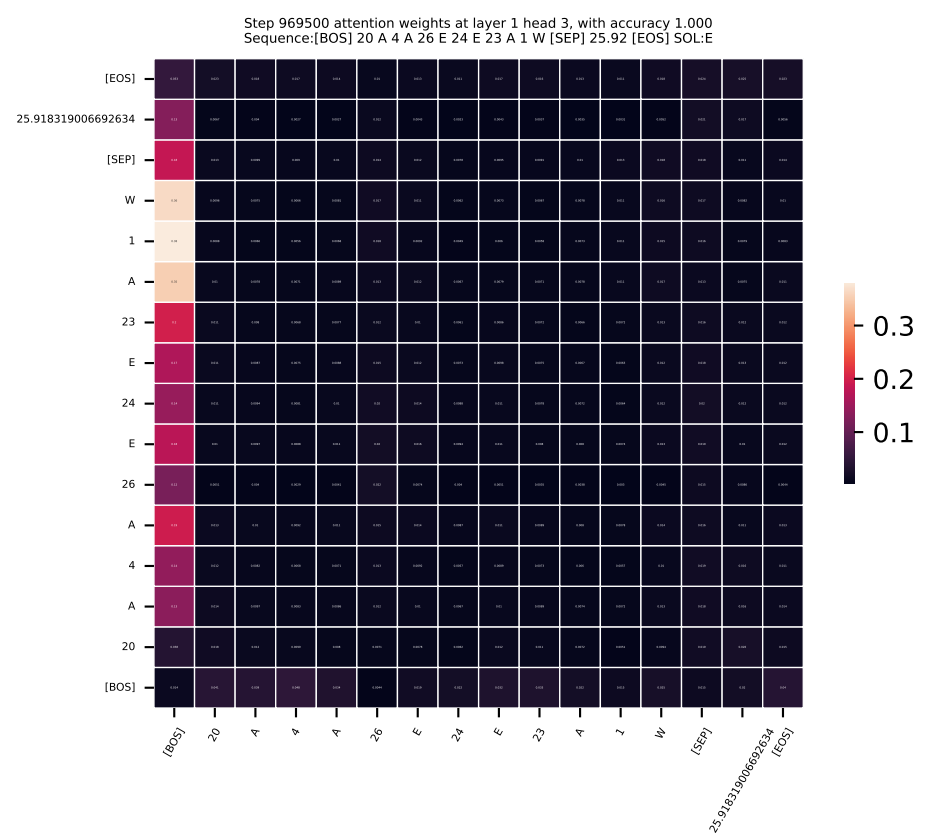

Step 969500 attention weights at layer 0 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

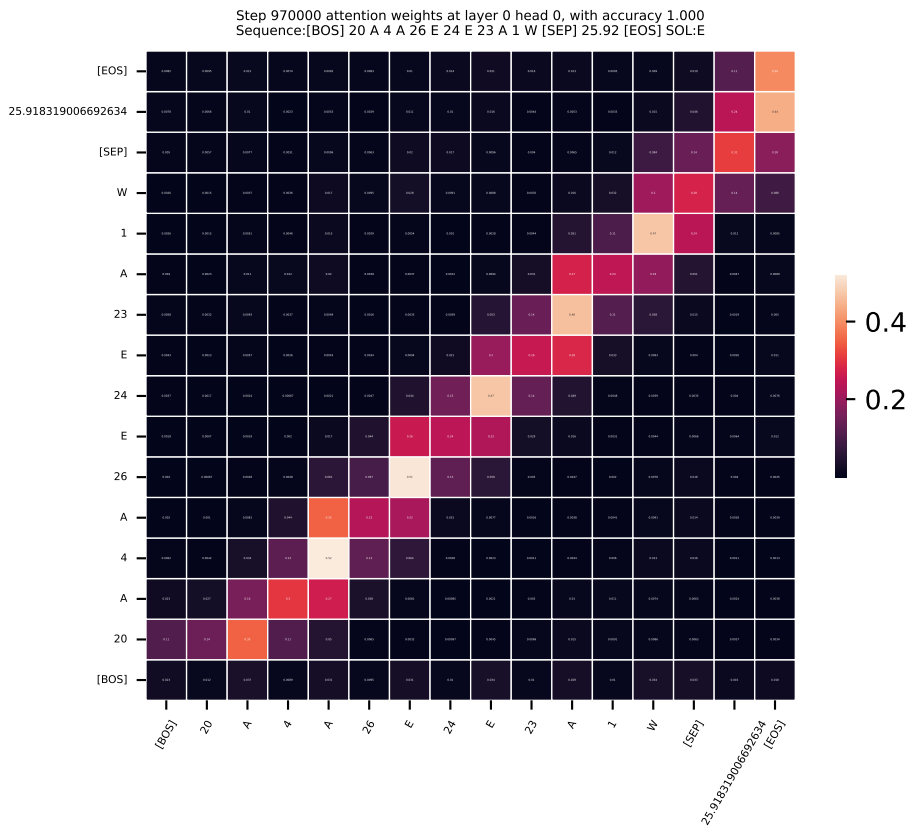

Step 969500 attention weights at layer 0 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

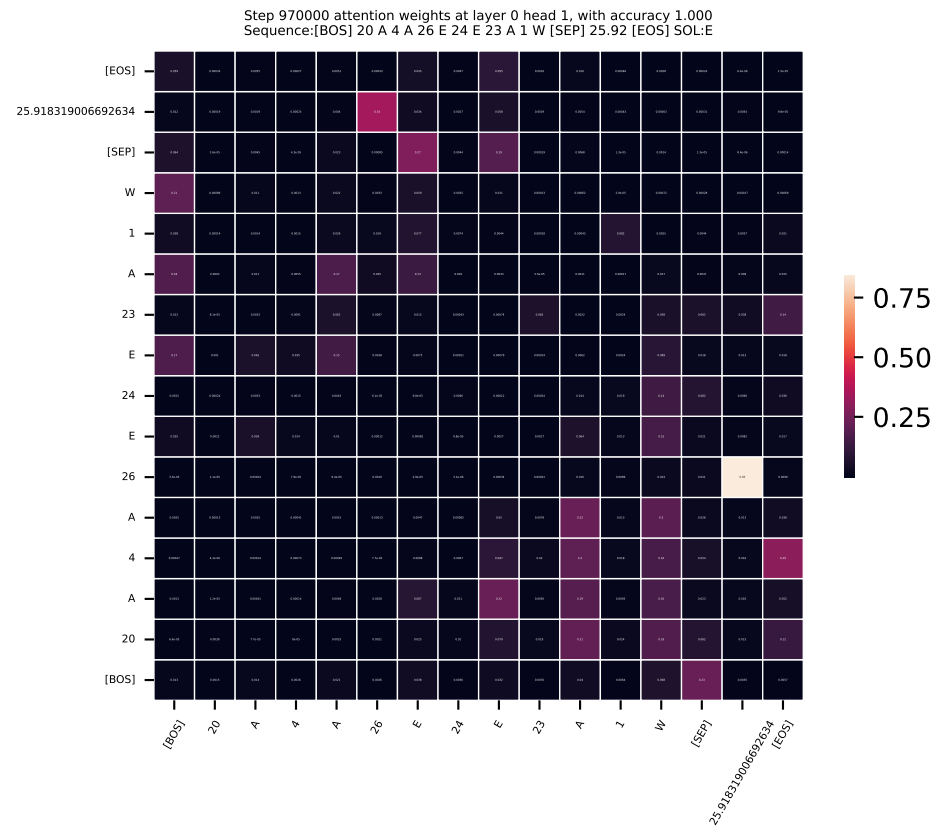

Step 969500 attention weights at layer 1 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

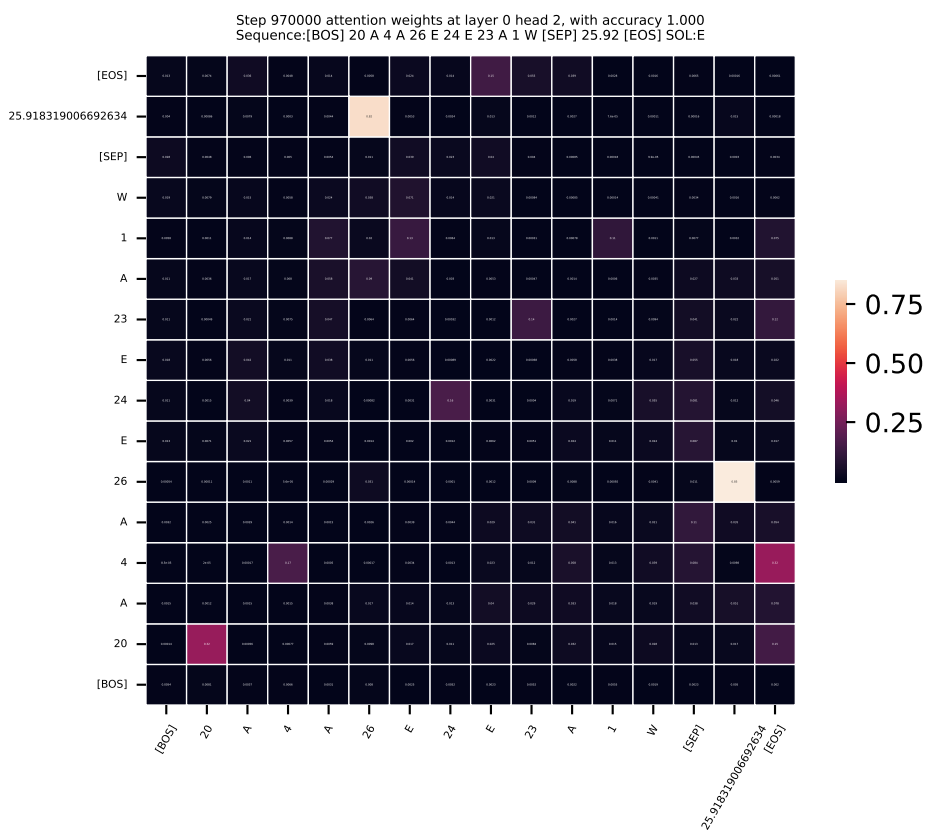

Step 969500 attention weights at layer 1 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

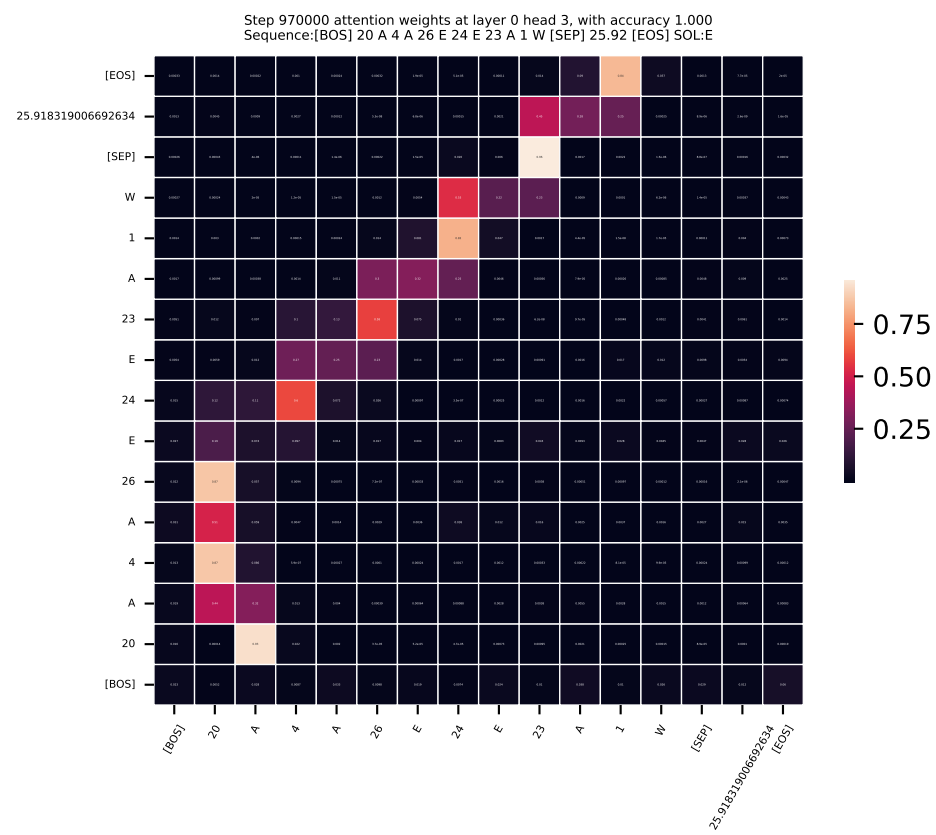

Step 969500 attention weights at layer 1 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

