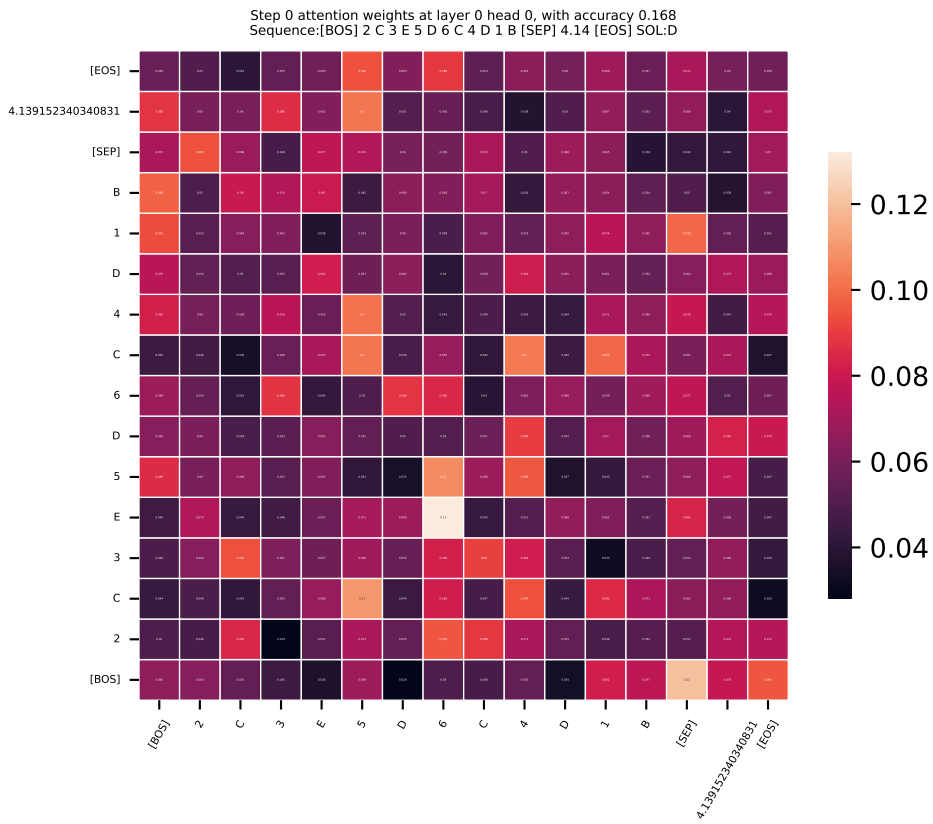
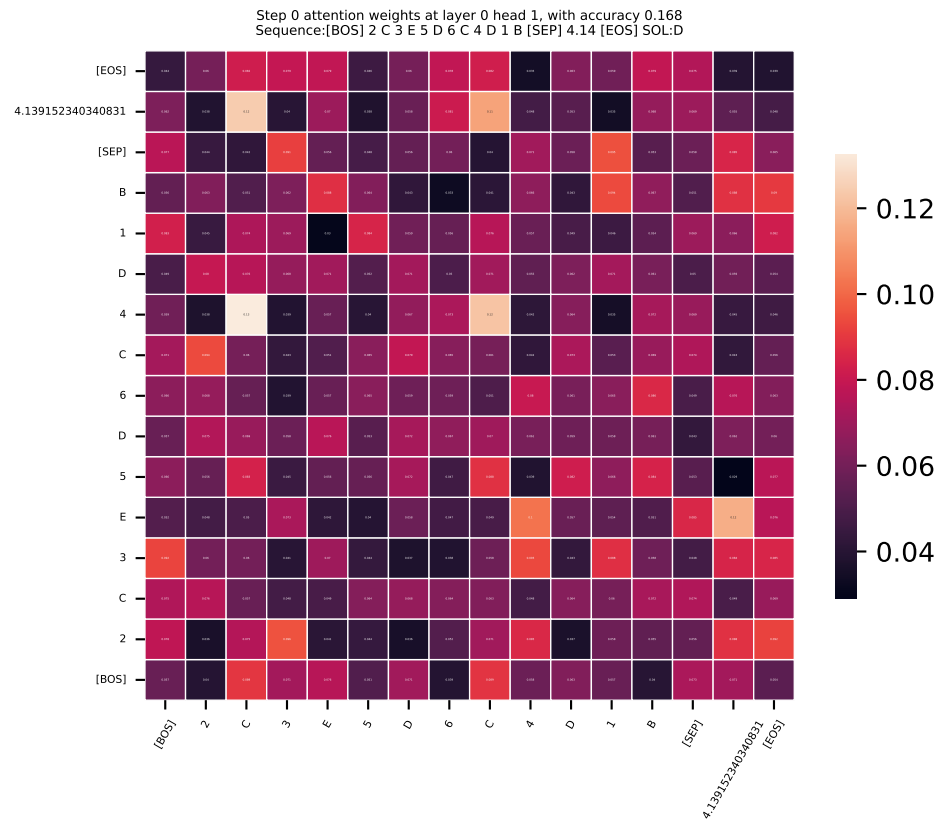
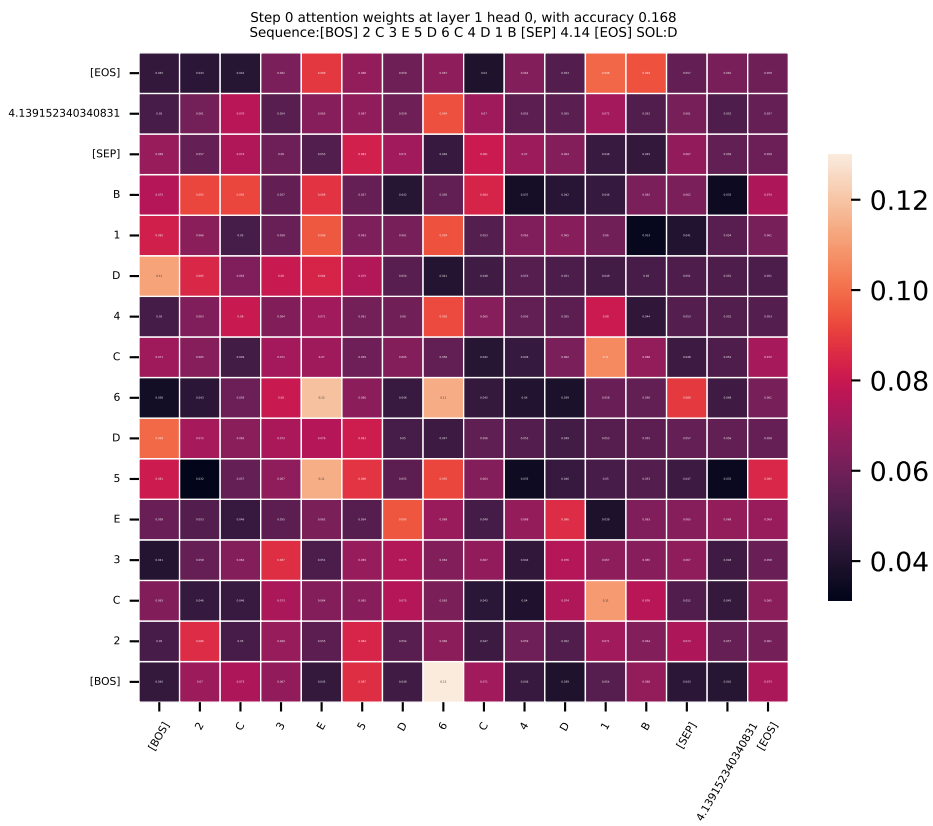
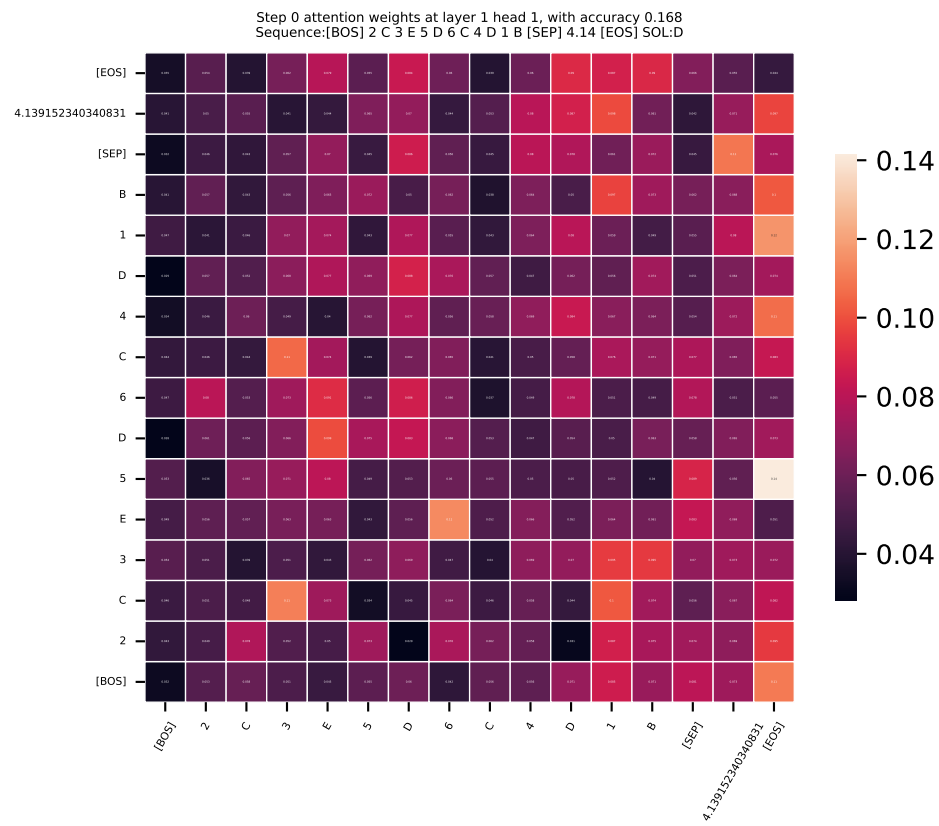
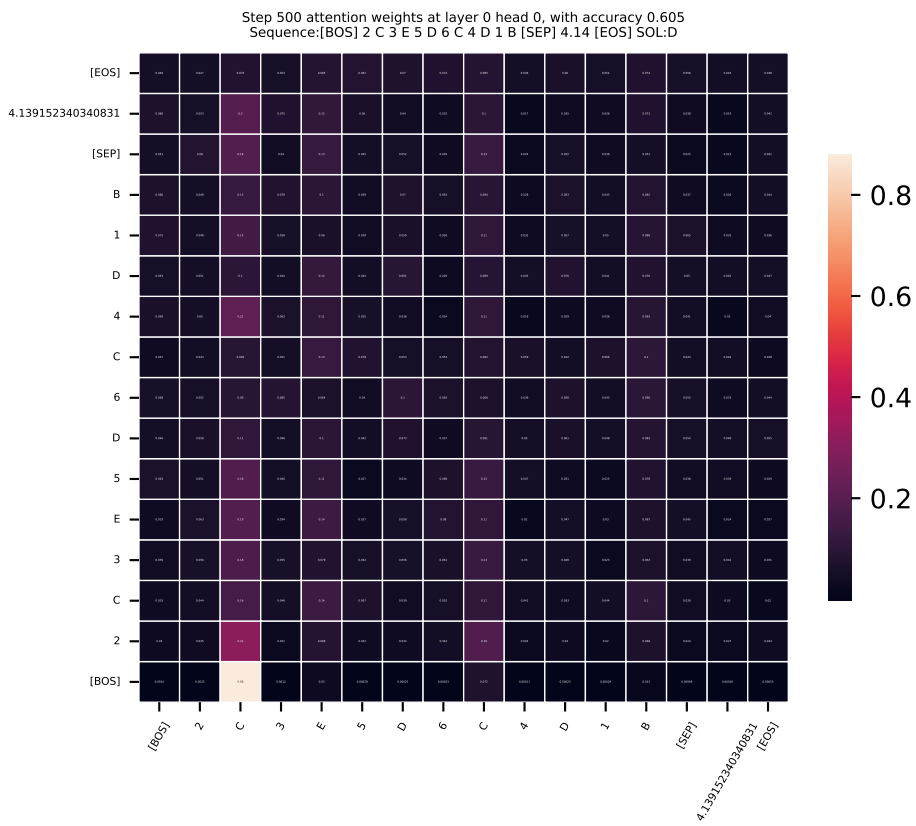
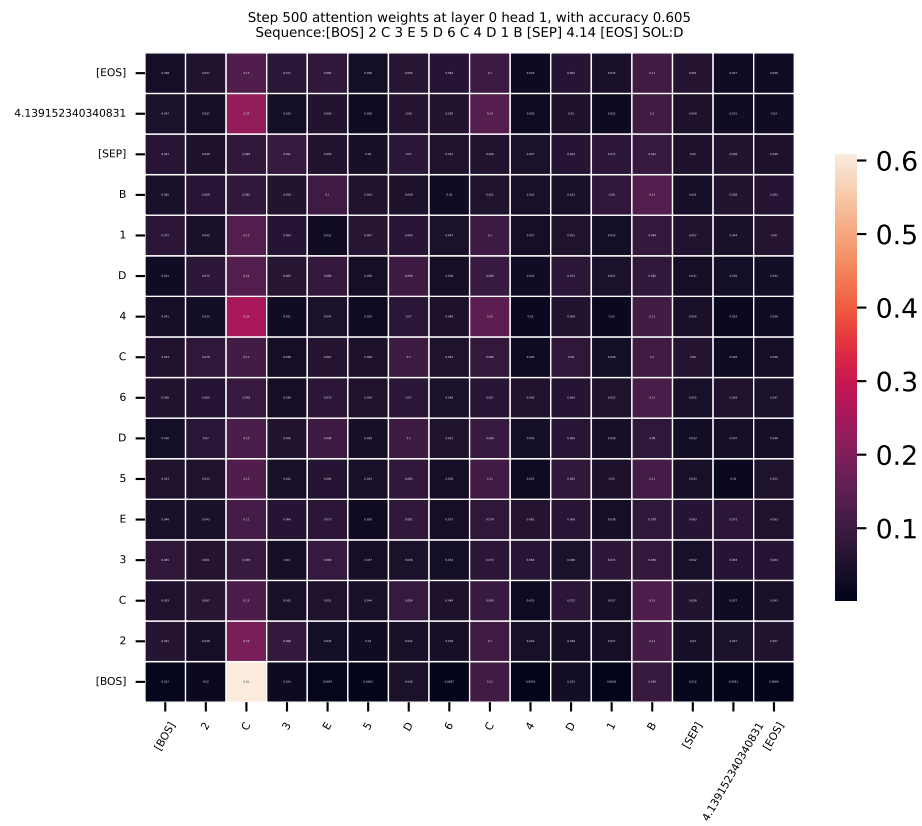

Step 0 attention weights at layer 0 head 0, with accuracy 0.168
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.12
0.10
0.08
0.06
0.04

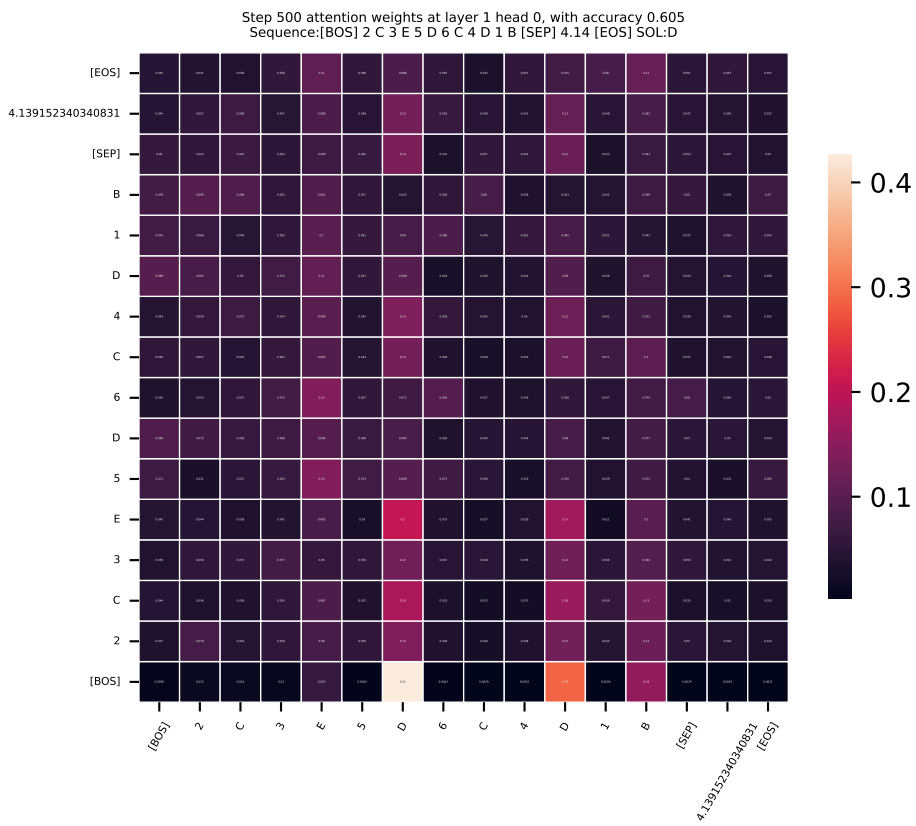

Step 0 attention weights at layer 0 head 1, with accuracy 0.168
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.12
0.10
0.08
0.06
0.04

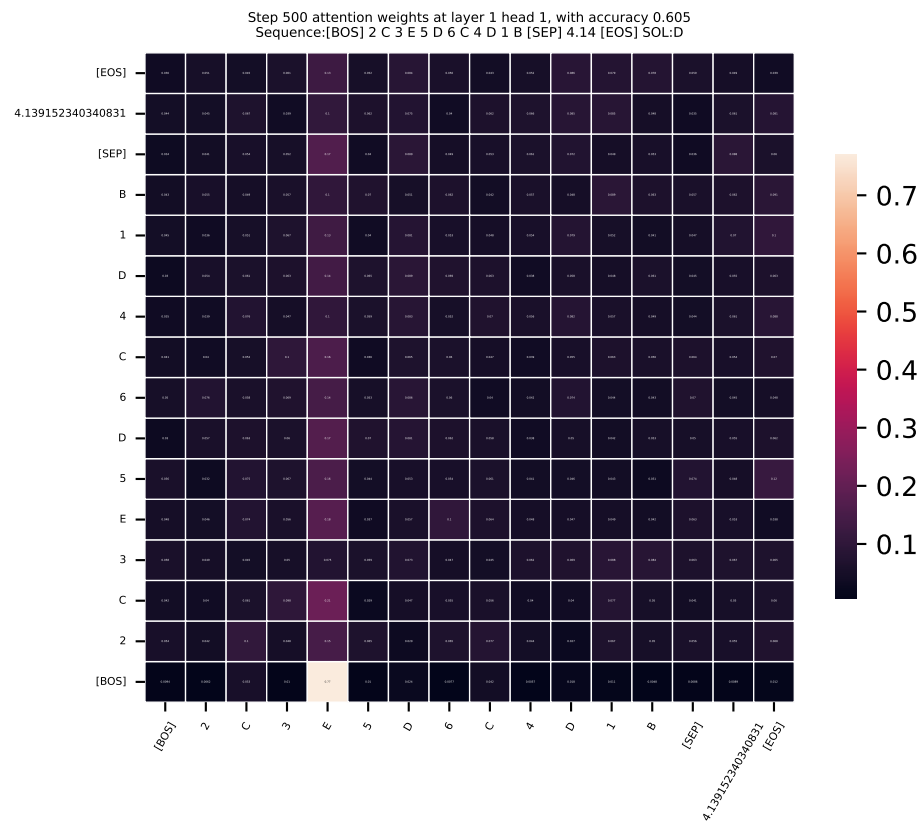

Step 0 attention weights at layer 1 head 0, with accuracy 0.168
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.12
0.10
0.08
0.06
0.04

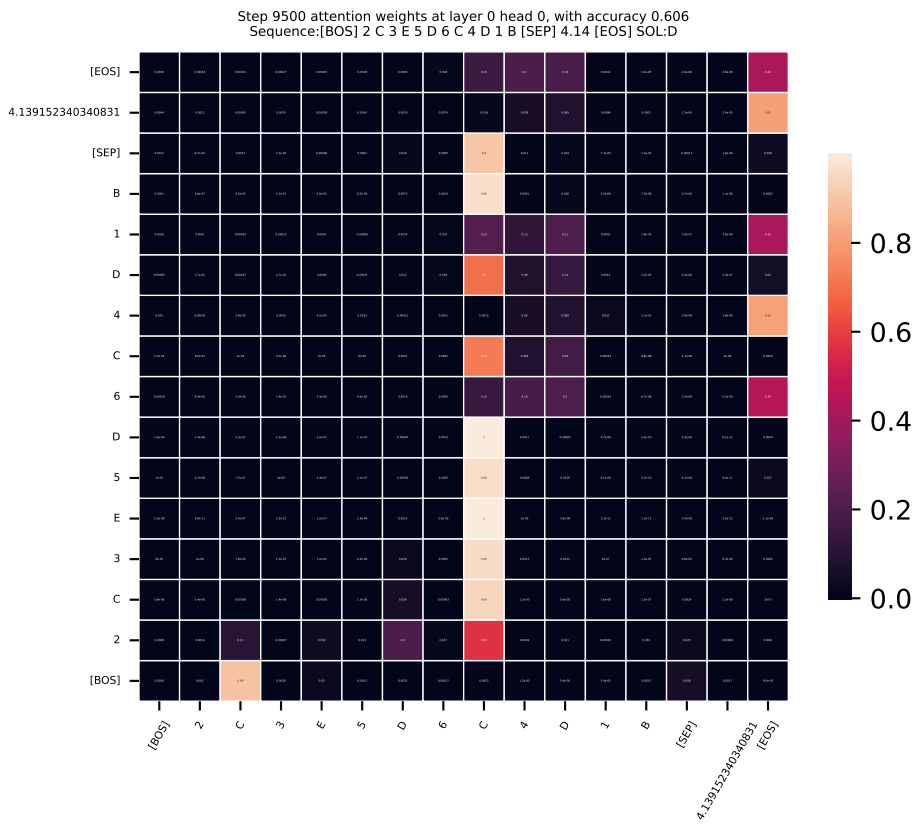

Step 0 attention weights at layer 1 head 1, with accuracy 0.168
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.14
0.12
0.10
0.08
0.06
0.04

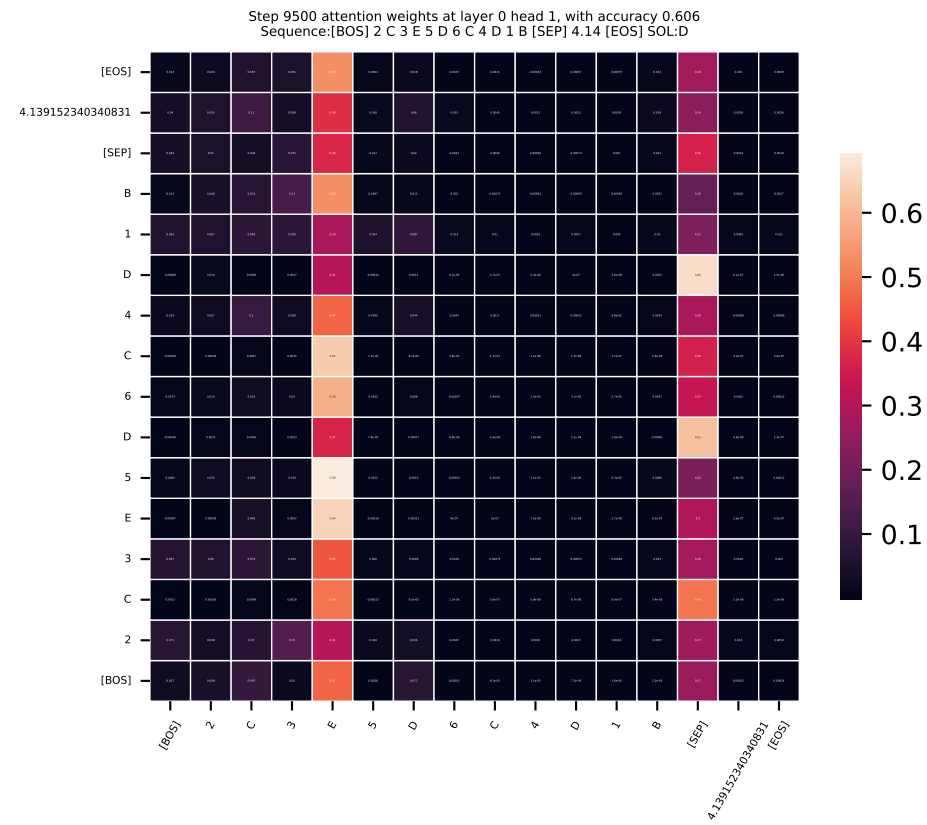

Step 500 attention weights at layer 0 head 0, with accuracy 0.605
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2

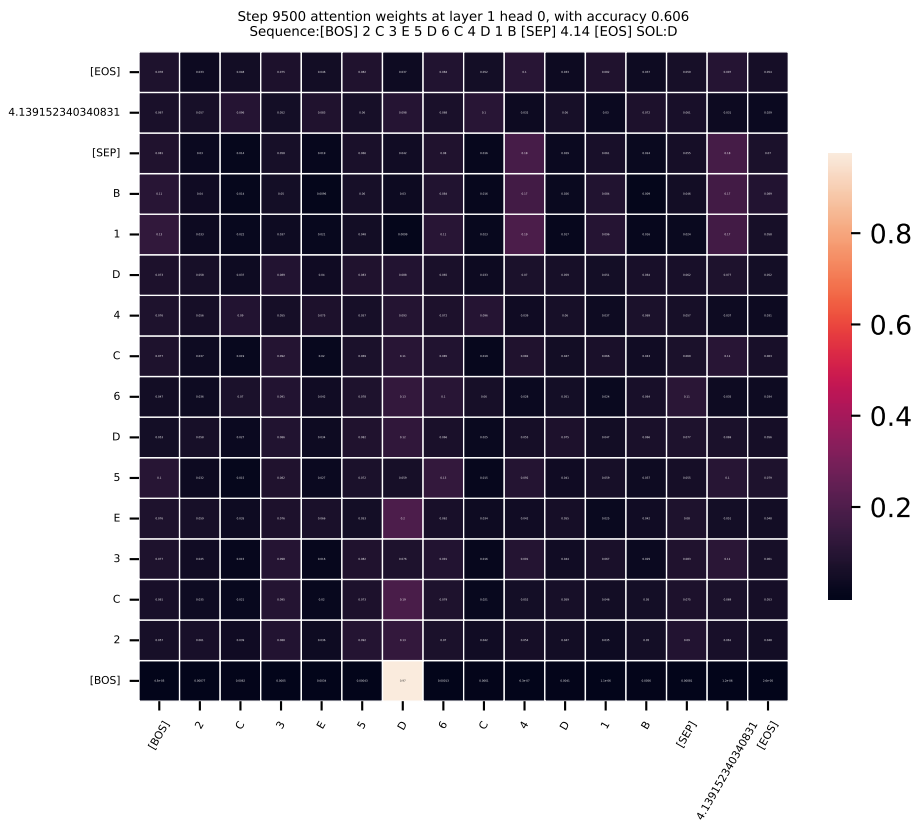

Step 500 attention weights at layer 0 head 1, with accuracy 0.605
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.6
0.5
0.4
0.3
0.2
0.1

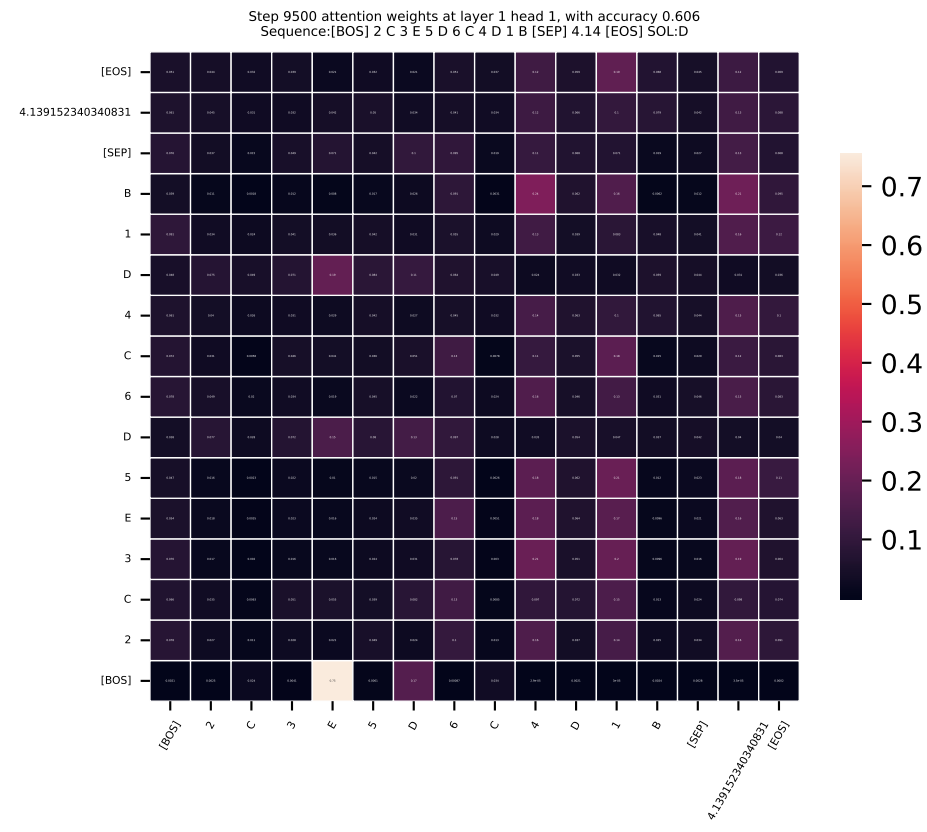

Step 500 attention weights at layer 1 head 0, with accuracy 0.605
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.4
0.3
0.2
0.1

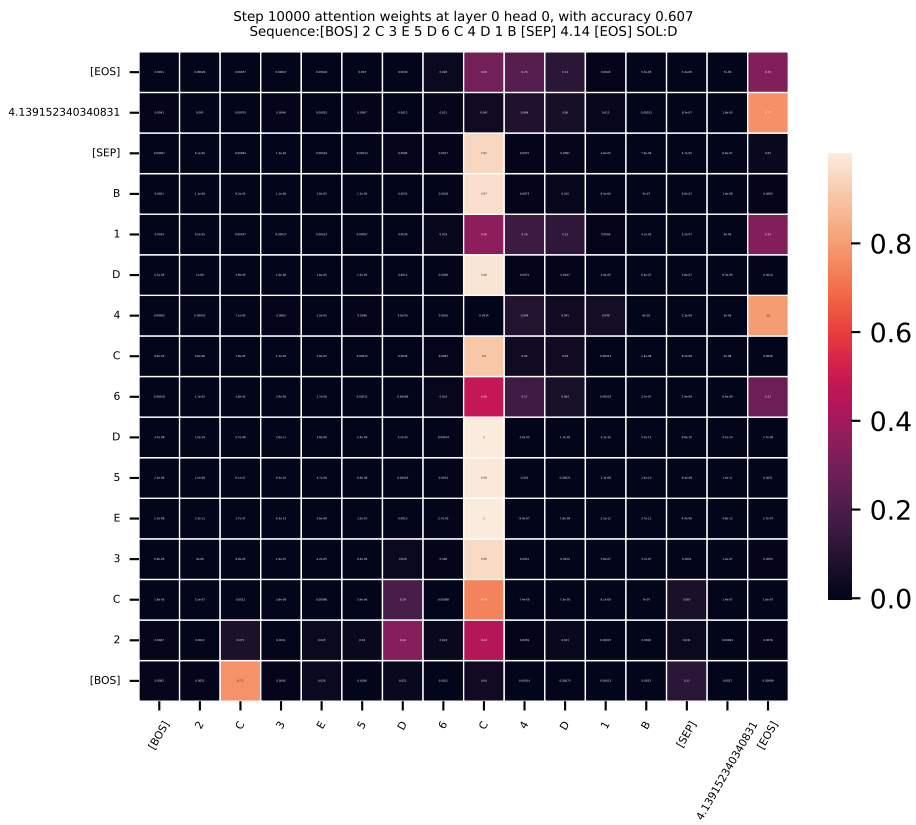

Step 500 attention weights at layer 1 head 1, with accuracy 0.605
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.7
0.6
0.5
0.4
0.3
0.2
0.1

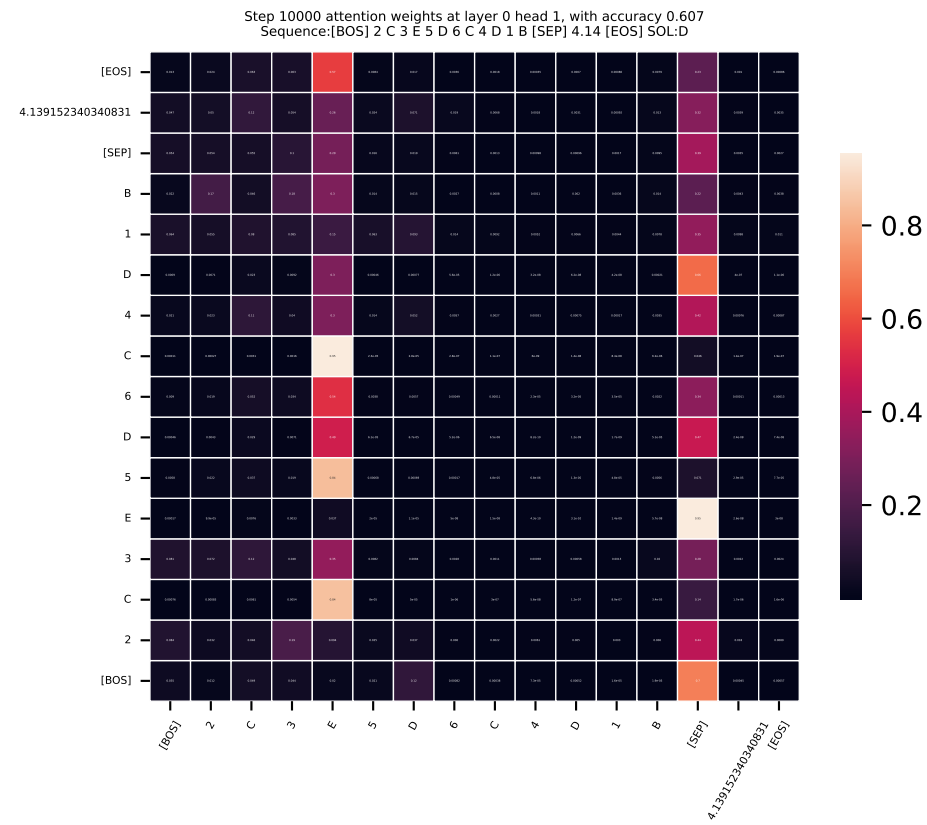

Step 9500 attention weights at layer 0 head 0, with accuracy 0.606
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2
0.0

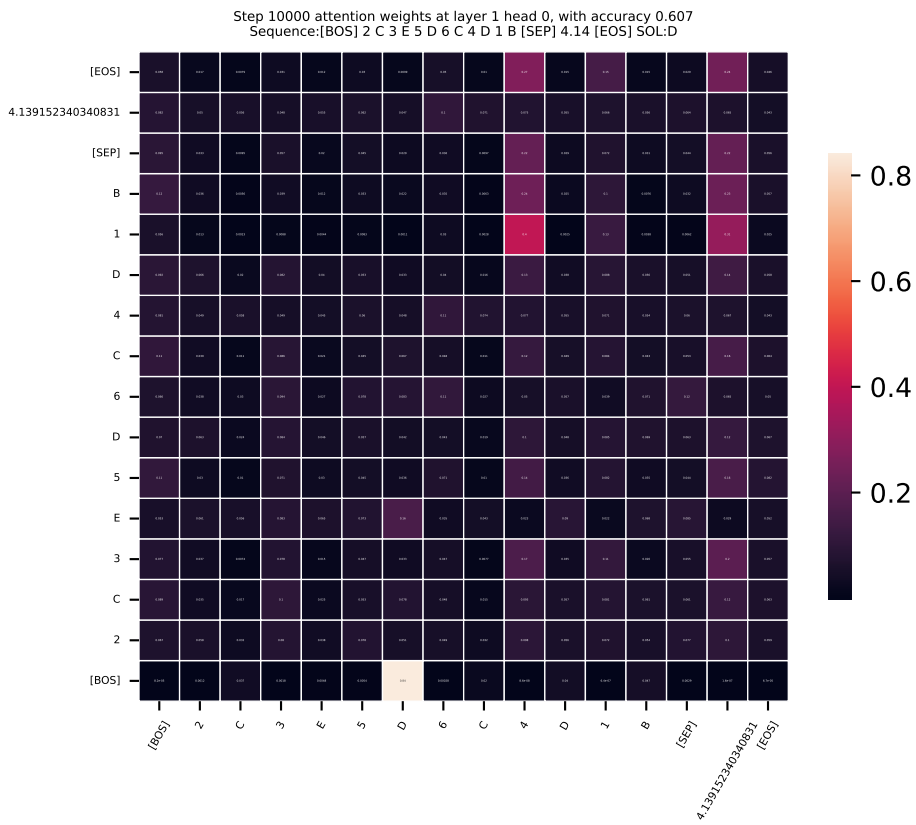

Step 9500 attention weights at layer 0 head 1, with accuracy 0.606
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.6
0.5
0.4
0.3
0.2
0.1

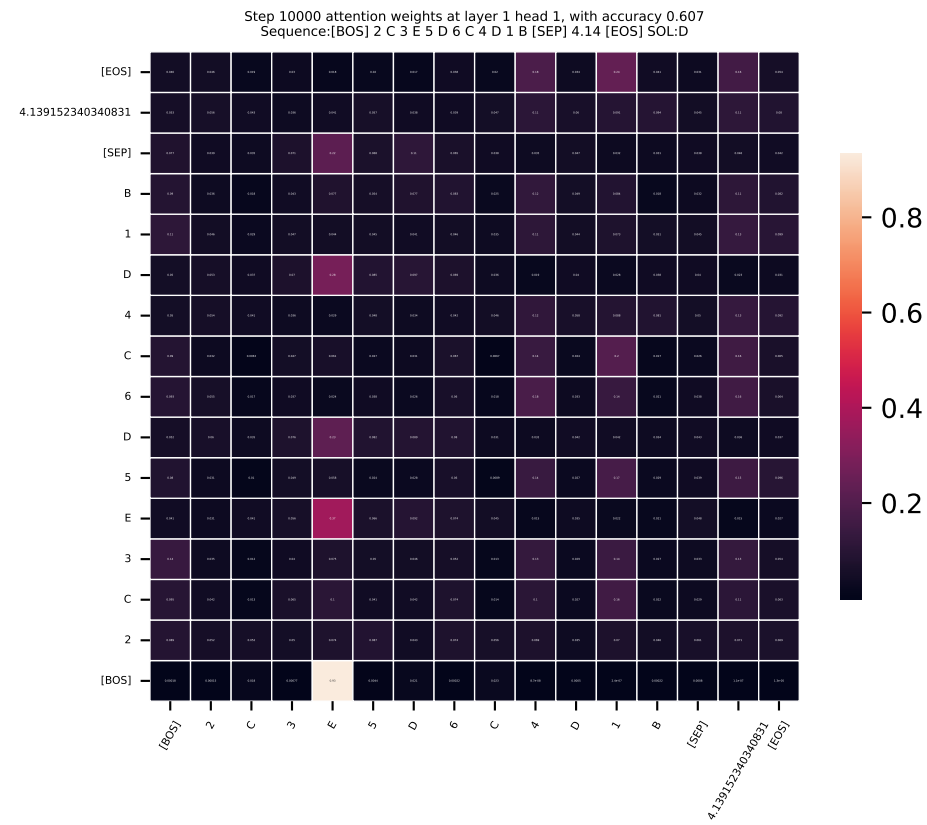

Step 9500 attention weights at layer 1 head 0, with accuracy 0.606
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.8
0.6
0.4
0.2

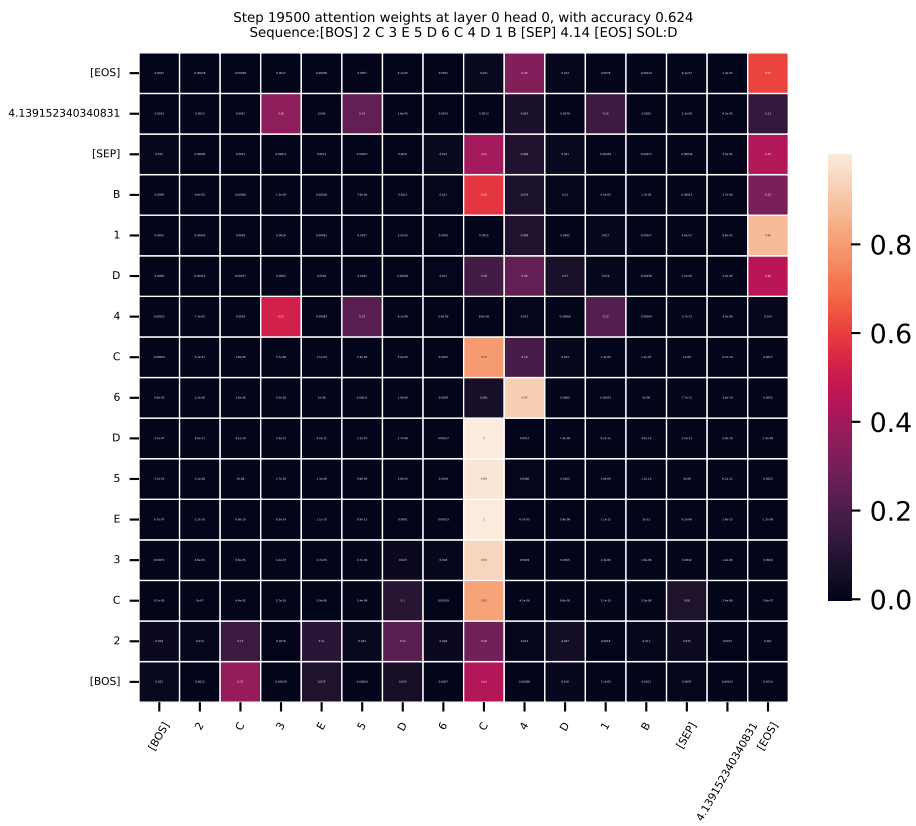

Step 9500 attention weights at layer 1 head 1, with accuracy 0.606
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.7
0.6
0.5
0.4
0.3
0.2
0.1