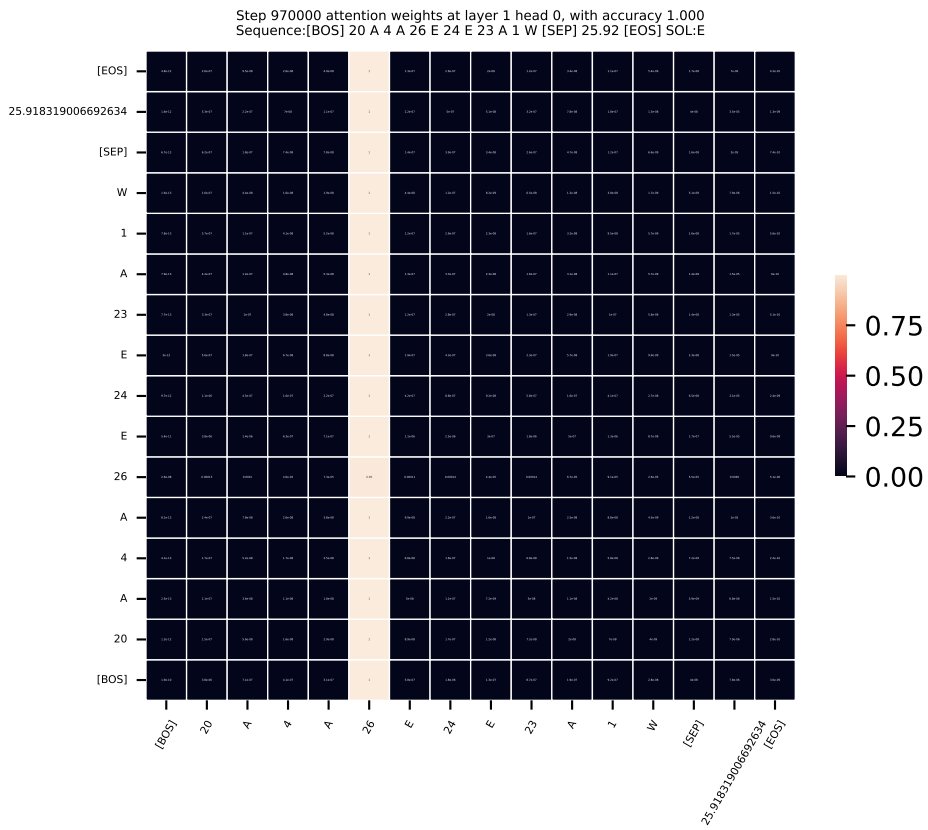

Step 969500 attention weights at layer 1 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

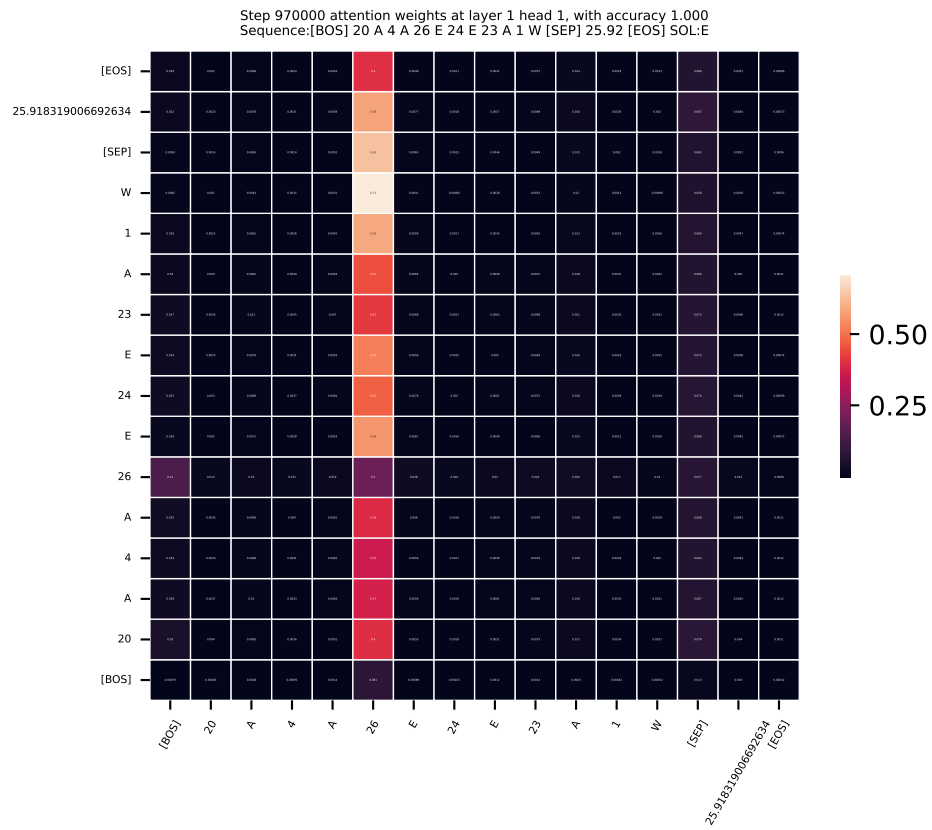

Step 970000 attention weights at layer 0 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

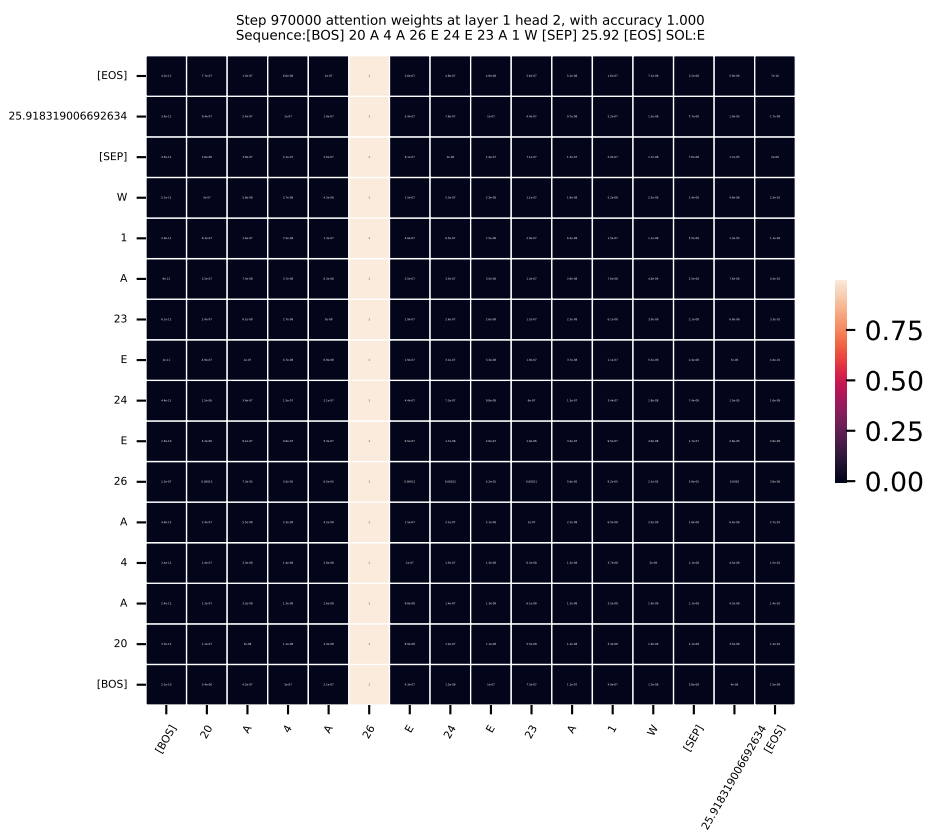

Step 970000 attention weights at layer 0 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

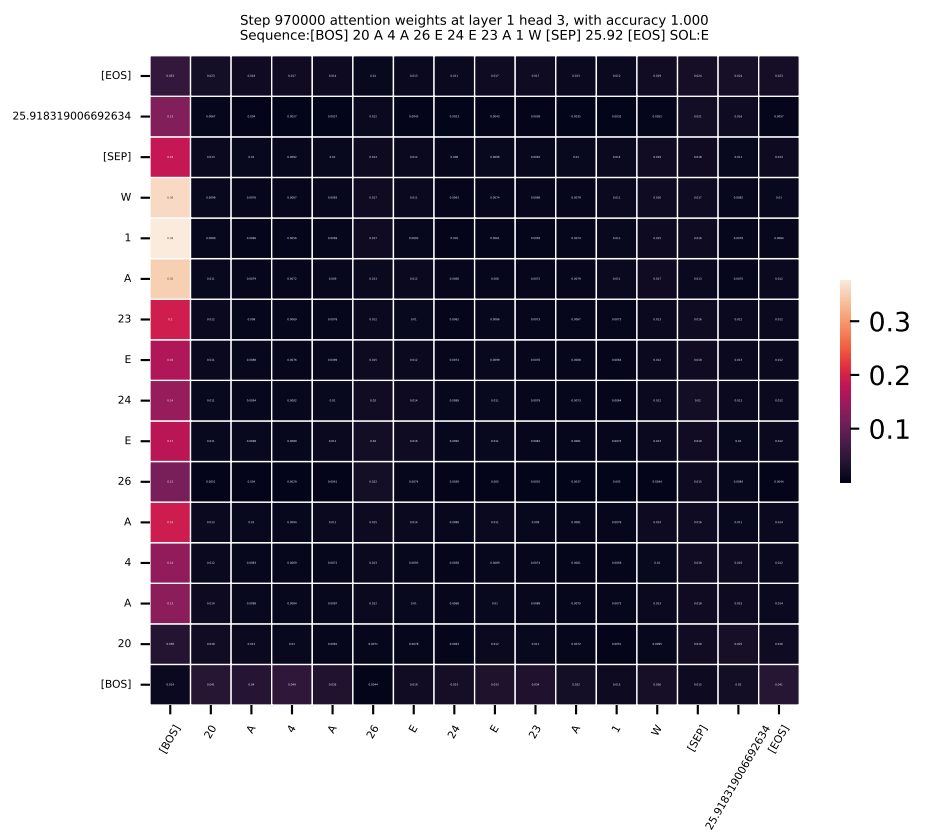

Step 970000 attention weights at layer 0 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

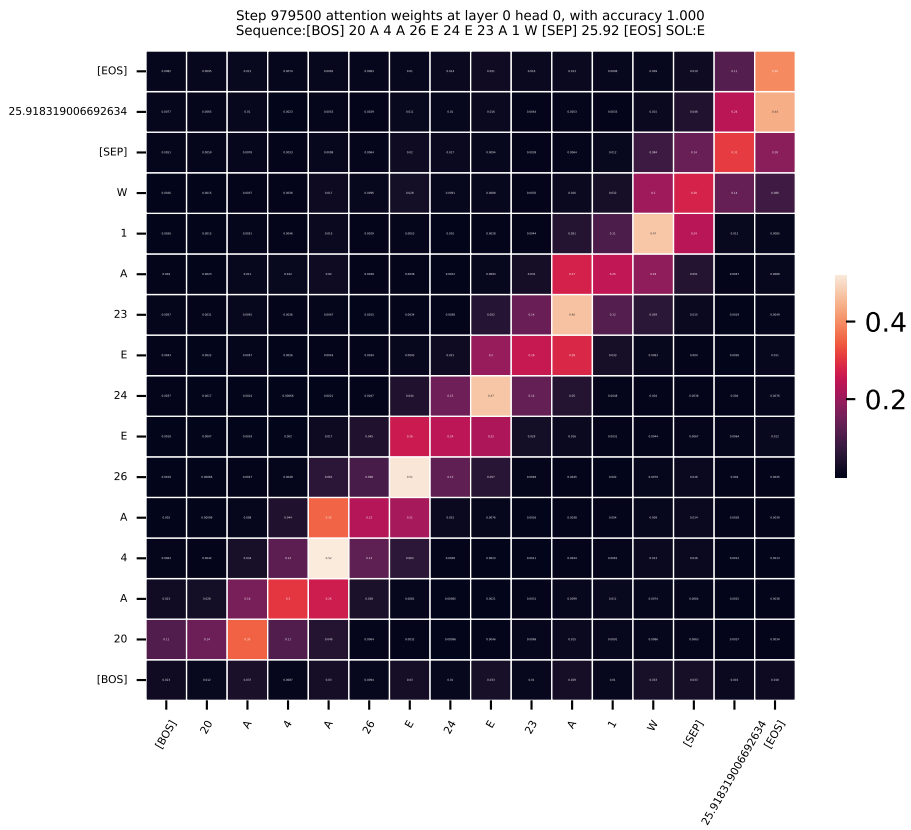

Step 970000 attention weights at layer 0 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

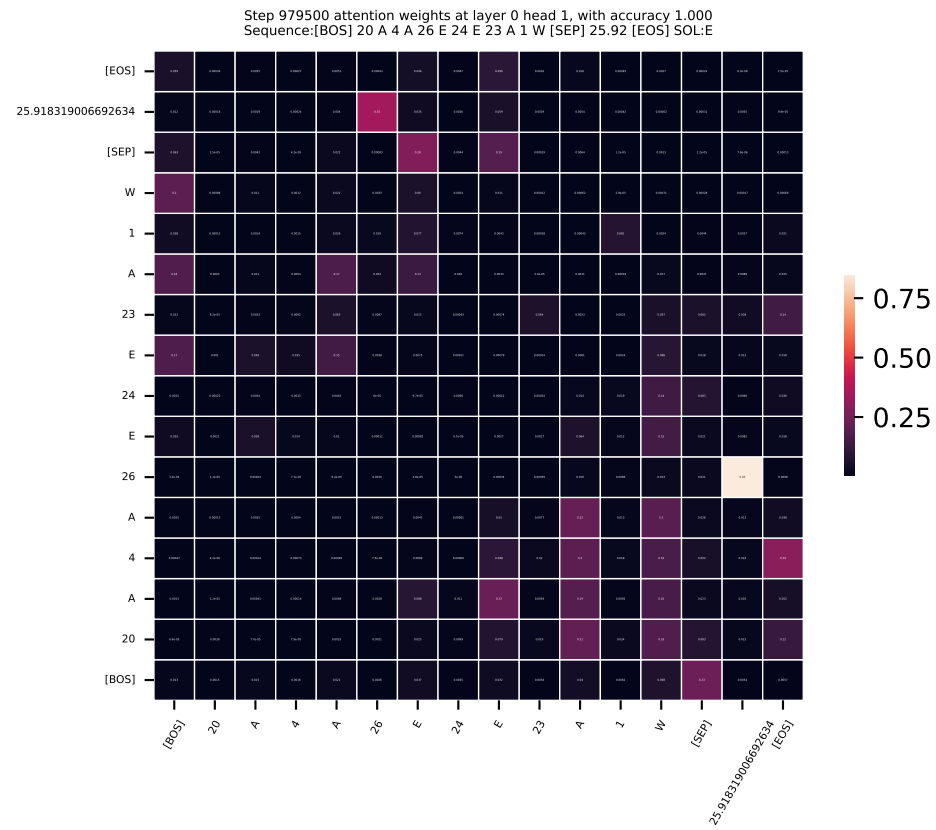

Step 970000 attention weights at layer 1 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

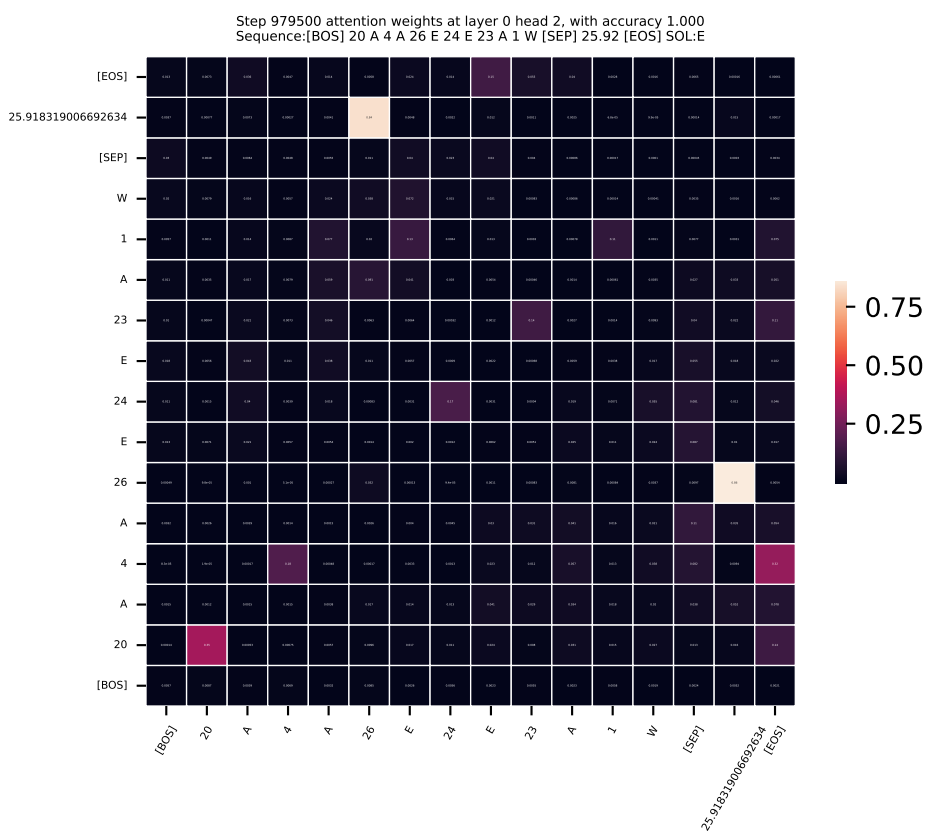

Step 970000 attention weights at layer 1 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

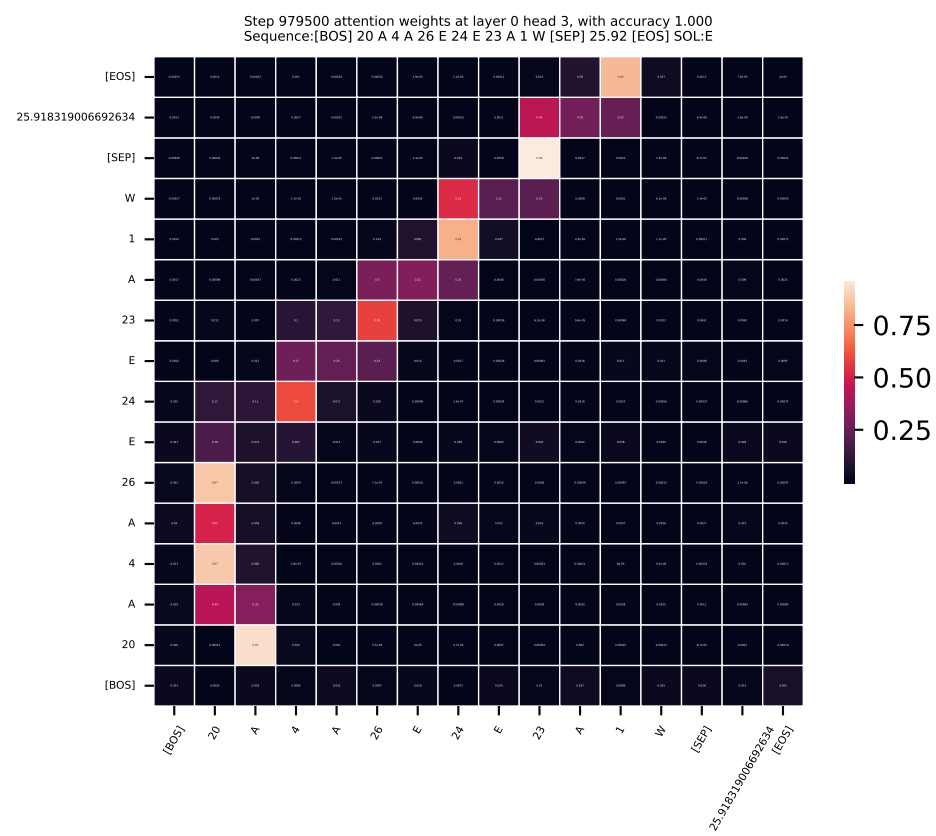

Step 970000 attention weights at layer 1 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

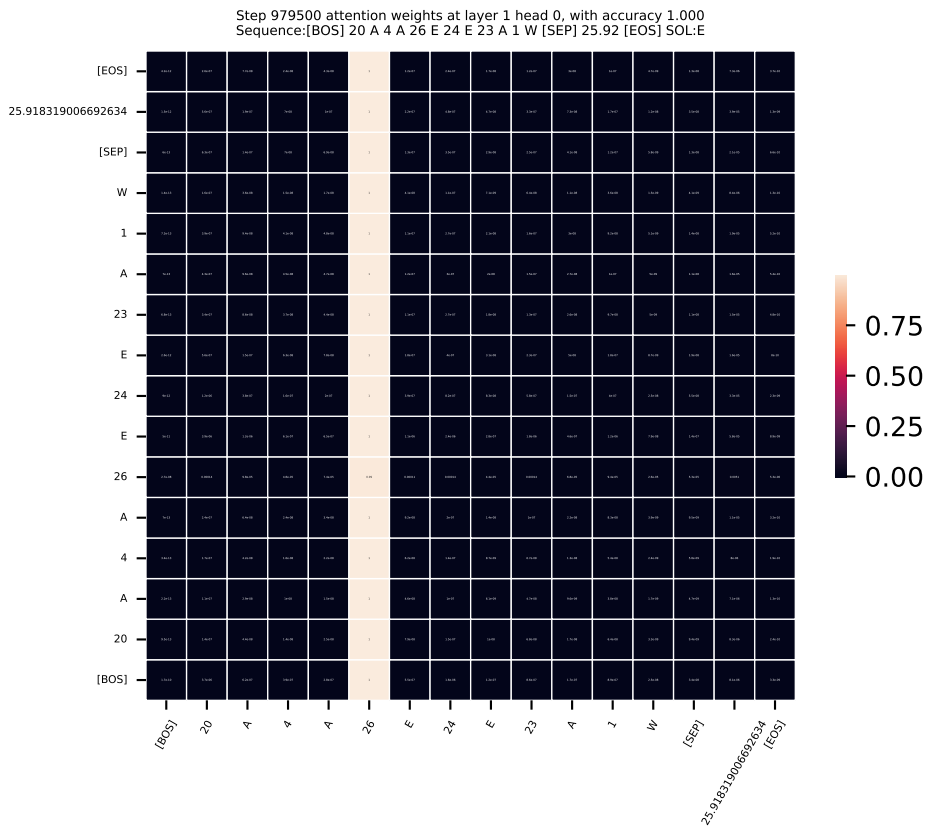

Step 970000 attention weights at layer 1 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

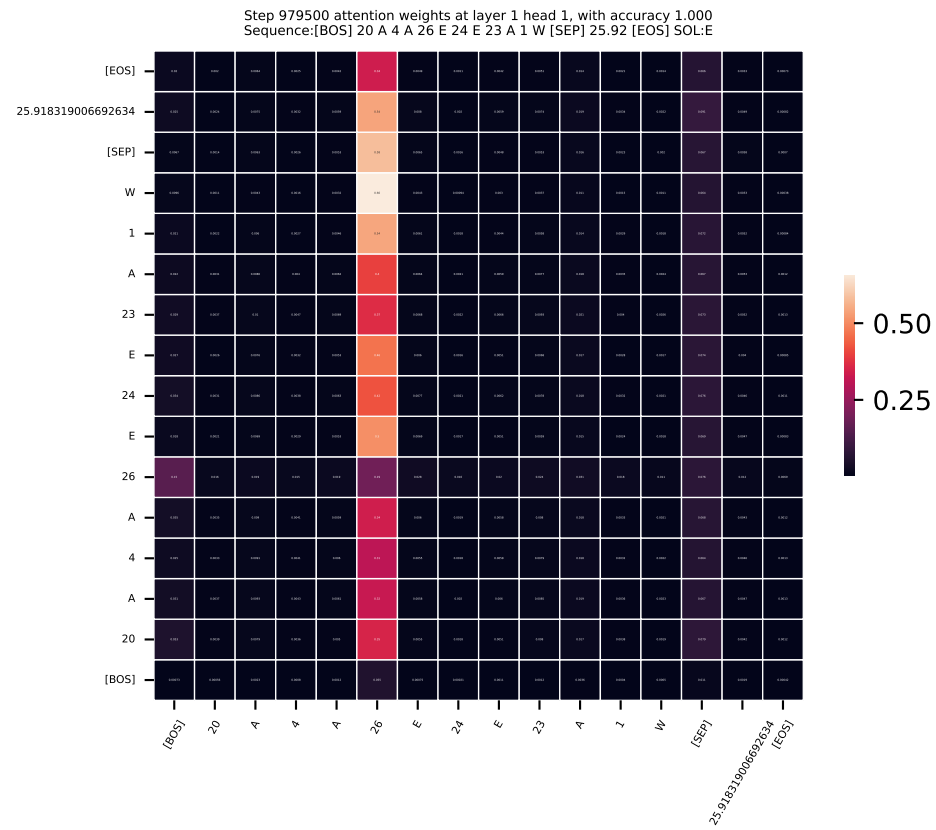

Step 979500 attention weights at layer 0 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