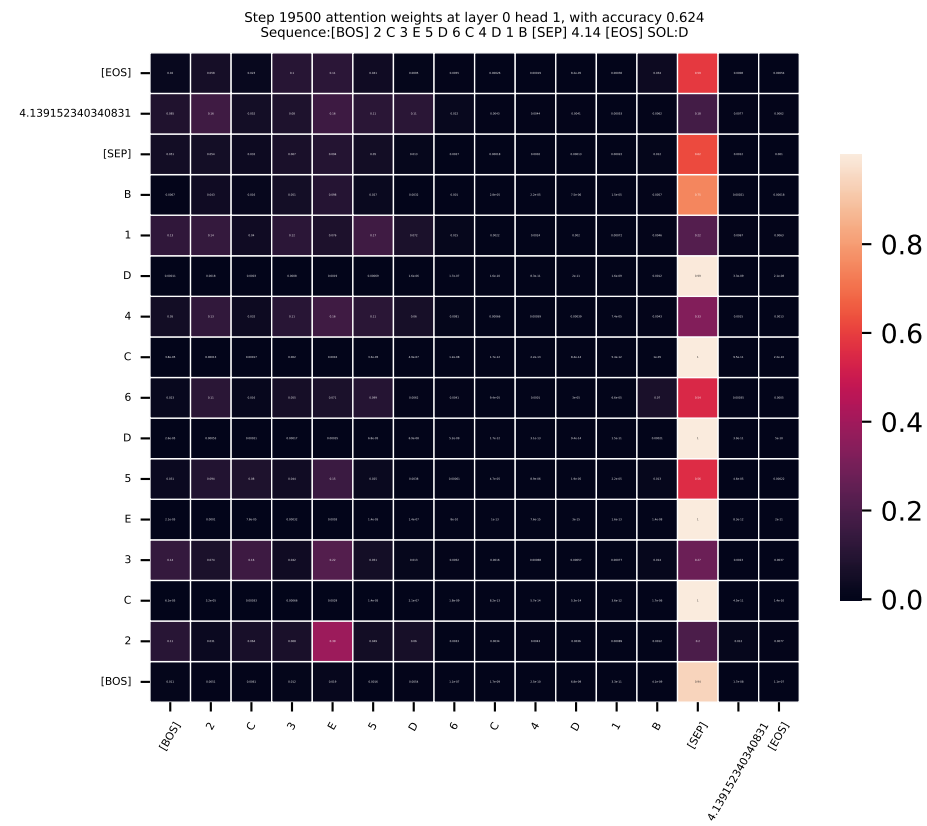

Step 10000 attention weights at layer 0 head 0, with accuracy 0.607
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.8
0.6
0.4
0.2
0.0

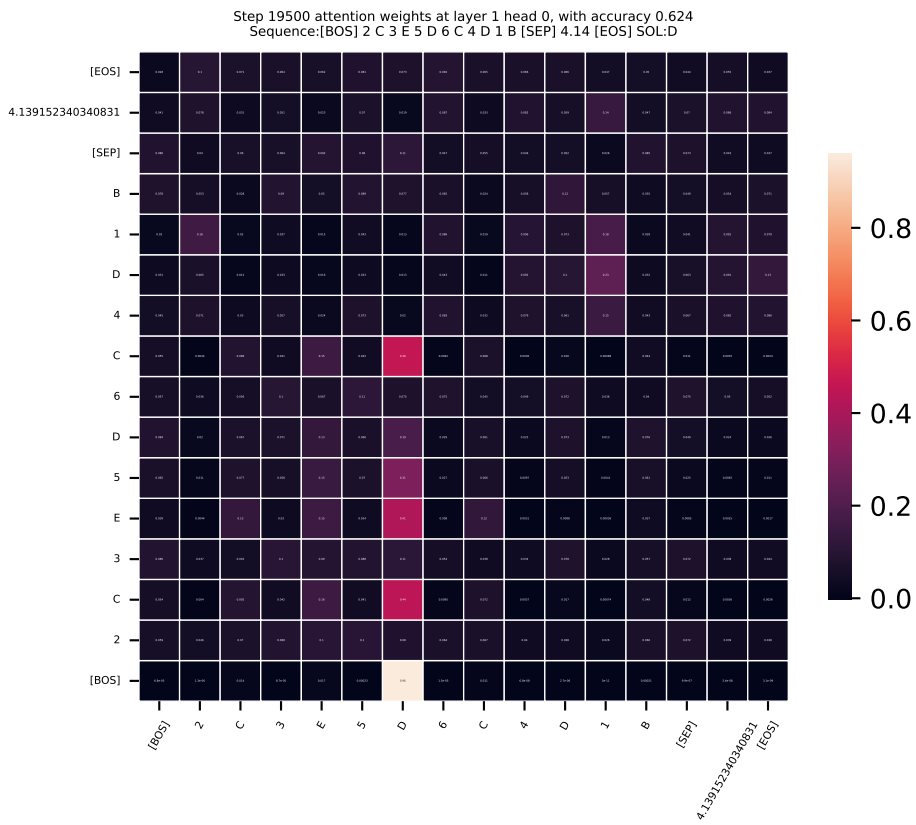

Step 10000 attention weights at layer 0 head 1, with accuracy 0.607
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.8
0.6
0.4
0.2

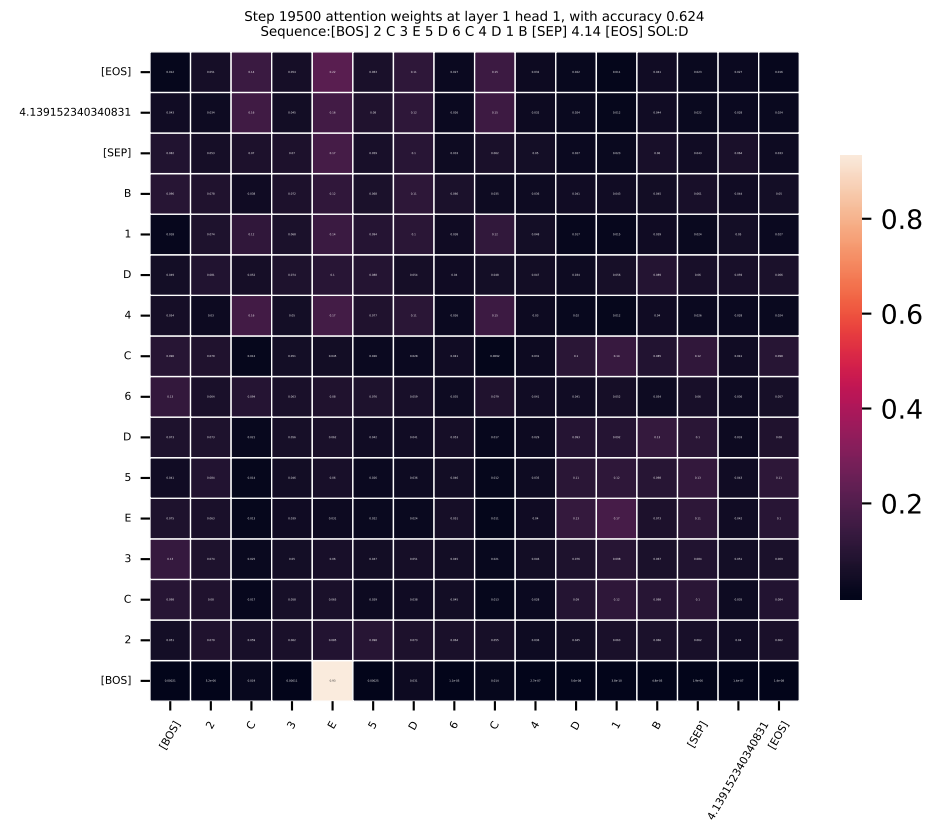

Step 10000 attention weights at layer 1 head 0, with accuracy 0.607
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2

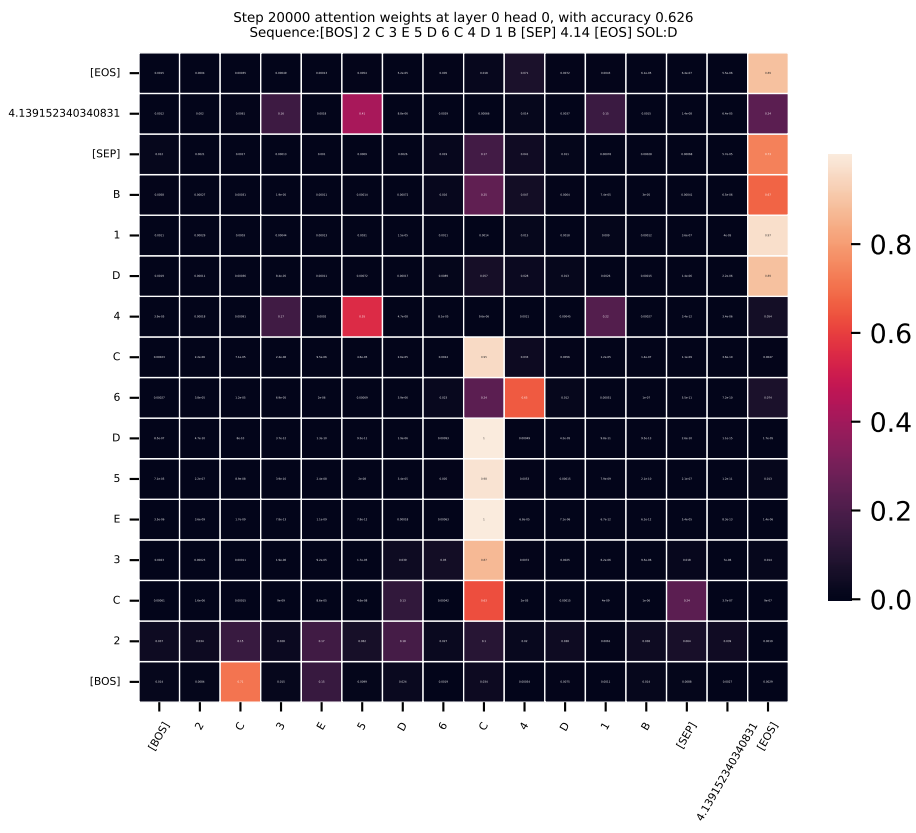

Step 10000 attention weights at layer 1 head 1, with accuracy 0.607
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2

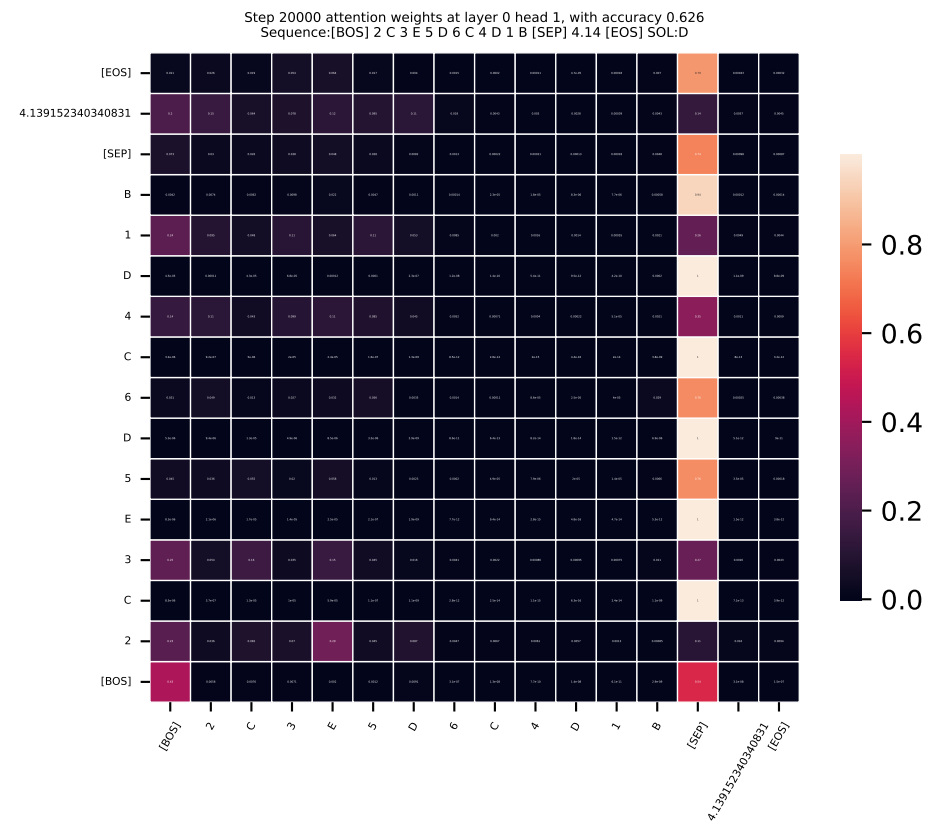

Step 19500 attention weights at layer 0 head 0, with accuracy 0.624
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.8
0.6
0.4
0.2
0.0

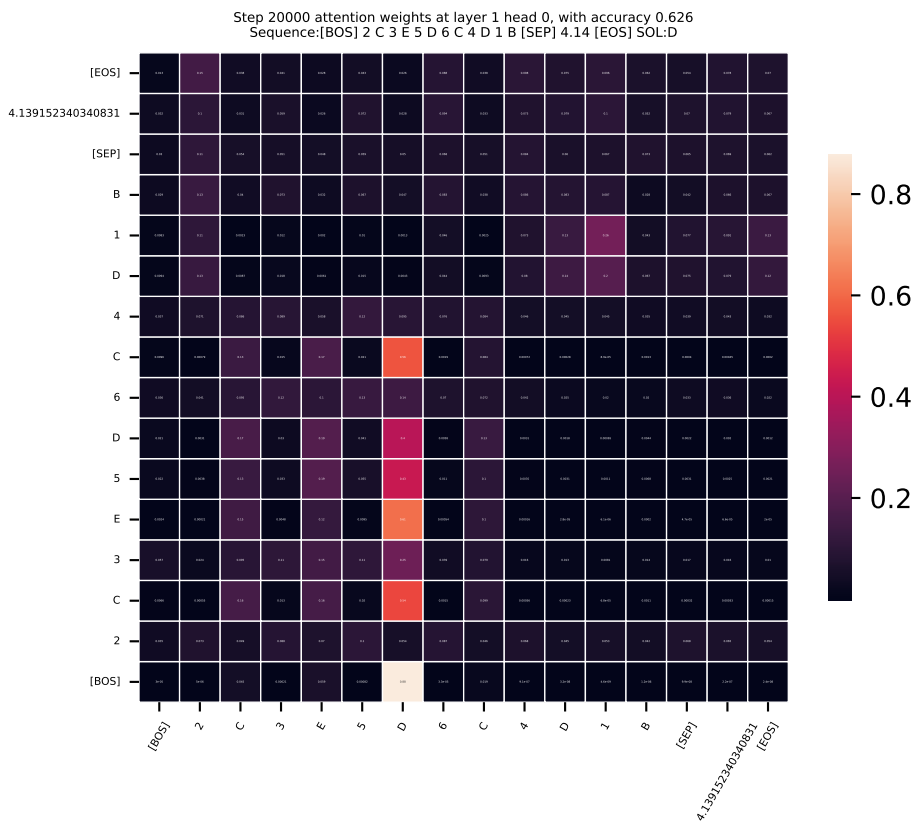

Step 19500 attention weights at layer 0 head 1, with accuracy 0.624
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.8
0.6
0.4
0.2
0.0

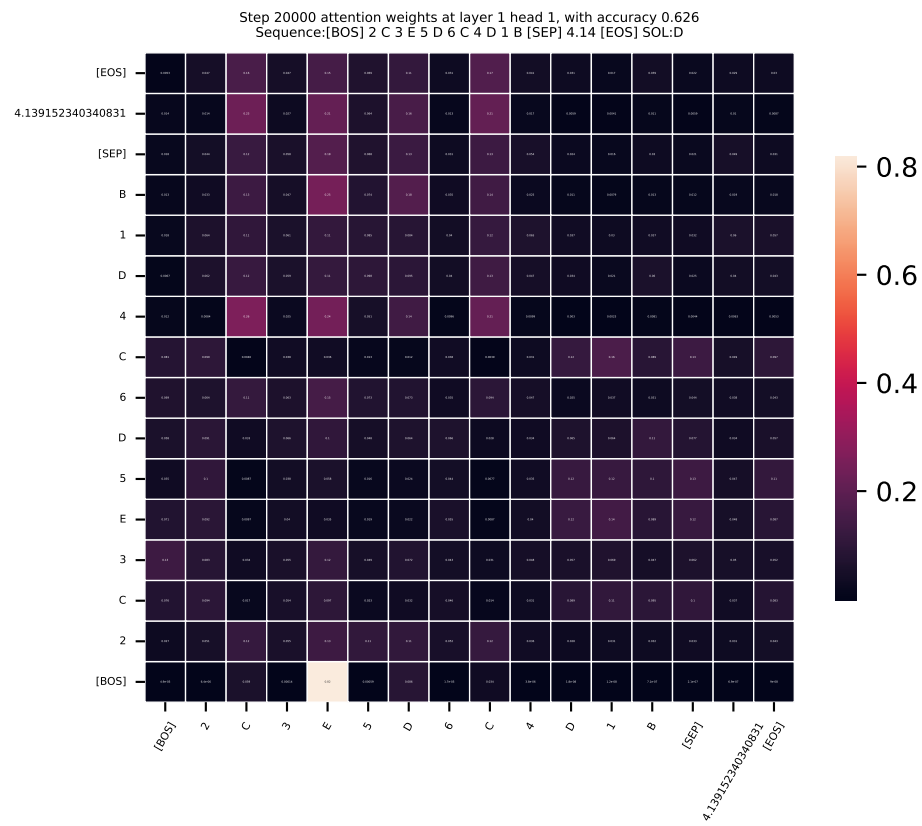

Step 19500 attention weights at layer 1 head 0, with accuracy 0.624
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.8
0.6
0.4
0.2
0.0

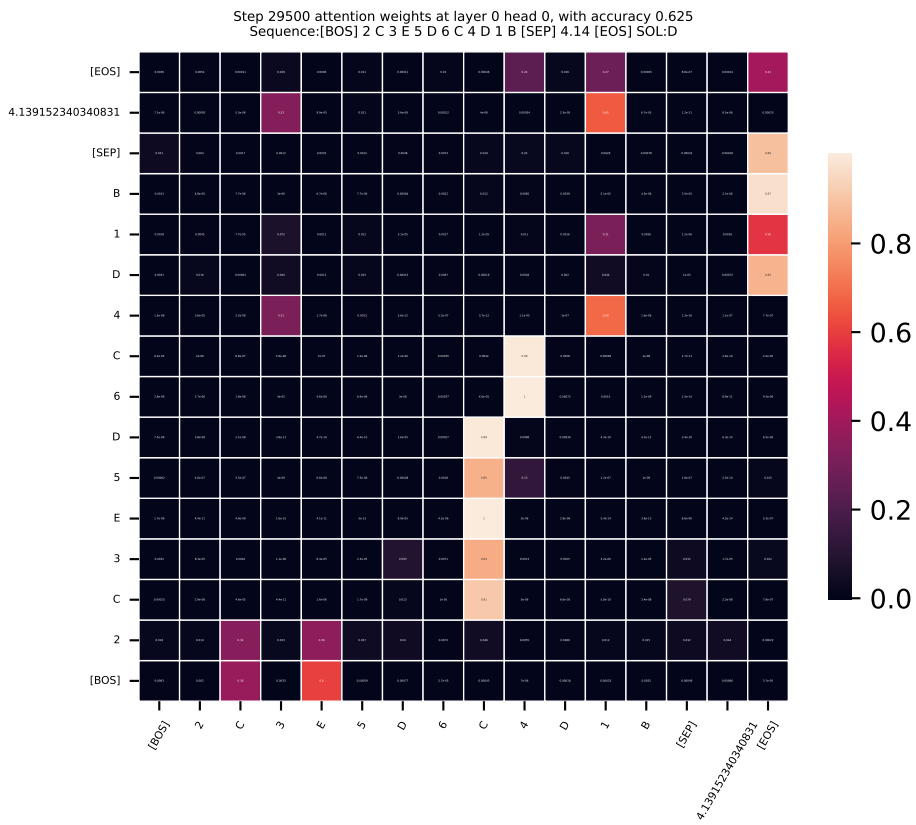

Step 19500 attention weights at layer 1 head 1, with accuracy 0.624
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.8
0.6
0.4
0.2

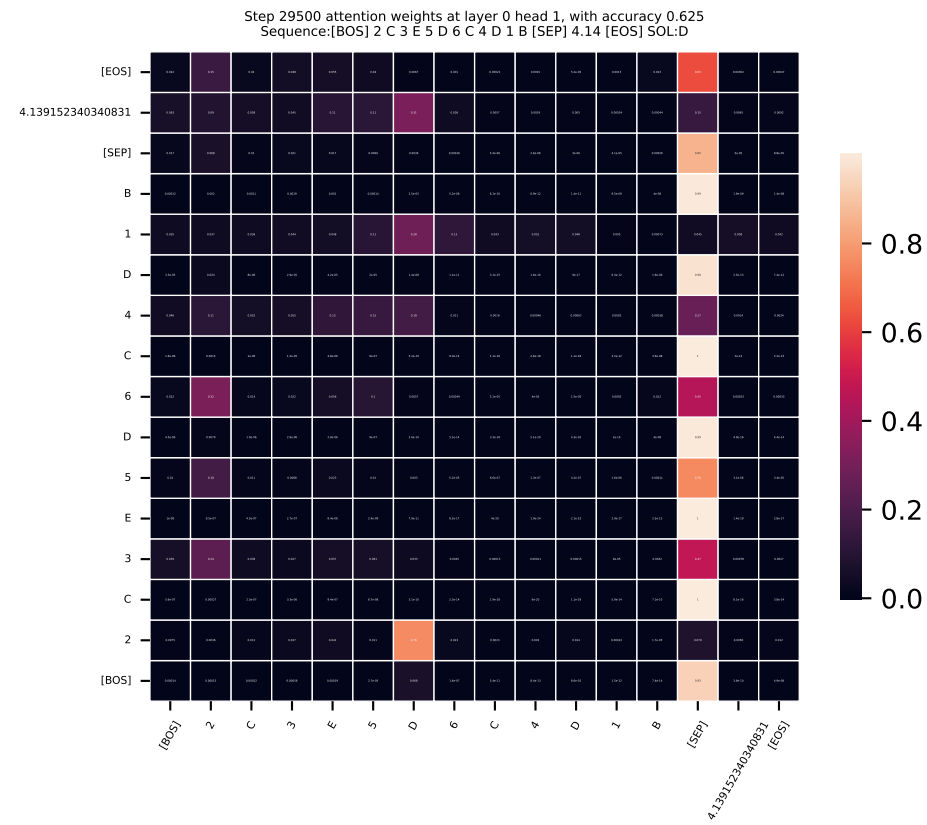

Step 20000 attention weights at layer 0 head 0, with accuracy 0.626
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.8
0.6
0.4
0.2
0.0

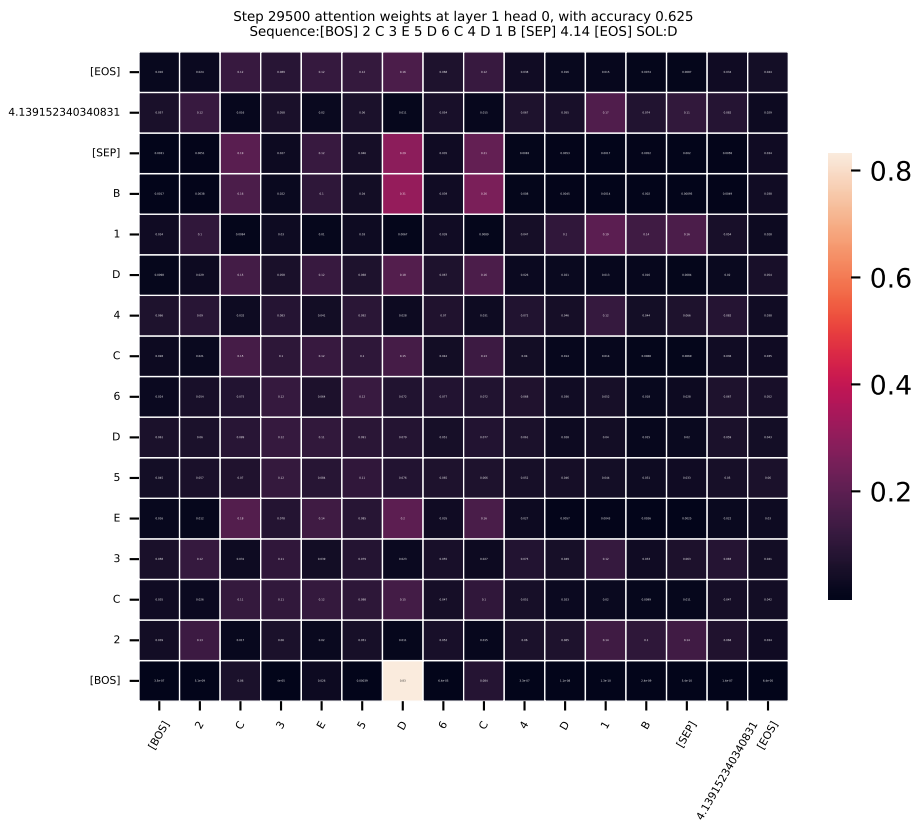

Step 20000 attention weights at layer 0 head 1, with accuracy 0.626
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.8
0.6
0.4
0.2
0.0

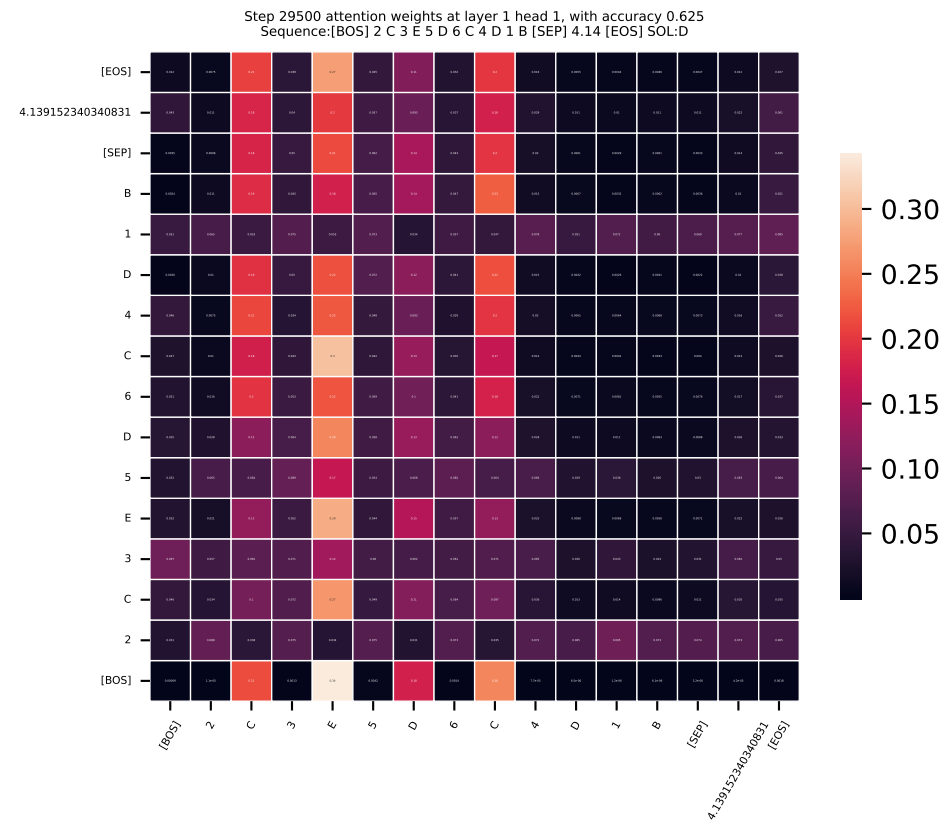

Step 20000 attention weights at layer 1 head 0, with accuracy 0.626
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2

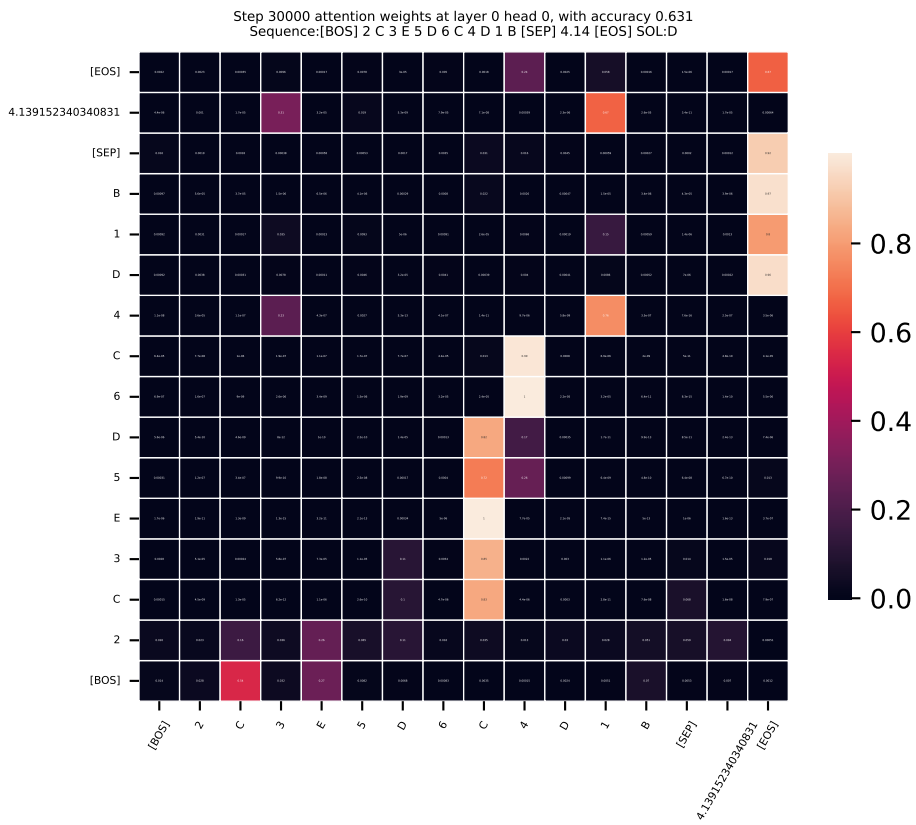

Step 20000 attention weights at layer 1 head 1, with accuracy 0.626
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2

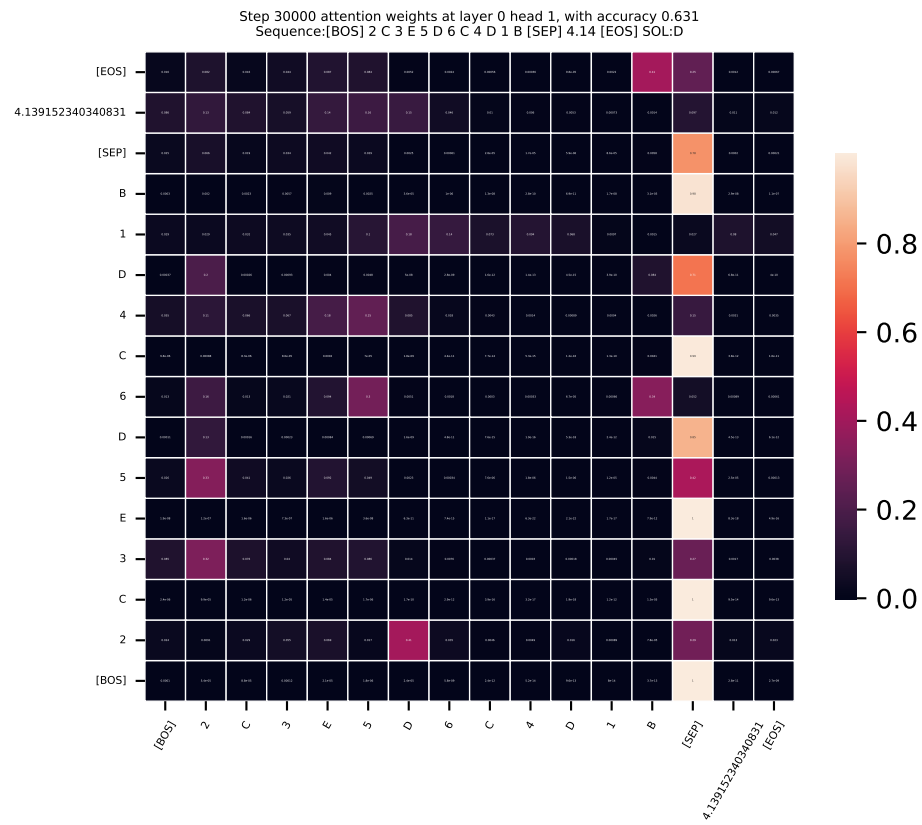

Step 29500 attention weights at layer 0 head 0, with accuracy 0.625
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2
0.0

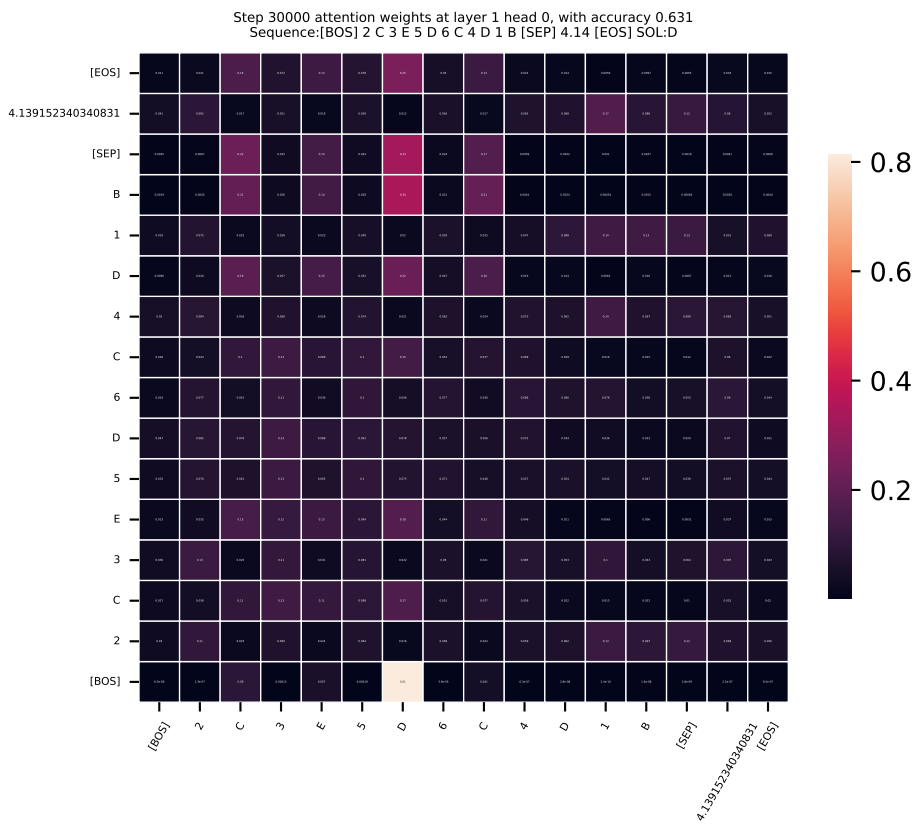

Step 29500 attention weights at layer 0 head 1, with accuracy 0.625
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2
0.0

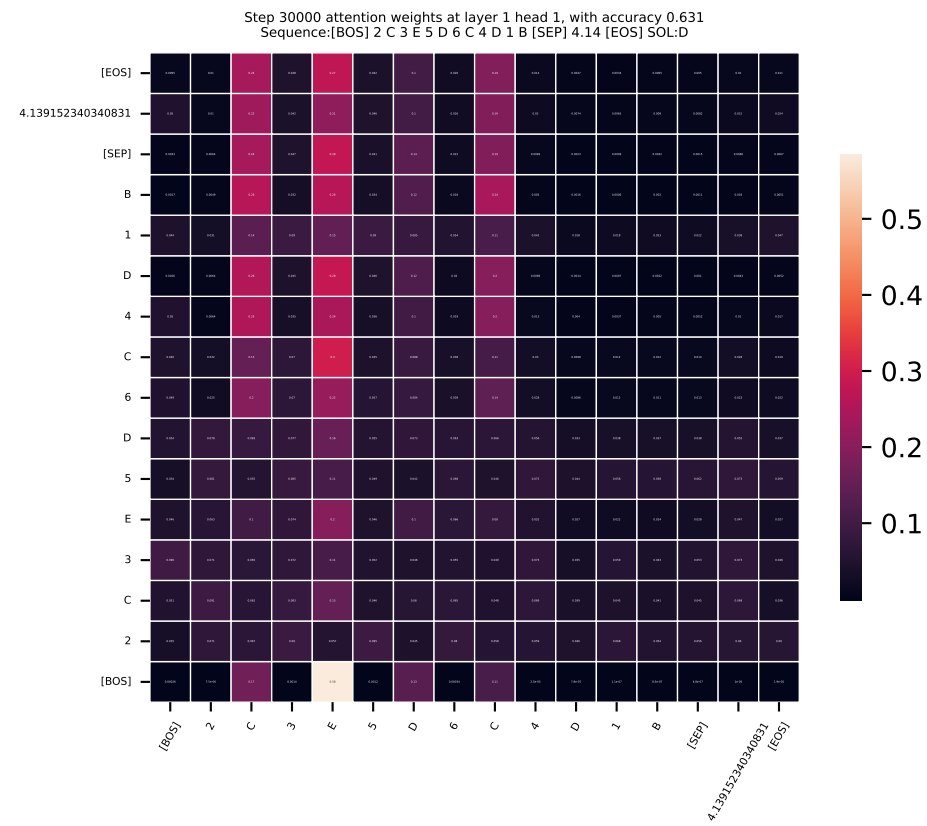

Step 29500 attention weights at layer 1 head 0, with accuracy 0.625
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.8
0.6
0.4
0.2

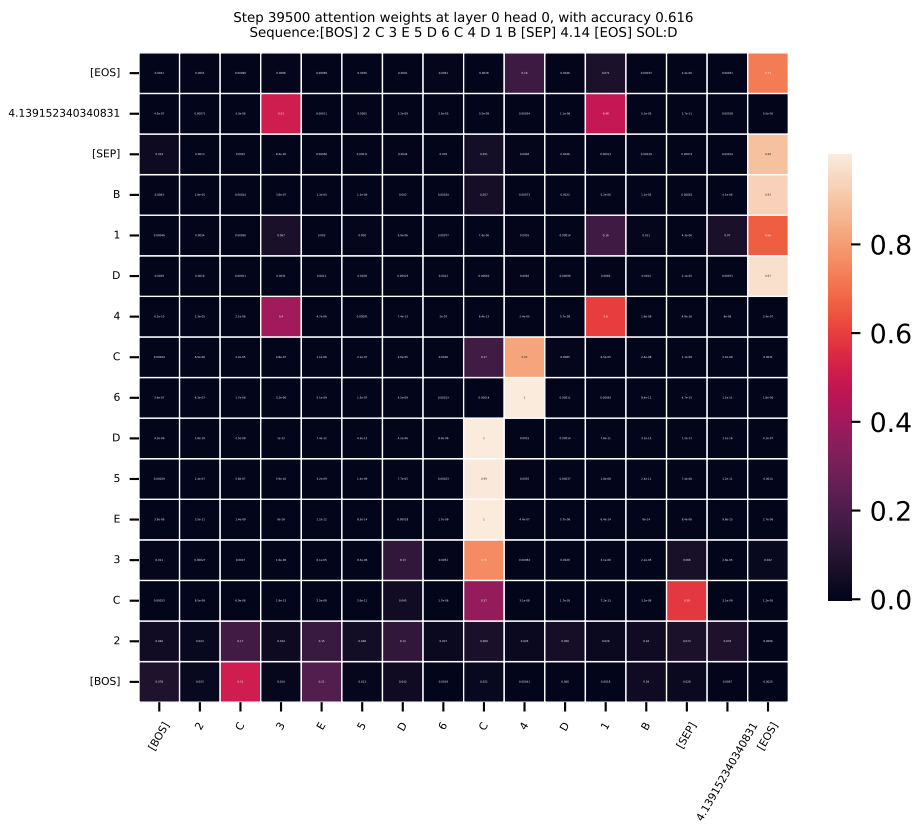

Step 29500 attention weights at layer 1 head 1, with accuracy 0.625
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.30
0.25
0.20
0.15
0.10
0.05