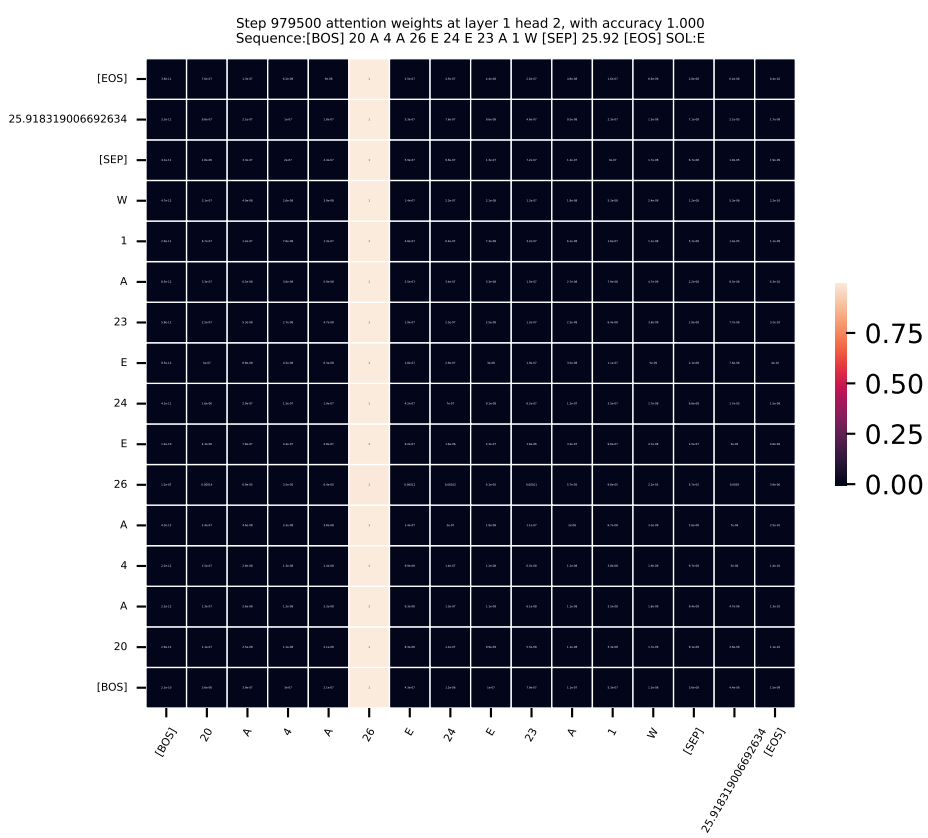

Step 979500 attention weights at layer 0 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

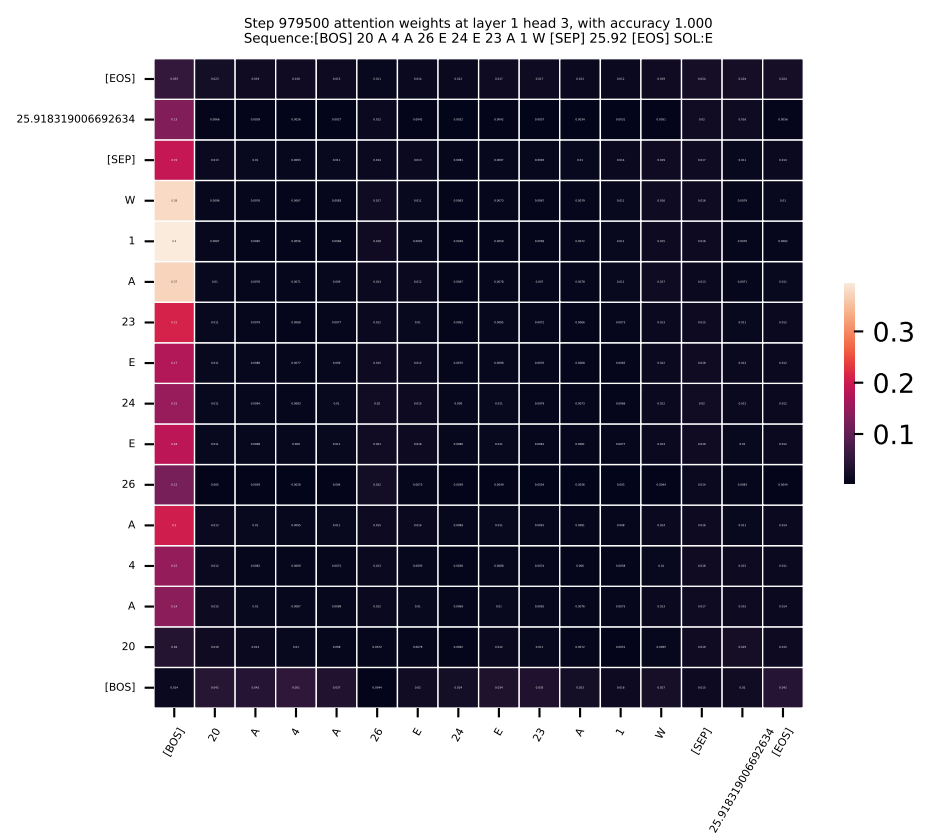

Step 979500 attention weights at layer 0 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

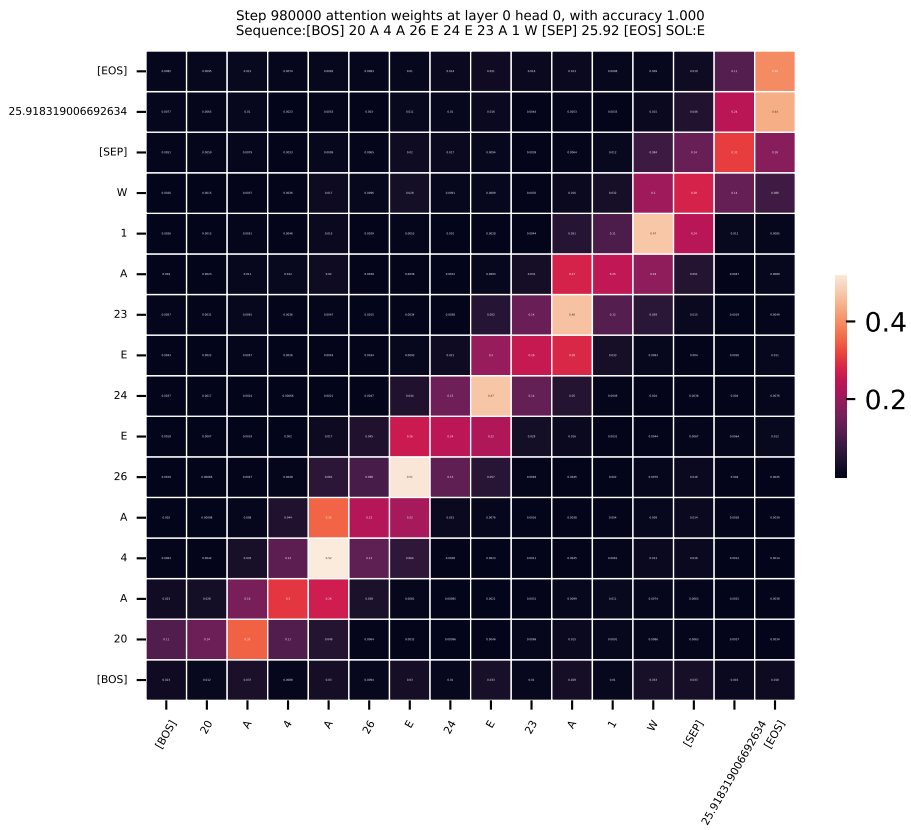

Step 979500 attention weights at layer 0 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

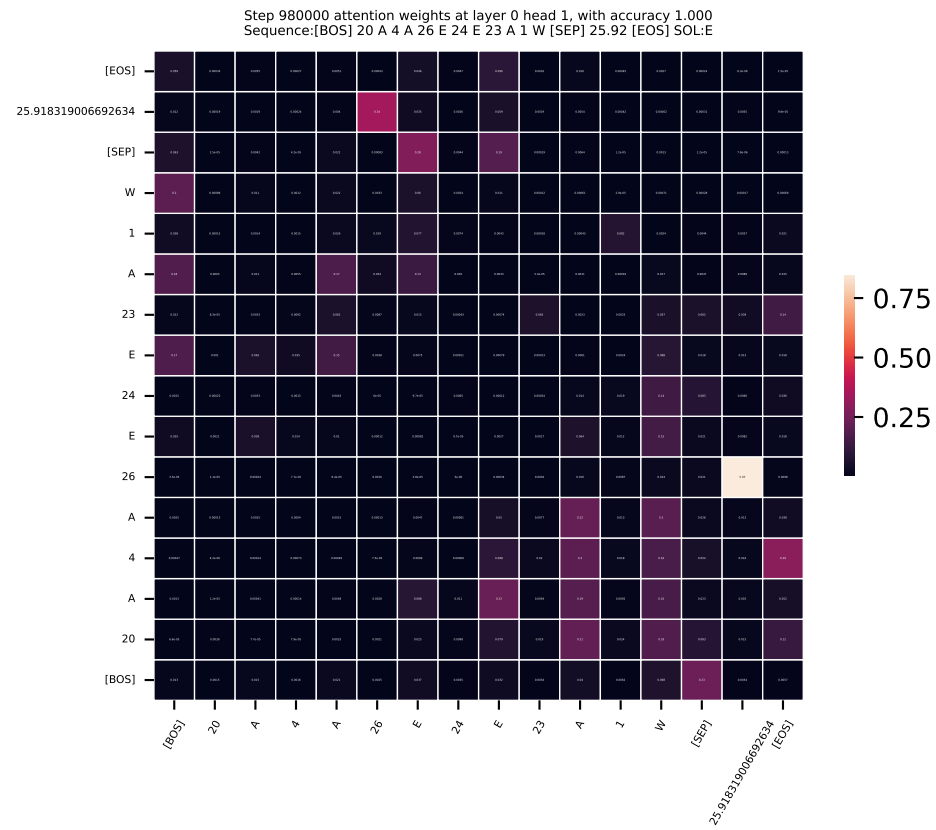

Step 979500 attention weights at layer 1 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

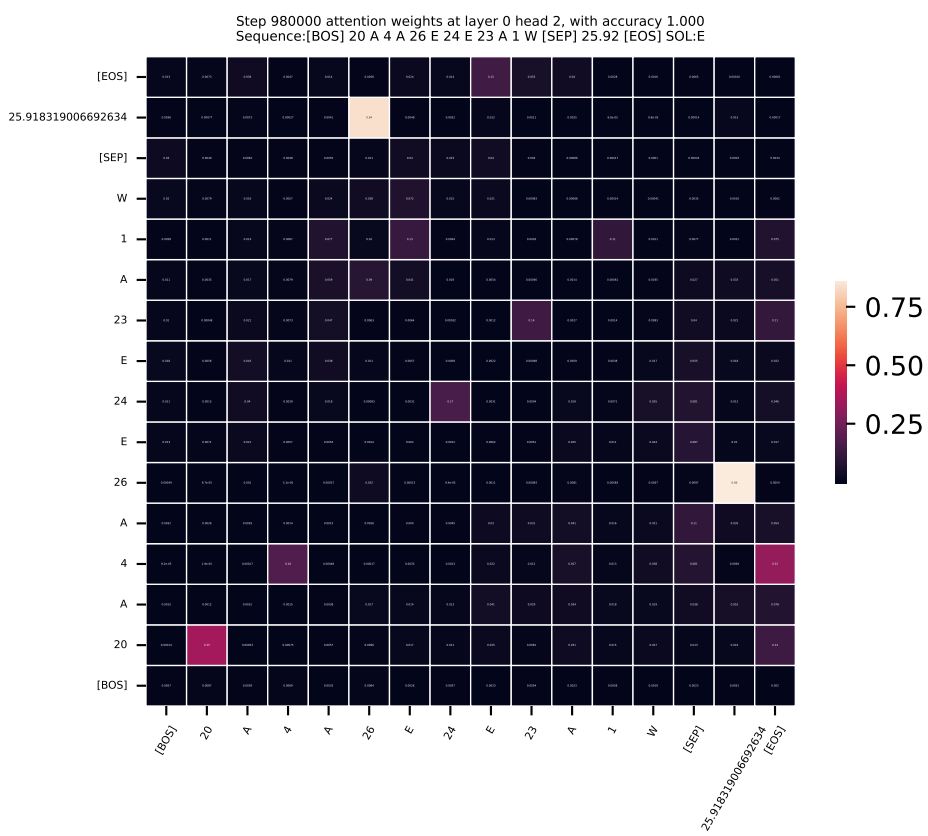

Step 979500 attention weights at layer 1 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

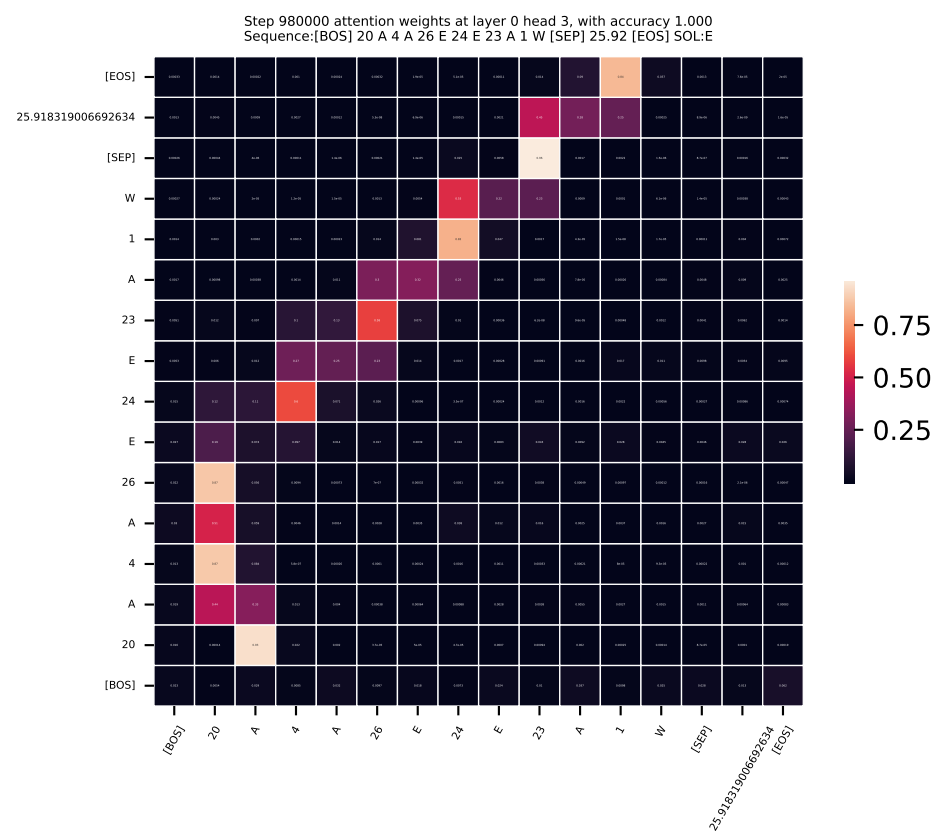

Step 979500 attention weights at layer 1 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

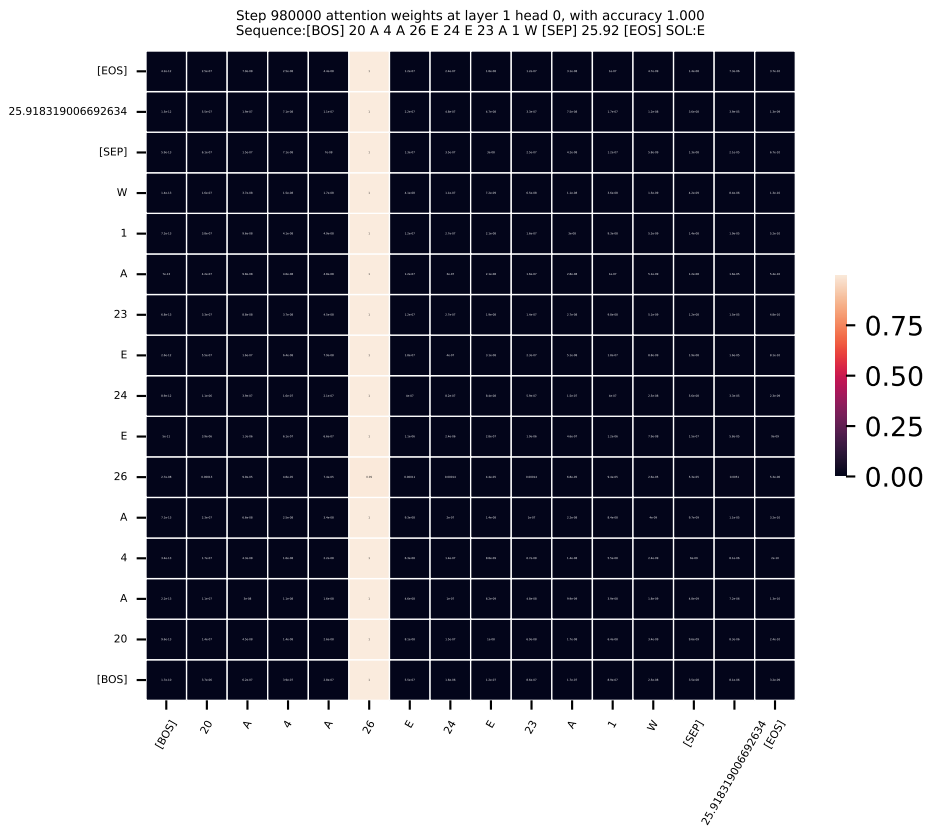

Step 979500 attention weights at layer 1 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

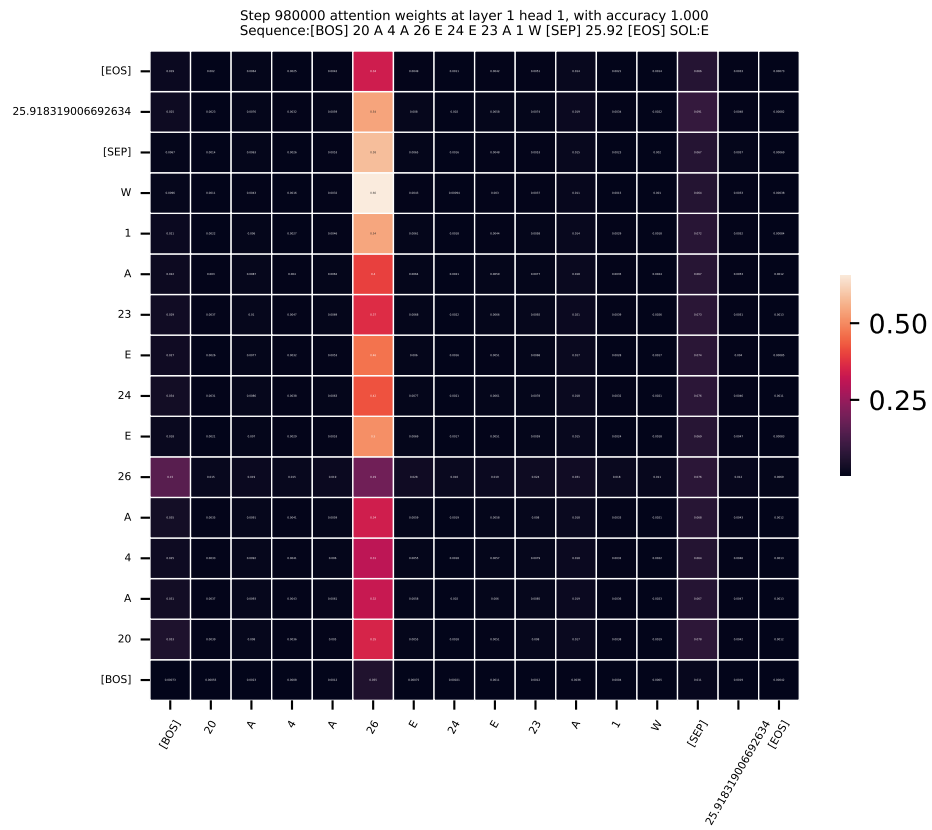

Step 980000 attention weights at layer 0 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

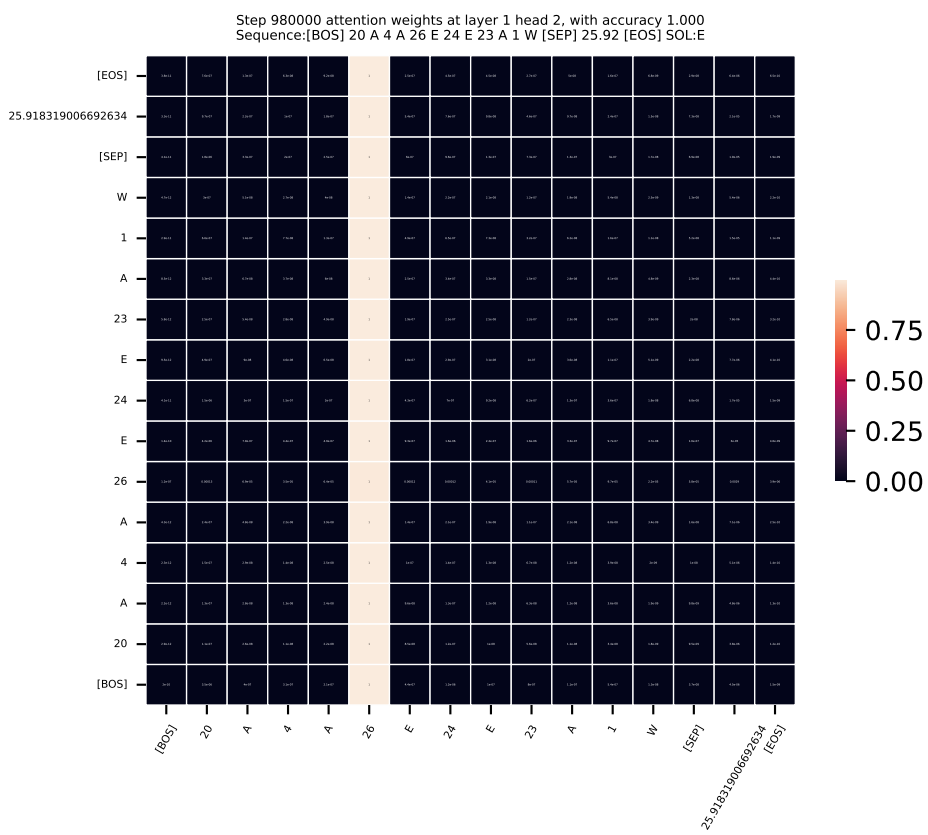

Step 980000 attention weights at layer 0 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

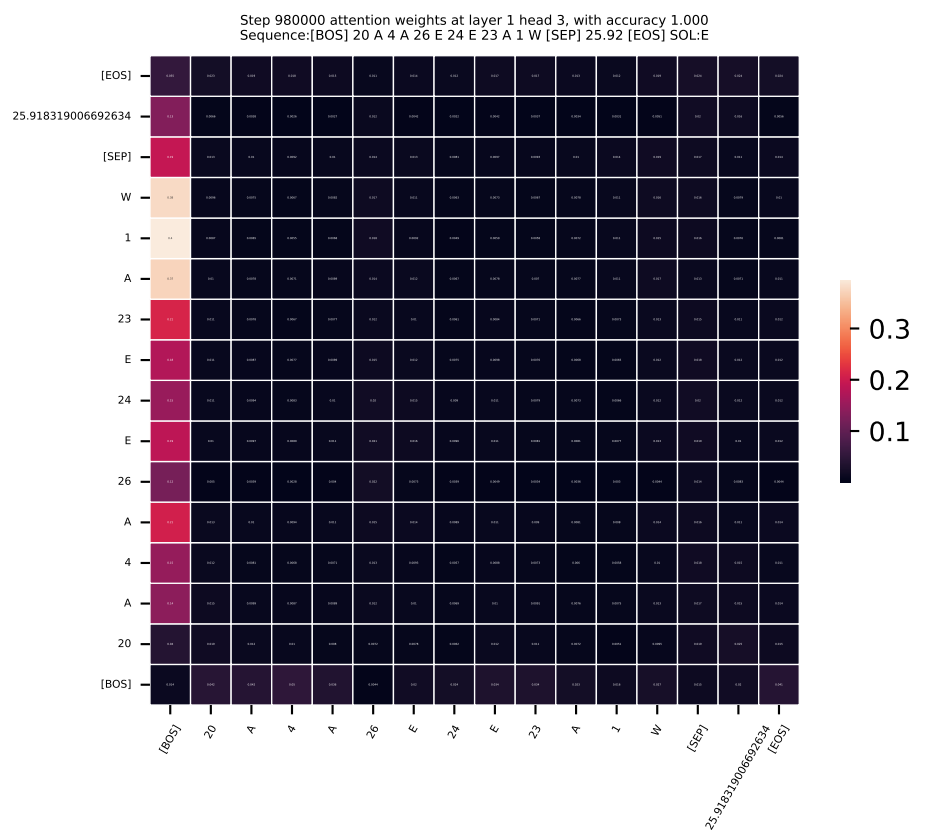

Step 980000 attention weights at layer 0 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

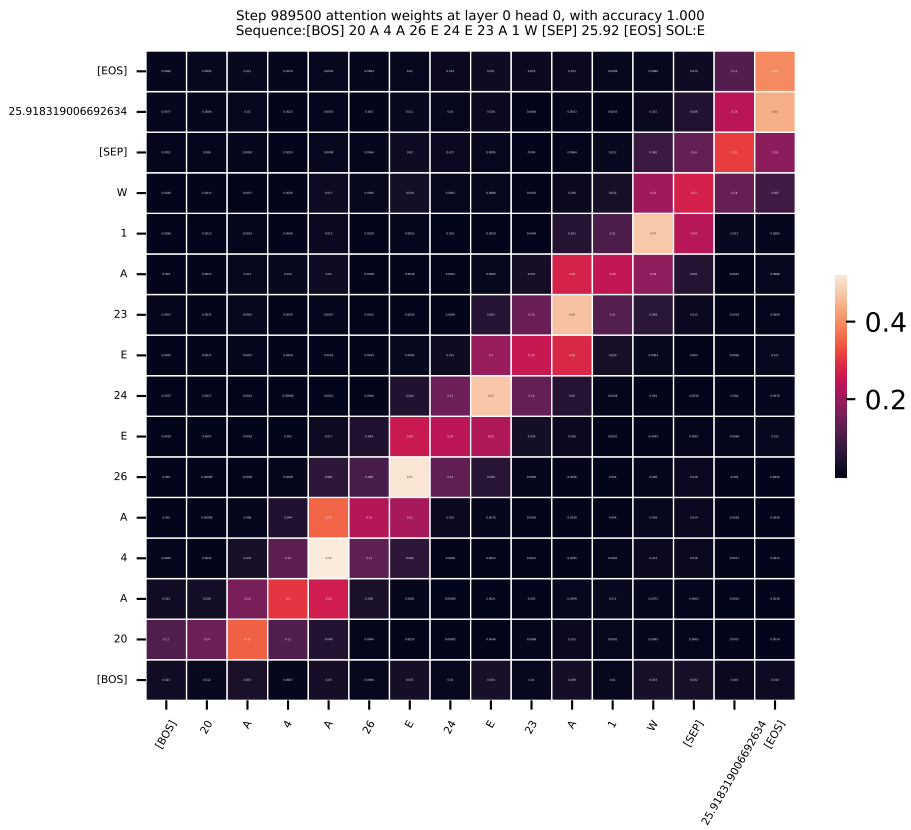

Step 980000 attention weights at layer 0 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