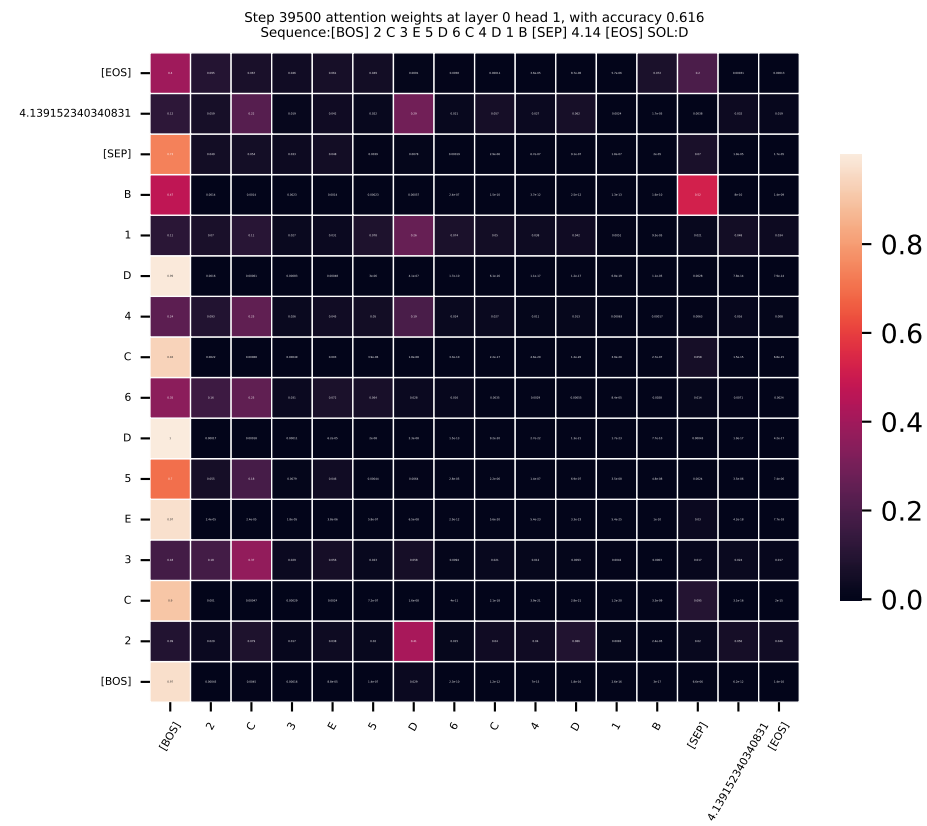

Step 30000 attention weights at layer 0 head 0, with accuracy 0.631
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2
0.0

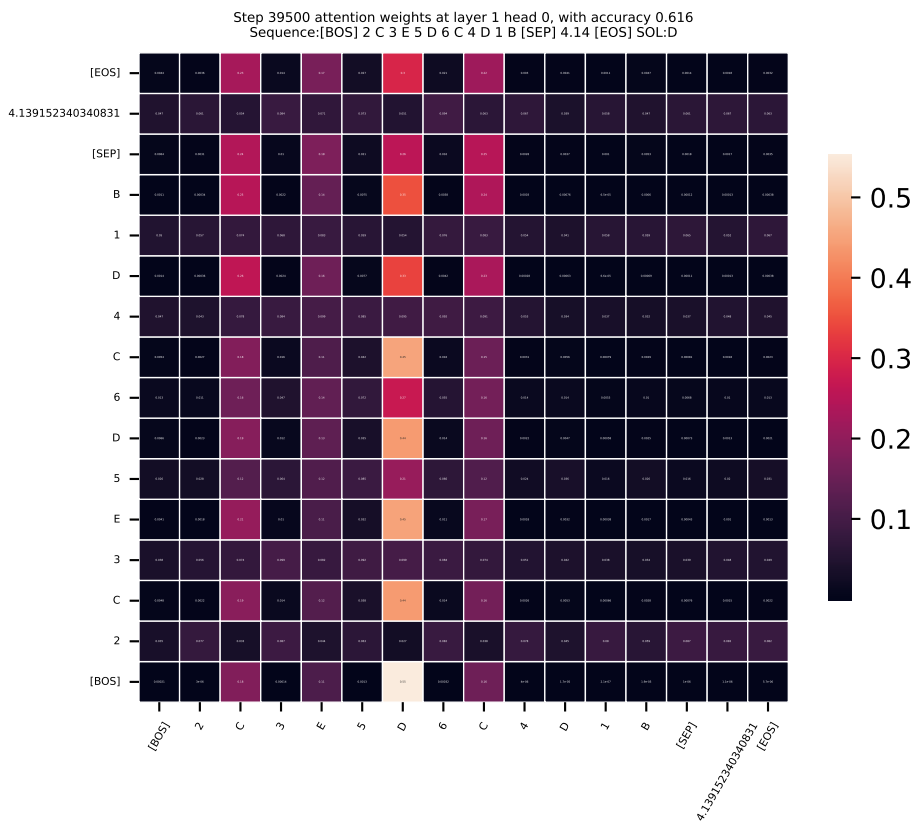

Step 30000 attention weights at layer 0 head 1, with accuracy 0.631
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2
0.0

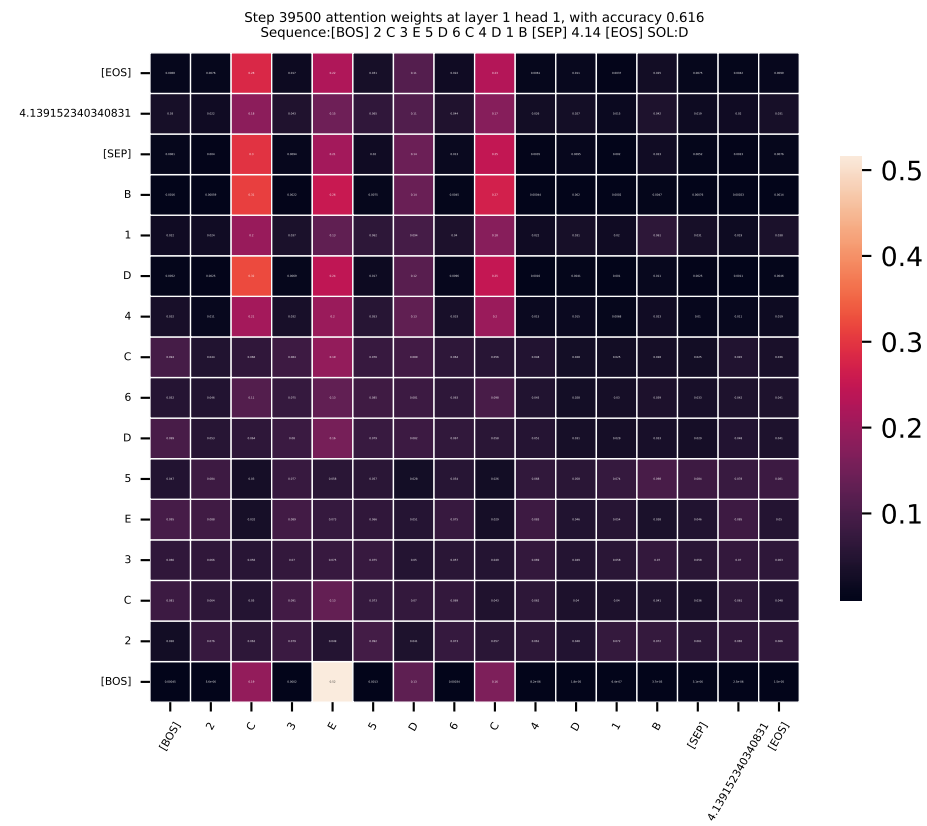

Step 30000 attention weights at layer 1 head 0, with accuracy 0.631
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2

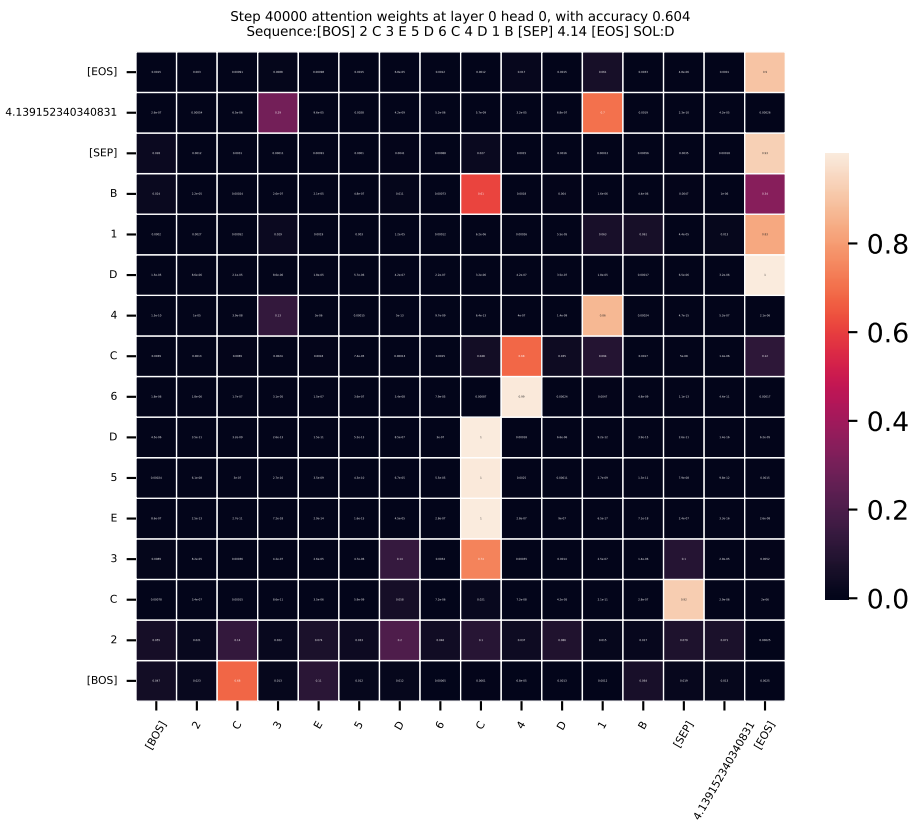

Step 30000 attention weights at layer 1 head 1, with accuracy 0.631
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.5
0.4
0.3
0.2
0.1

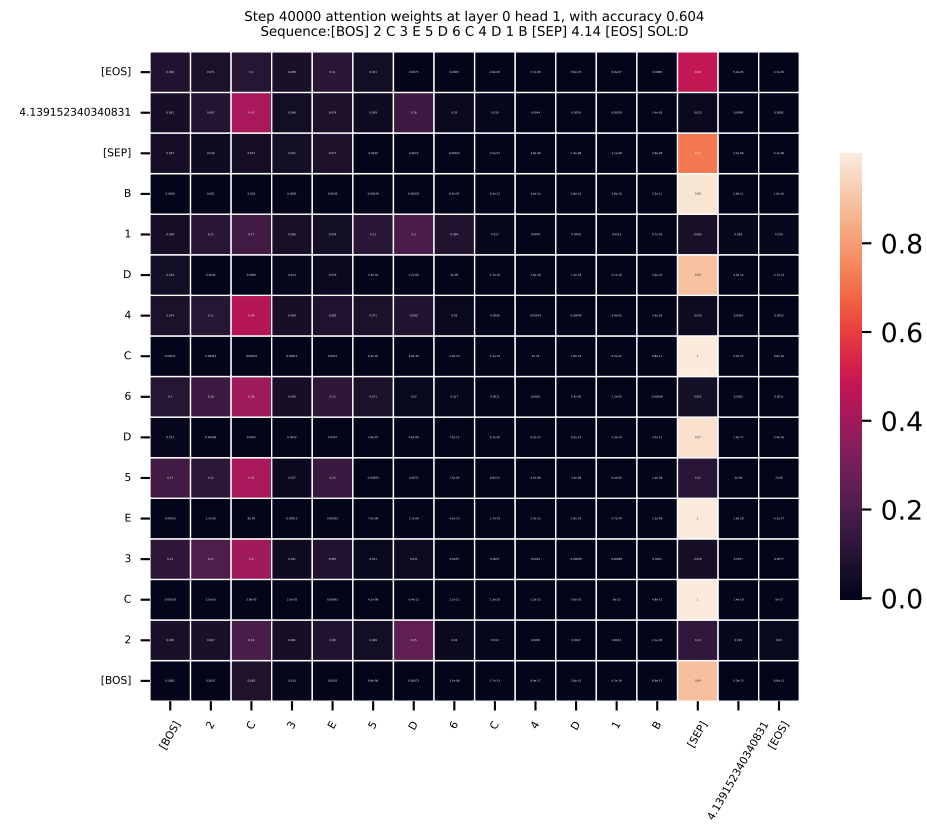

Step 39500 attention weights at layer 0 head 0, with accuracy 0.616
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2
0.0

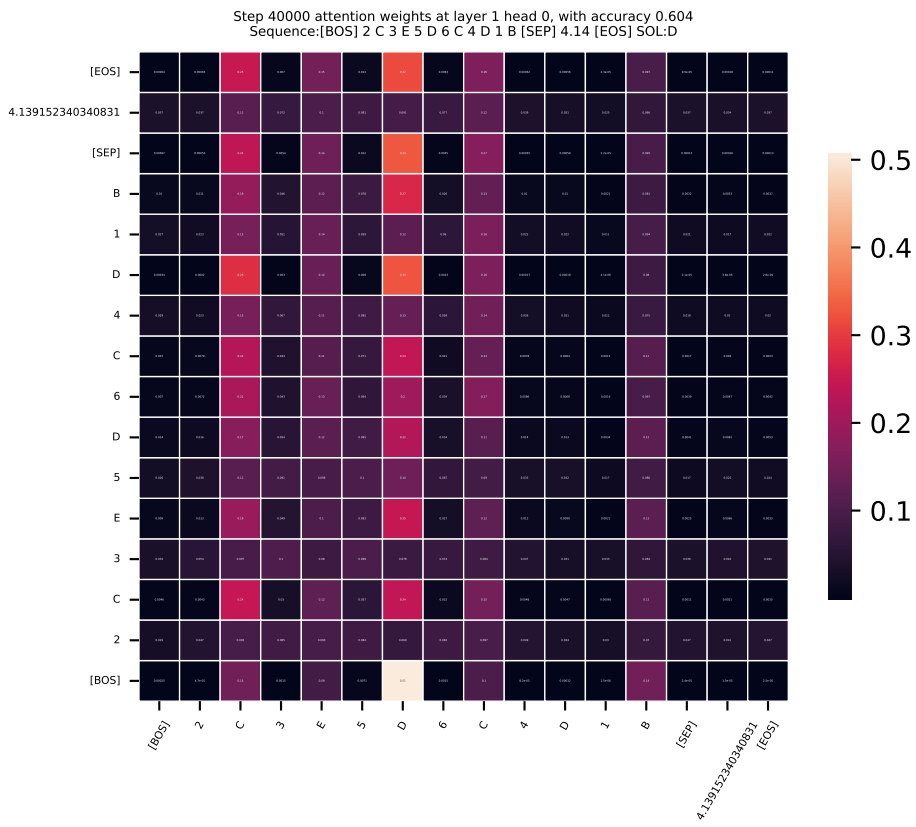

Step 39500 attention weights at layer 0 head 1, with accuracy 0.616
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2
0.0

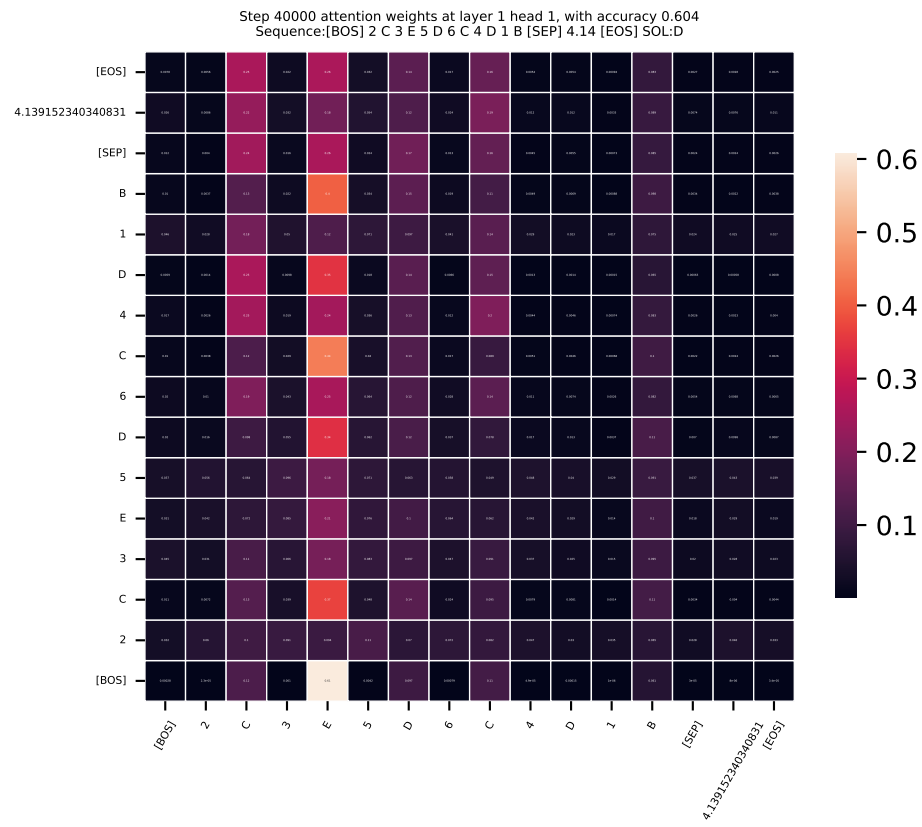

Step 39500 attention weights at layer 1 head 0, with accuracy 0.616
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.5
0.4
0.3
0.2
0.1

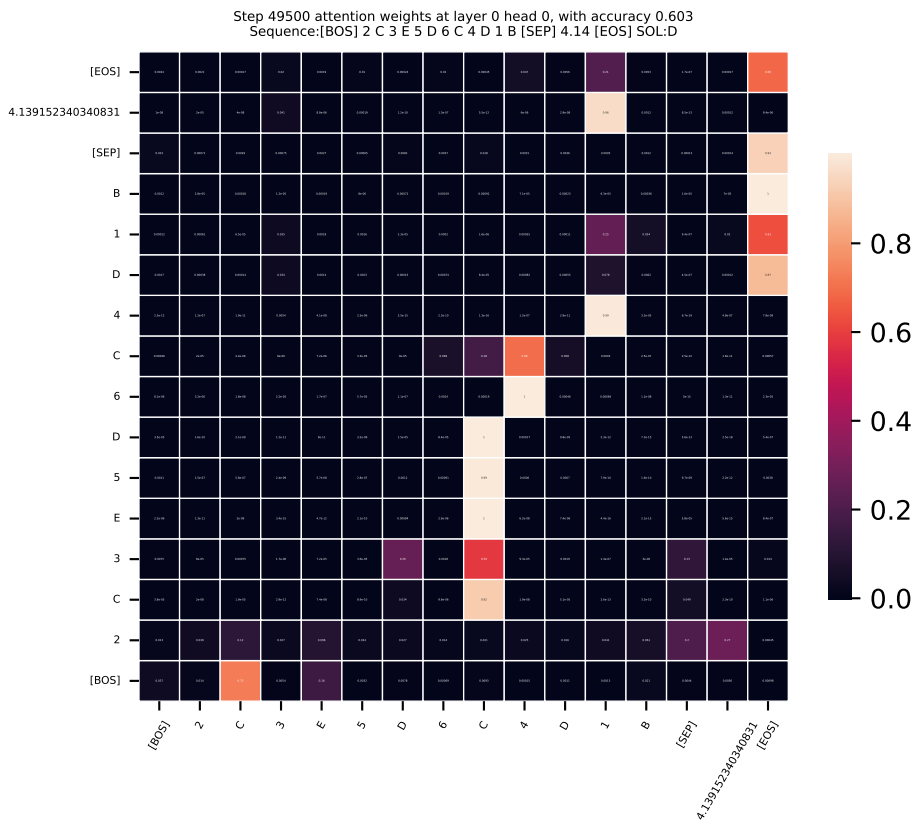

Step 39500 attention weights at layer 1 head 1, with accuracy 0.616
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.5
0.4
0.3
0.2
0.1

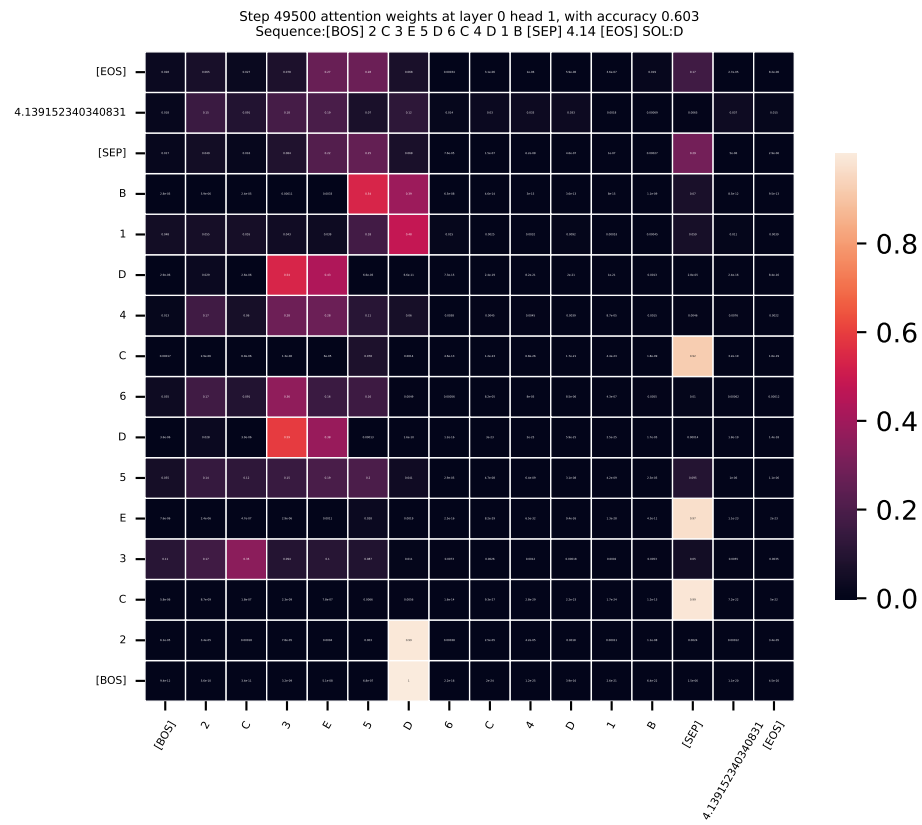

Step 40000 attention weights at layer 0 head 0, with accuracy 0.604
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2
0.0

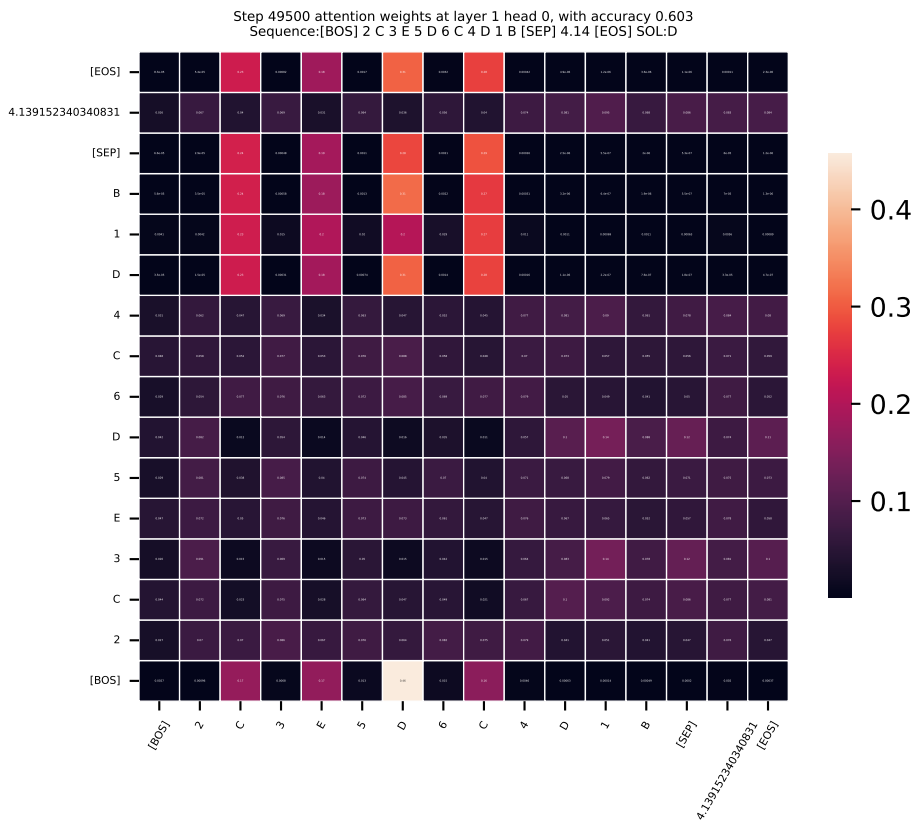

Step 40000 attention weights at layer 0 head 1, with accuracy 0.604
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2
0.0

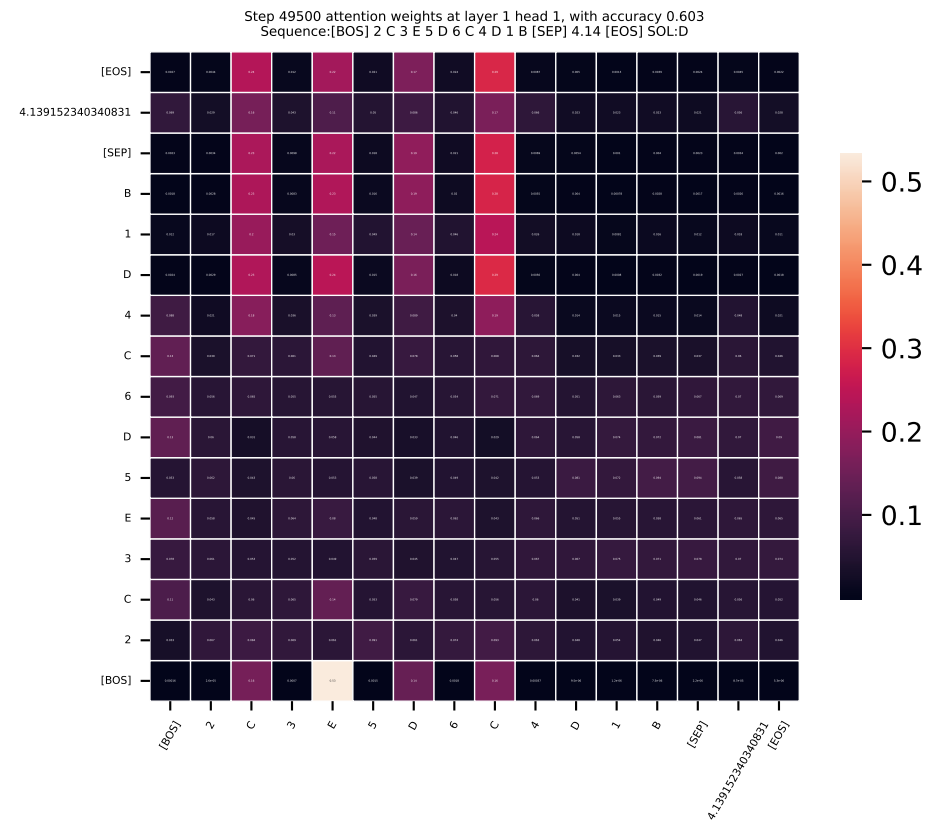

Step 40000 attention weights at layer 1 head 0, with accuracy 0.604
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.5
0.4
0.3
0.2
0.1

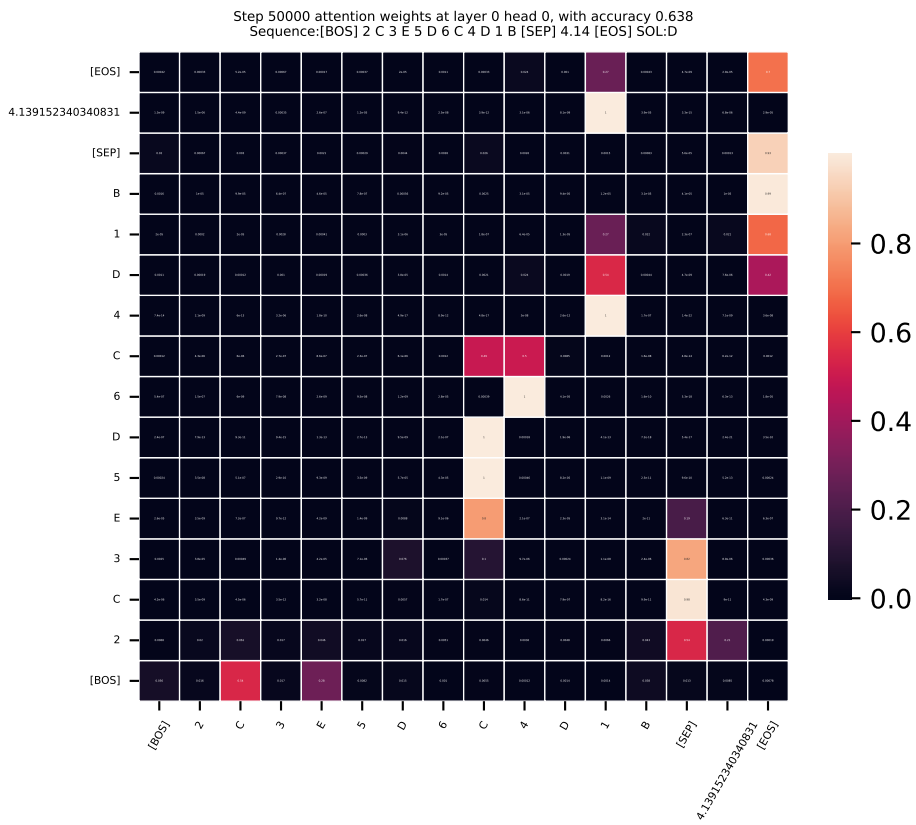

Step 40000 attention weights at layer 1 head 1, with accuracy 0.604
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.6
0.5
0.4
0.3
0.2
0.1

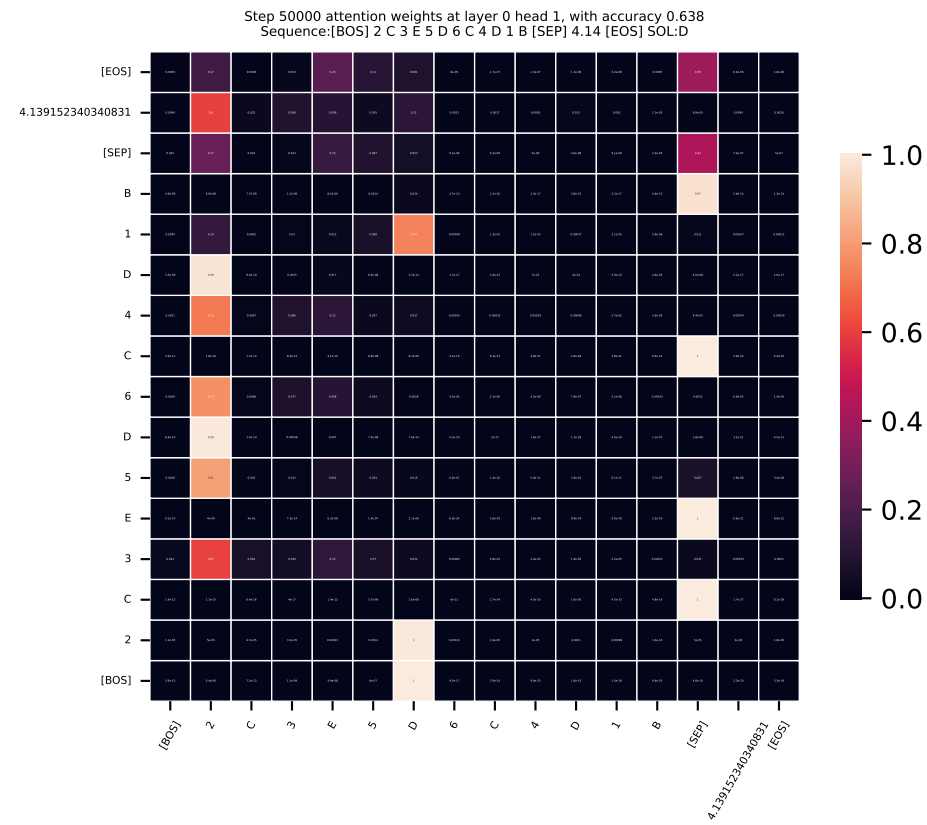

Step 49500 attention weights at layer 0 head 0, with accuracy 0.603
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2
0.0

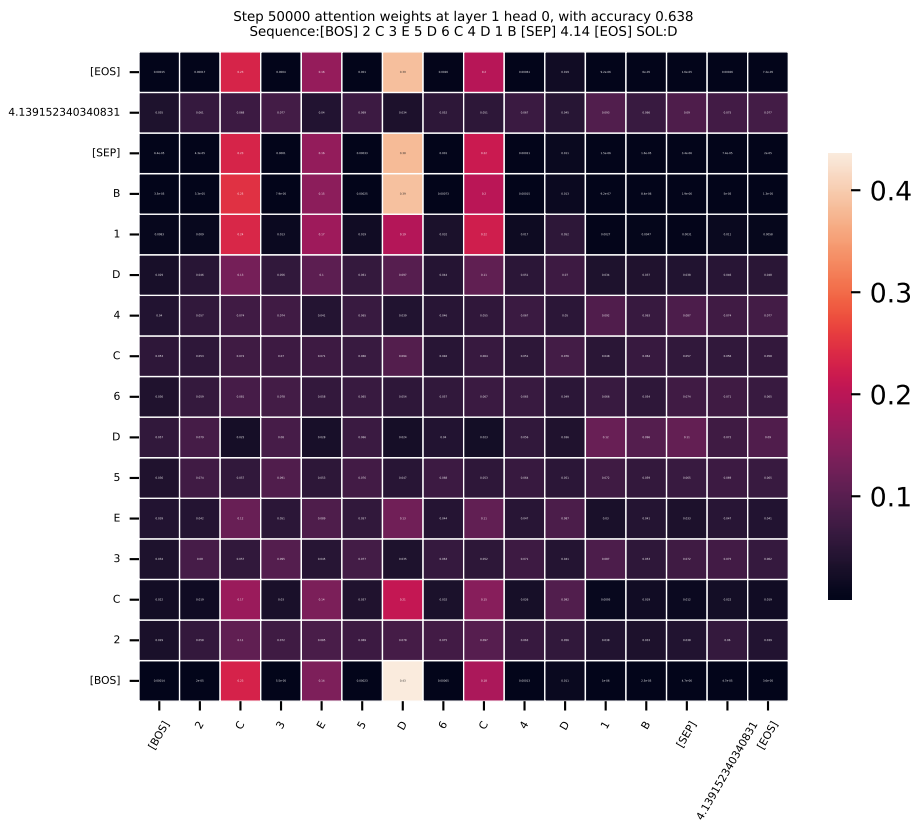

Step 49500 attention weights at layer 0 head 1, with accuracy 0.603
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2
0.0

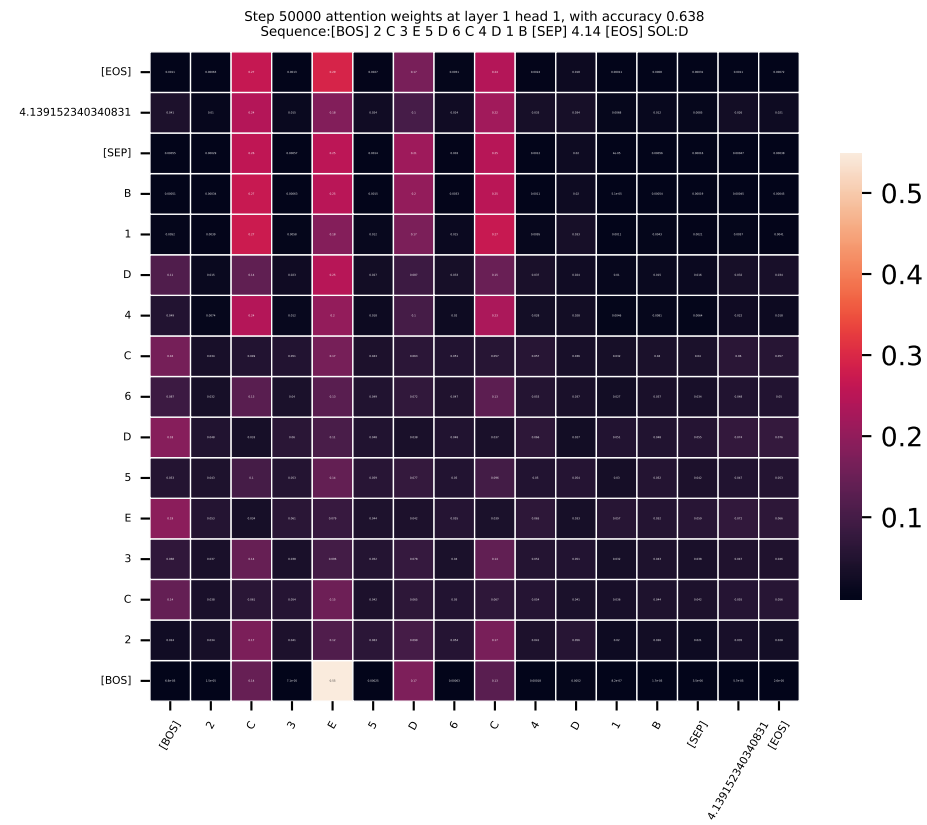

Step 49500 attention weights at layer 1 head 0, with accuracy 0.603
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.4
0.3
0.2
0.1

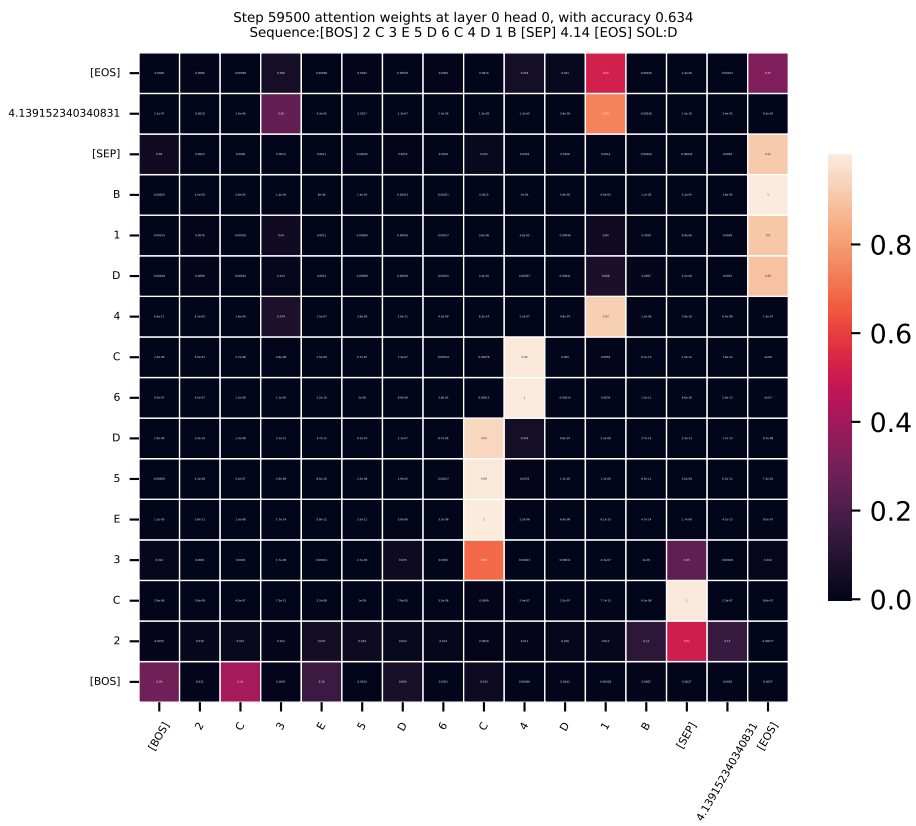

Step 49500 attention weights at layer 1 head 1, with accuracy 0.603
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.5
0.4
0.3
0.2
0.1