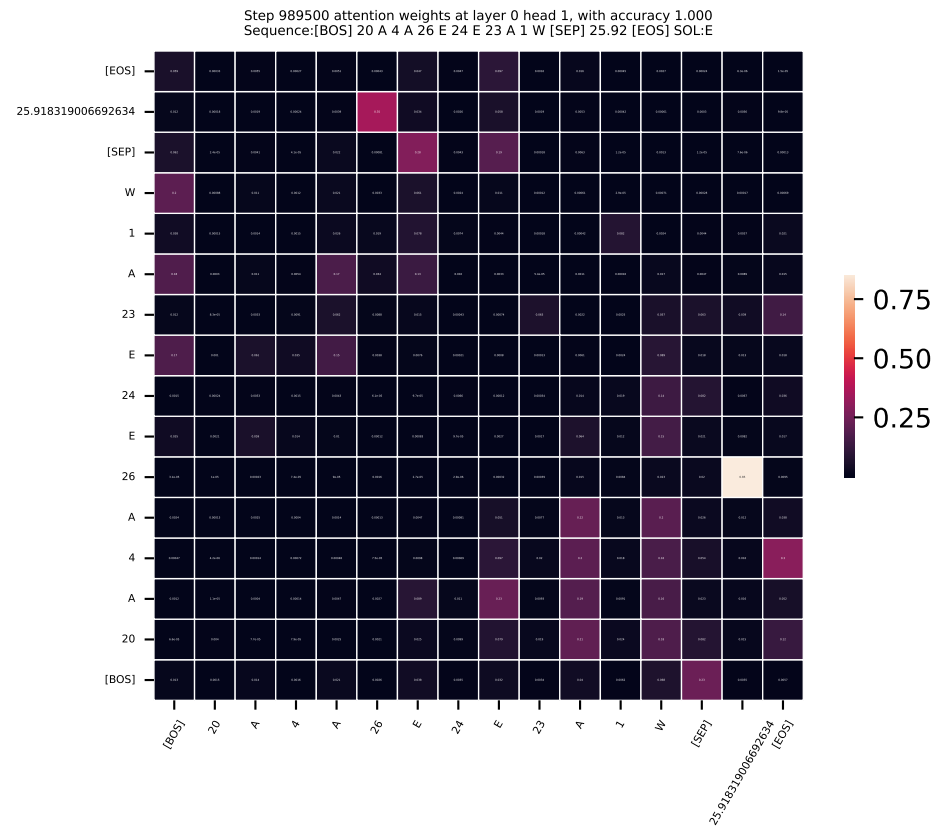

Step 980000 attention weights at layer 1 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

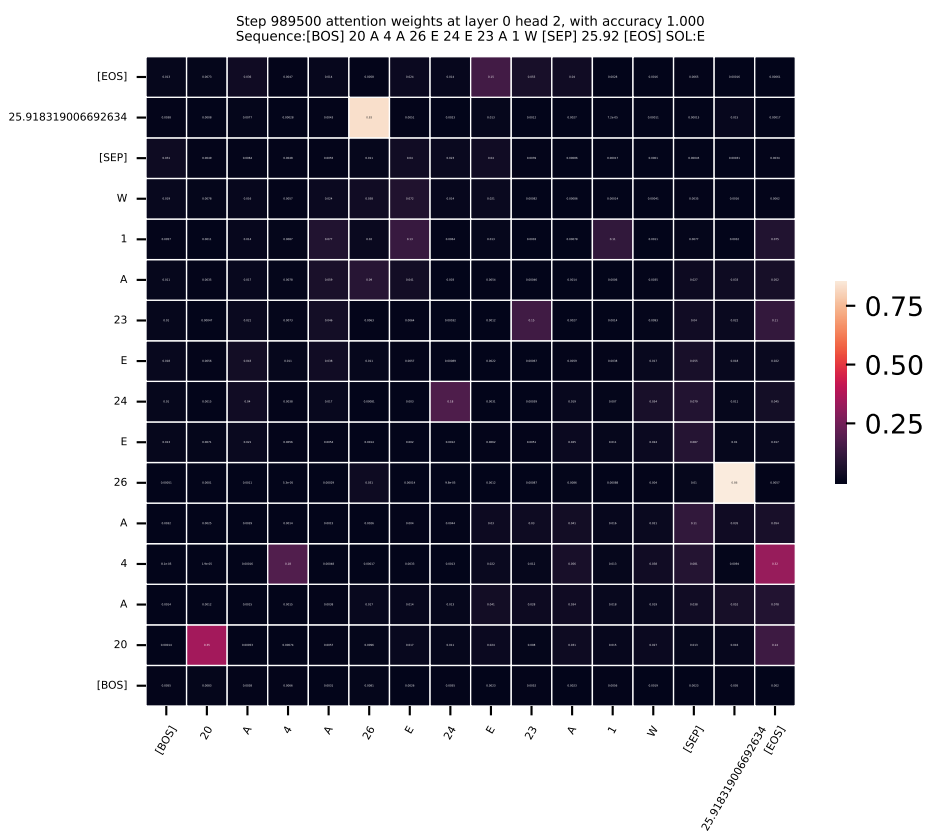

Step 980000 attention weights at layer 1 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

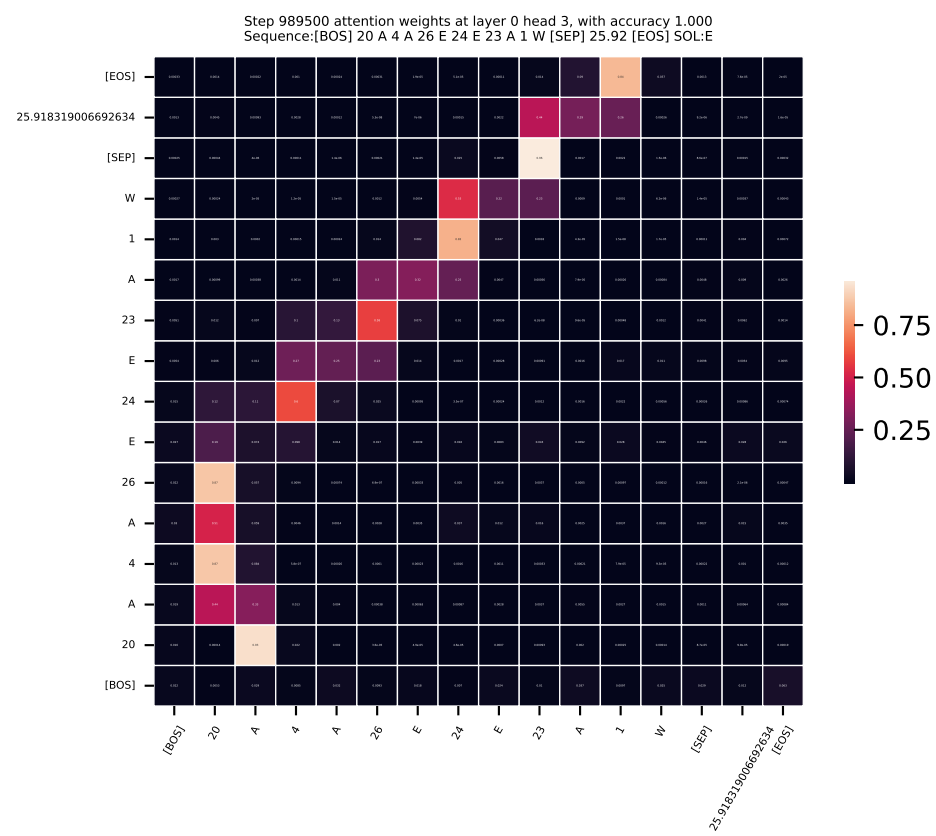

Step 980000 attention weights at layer 1 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

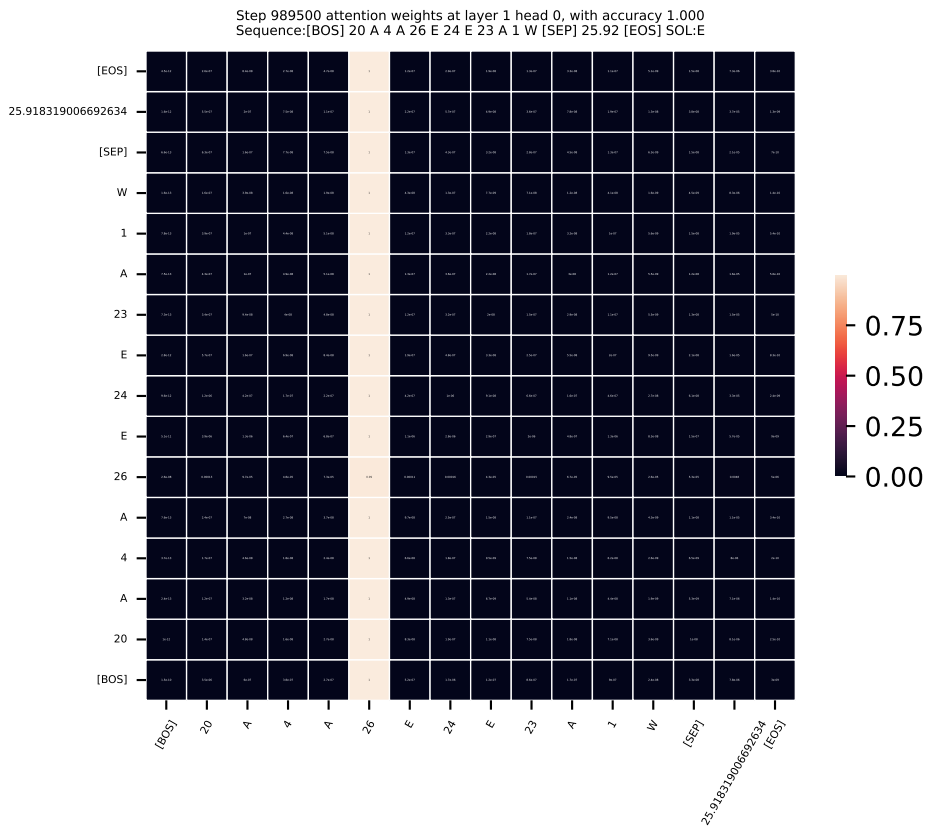

Step 980000 attention weights at layer 1 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

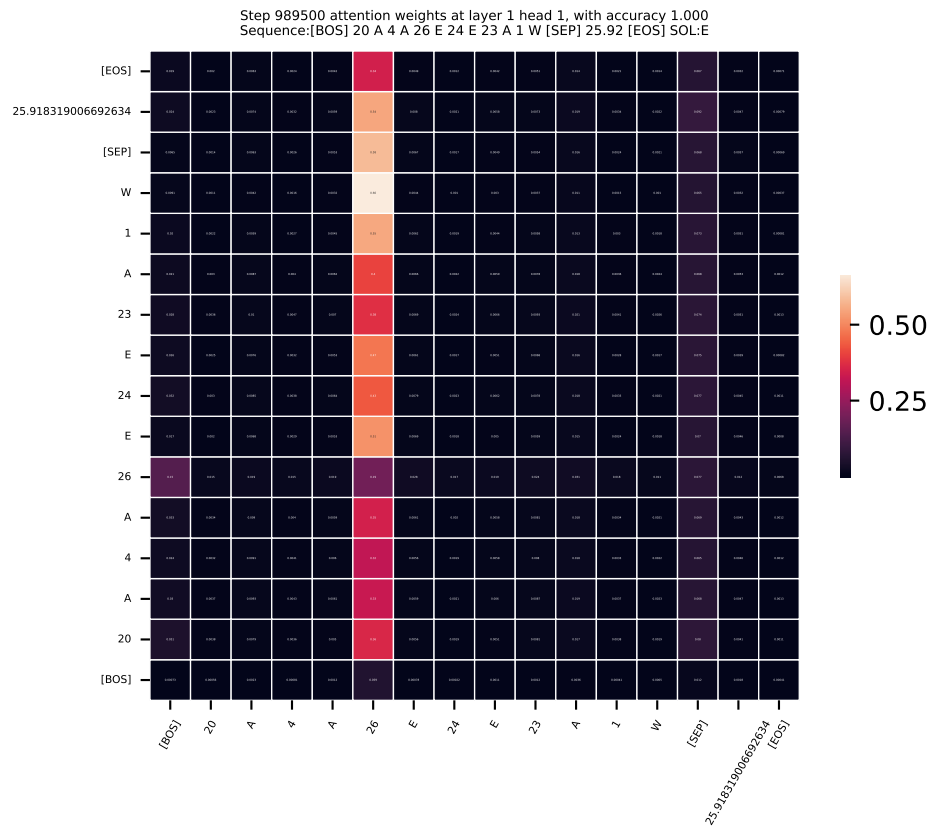

Step 989500 attention weights at layer 0 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