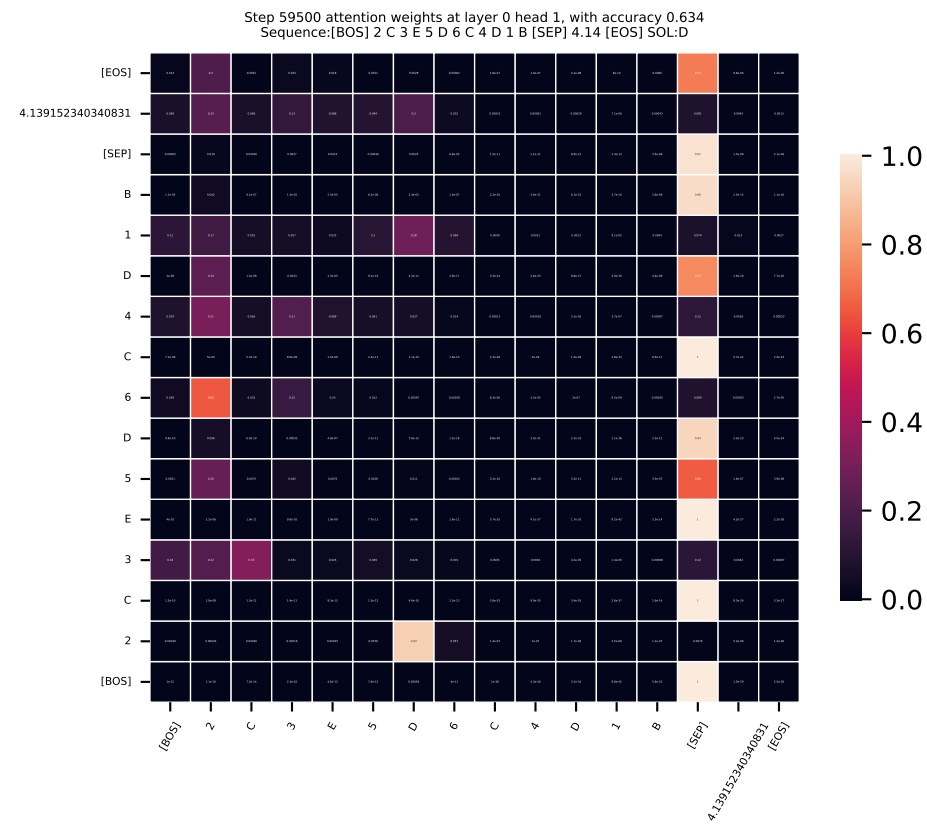

Step 50000 attention weights at layer 0 head 0, with accuracy 0.638
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2
0.0

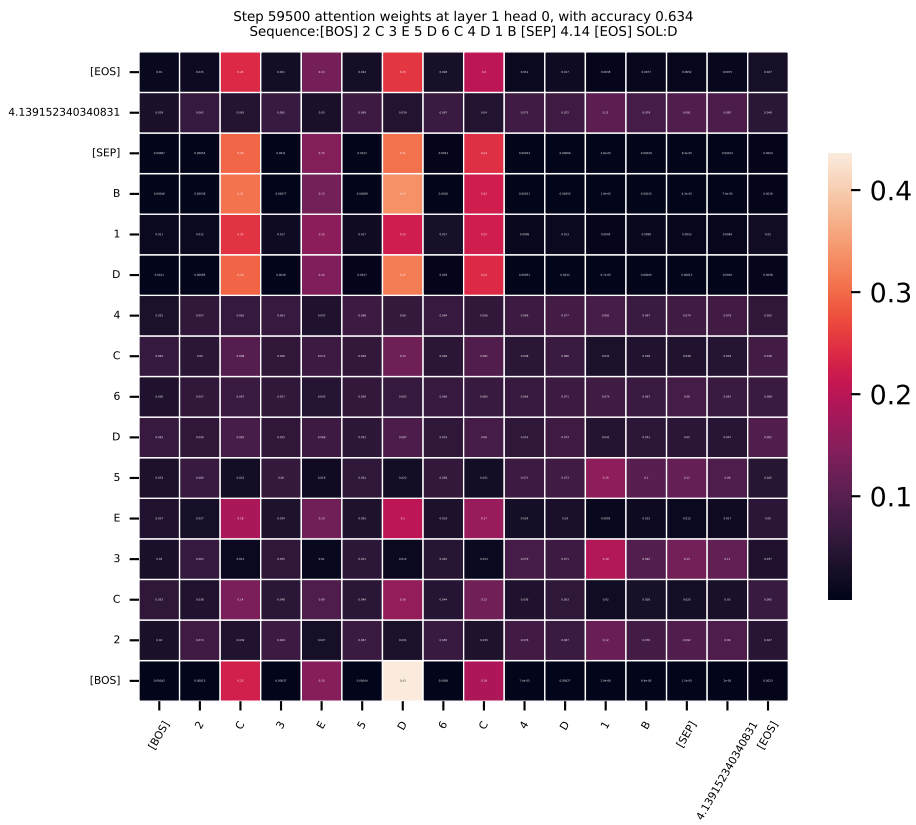

Step 50000 attention weights at layer 0 head 1, with accuracy 0.638
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

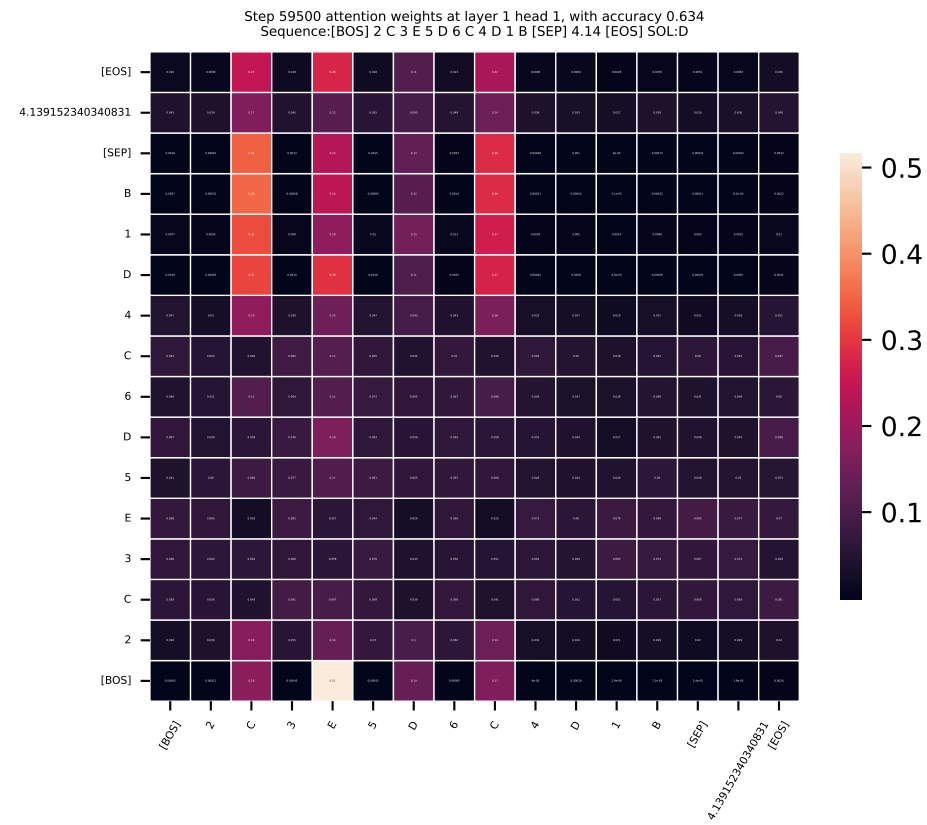

Step 50000 attention weights at layer 1 head 0, with accuracy 0.638
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.4
0.3
0.2
0.1

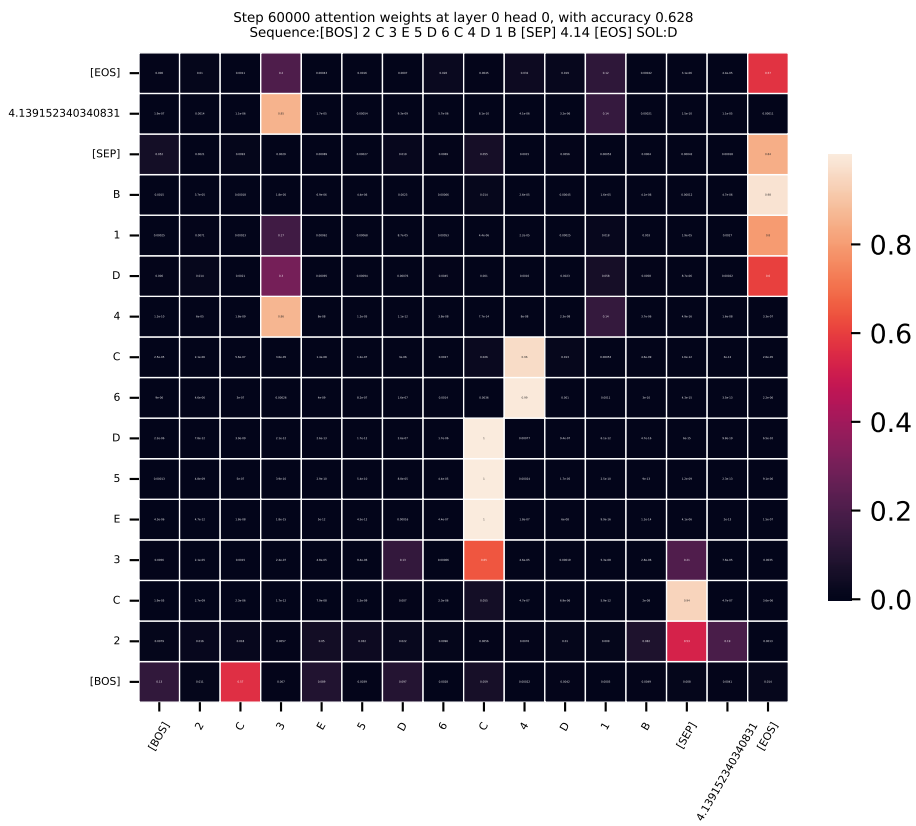

Step 50000 attention weights at layer 1 head 1, with accuracy 0.638
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.5
0.4
0.3
0.2
0.1

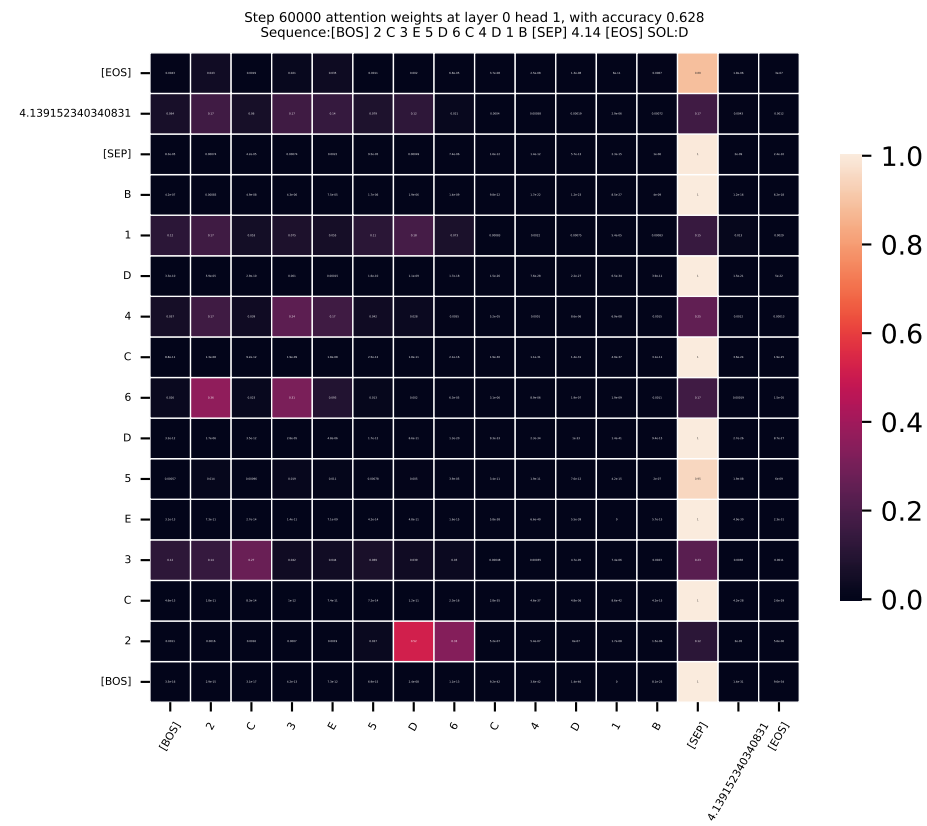

Step 59500 attention weights at layer 0 head 0, with accuracy 0.634
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2
0.0

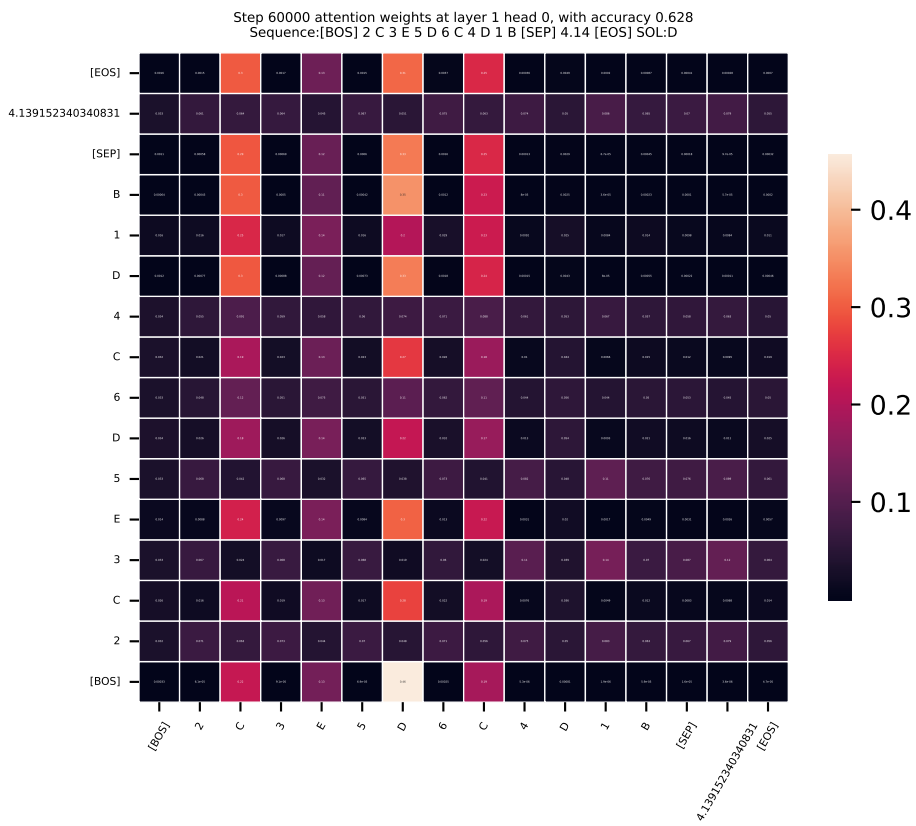

Step 59500 attention weights at layer 0 head 1, with accuracy 0.634
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

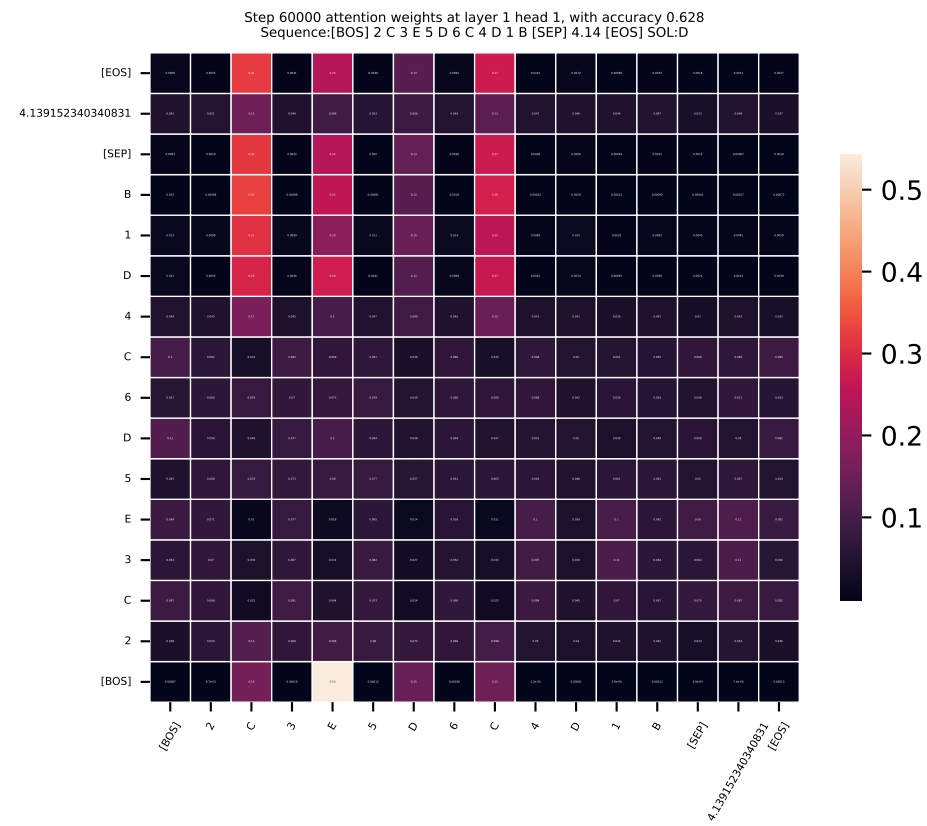

Step 59500 attention weights at layer 1 head 0, with accuracy 0.634
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.4
0.3
0.2
0.1

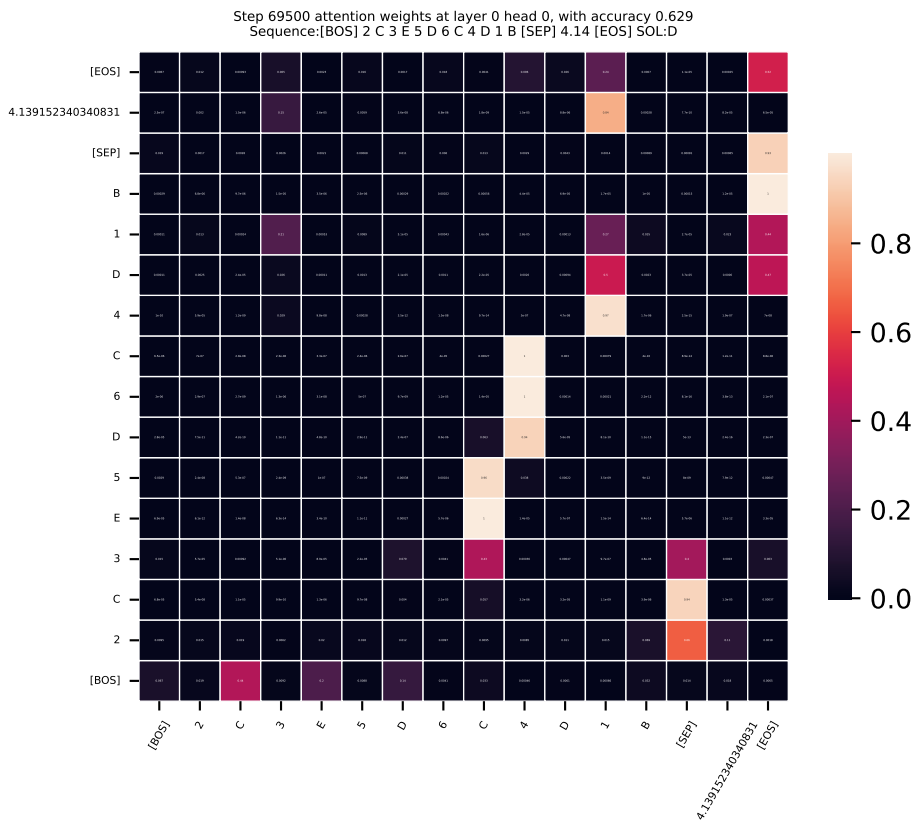

Step 59500 attention weights at layer 1 head 1, with accuracy 0.634
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.5
0.4
0.3
0.2
0.1

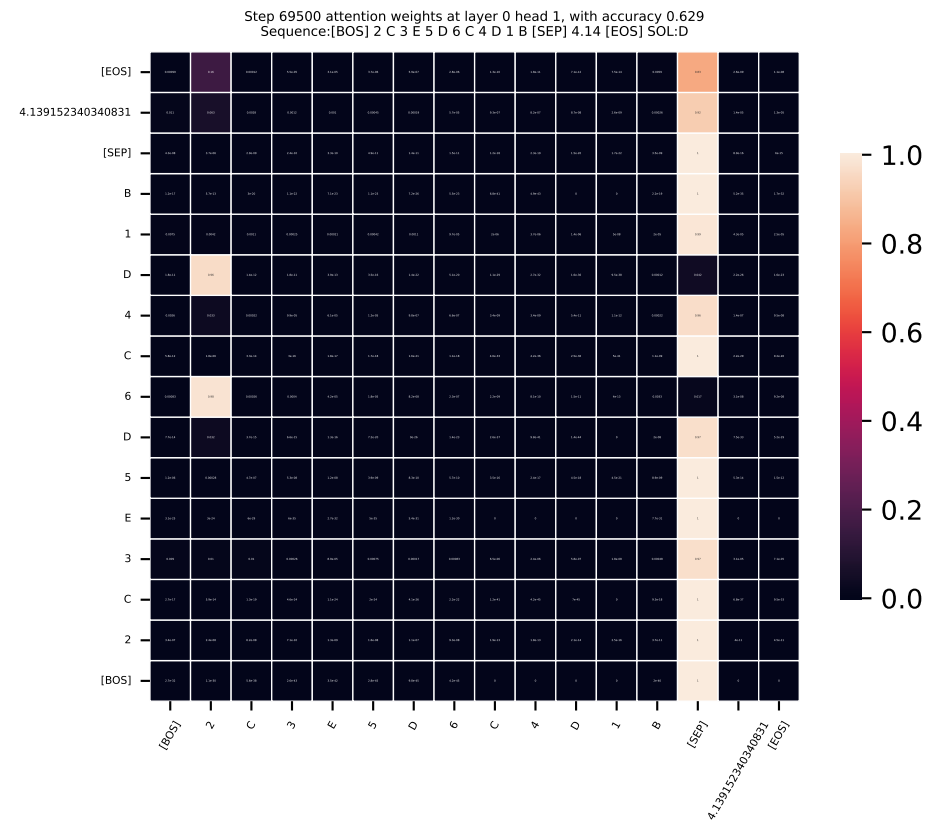

Step 60000 attention weights at layer 0 head 0, with accuracy 0.628
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2
0.0

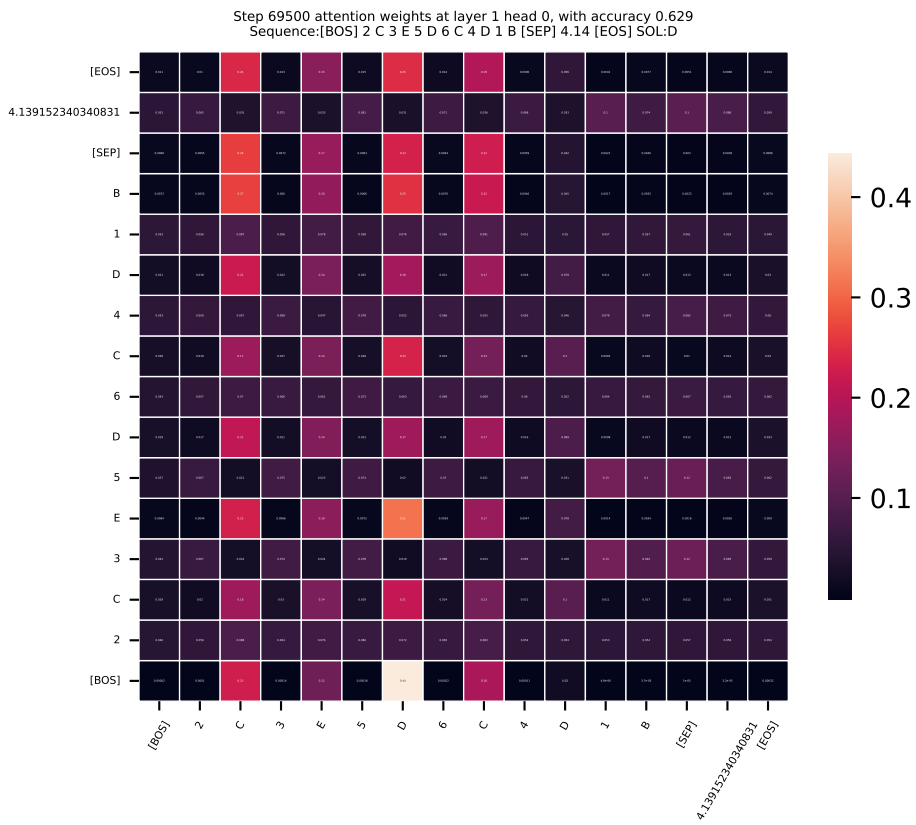

Step 60000 attention weights at layer 0 head 1, with accuracy 0.628
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

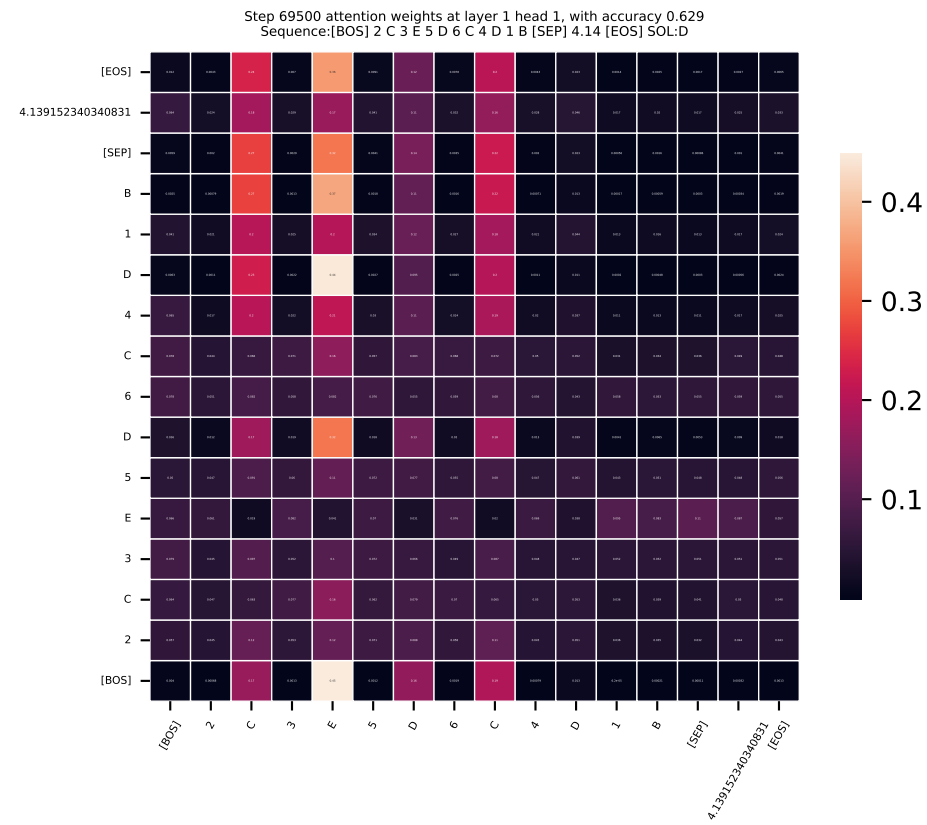

Step 60000 attention weights at layer 1 head 0, with accuracy 0.628
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.4
0.3
0.2
0.1

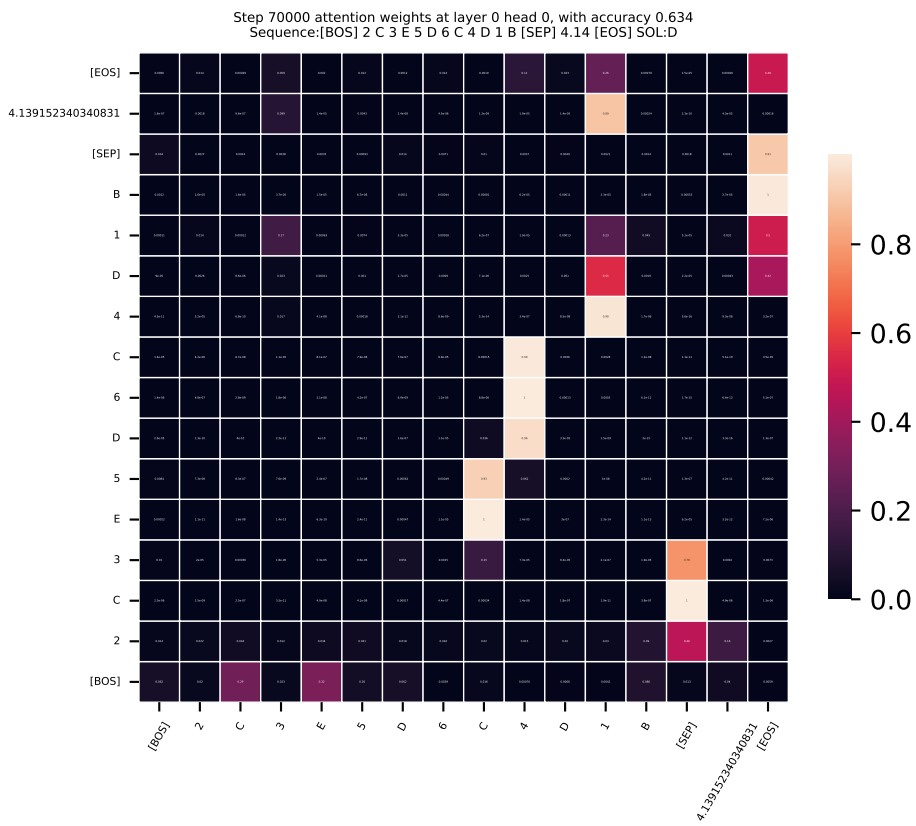

Step 60000 attention weights at layer 1 head 1, with accuracy 0.628
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.5
0.4
0.3
0.2
0.1

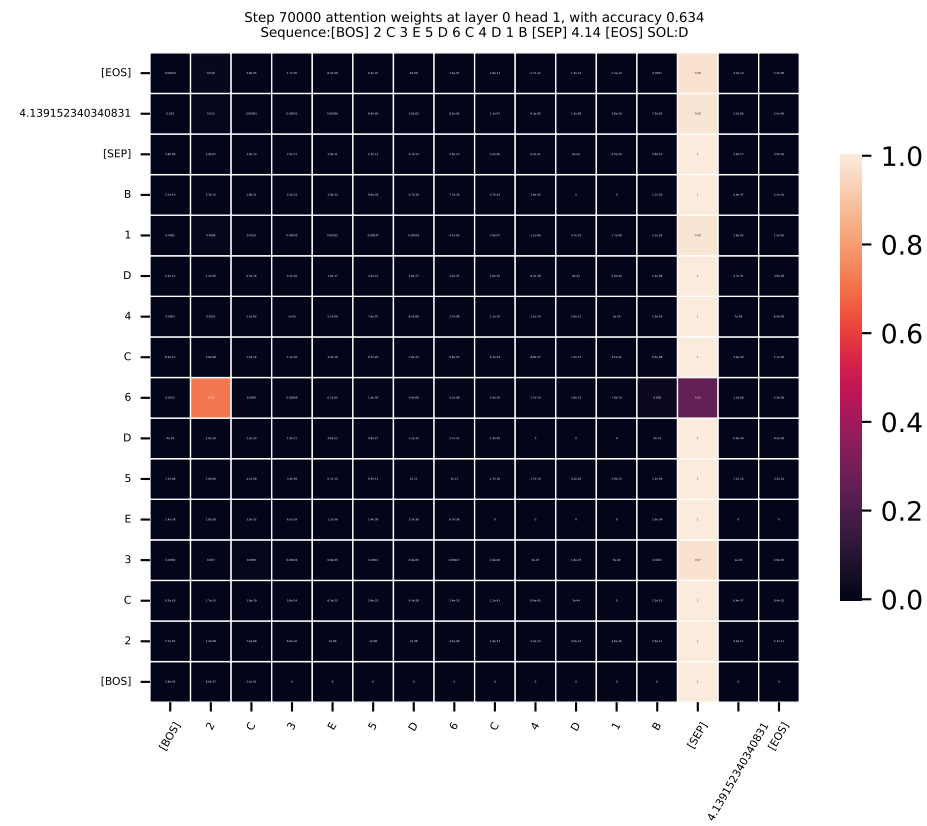

Step 69500 attention weights at layer 0 head 0, with accuracy 0.629
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2
0.0

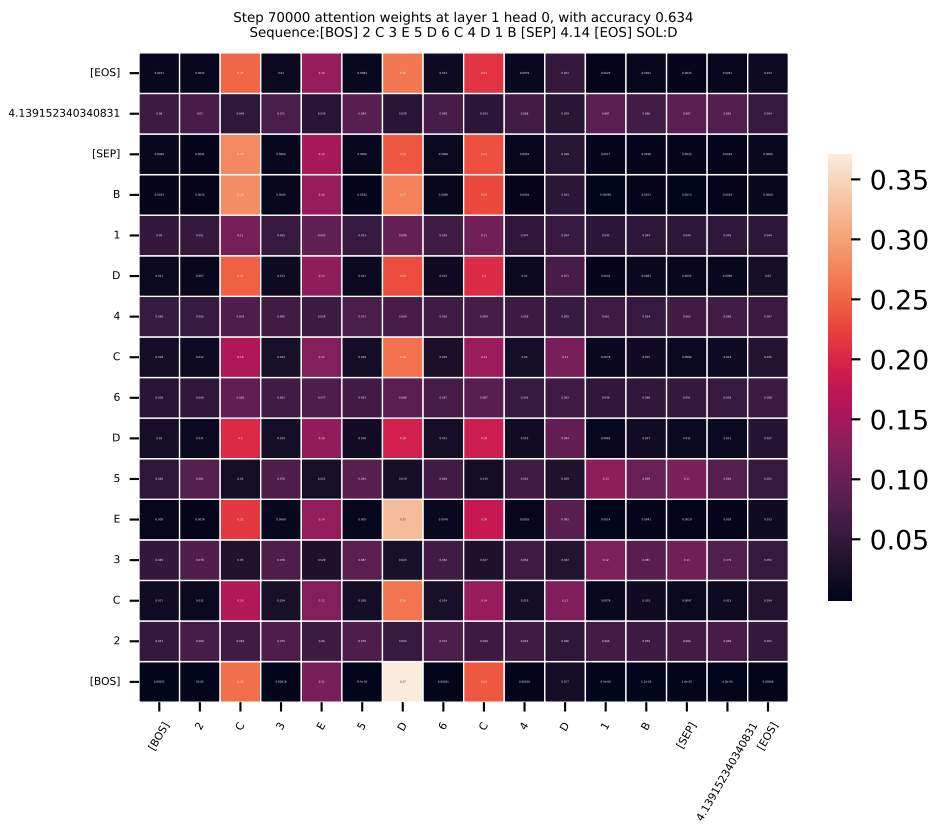

Step 69500 attention weights at layer 0 head 1, with accuracy 0.629
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

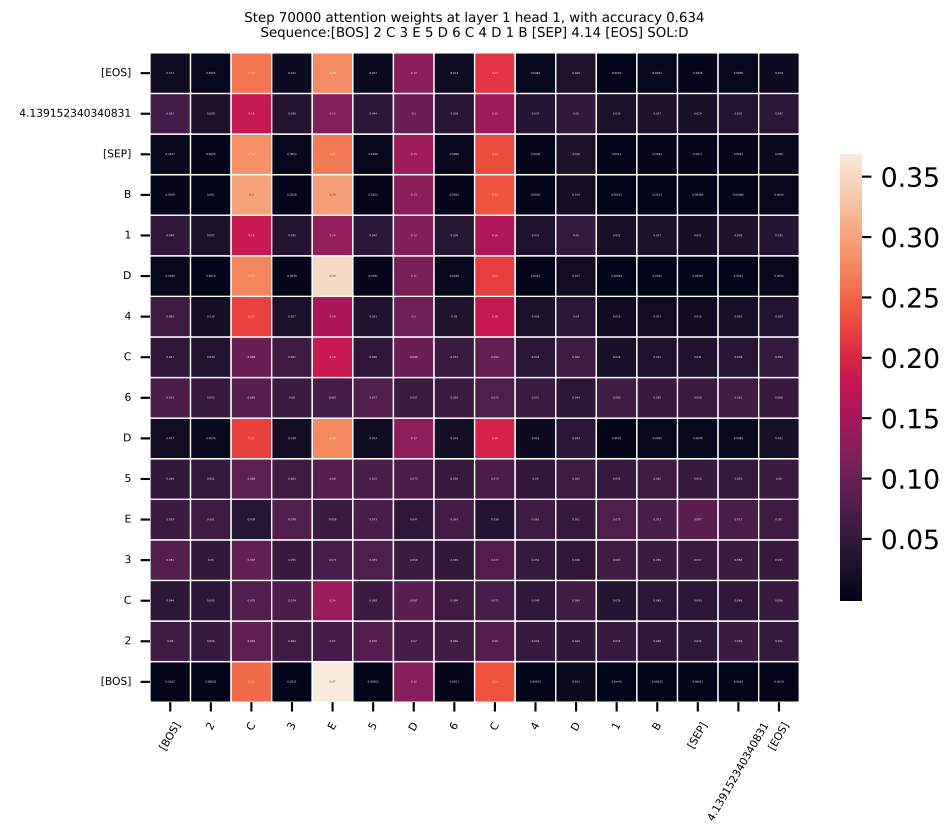

Step 69500 attention weights at layer 1 head 0, with accuracy 0.629
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.4
0.3
0.2
0.1

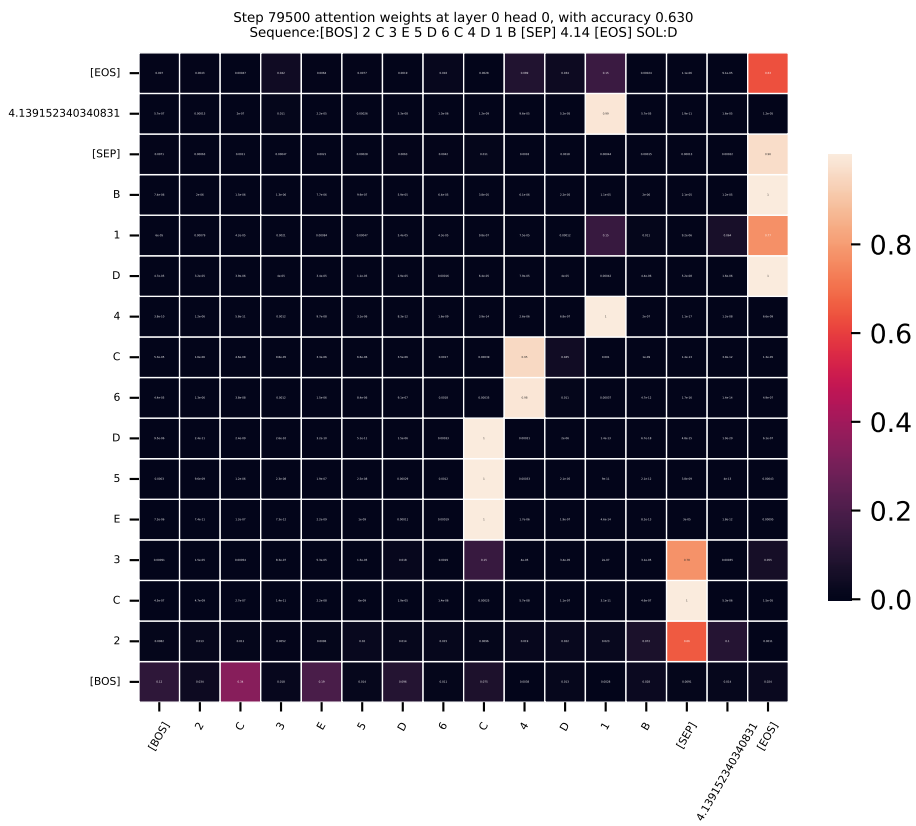

Step 69500 attention weights at layer 1 head 1, with accuracy 0.629
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.4
0.3
0.2
0.1

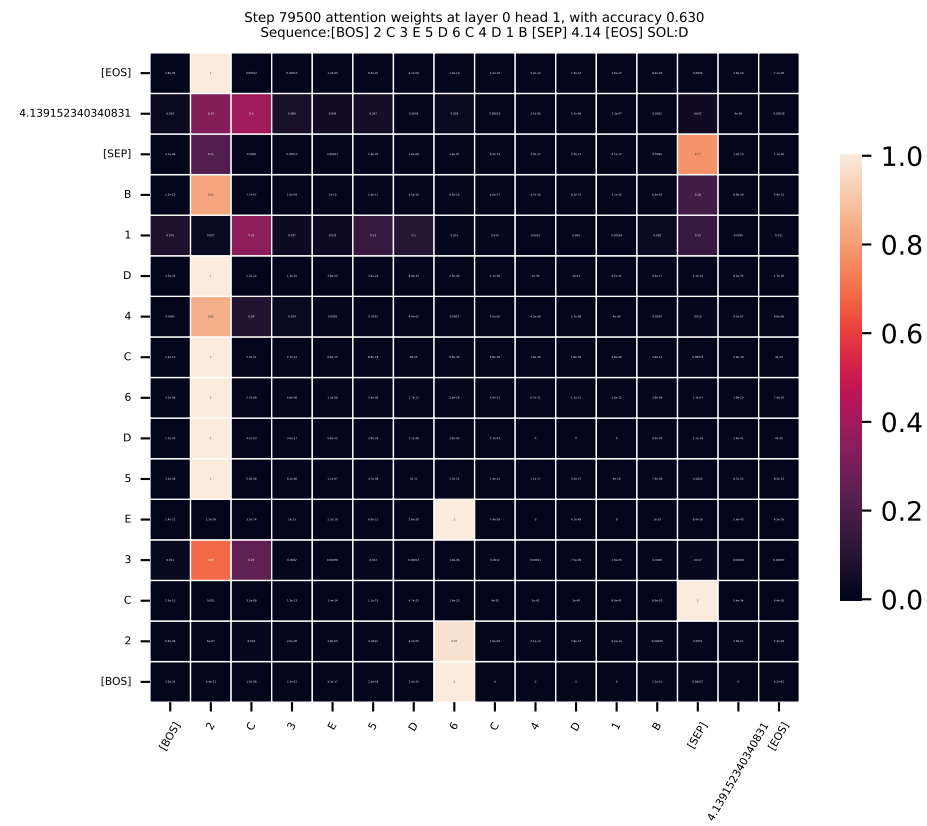

Step 70000 attention weights at layer 0 head 0, with accuracy 0.634
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.8
0.6
0.4
0.2
0.0

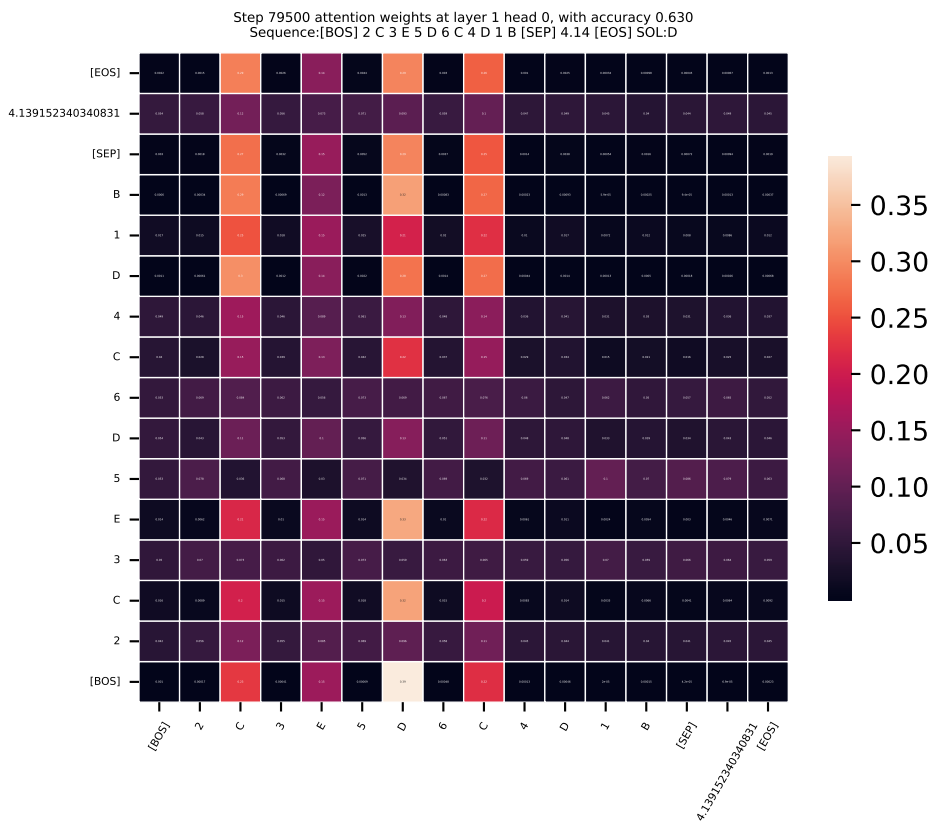

Step 70000 attention weights at layer 0 head 1, with accuracy 0.634
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
1.0
0.8
0.6
0.4
0.2
0.0

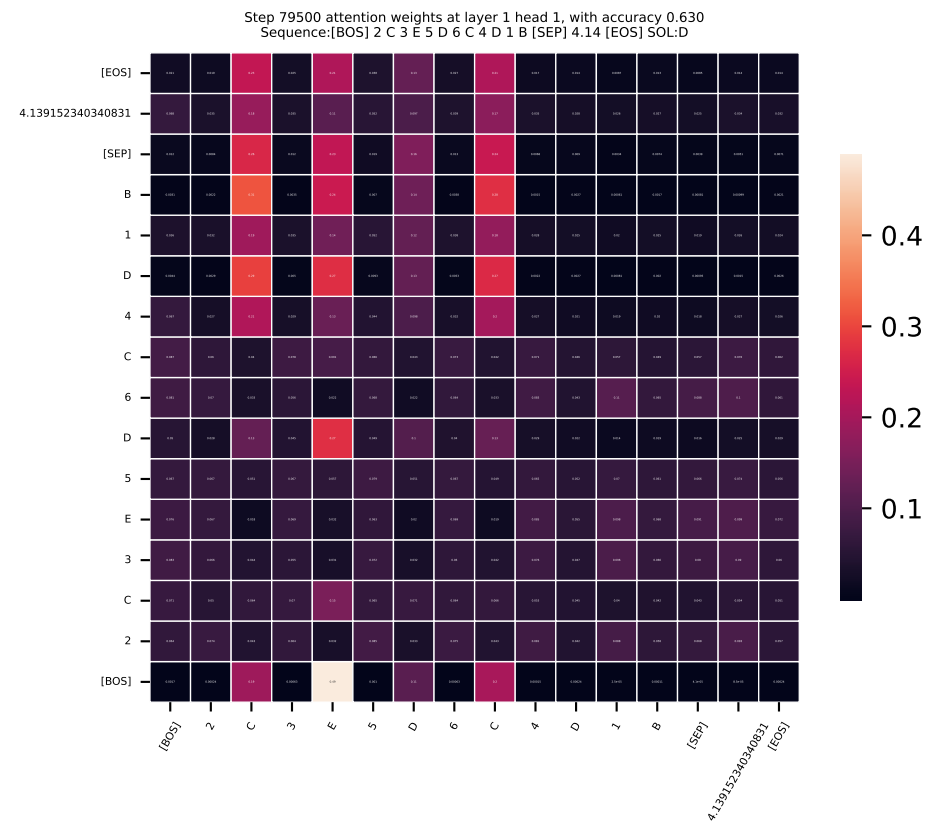

Step 70000 attention weights at layer 1 head 0, with accuracy 0.634
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.35
0.30
0.25
0.20
0.15
0.10
0.05

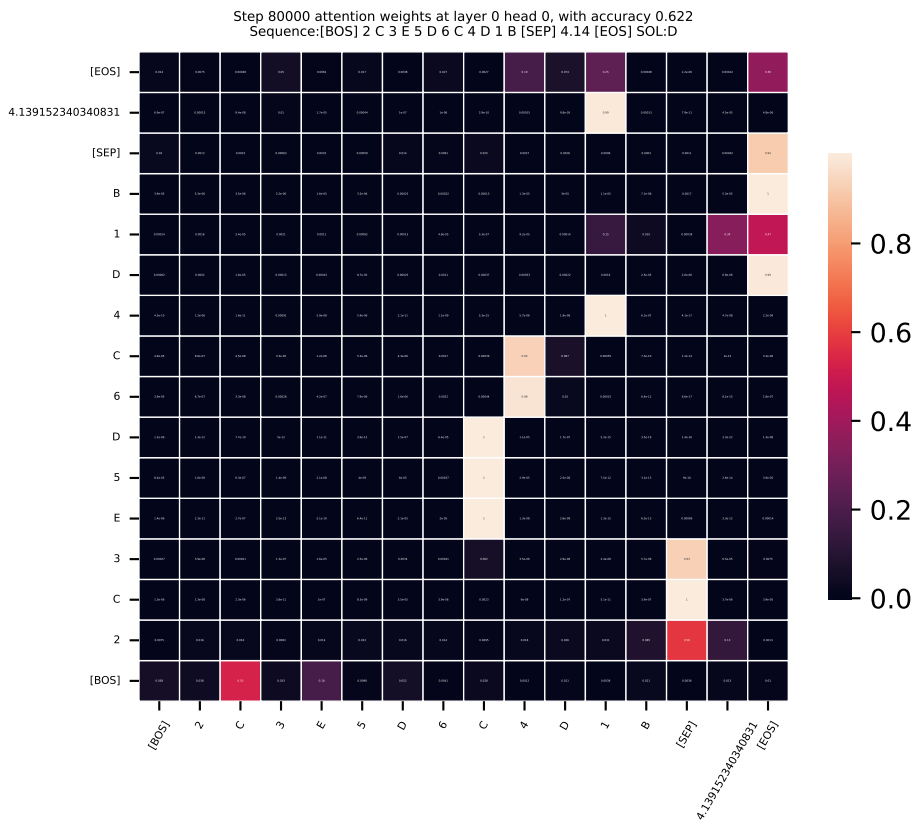

Step 70000 attention weights at layer 1 head 1, with accuracy 0.634
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.35
0.30
0.25
0.20
0.15
0.10
0.05

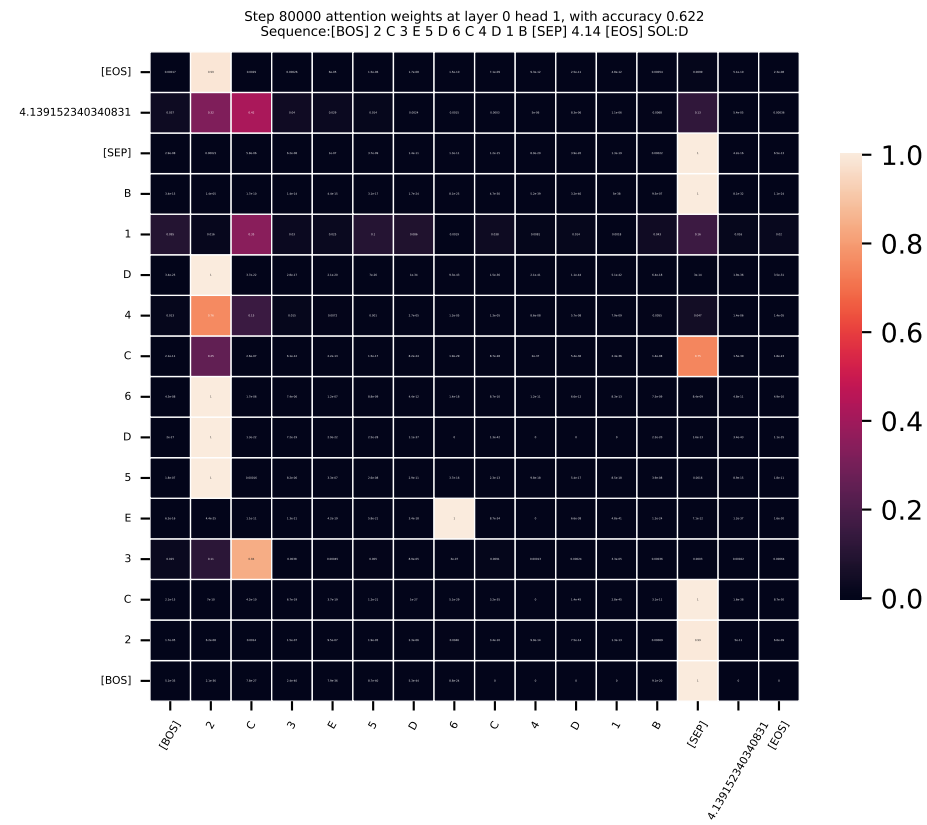

Step 79500 attention weights at layer 0 head 0, with accuracy 0.630
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.8
0.6
0.4
0.2
0.0

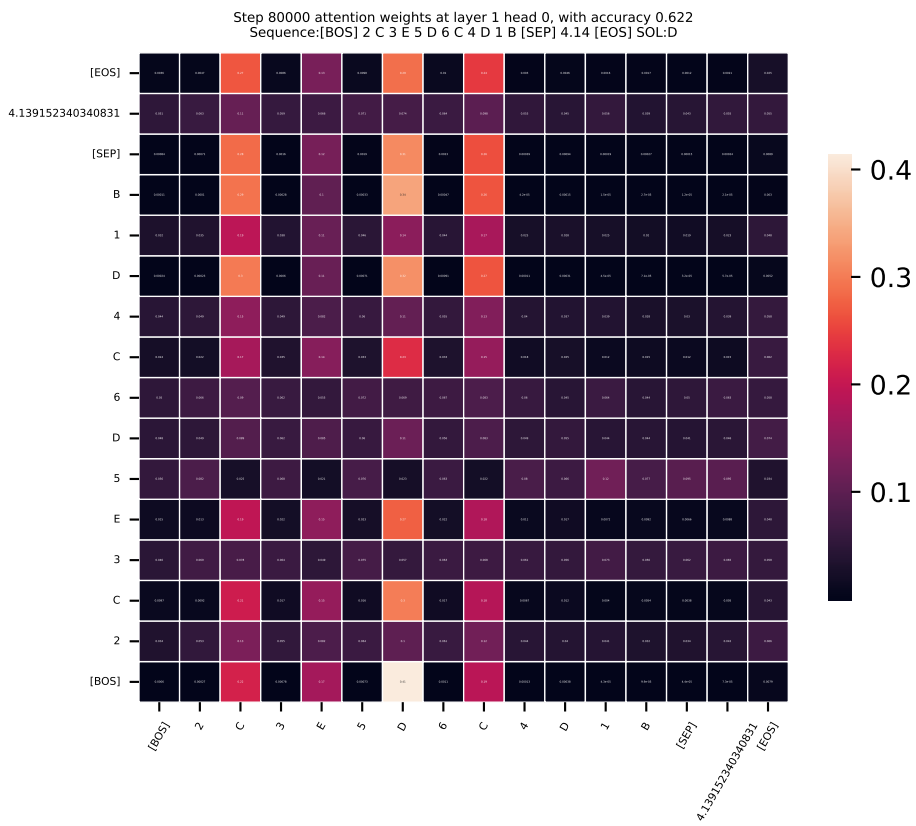

Step 79500 attention weights at layer 0 head 1, with accuracy 0.630
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
1.0
0.8
0.6
0.4
0.2
0.0

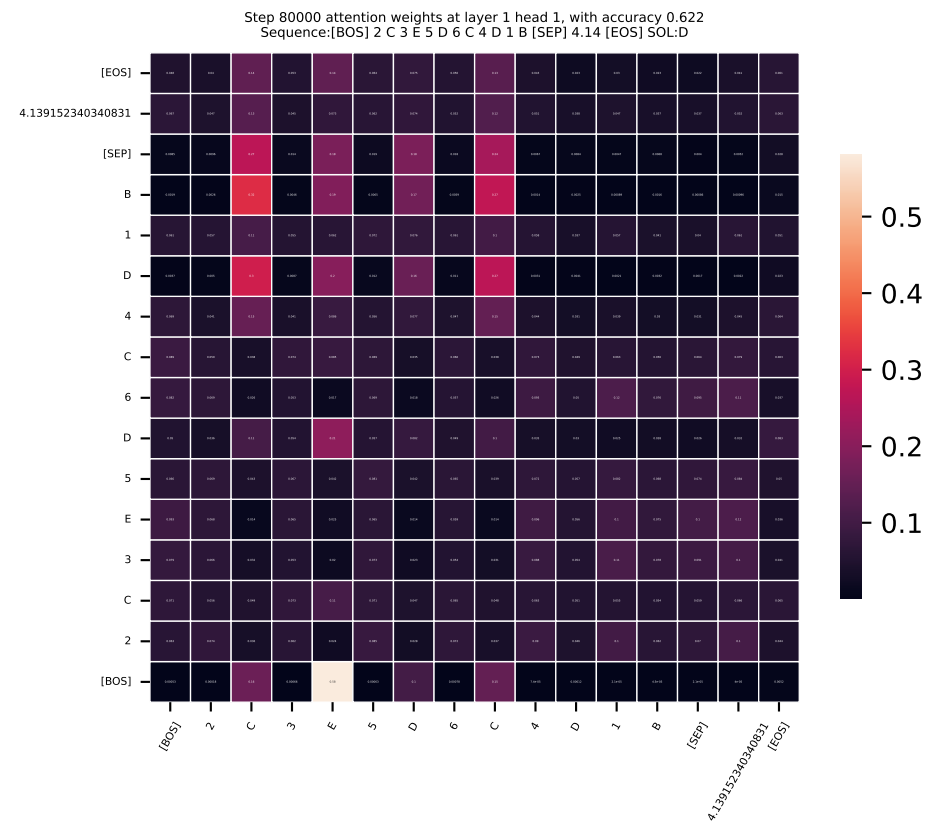

Step 79500 attention weights at layer 1 head 0, with accuracy 0.630
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.35
0.30
0.25
0.20
0.15
0.10
0.05

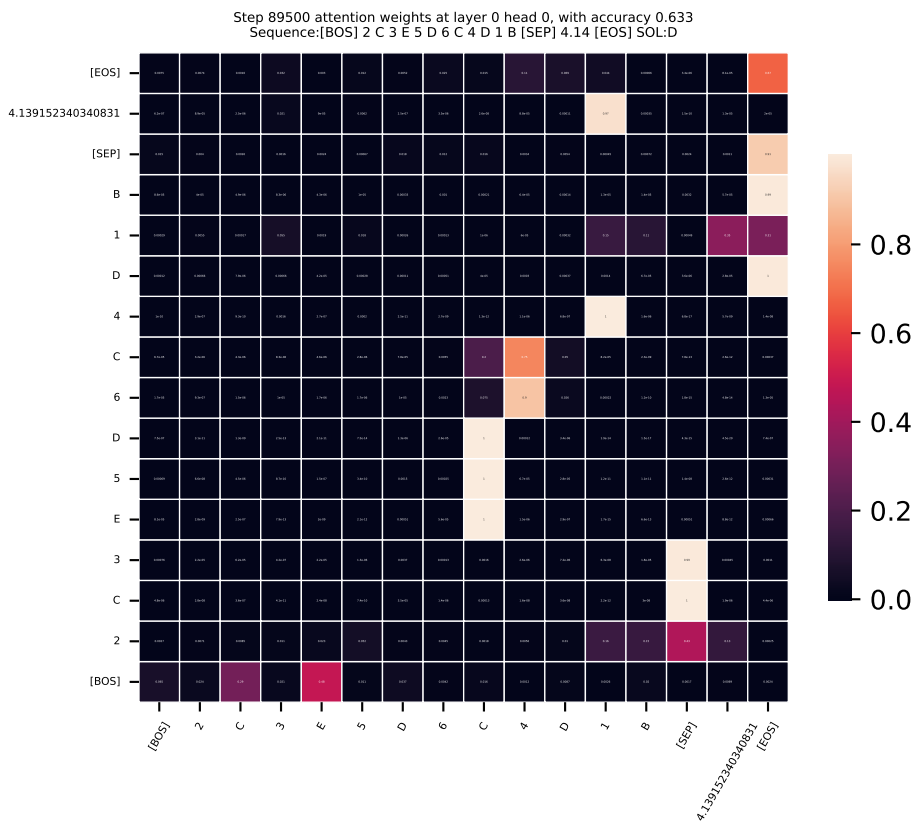

Step 79500 attention weights at layer 1 head 1, with accuracy 0.630
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.4
0.3
0.2
0.1

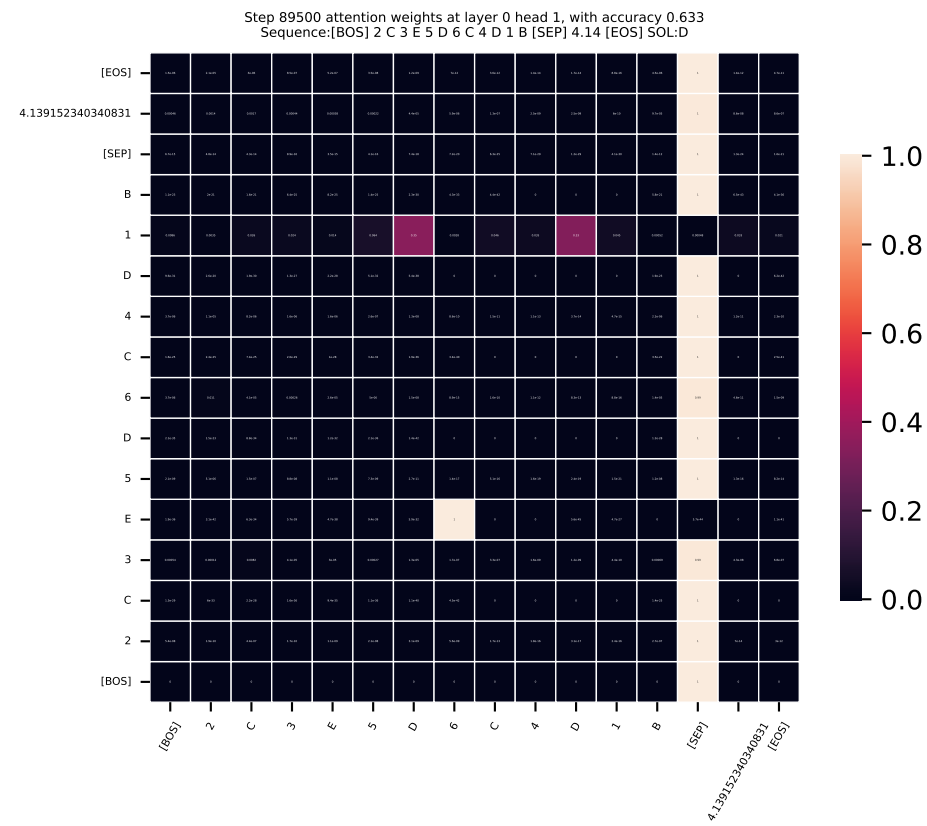

Step 80000 attention weights at layer 0 head 0, with accuracy 0.622
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2
0.0

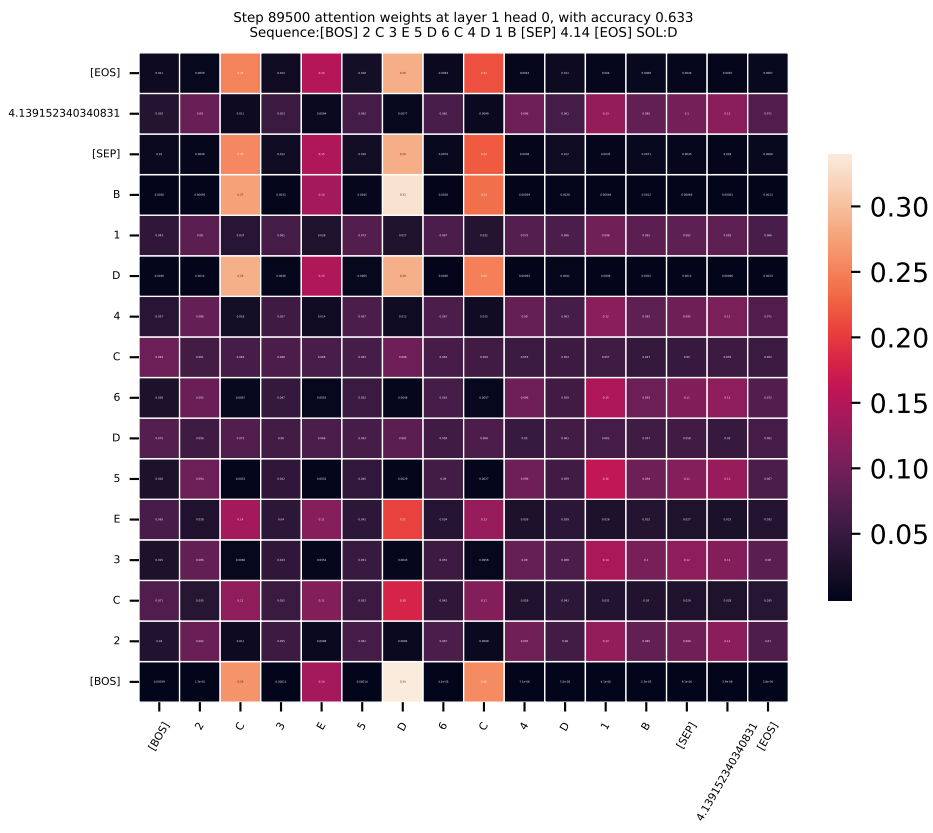

Step 80000 attention weights at layer 0 head 1, with accuracy 0.622
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

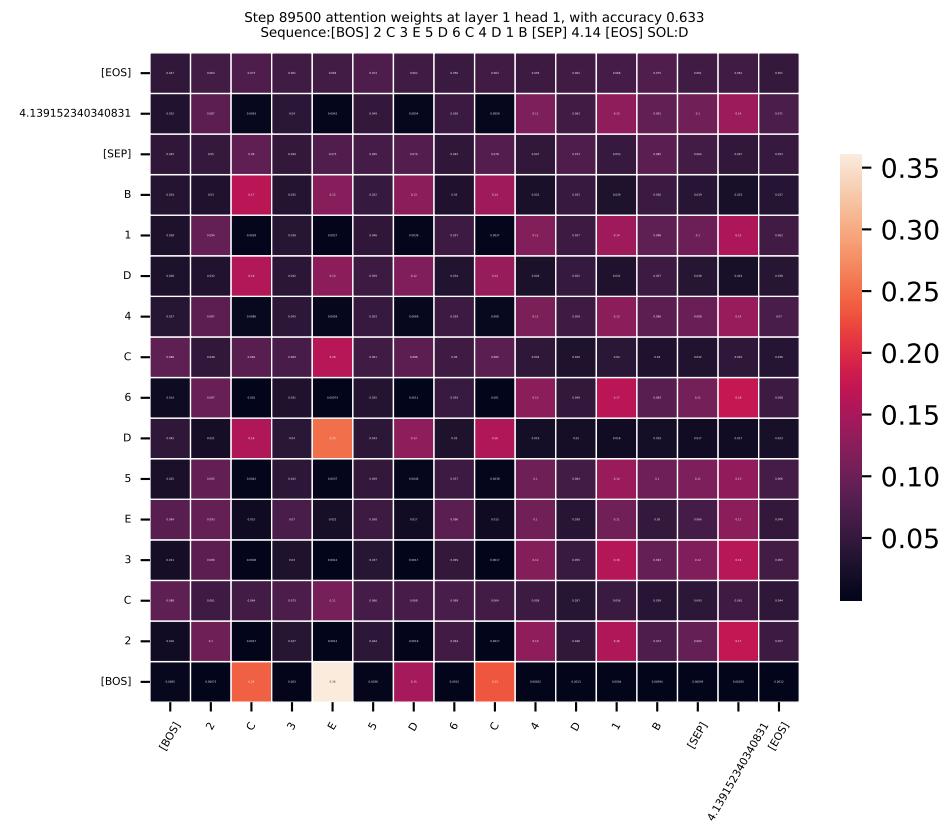

Step 80000 attention weights at layer 1 head 0, with accuracy 0.622
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.4
0.3
0.2
0.1

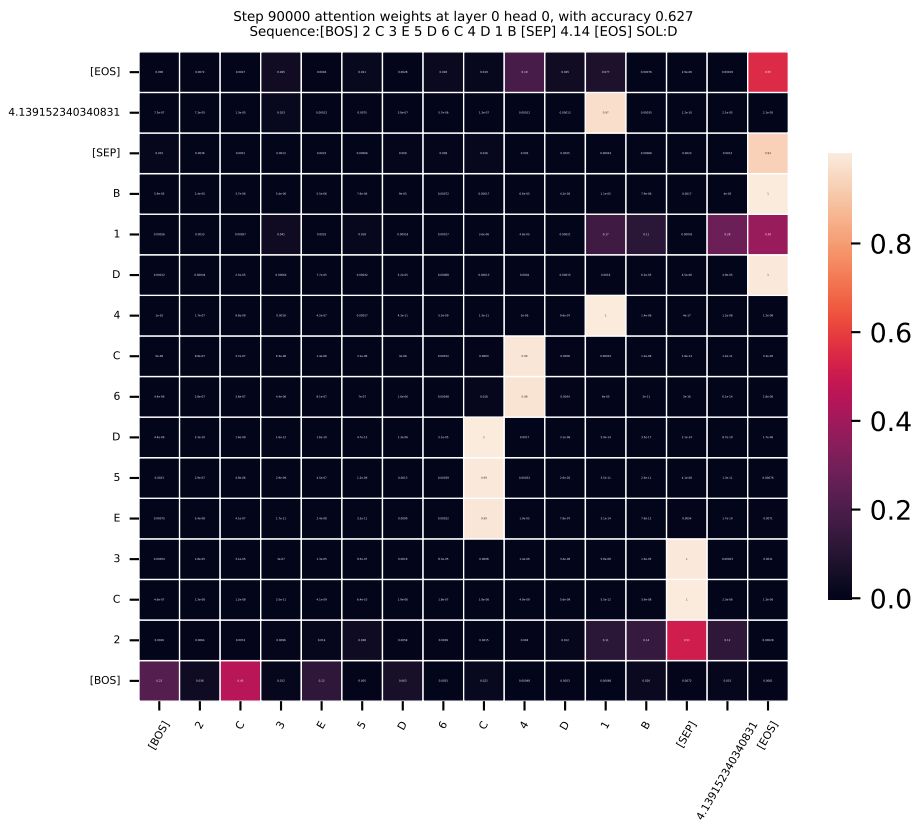

Step 80000 attention weights at layer 1 head 1, with accuracy 0.622
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.5
0.4
0.3
0.2
0.1

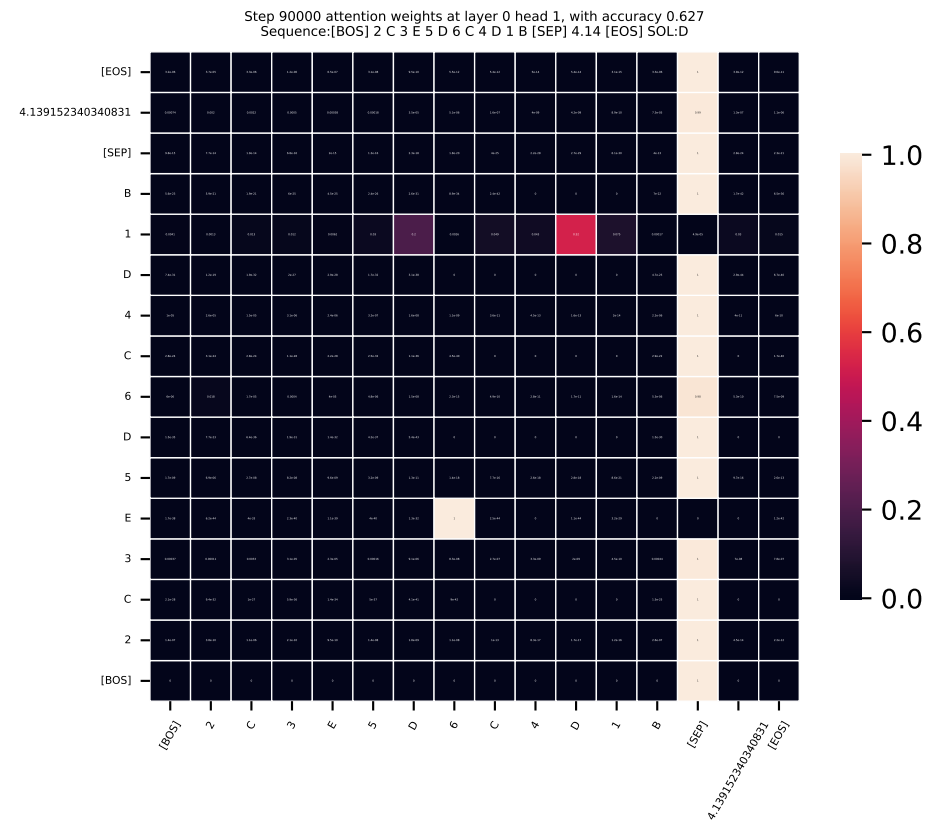

Step 89500 attention weights at layer 0 head 0, with accuracy 0.633
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2
0.0

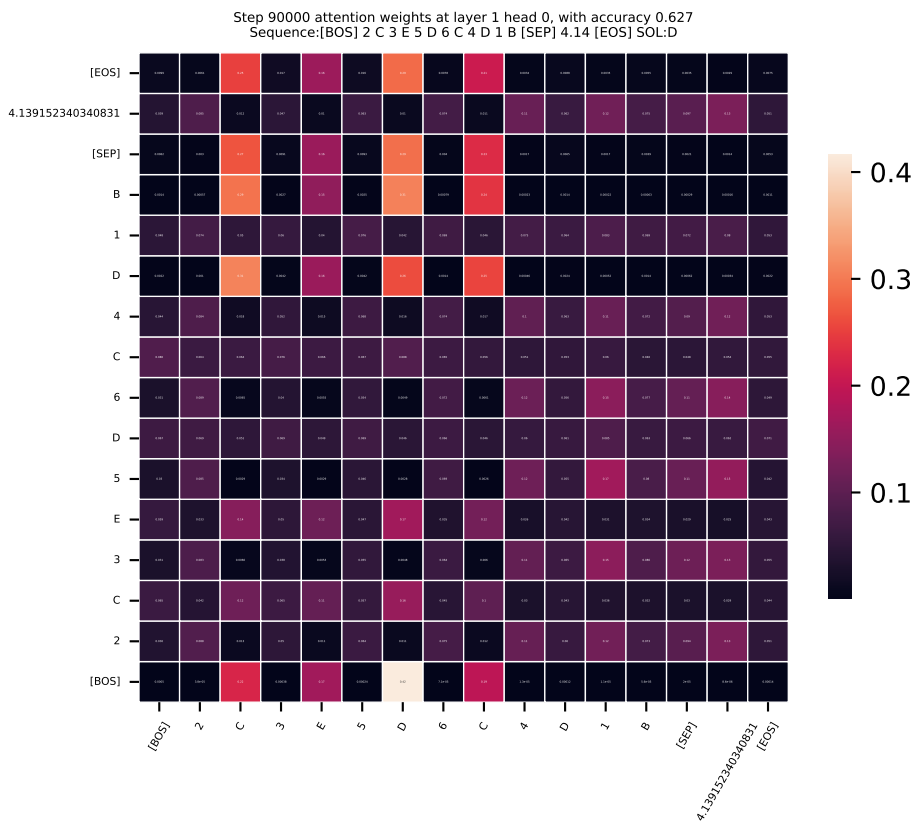

Step 89500 attention weights at layer 0 head 1, with accuracy 0.633
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

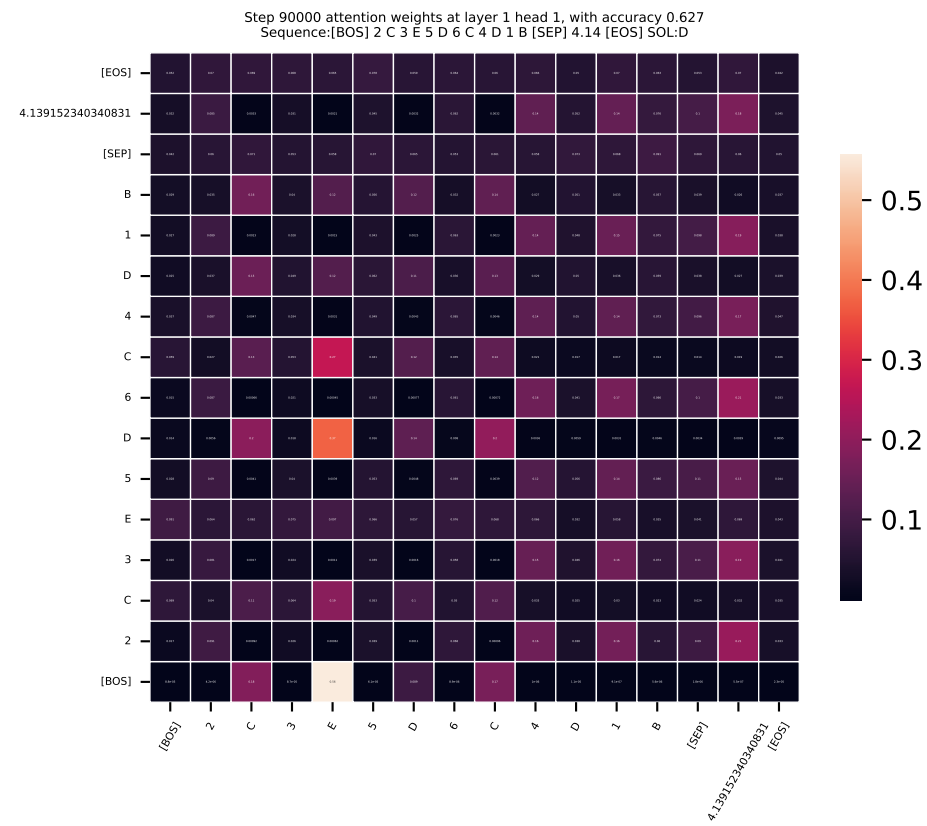

Step 89500 attention weights at layer 1 head 0, with accuracy 0.633
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.30
0.25
0.20
0.15
0.10
0.05

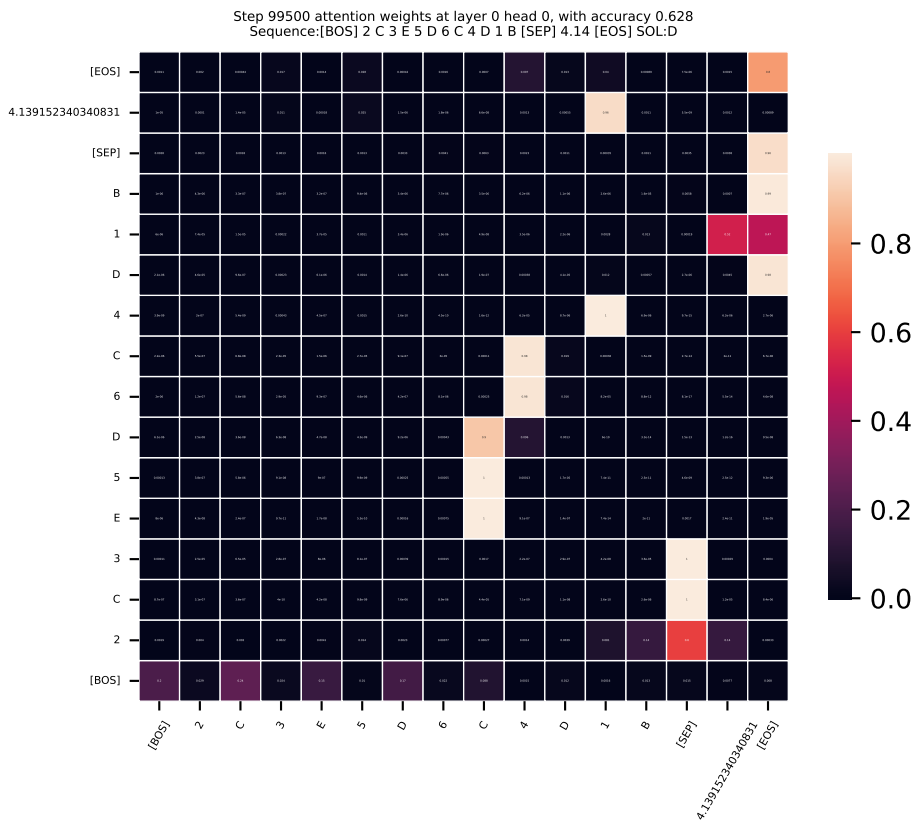

Step 89500 attention weights at layer 1 head 1, with accuracy 0.633
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.35
0.30
0.25
0.20
0.15
0.10
0.05

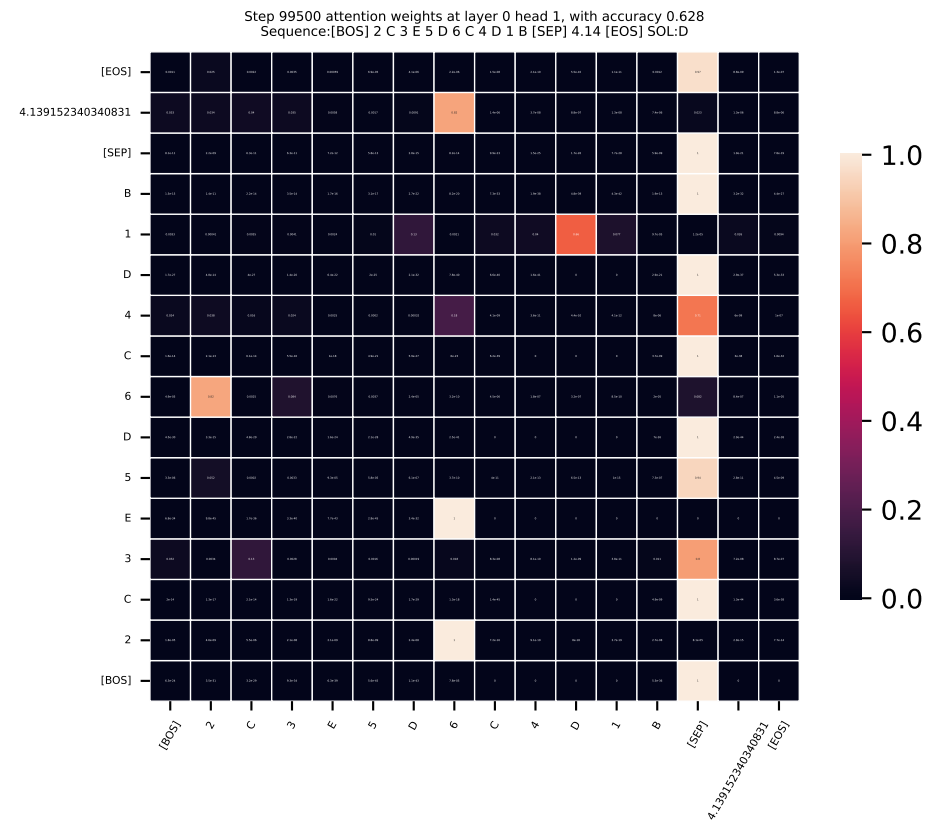

Step 90000 attention weights at layer 0 head 0, with accuracy 0.627
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2
0.0