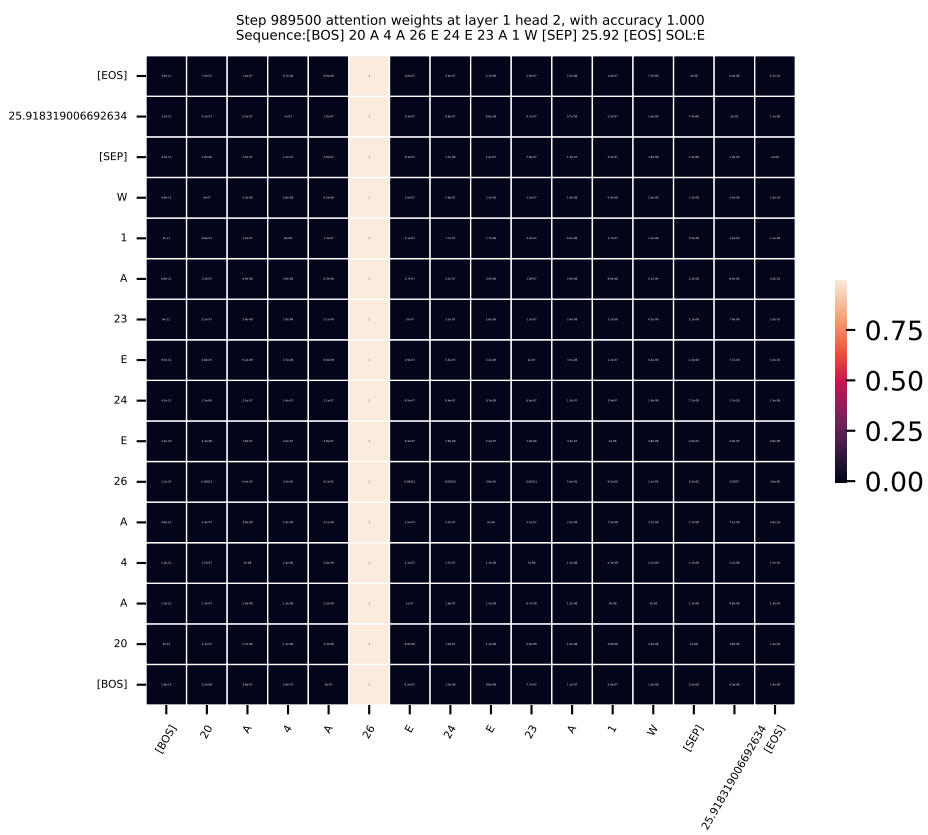

Step 989500 attention weights at layer 0 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

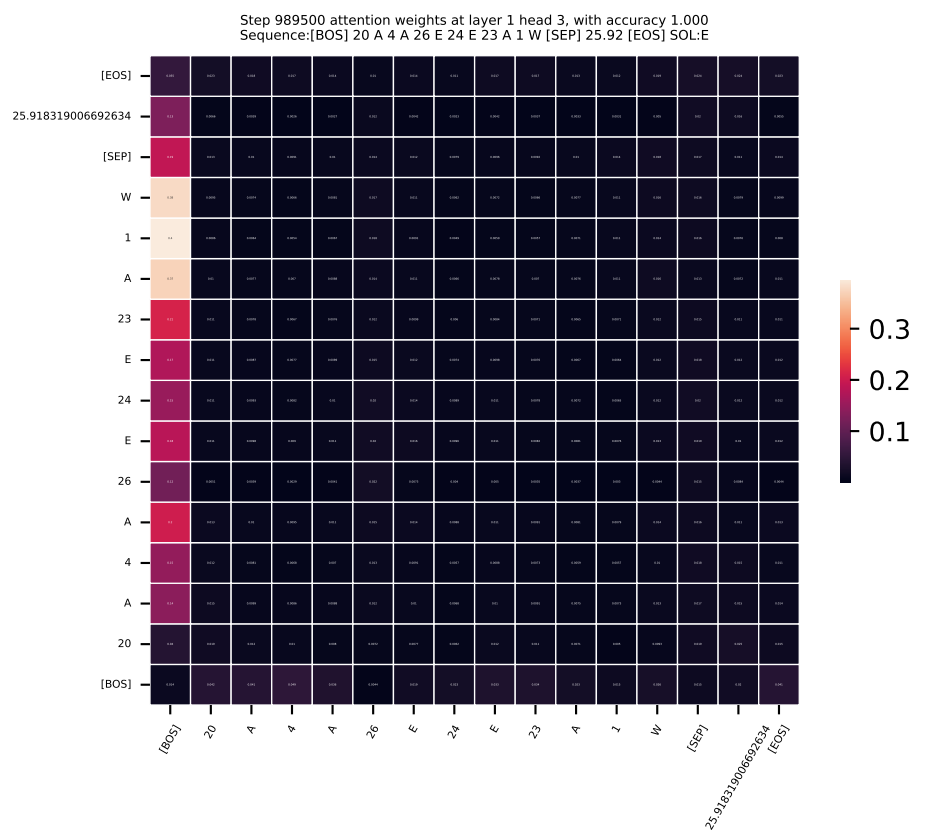

Step 989500 attention weights at layer 0 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

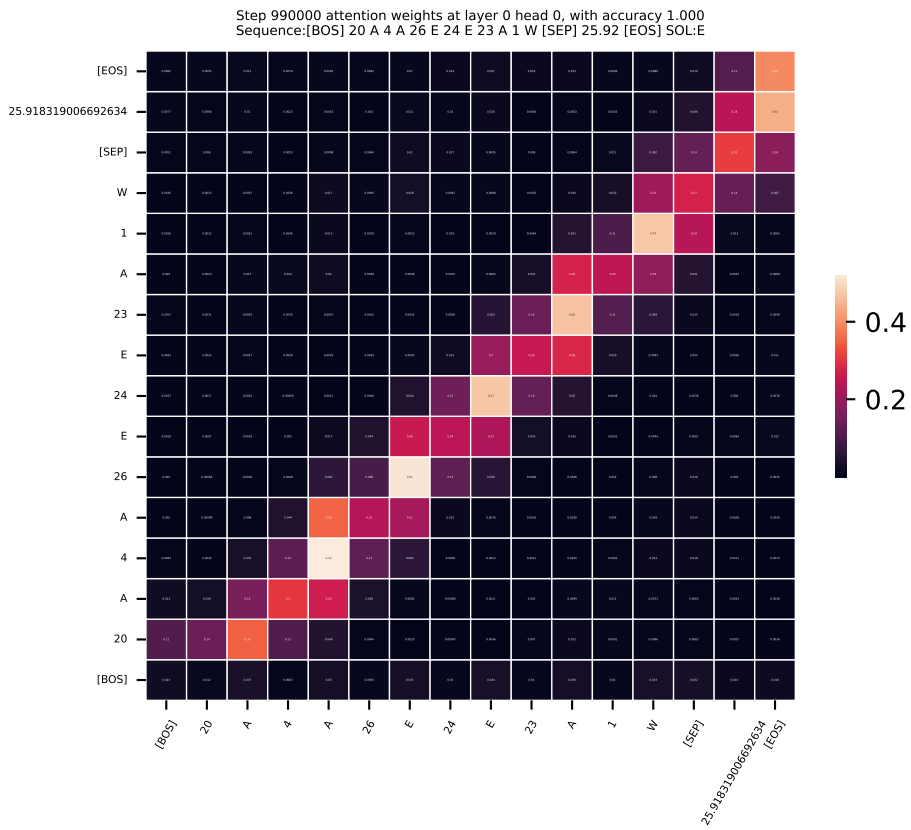

Step 989500 attention weights at layer 0 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

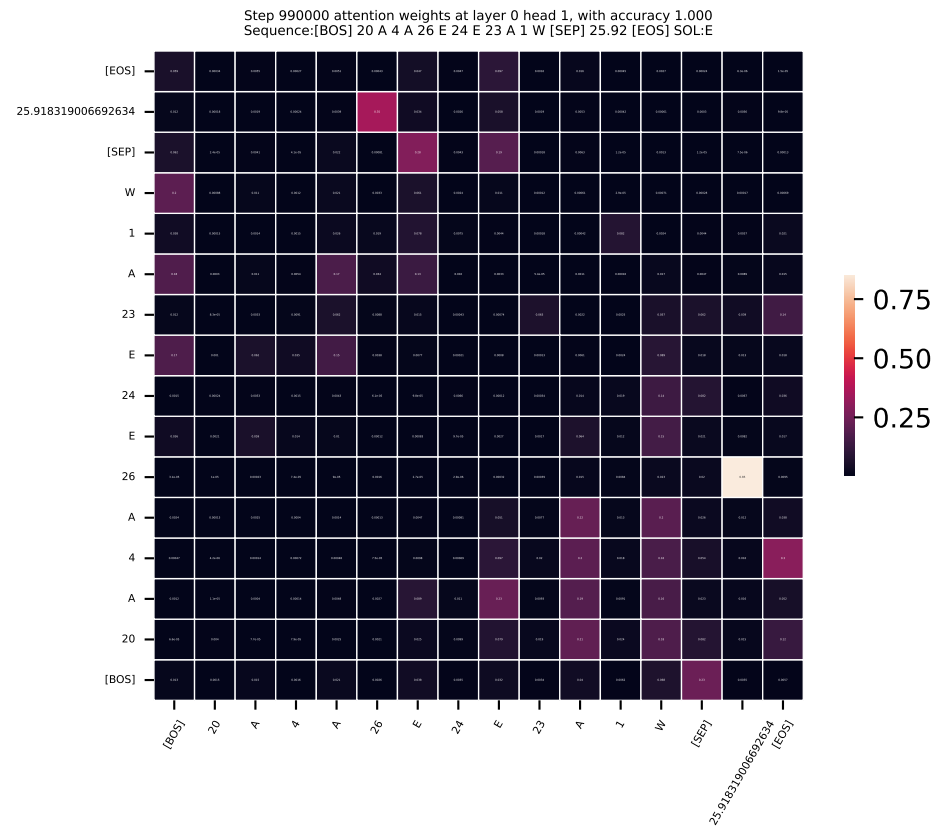

Step 989500 attention weights at layer 1 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

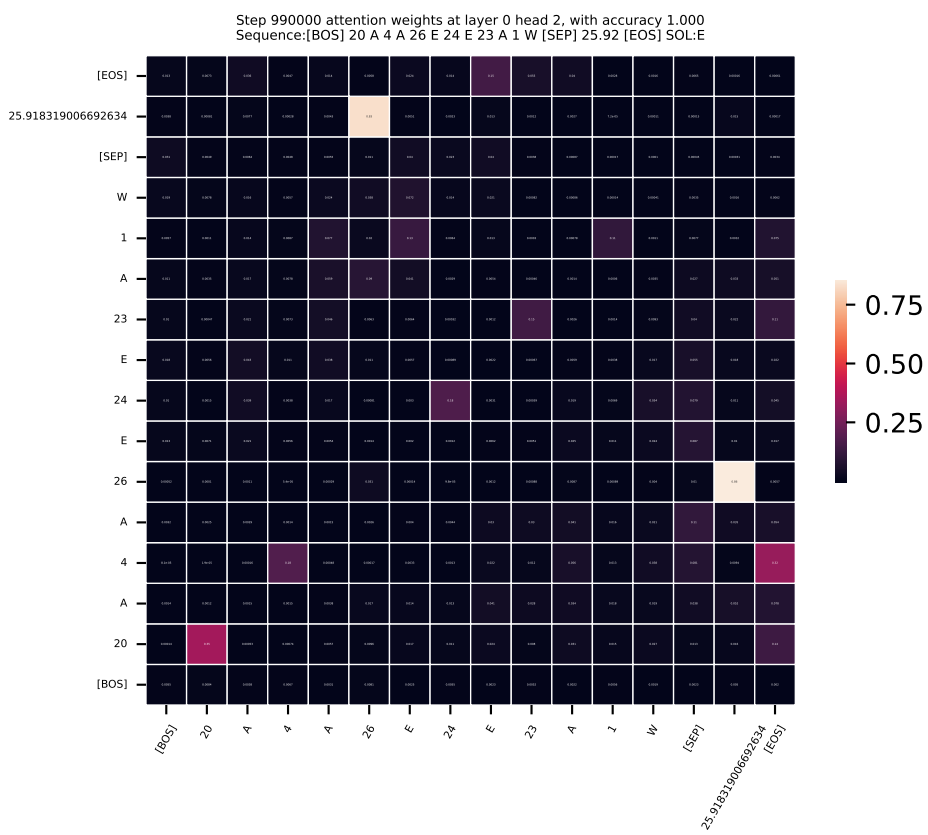

Step 989500 attention weights at layer 1 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25