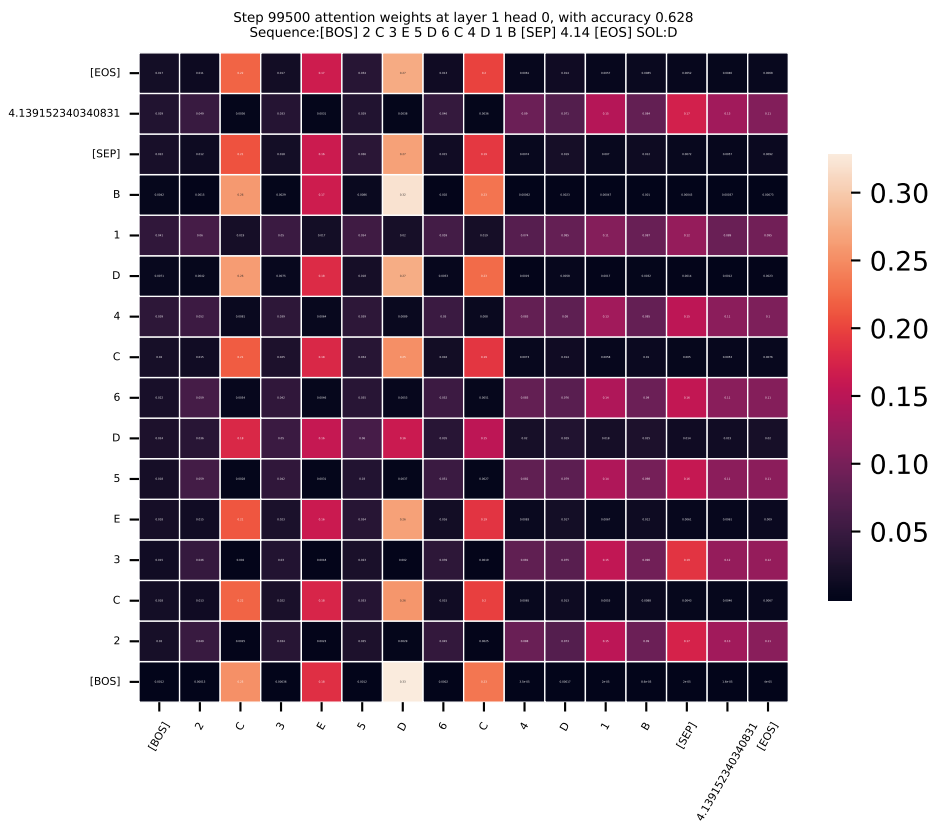

Step 90000 attention weights at layer 0 head 1, with accuracy 0.627
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

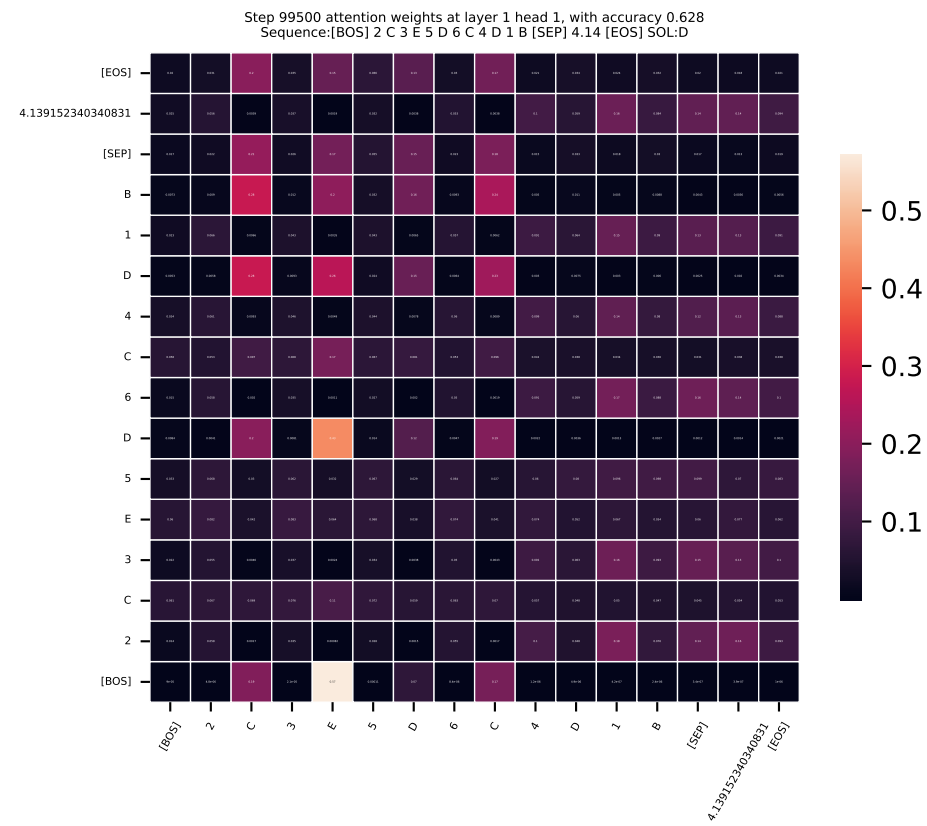

Step 90000 attention weights at layer 1 head 0, with accuracy 0.627
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.4
0.3
0.2
0.1

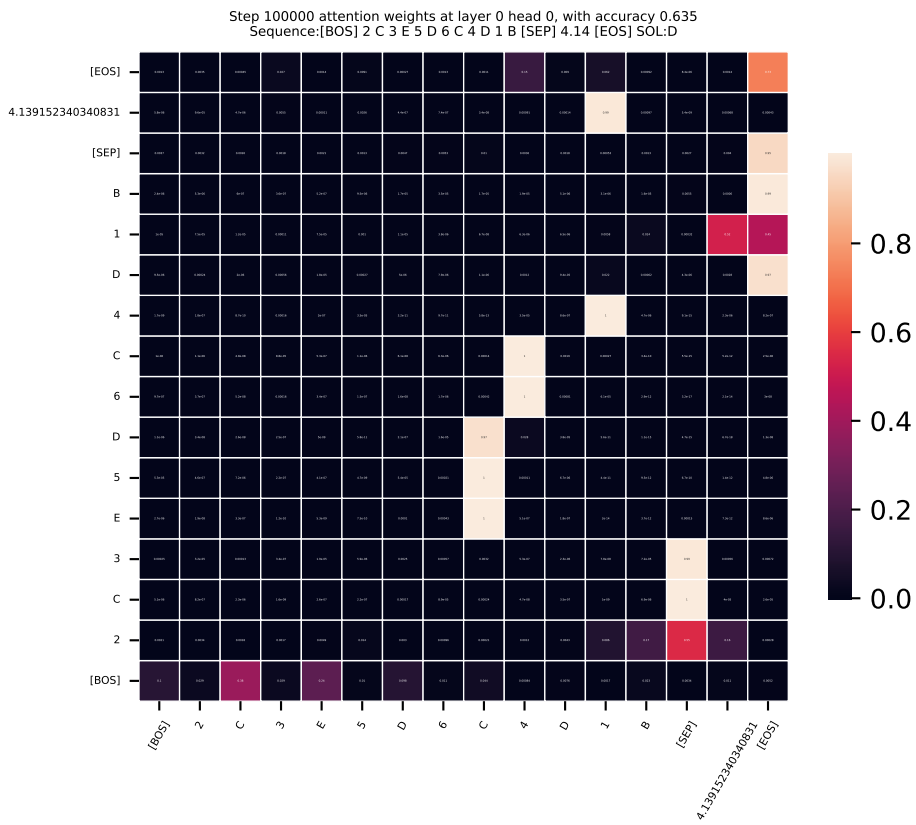

Step 90000 attention weights at layer 1 head 1, with accuracy 0.627
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.5
0.4
0.3
0.2
0.1

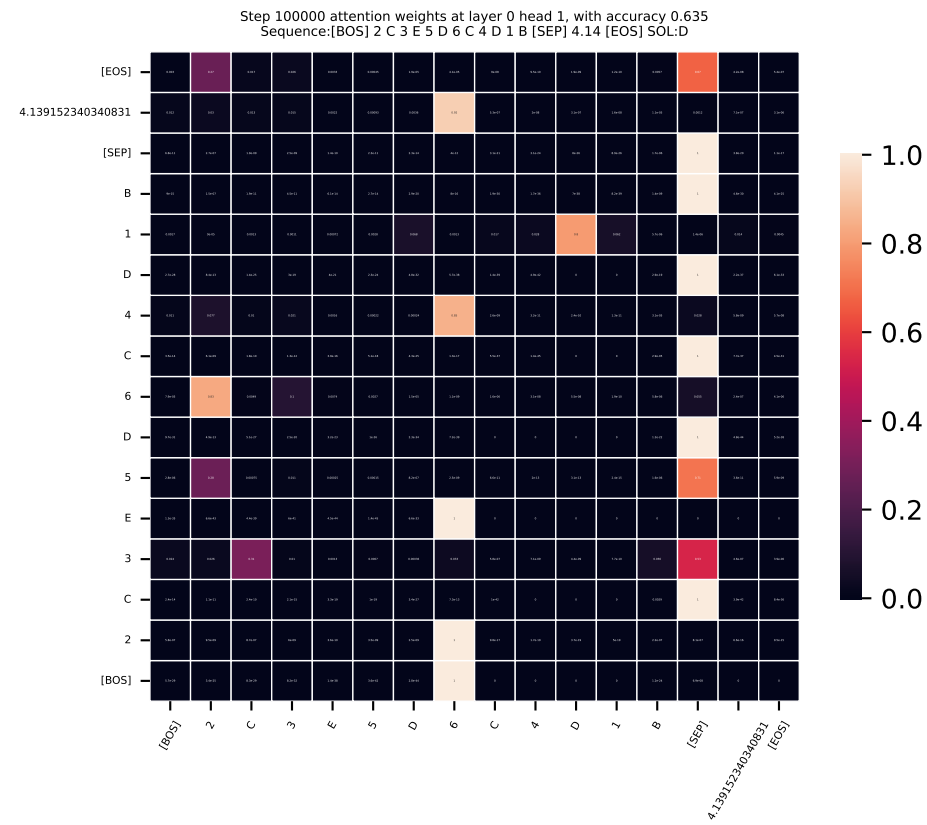

Step 99500 attention weights at layer 0 head 0, with accuracy 0.628
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2
0.0

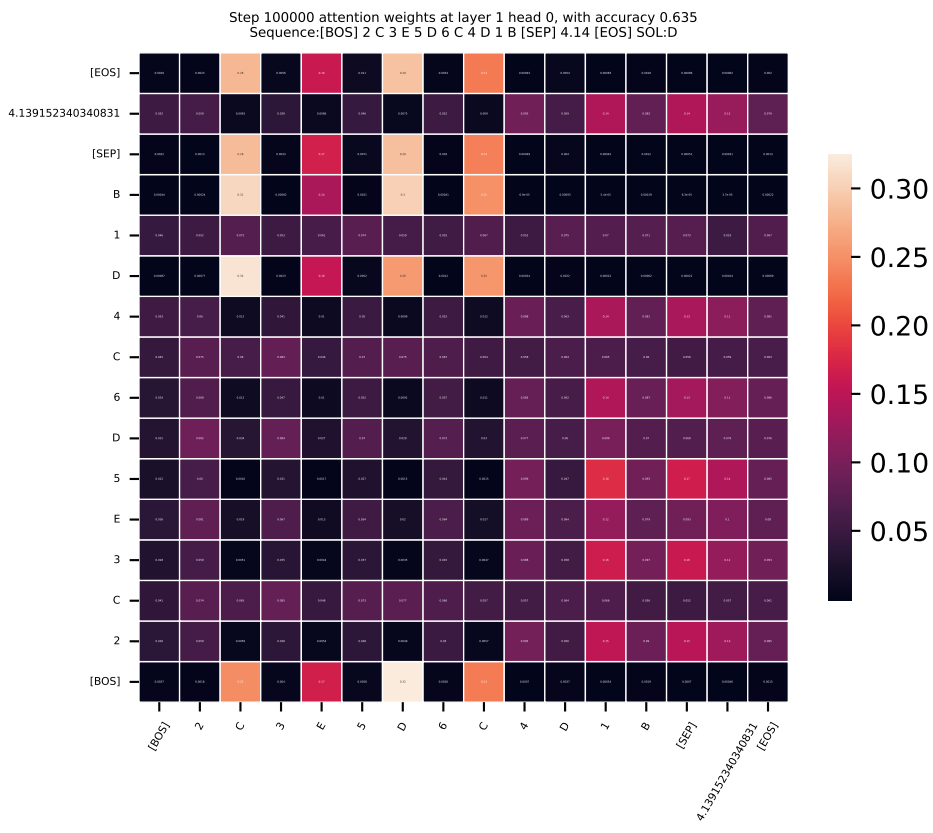

Step 99500 attention weights at layer 0 head 1, with accuracy 0.628
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

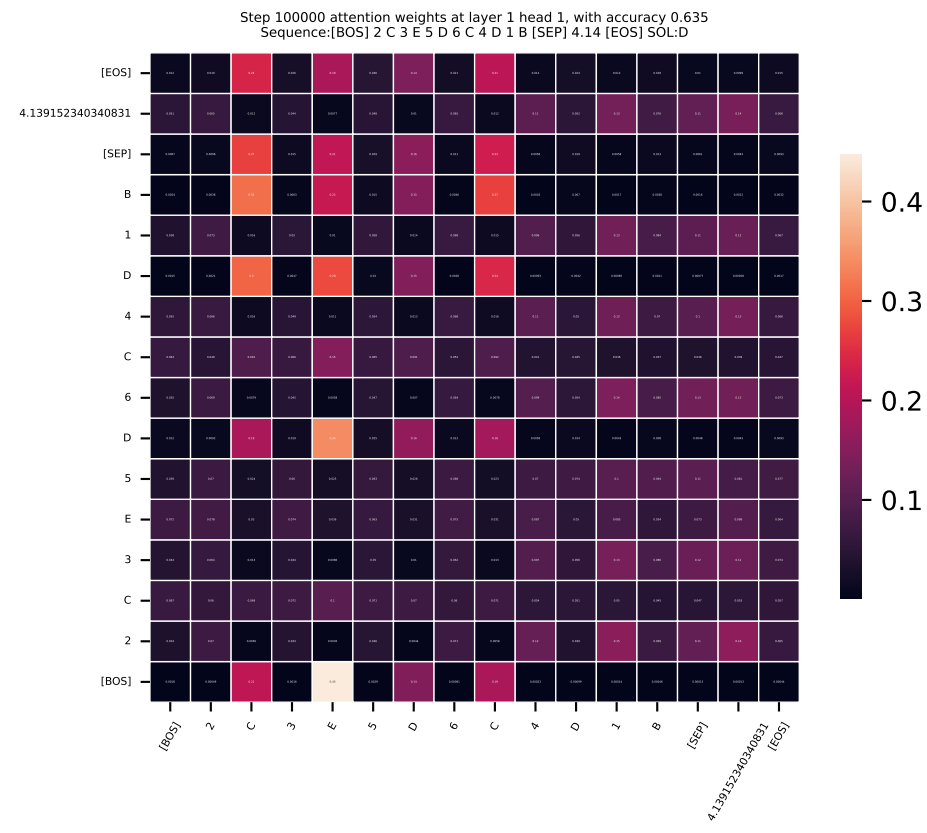

Step 99500 attention weights at layer 1 head 0, with accuracy 0.628
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.30
0.25
0.20
0.15
0.10
0.05

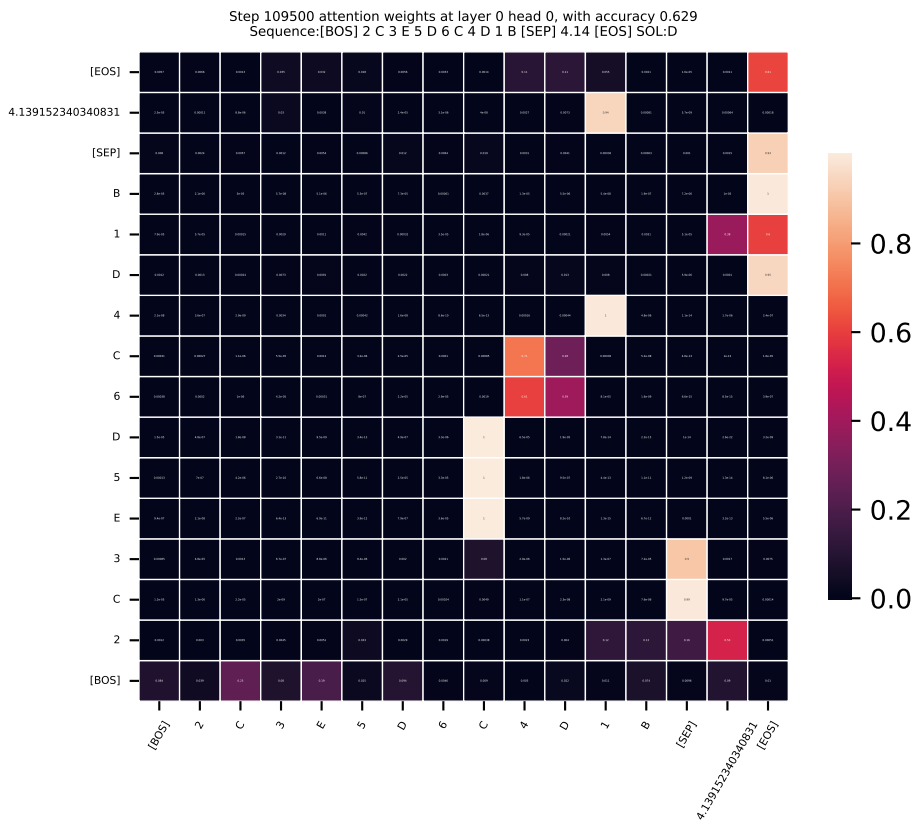

Step 99500 attention weights at layer 1 head 1, with accuracy 0.628
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.5
0.4
0.3
0.2
0.1

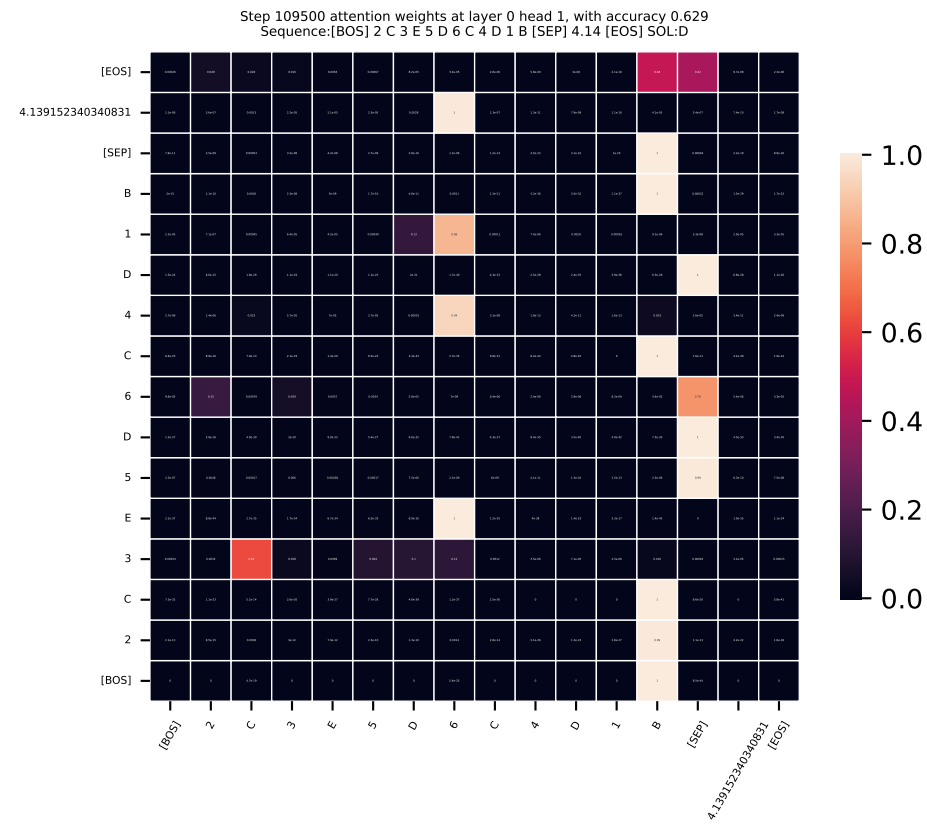

Step 100000 attention weights at layer 0 head 0, with accuracy 0.635
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2
0.0

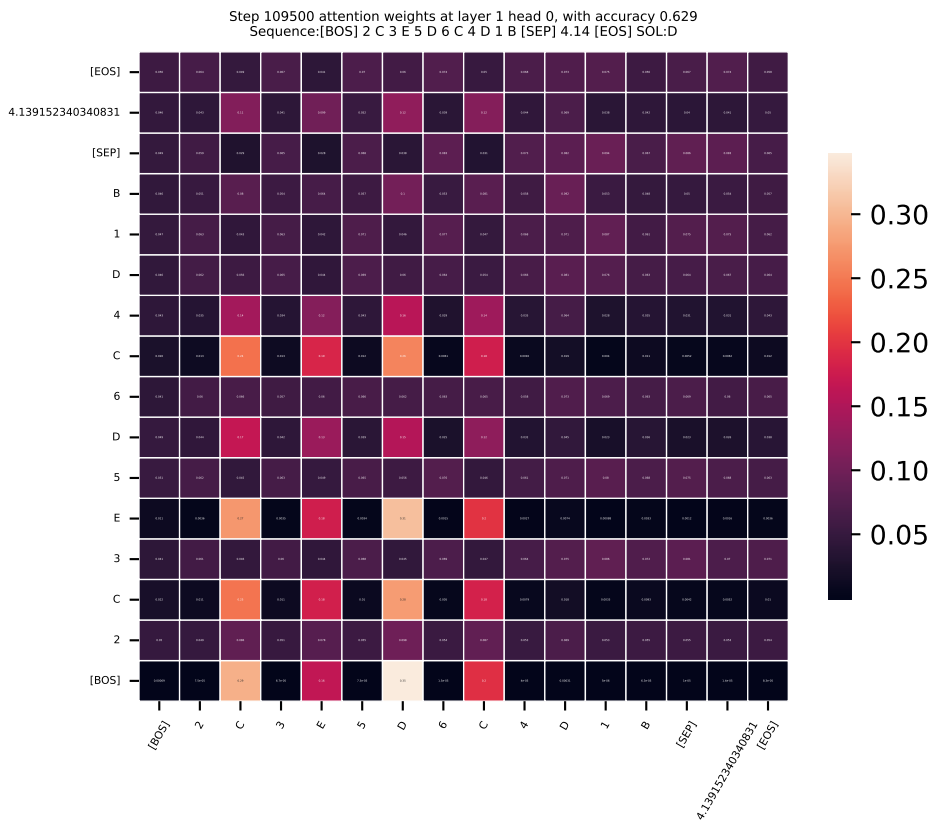

Step 100000 attention weights at layer 0 head 1, with accuracy 0.635
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

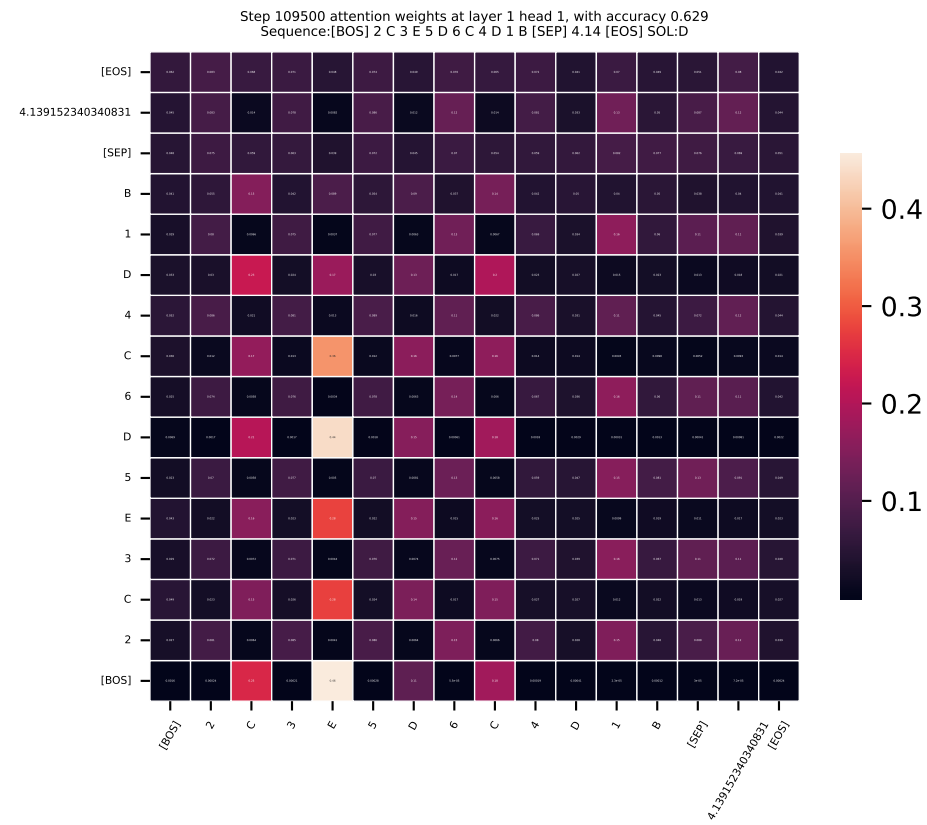

Step 100000 attention weights at layer 1 head 0, with accuracy 0.635
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.30
0.25
0.20
0.15
0.10
0.05

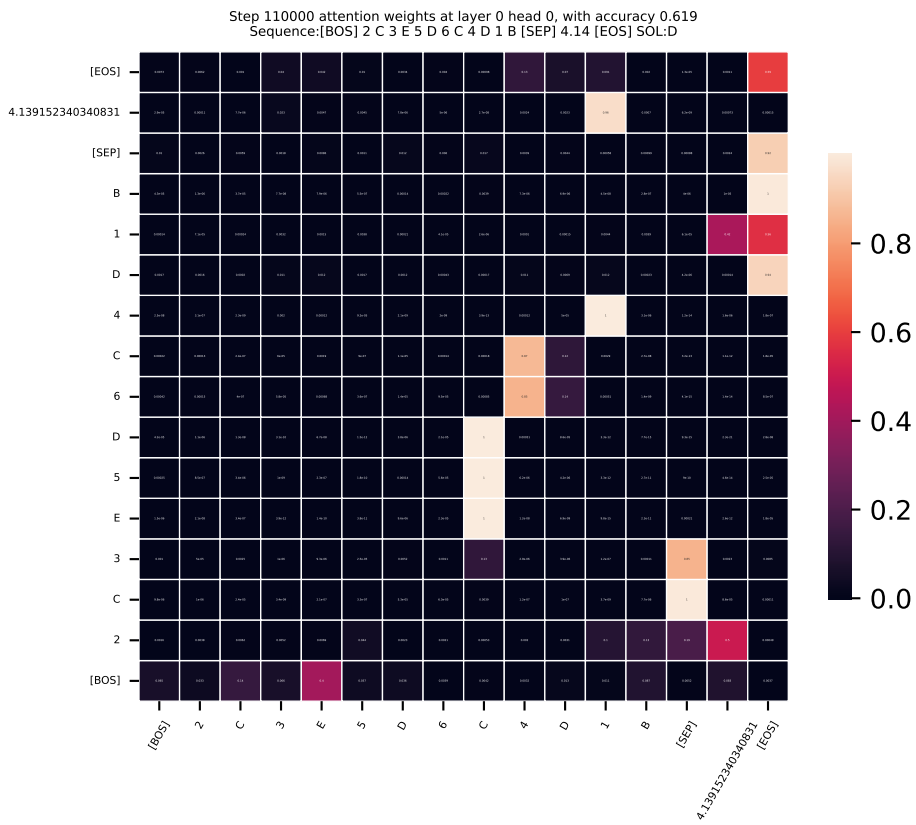

Step 100000 attention weights at layer 1 head 1, with accuracy 0.635
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.4
0.3
0.2
0.1

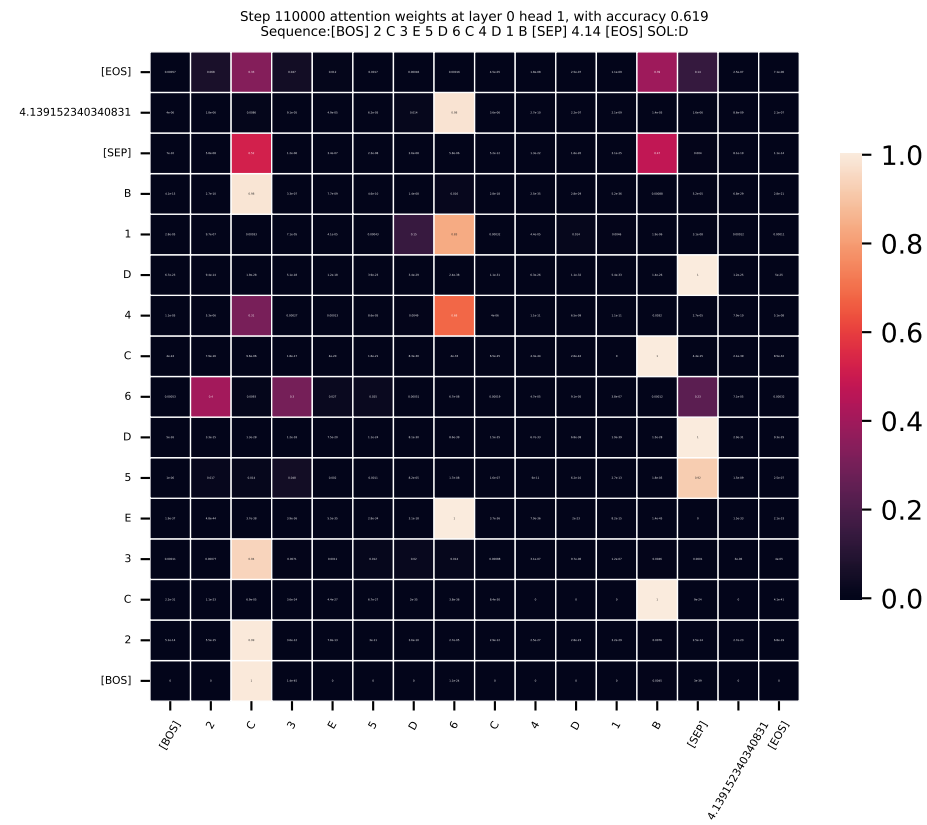

Step 109500 attention weights at layer 0 head 0, with accuracy 0.629
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2
0.0

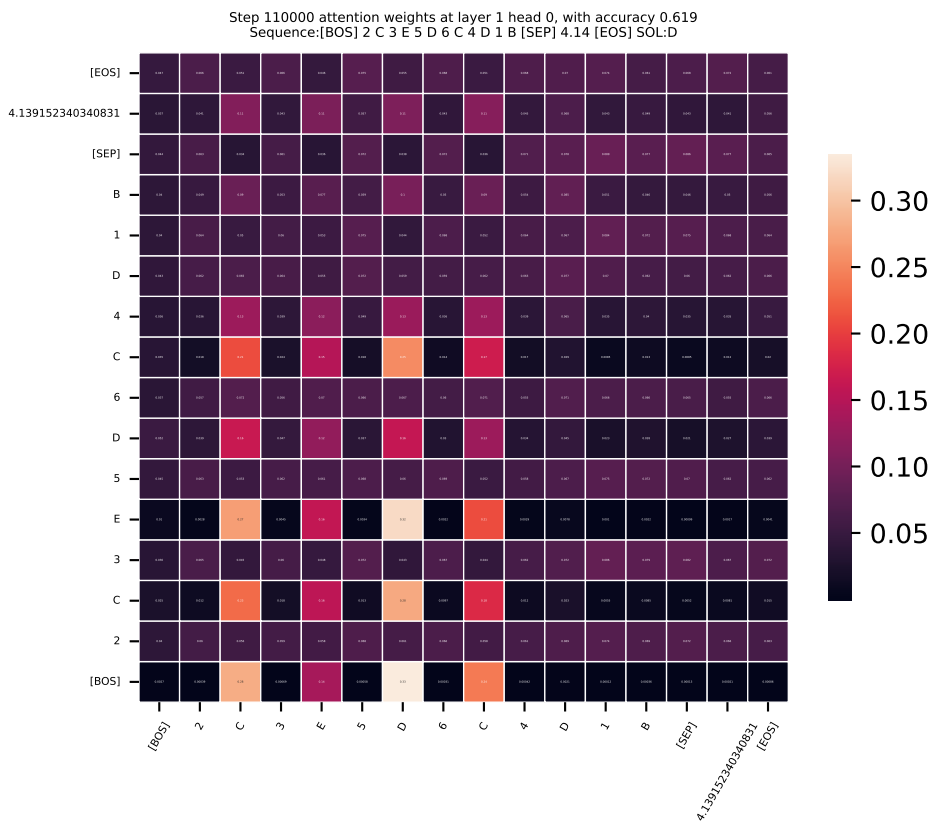

Step 109500 attention weights at layer 0 head 1, with accuracy 0.629
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

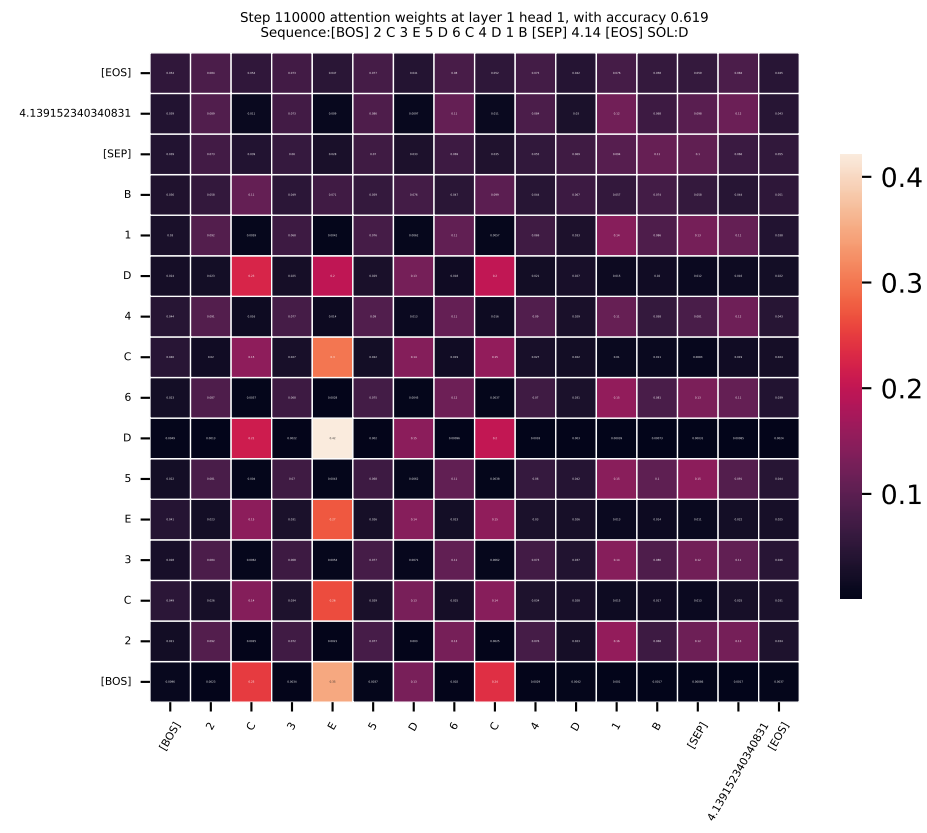

Step 109500 attention weights at layer 1 head 0, with accuracy 0.629
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.30
0.25
0.20
0.15
0.10
0.05

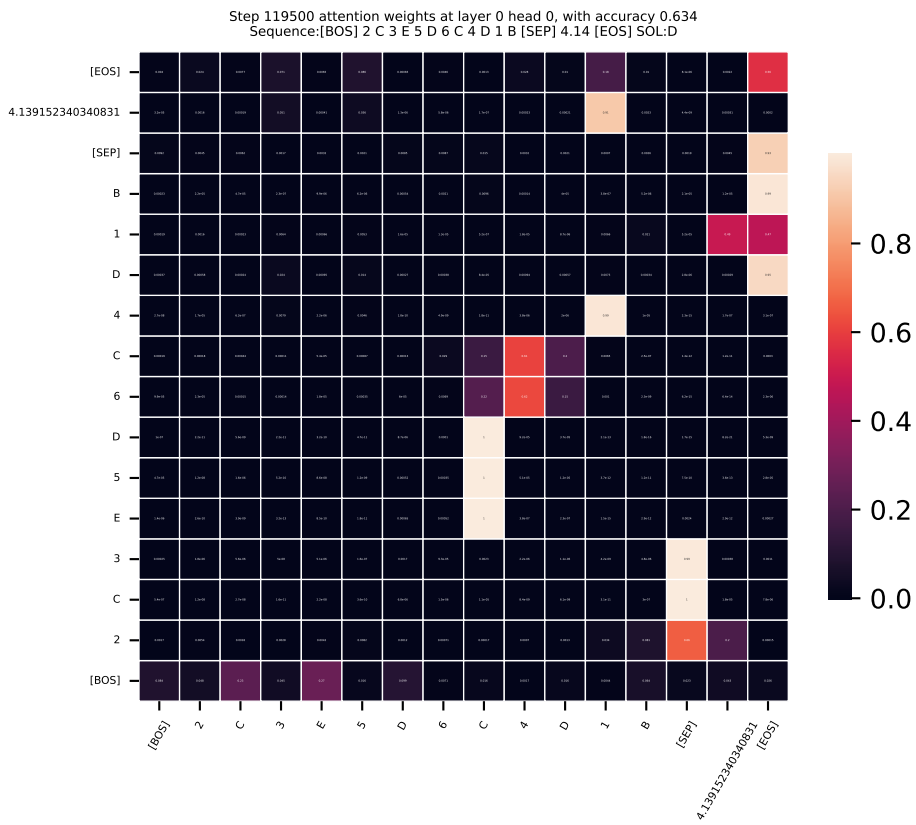

Step 109500 attention weights at layer 1 head 1, with accuracy 0.629
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.4
0.3
0.2
0.1

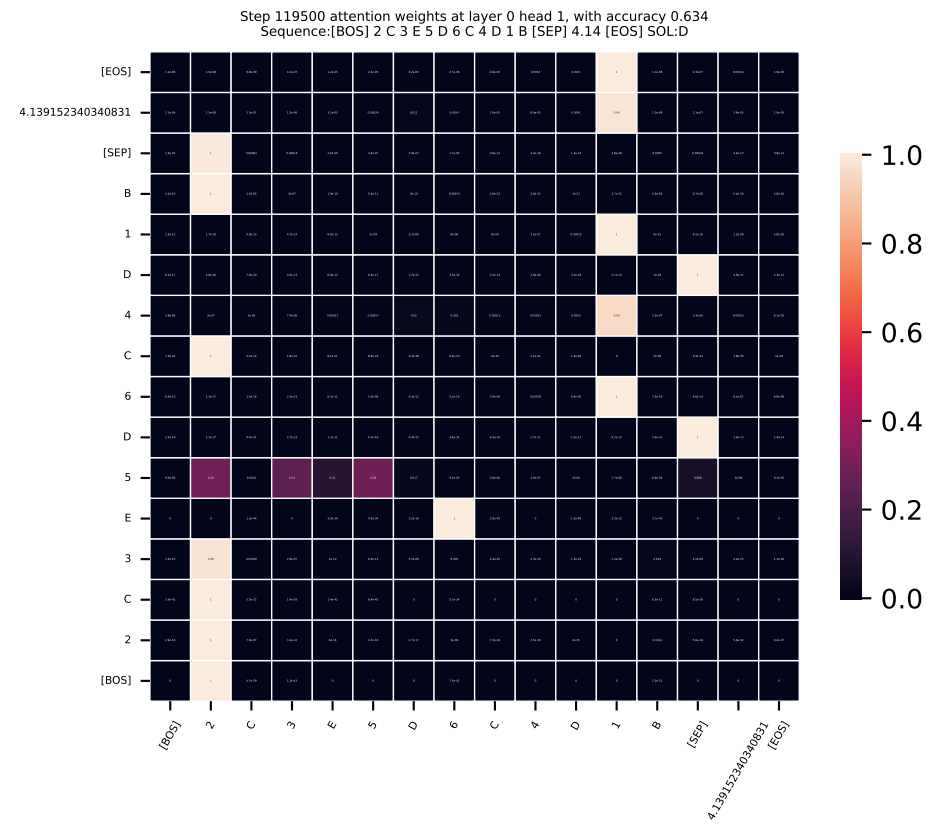

Step 110000 attention weights at layer 0 head 0, with accuracy 0.619
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2
0.0