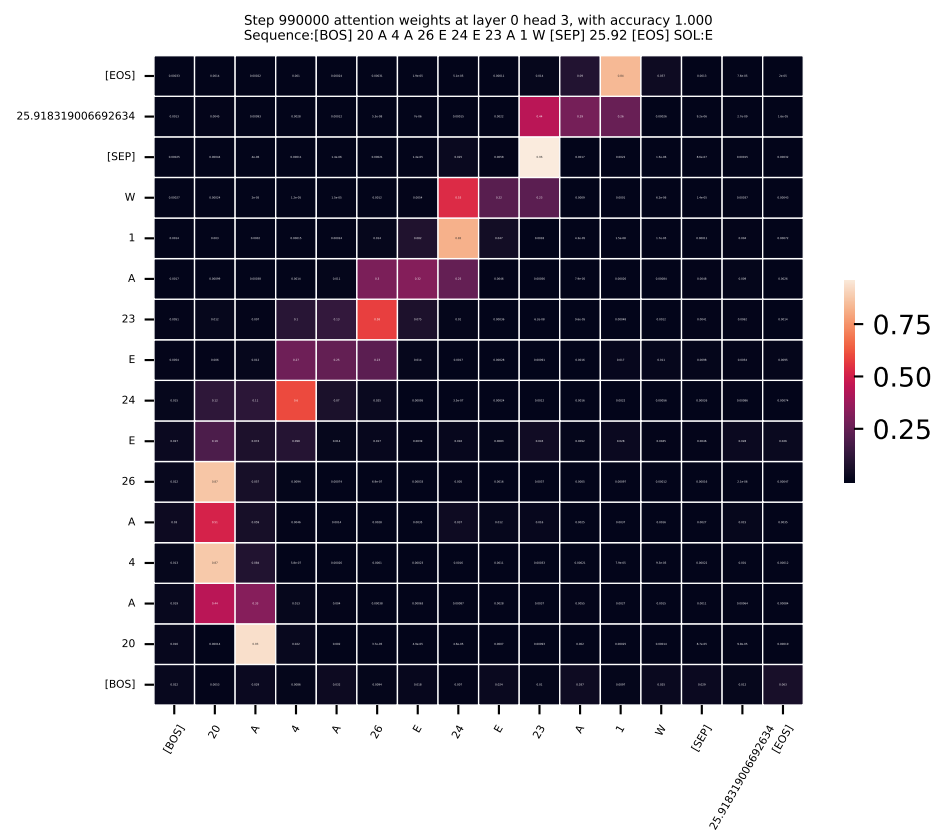

Step 989500 attention weights at layer 1 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00

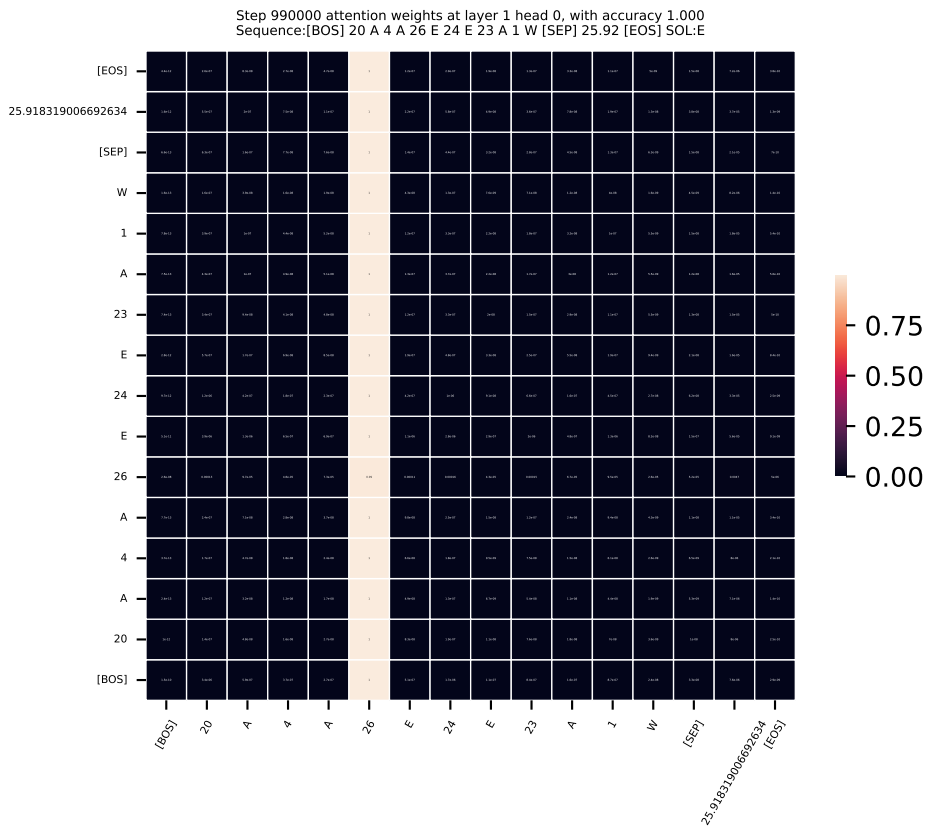

Step 989500 attention weights at layer 1 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1

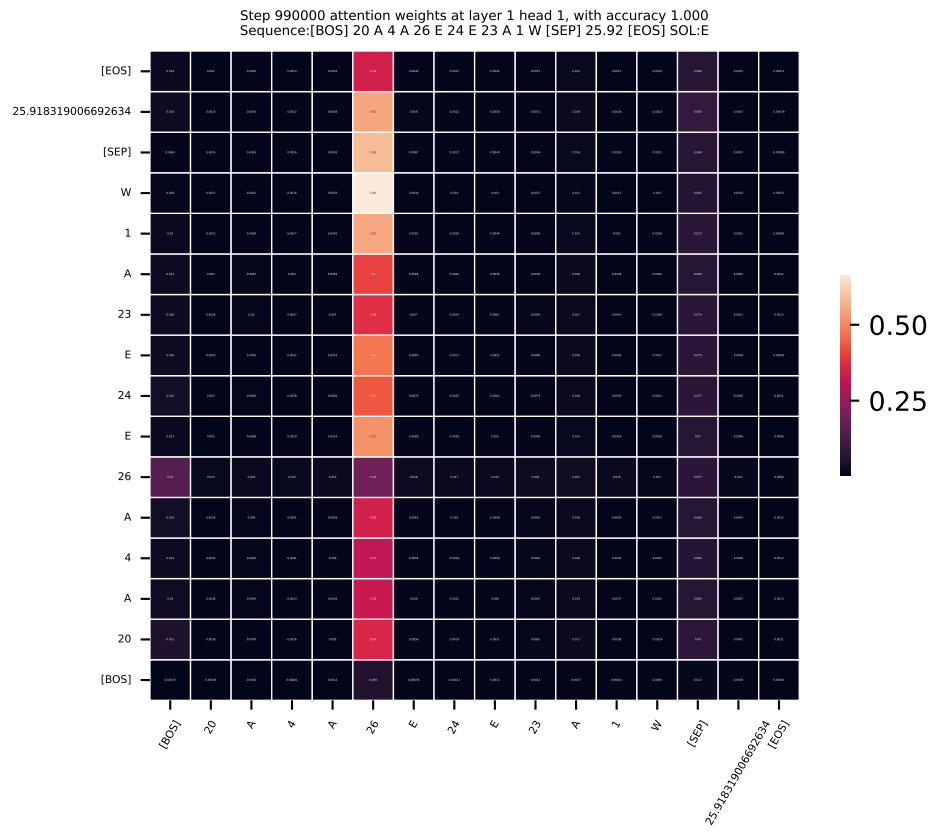

Step 990000 attention weights at layer 0 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.4
0.2

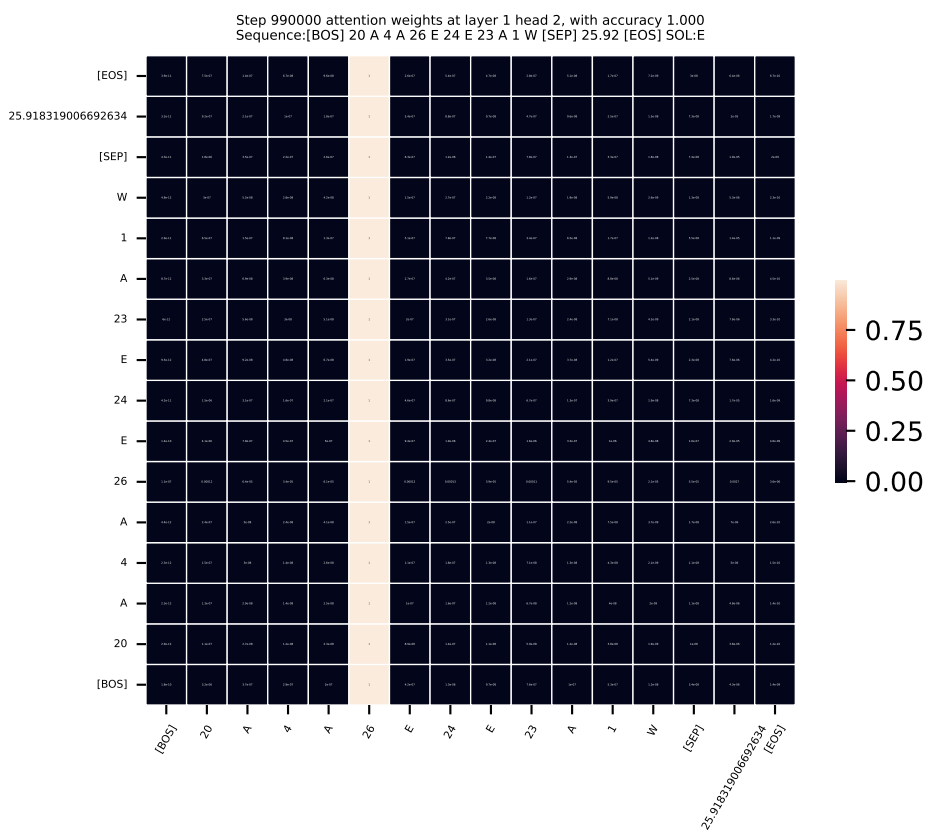

Step 990000 attention weights at layer 0 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25

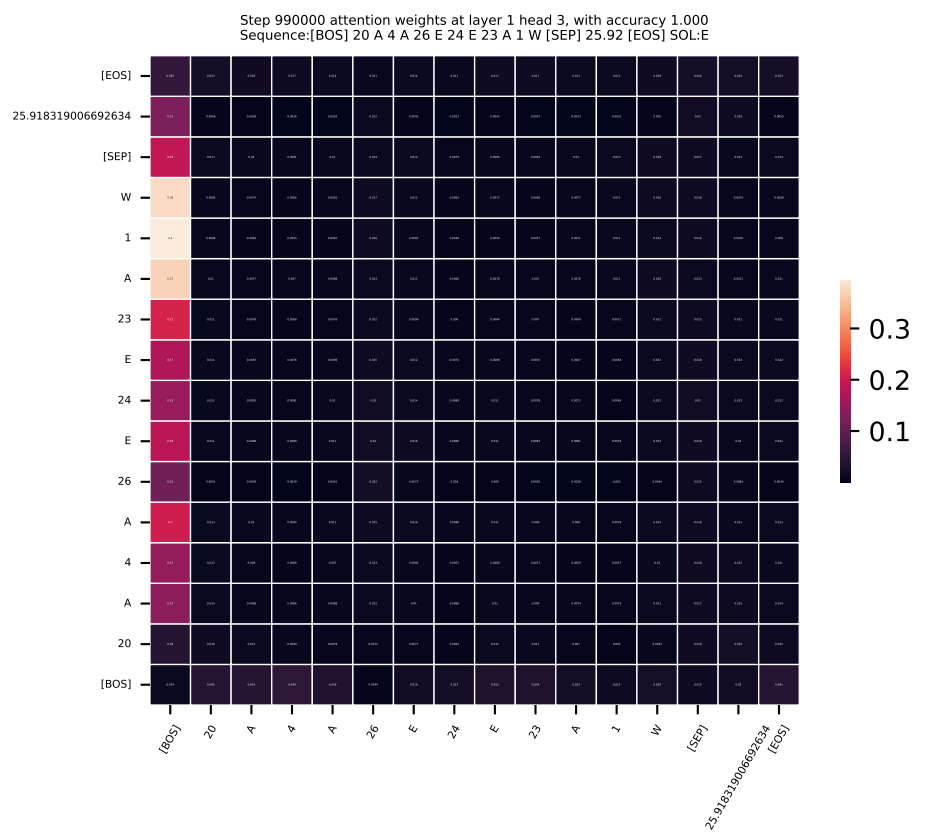

Step 990000 attention weights at layer 0 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25


Step 990000 attention weights at layer 0 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
0.75
0.50
0.25


Step 990000 attention weights at layer 1 head 0, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00


Step 990000 attention weights at layer 1 head 1, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.50
0.25


Step 990000 attention weights at layer 1 head 2, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.75
0.50
0.25
0.00


Step 990000 attention weights at layer 1 head 3, with accuracy 1.000
Sequence:[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.92 [EOS] SOL:E
[EOS]
25.918319006692634
[SEP]
W
1
A
23
E
24
E
26
A
4
A
20
[BOS]
[BOS] 20 A 4 A 26 E 24 E 23 A 1 W [SEP] 25.918319006692634 [EOS]
0.3
0.2
0.1