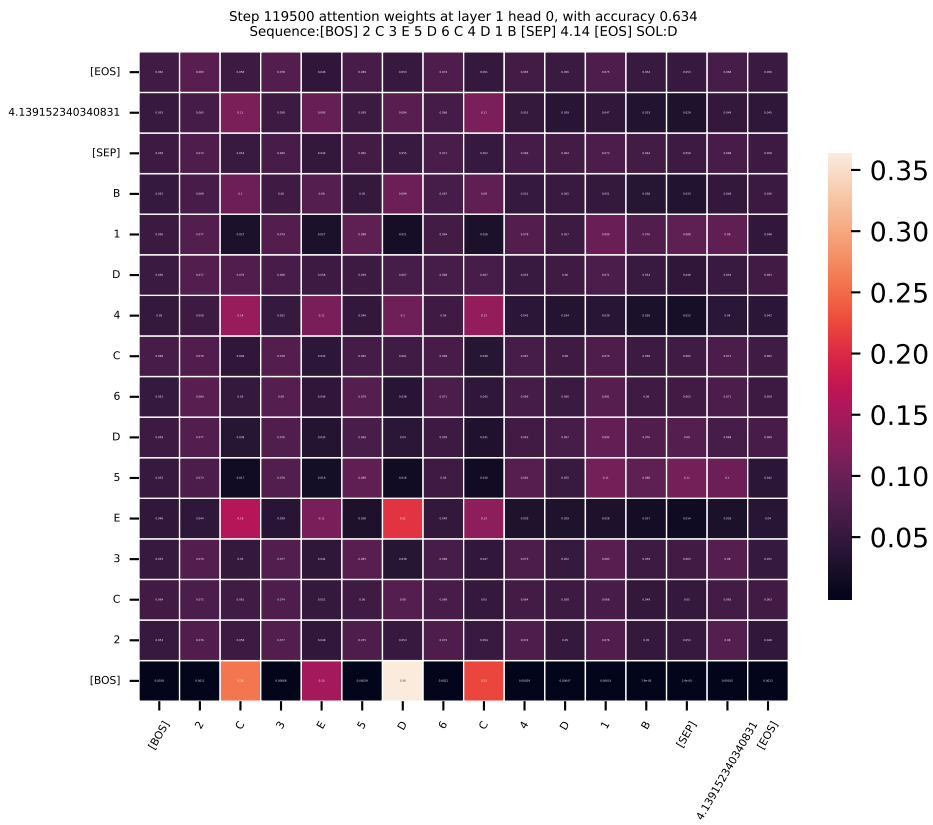

Step 110000 attention weights at layer 0 head 1, with accuracy 0.619
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

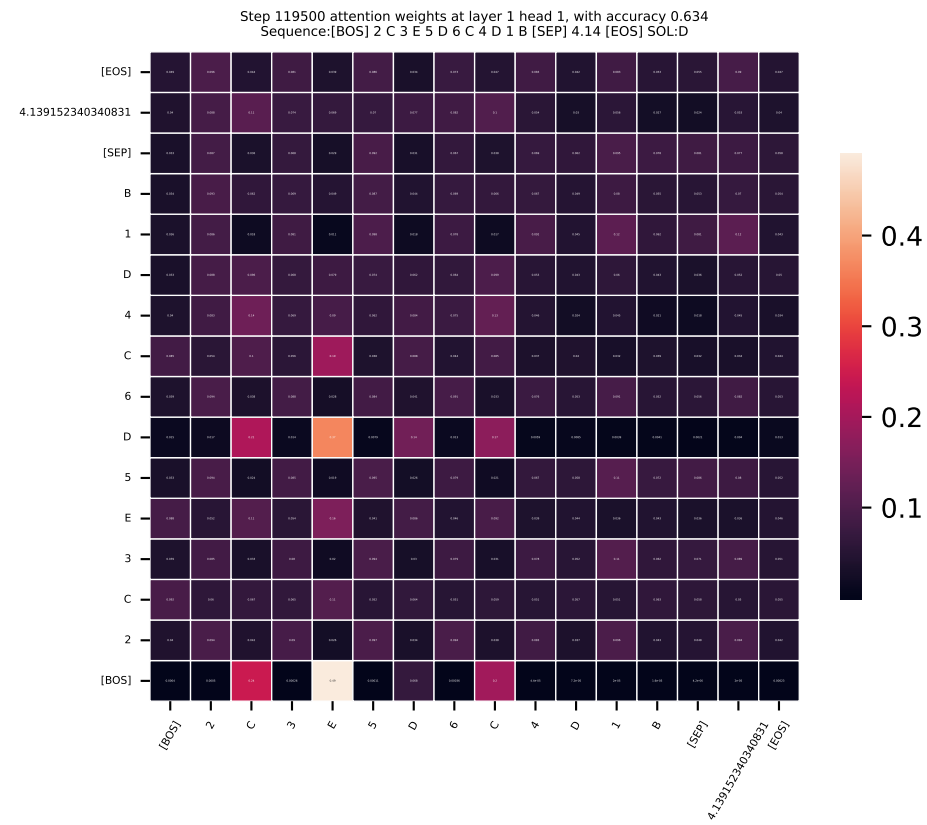

Step 110000 attention weights at layer 1 head 0, with accuracy 0.619
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.30
0.25
0.20
0.15
0.10
0.05

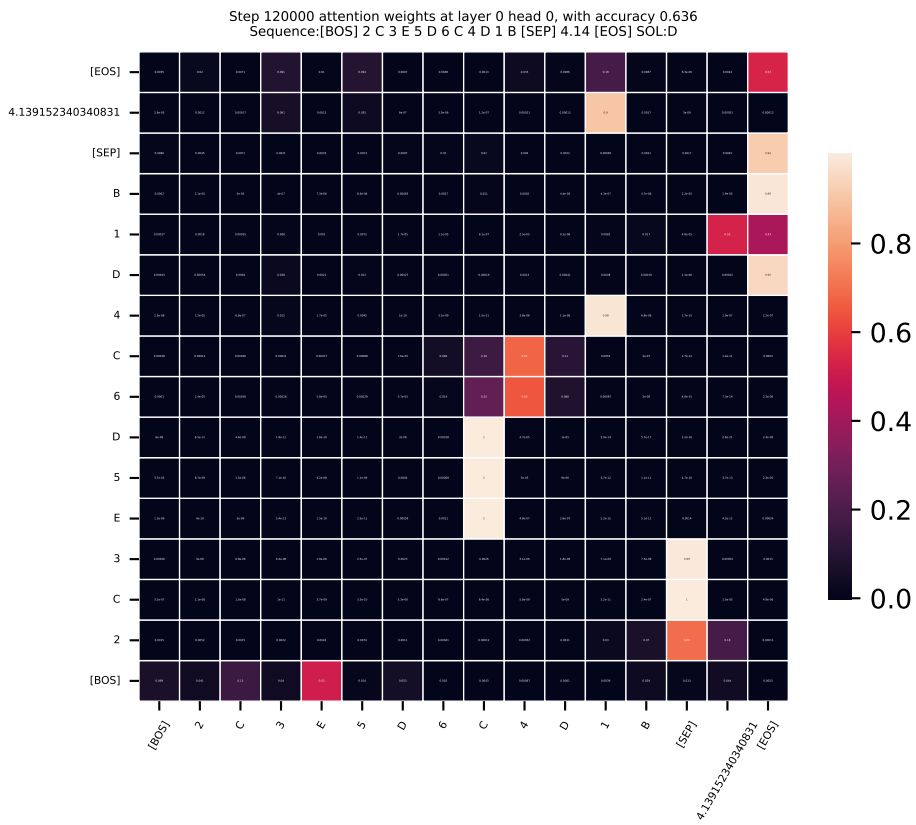

Step 110000 attention weights at layer 1 head 1, with accuracy 0.619
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.4
0.3
0.2
0.1

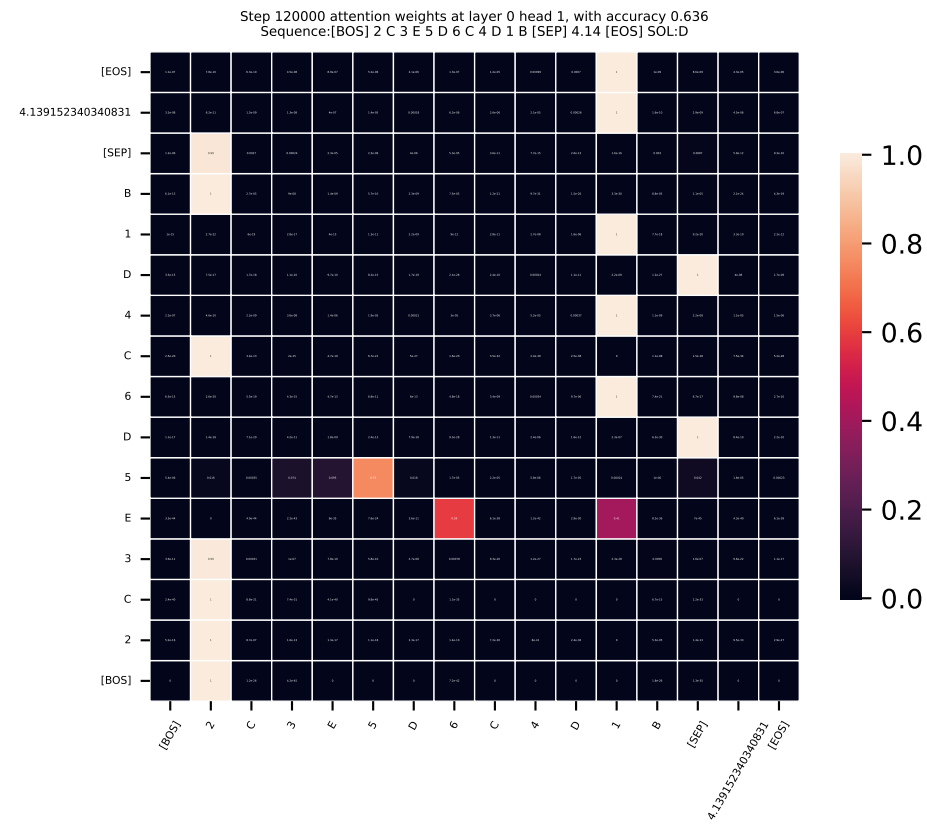

Step 119500 attention weights at layer 0 head 0, with accuracy 0.634
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.8
0.6
0.4
0.2
0.0

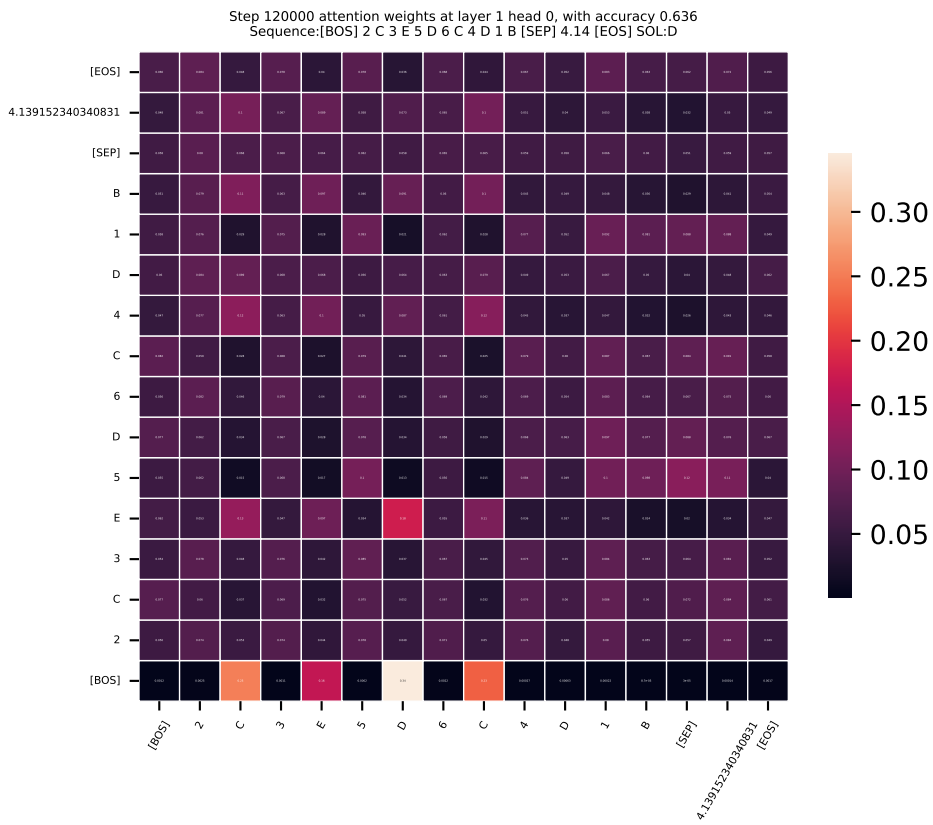

Step 119500 attention weights at layer 0 head 1, with accuracy 0.634
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
1.0
0.8
0.6
0.4
0.2
0.0

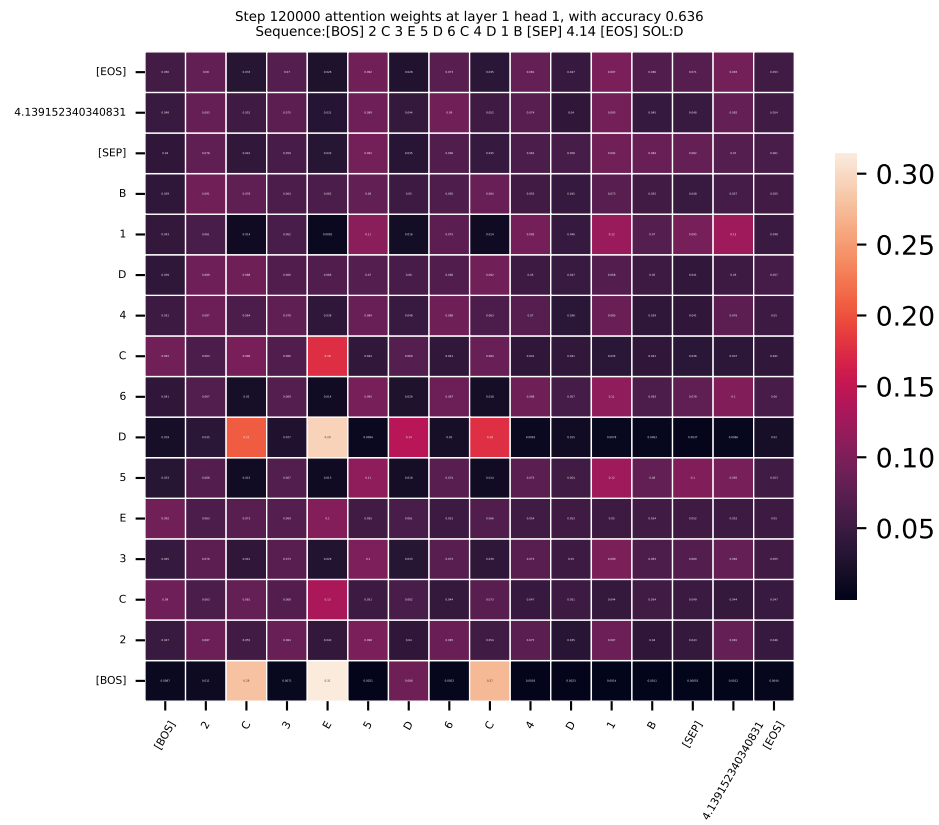

Step 119500 attention weights at layer 1 head 0, with accuracy 0.634
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.35
0.30
0.25
0.20
0.15
0.10
0.05

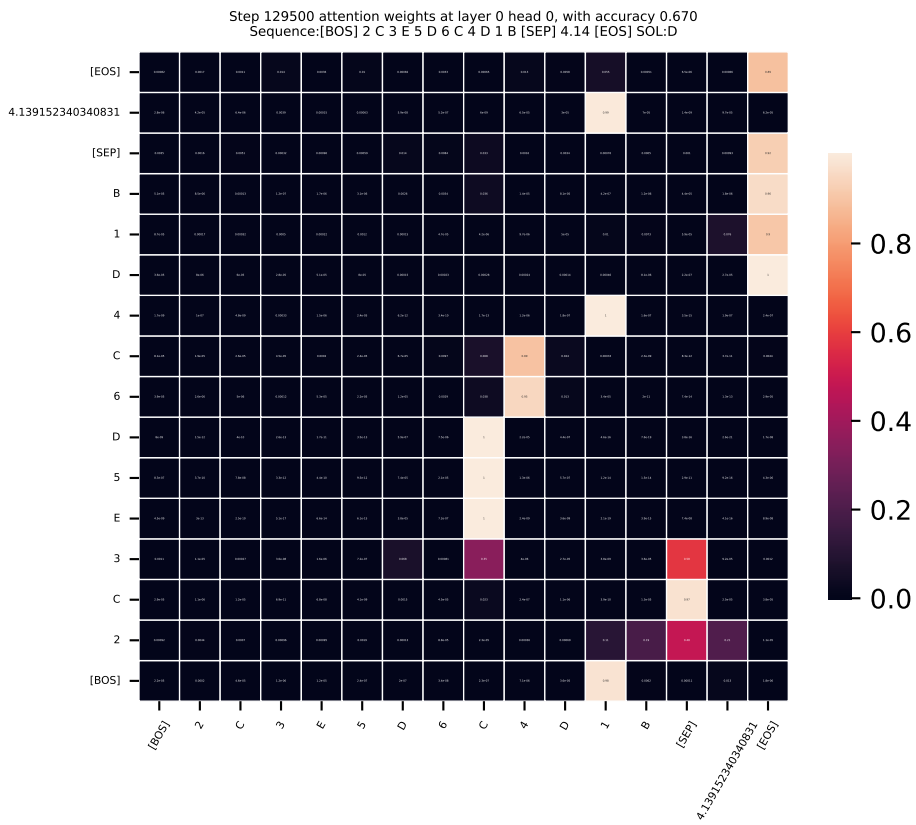

Step 119500 attention weights at layer 1 head 1, with accuracy 0.634
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.4
0.3
0.2
0.1

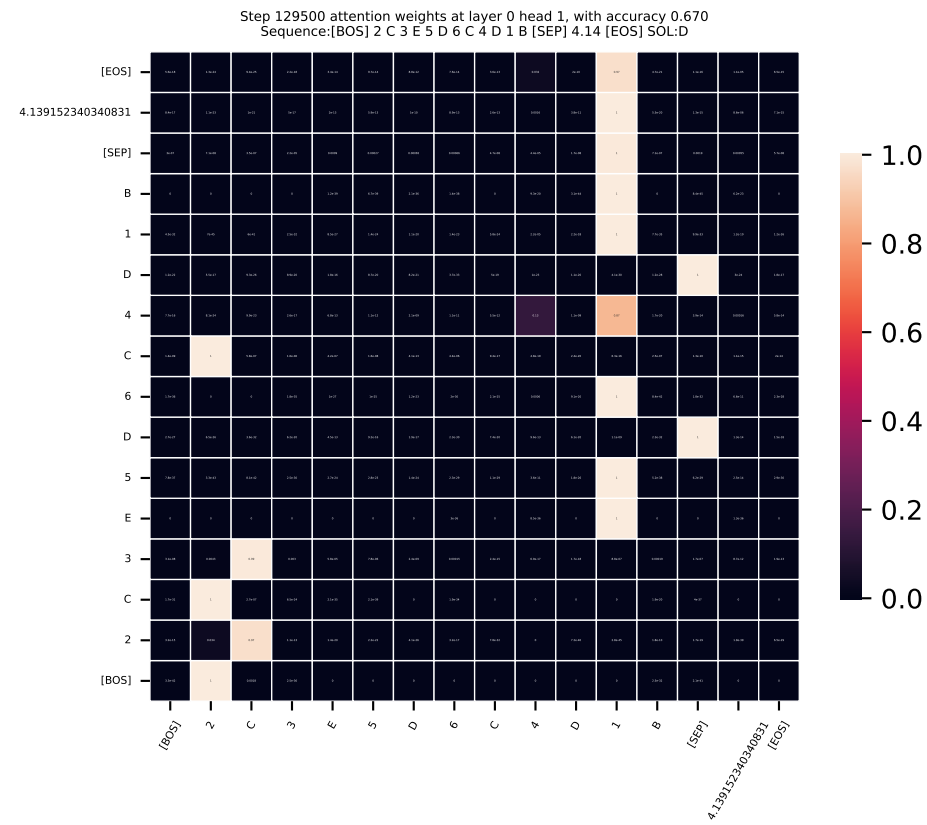

Step 120000 attention weights at layer 0 head 0, with accuracy 0.636
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2
0.0

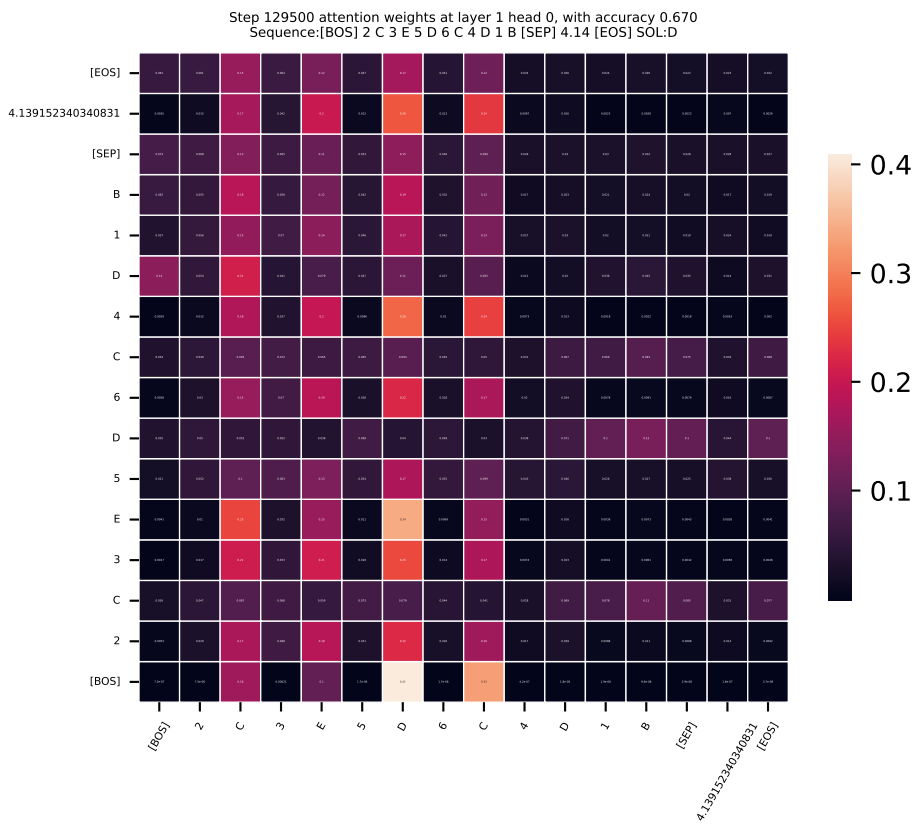

Step 120000 attention weights at layer 0 head 1, with accuracy 0.636
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

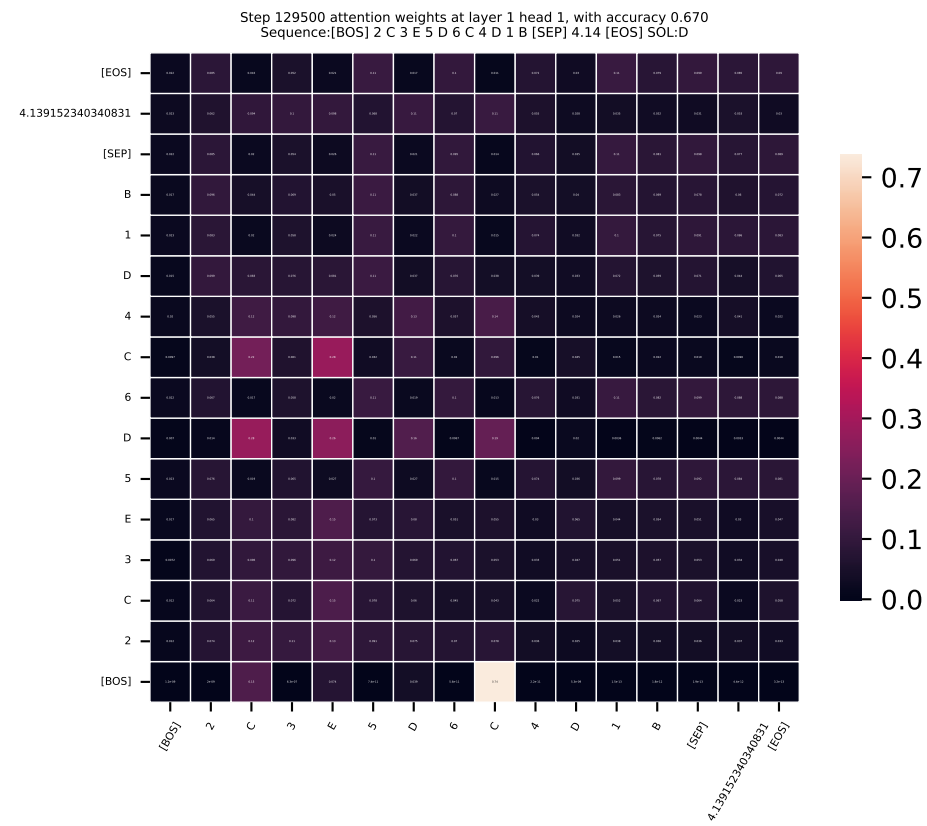

Step 120000 attention weights at layer 1 head 0, with accuracy 0.636
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.30
0.25
0.20
0.15
0.10
0.05

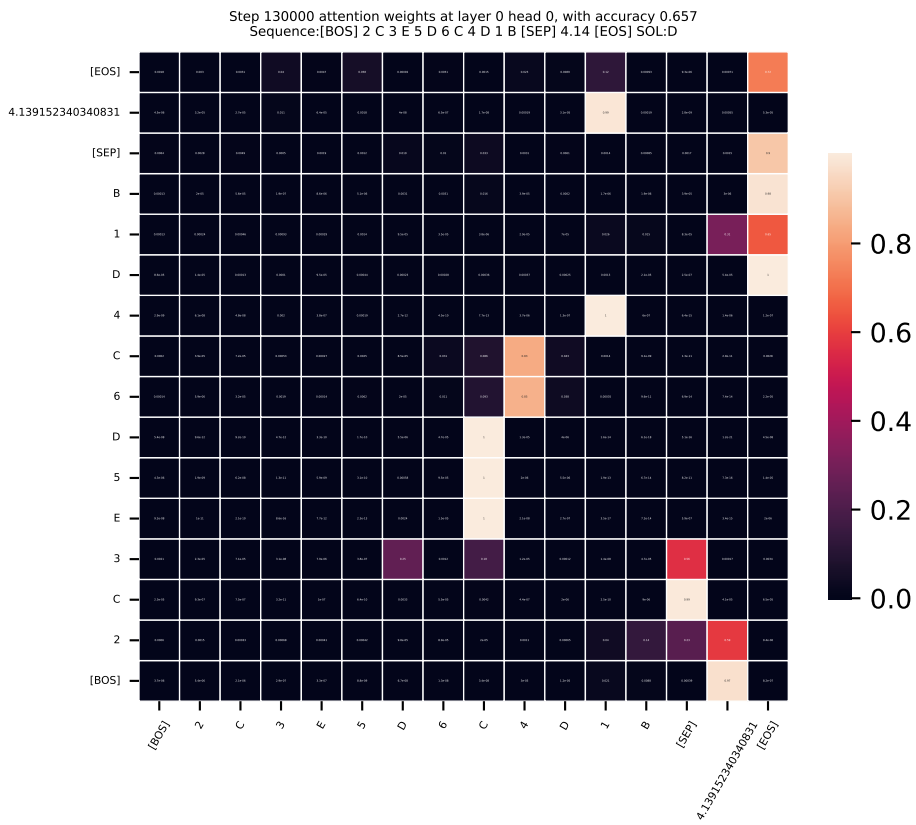

Step 120000 attention weights at layer 1 head 1, with accuracy 0.636
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.30
0.25
0.20
0.15
0.10
0.05

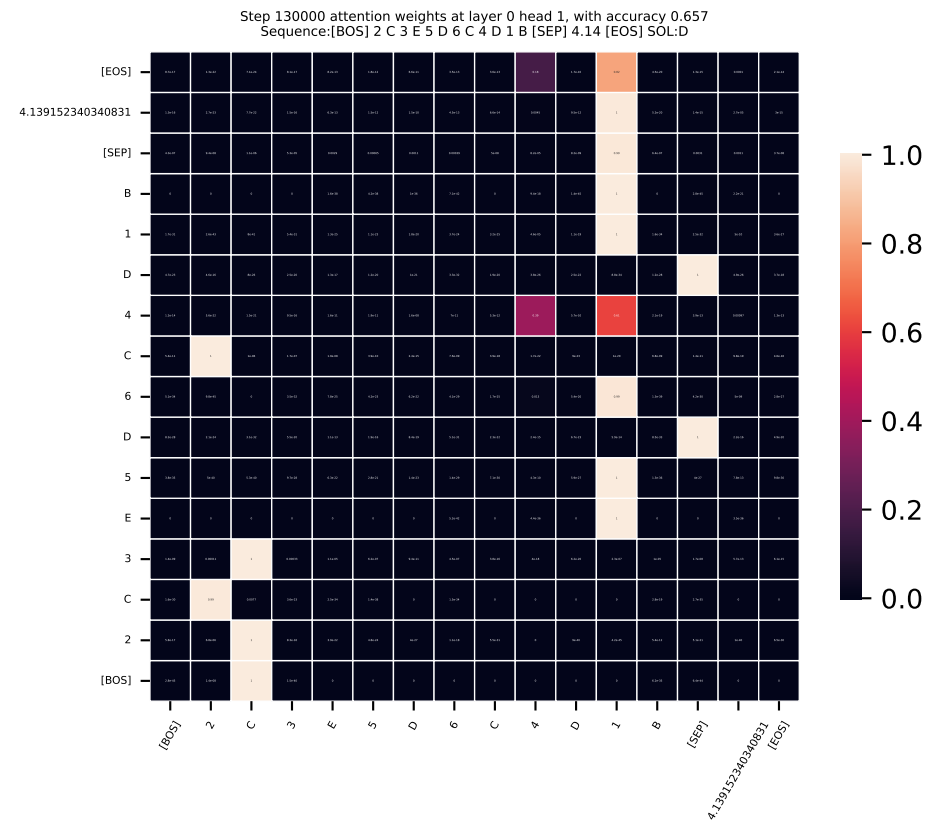

Step 129500 attention weights at layer 0 head 0, with accuracy 0.670
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.8
0.6
0.4
0.2
0.0

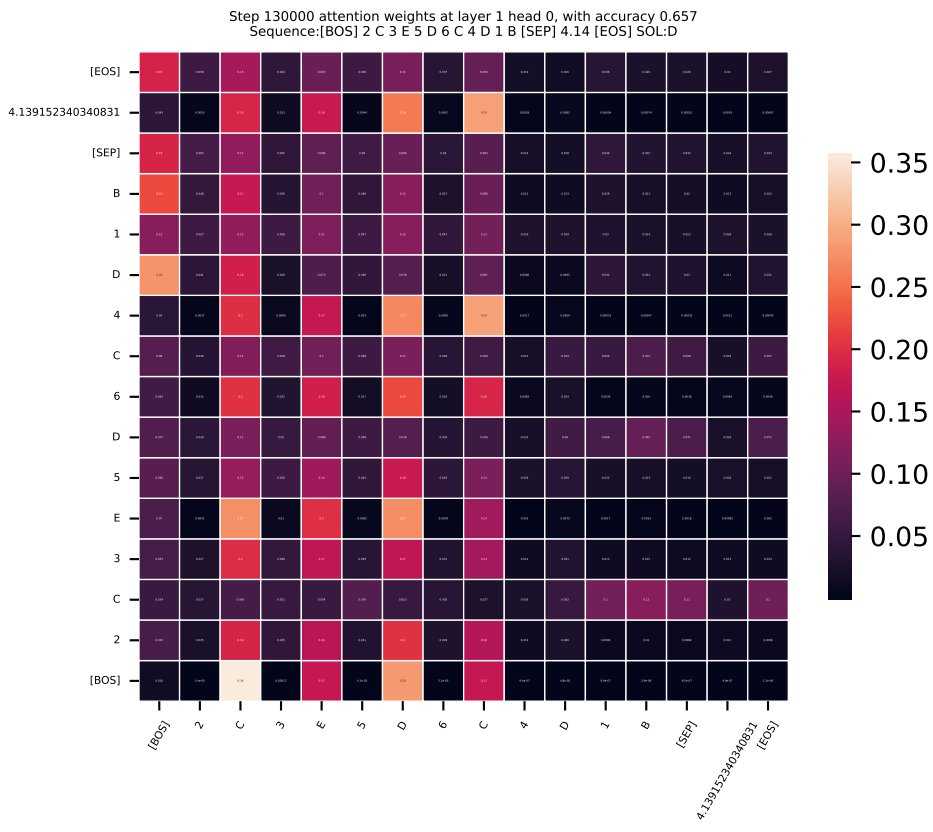

Step 129500 attention weights at layer 0 head 1, with accuracy 0.670
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
1.0
0.8
0.6
0.4
0.2
0.0

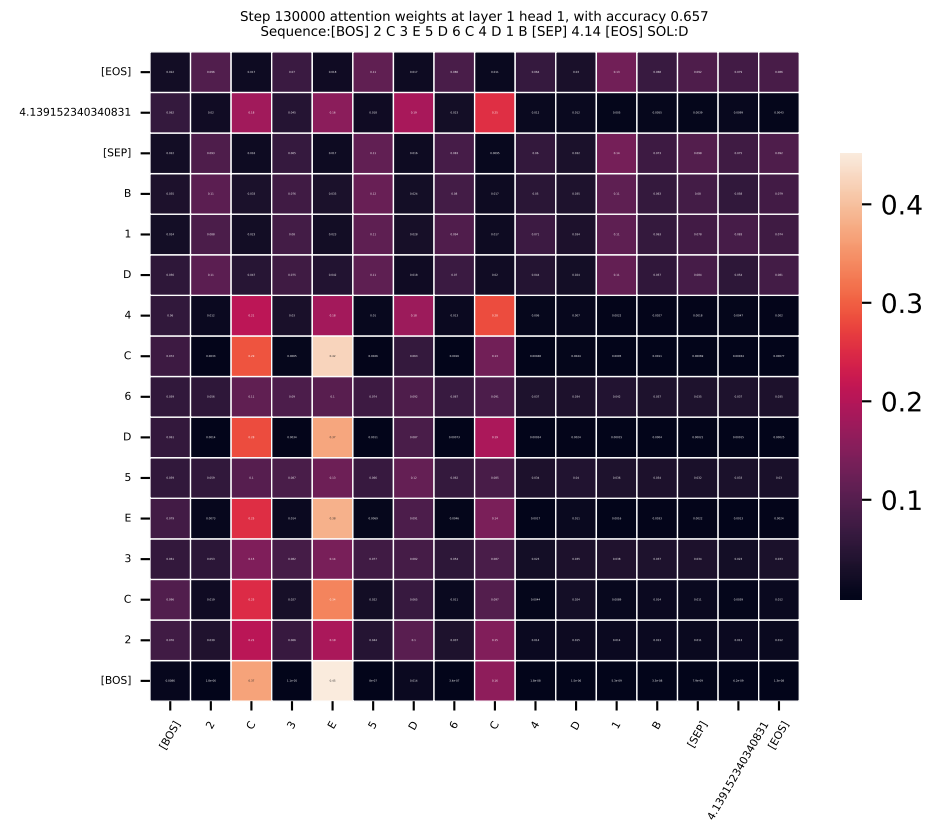

Step 129500 attention weights at layer 1 head 0, with accuracy 0.670
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.4
0.3
0.2
0.1

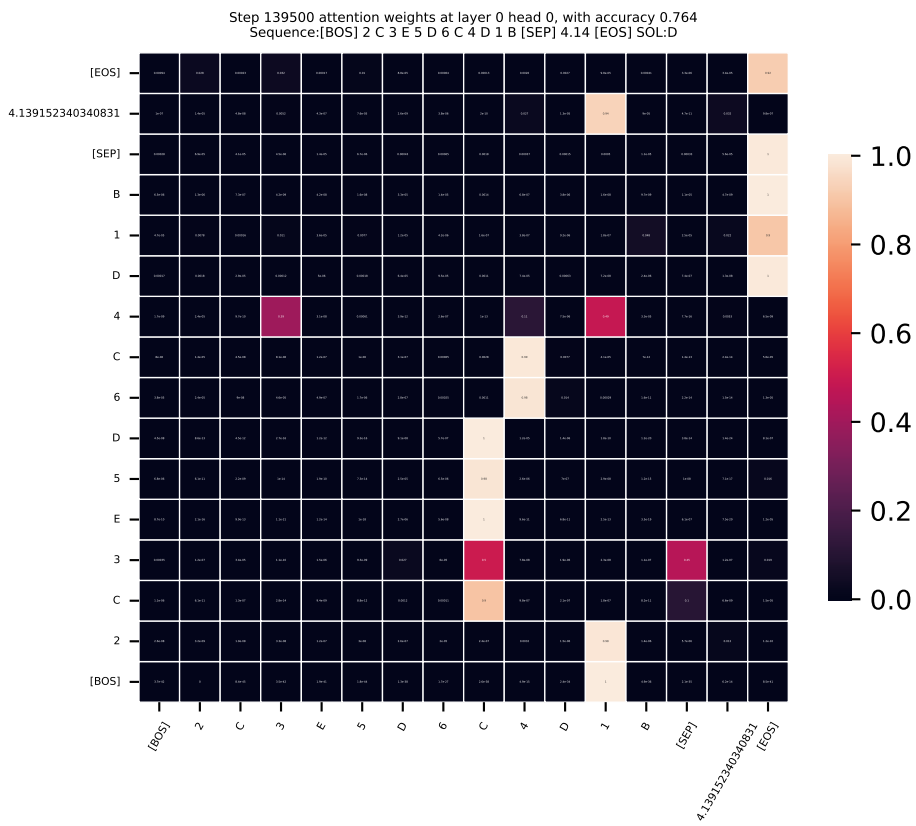

Step 129500 attention weights at layer 1 head 1, with accuracy 0.670
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0

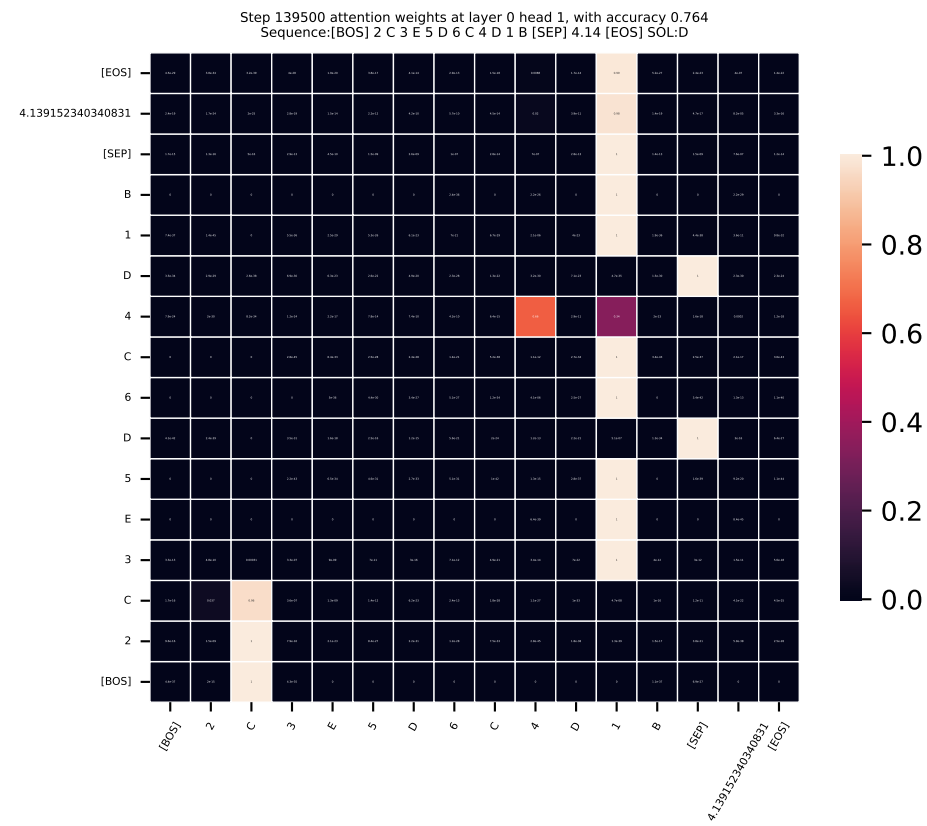

Step 130000 attention weights at layer 0 head 0, with accuracy 0.657
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2
0.0

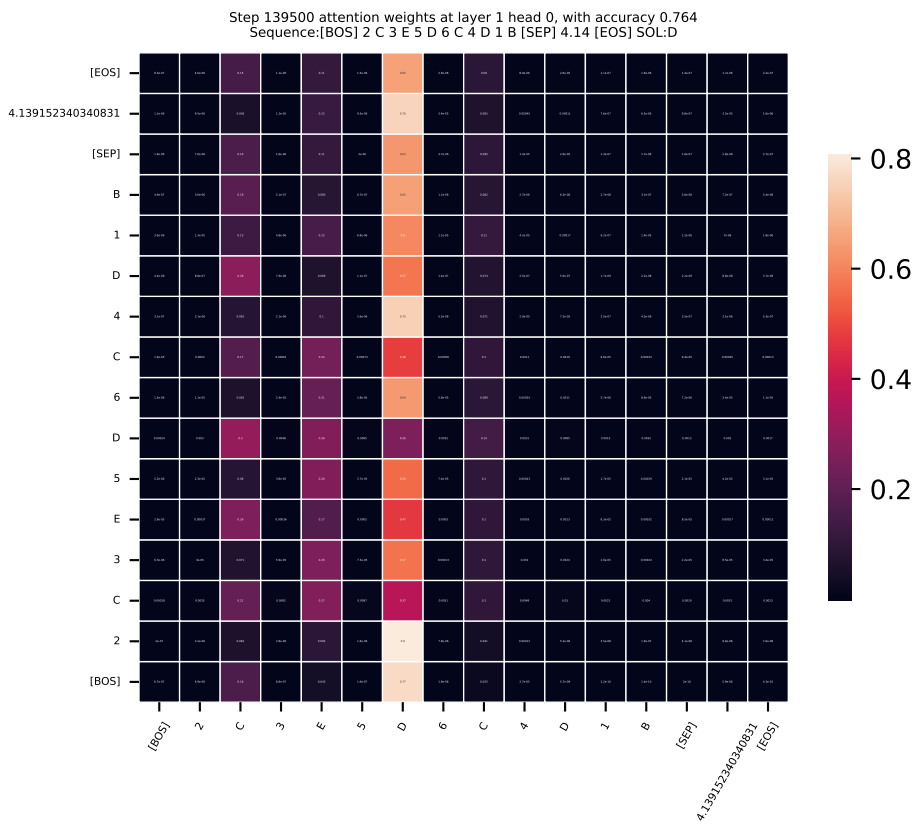

Step 130000 attention weights at layer 0 head 1, with accuracy 0.657
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

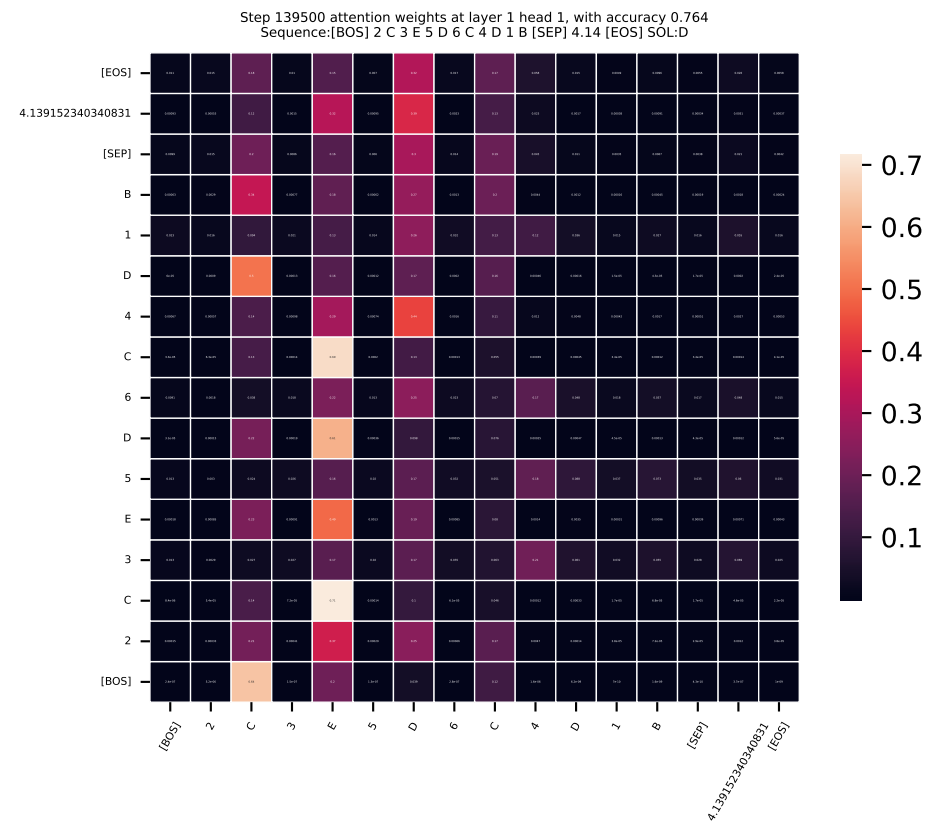

Step 130000 attention weights at layer 1 head 0, with accuracy 0.657
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.35
0.30
0.25
0.20
0.15
0.10
0.05

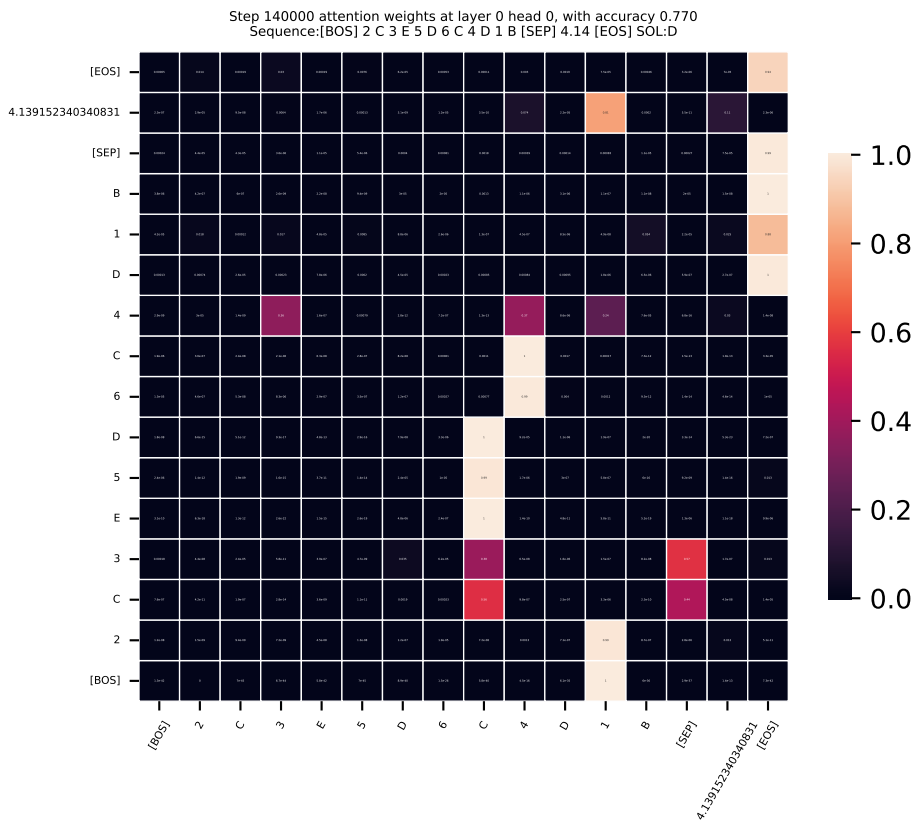

Step 130000 attention weights at layer 1 head 1, with accuracy 0.657
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.4
0.3
0.2
0.1

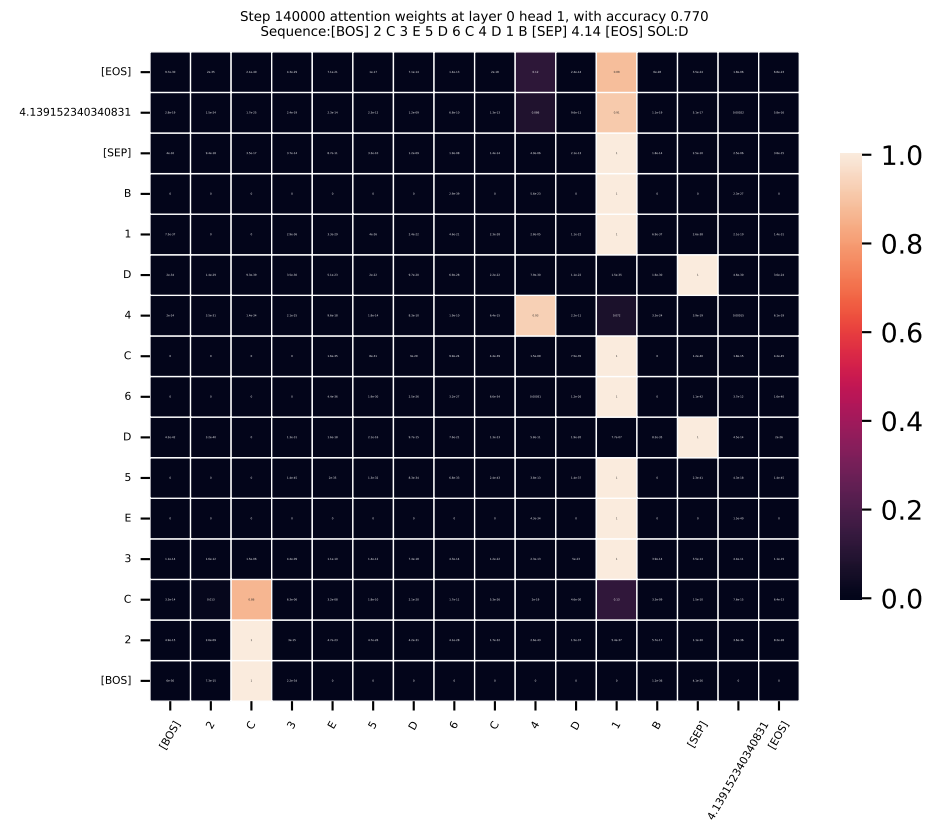

Step 139500 attention weights at layer 0 head 0, with accuracy 0.764
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

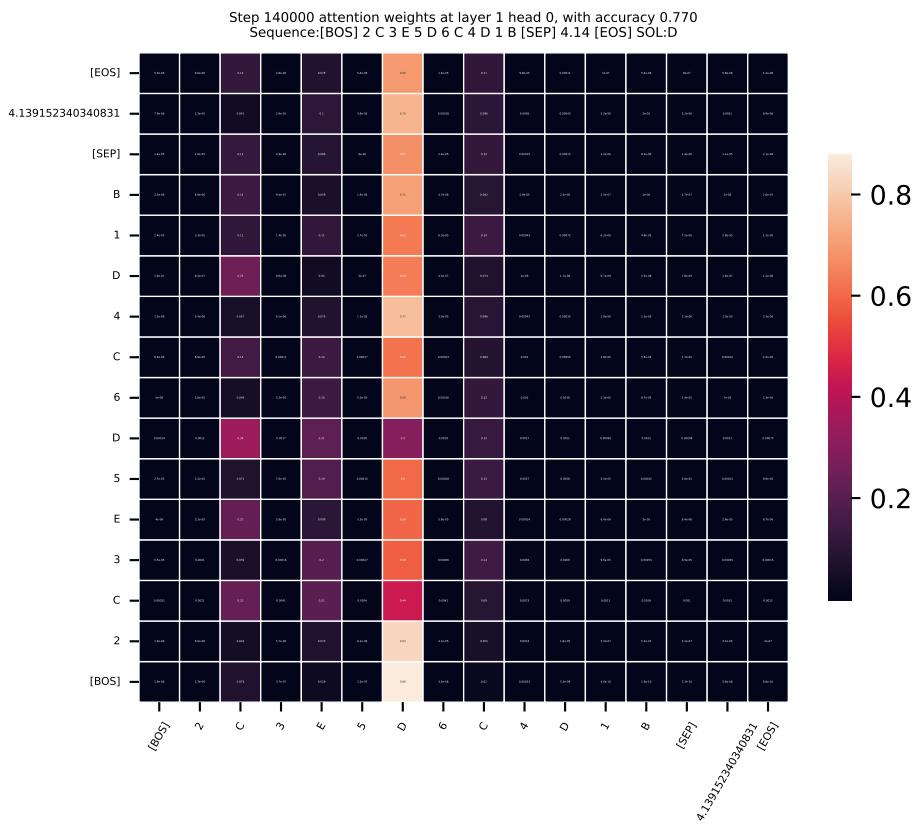

Step 139500 attention weights at layer 0 head 1, with accuracy 0.764
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

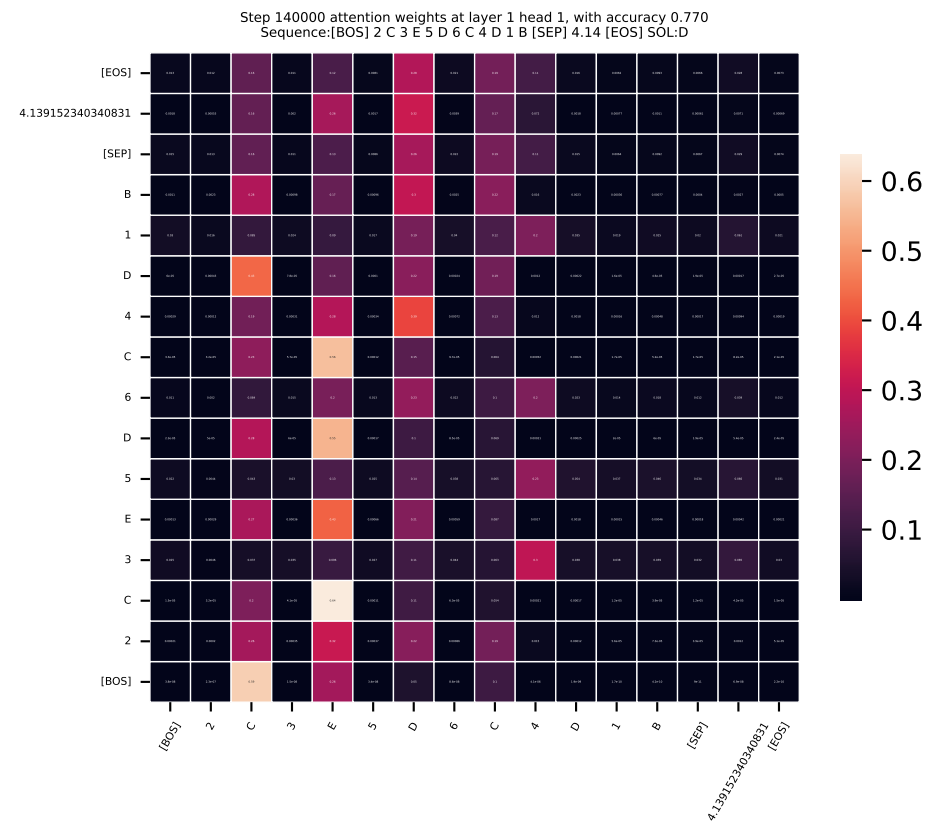

Step 139500 attention weights at layer 1 head 0, with accuracy 0.764
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.8
0.6
0.4
0.2

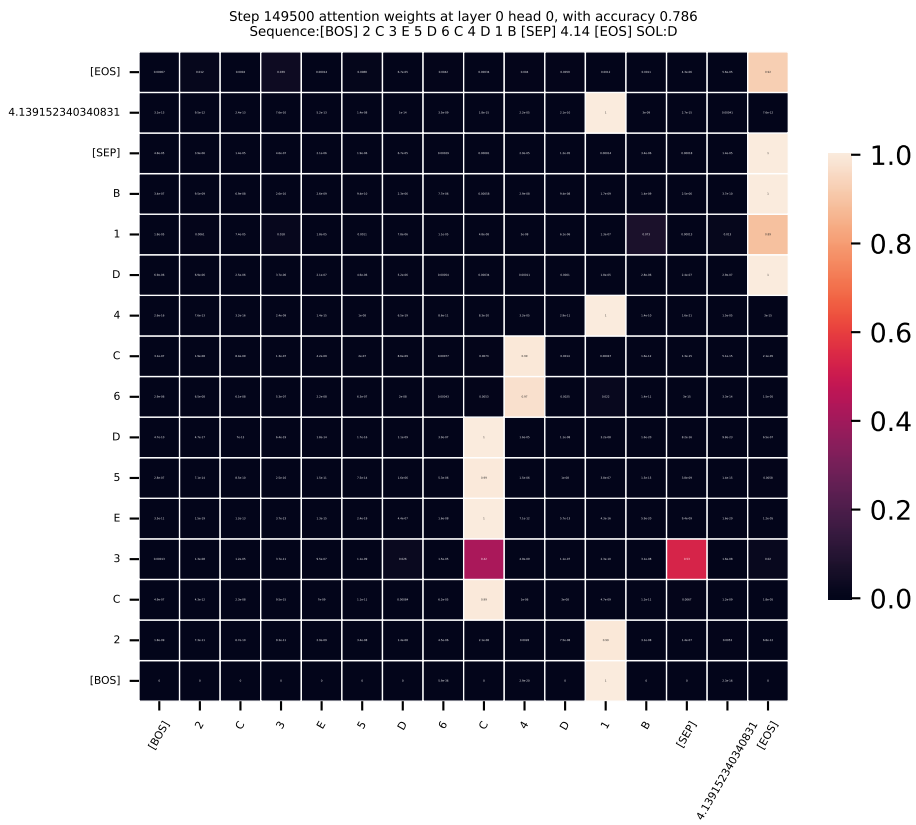

Step 139500 attention weights at layer 1 head 1, with accuracy 0.764
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.7
0.6
0.5
0.4
0.3
0.2
0.1

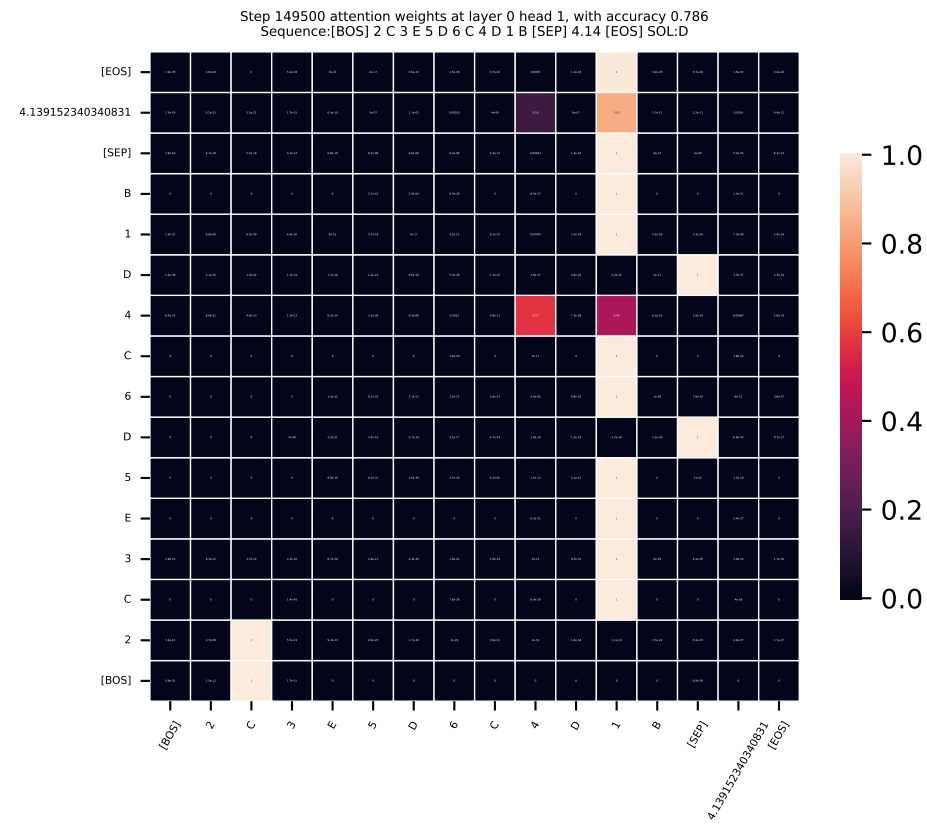

Step 140000 attention weights at layer 0 head 0, with accuracy 0.770
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
1.0
0.8
0.6
0.4
0.2
0.0

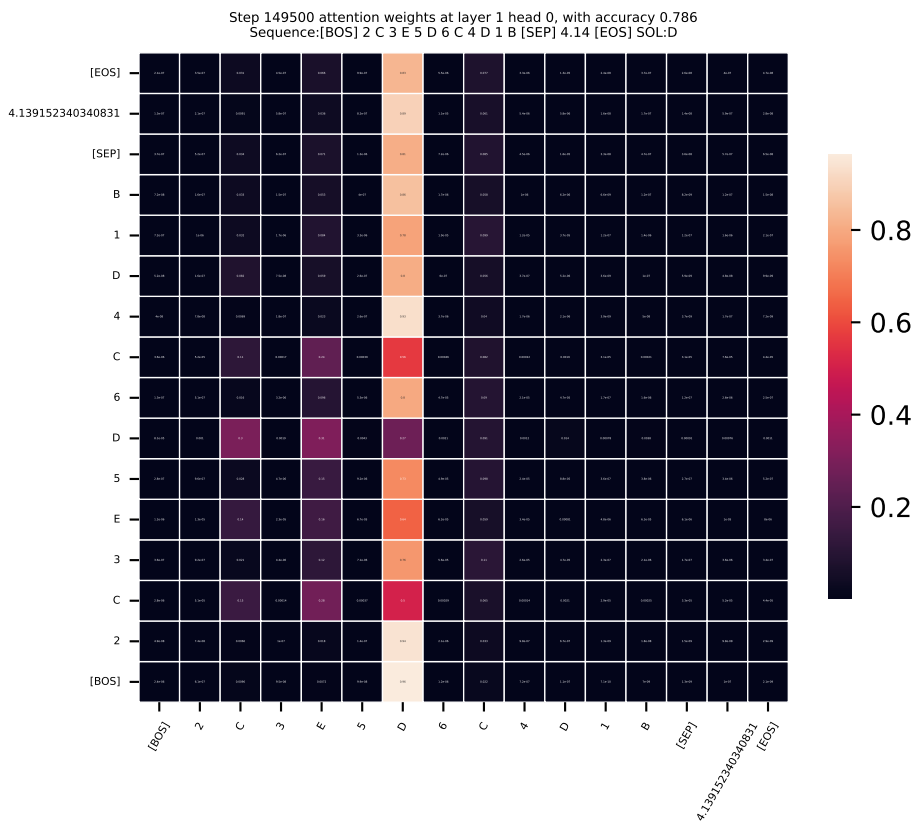

Step 140000 attention weights at layer 0 head 1, with accuracy 0.770
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
1.0
0.8
0.6
0.4
0.2
0.0

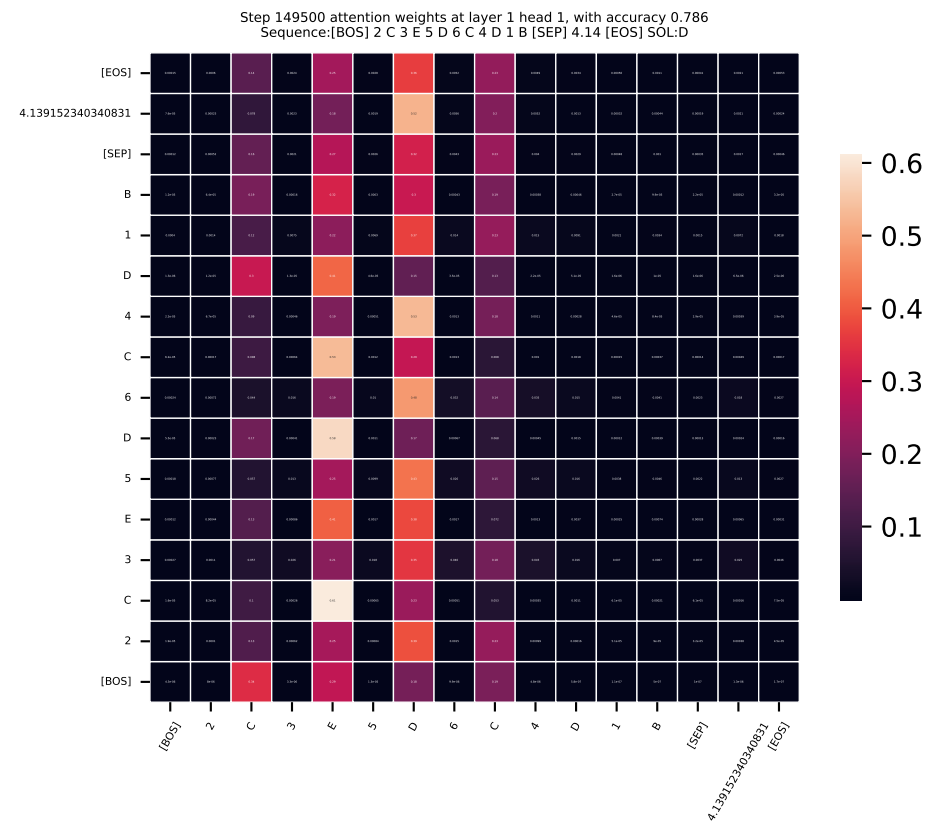

Step 140000 attention weights at layer 1 head 0, with accuracy 0.770
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2

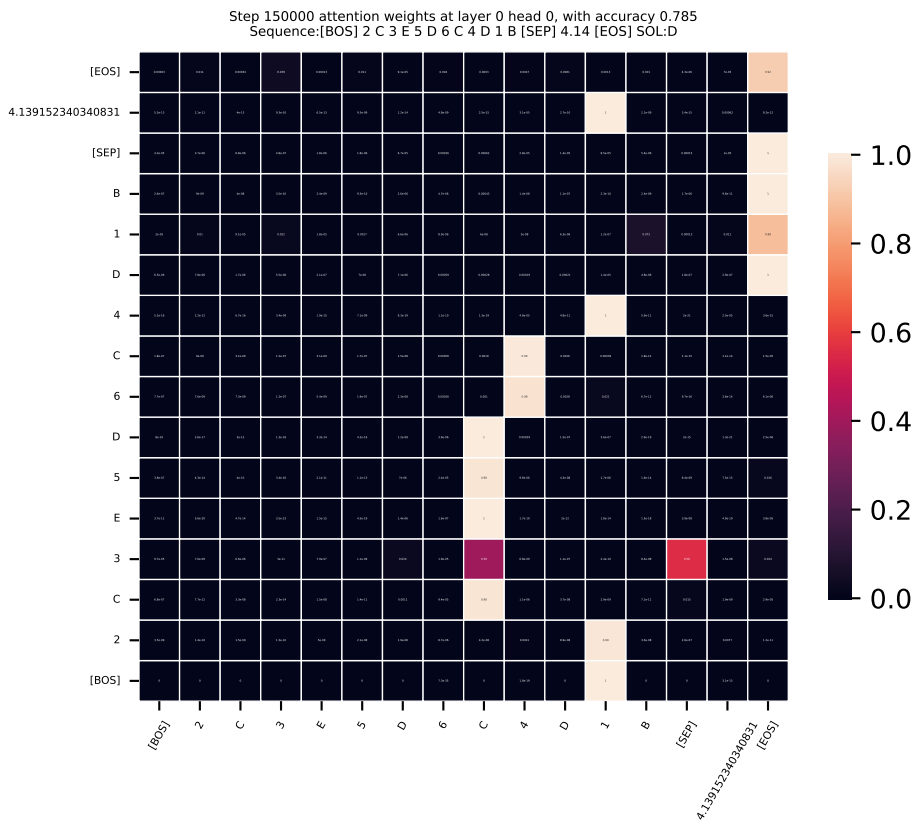

Step 140000 attention weights at layer 1 head 1, with accuracy 0.770
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.6
0.5
0.4
0.3
0.2
0.1

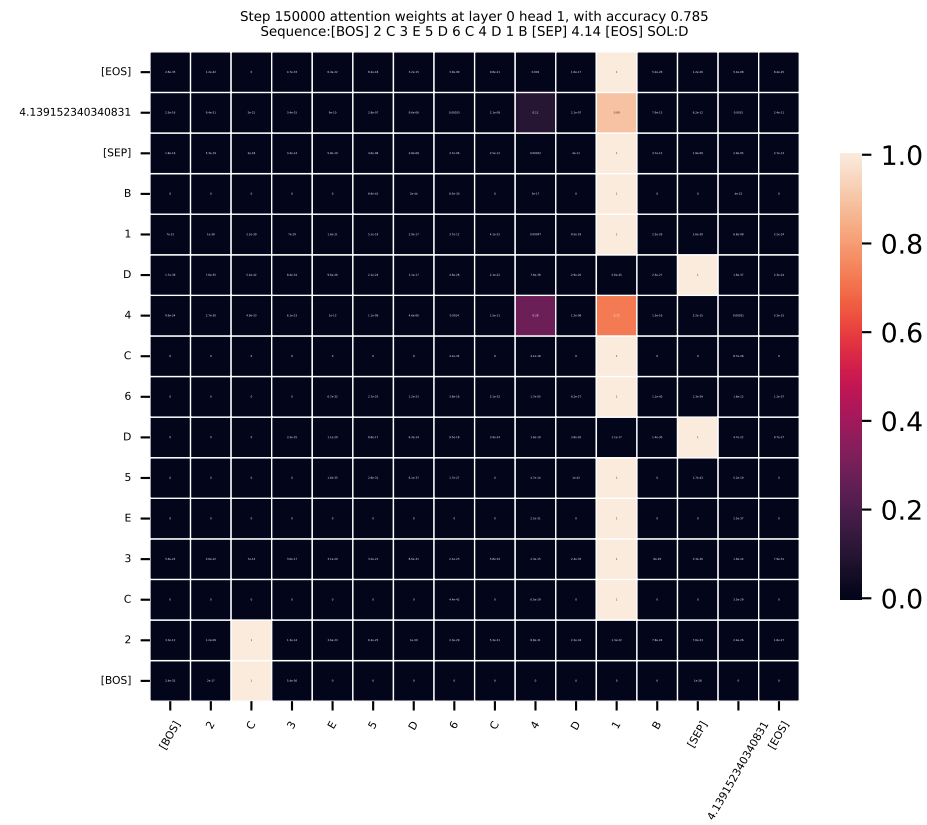

Step 149500 attention weights at layer 0 head 0, with accuracy 0.786
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
1.0
0.8
0.6
0.4
0.2
0.0

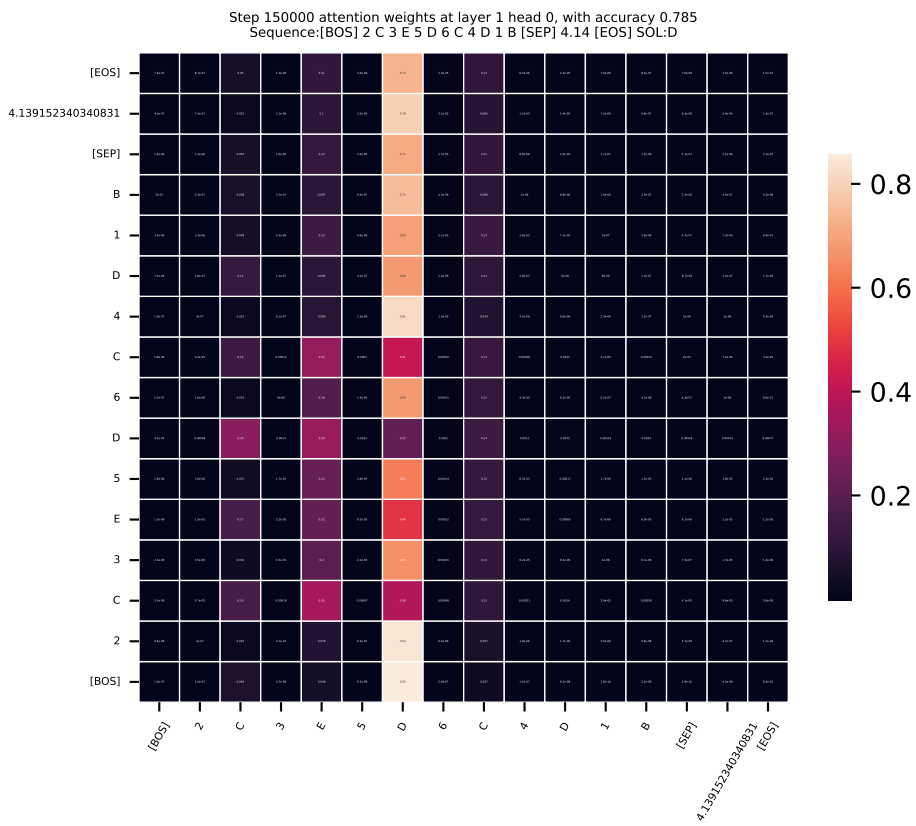

Step 149500 attention weights at layer 0 head 1, with accuracy 0.786
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
1.0
0.8
0.6
0.4
0.2
0.0

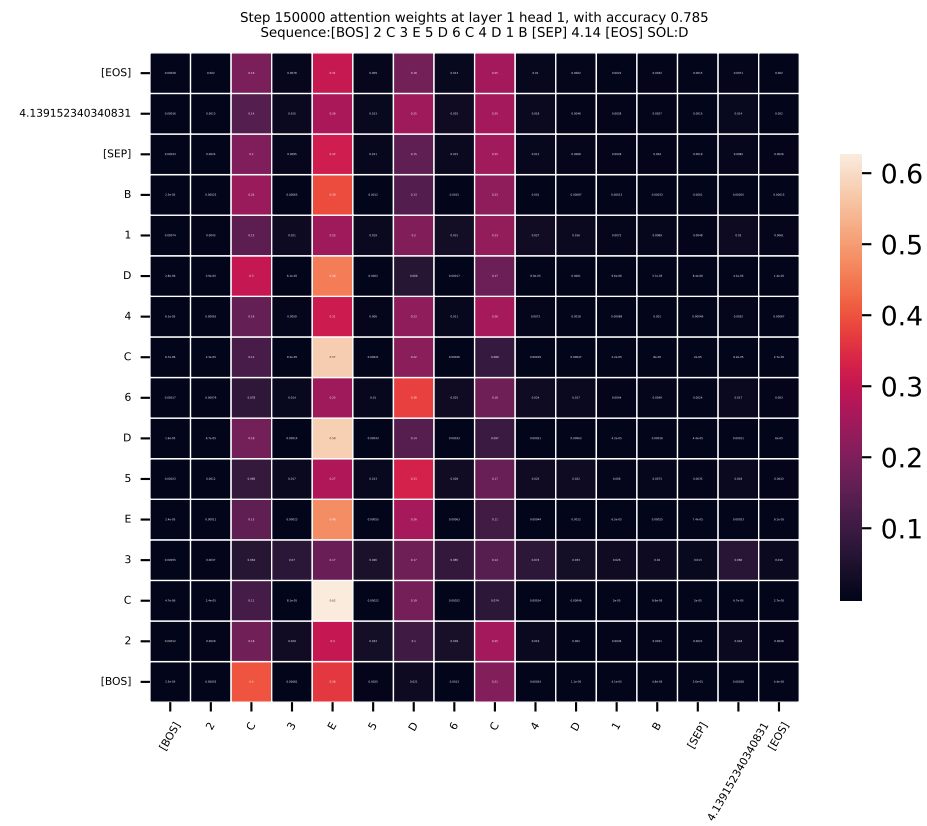

Step 149500 attention weights at layer 1 head 0, with accuracy 0.786
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.8
0.6
0.4
0.2

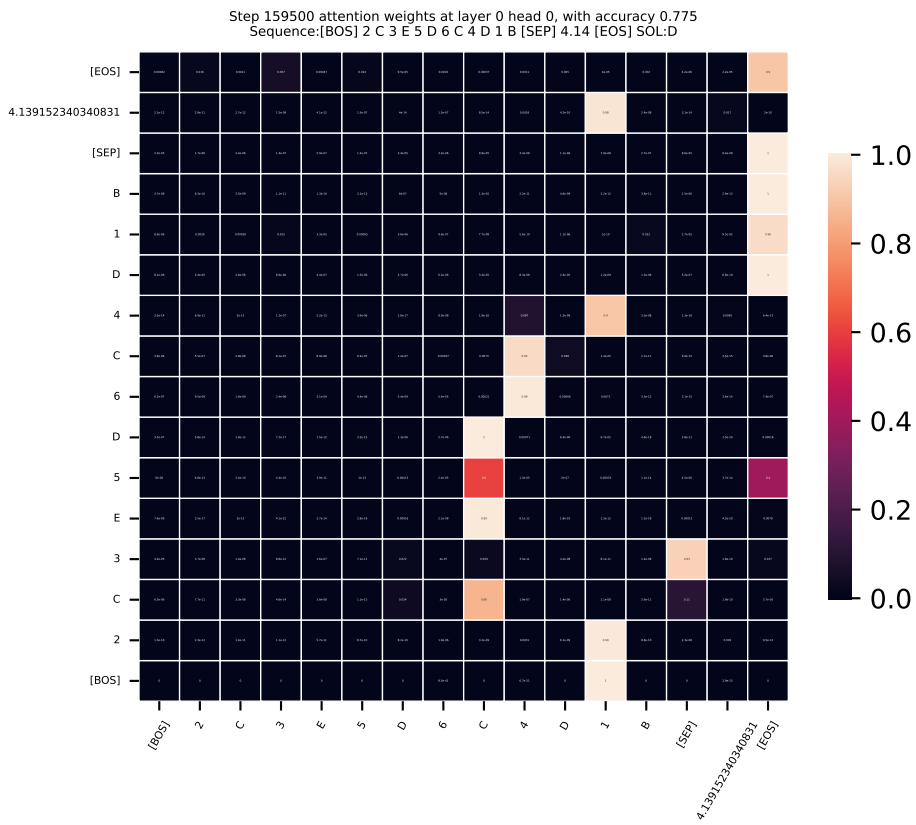

Step 149500 attention weights at layer 1 head 1, with accuracy 0.786
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.6
0.5
0.4
0.3
0.2
0.1

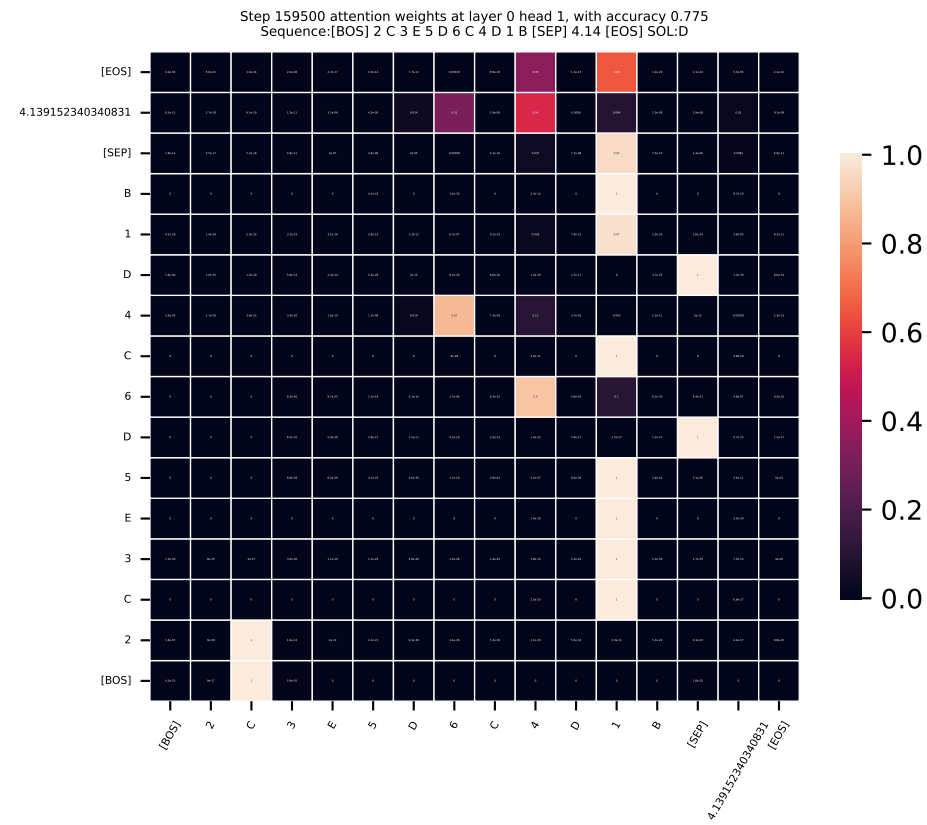

Step 150000 attention weights at layer 0 head 0, with accuracy 0.785
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
1.0
0.8
0.6
0.4
0.2
0.0

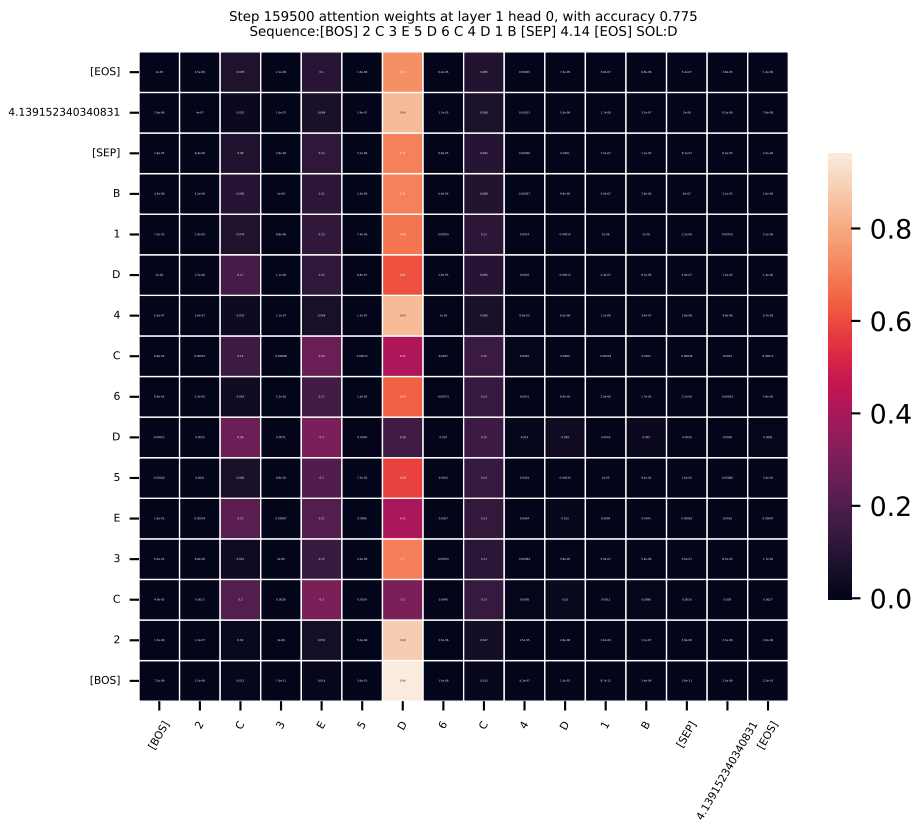

Step 150000 attention weights at layer 0 head 1, with accuracy 0.785
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
1.0
0.8
0.6
0.4
0.2
0.0

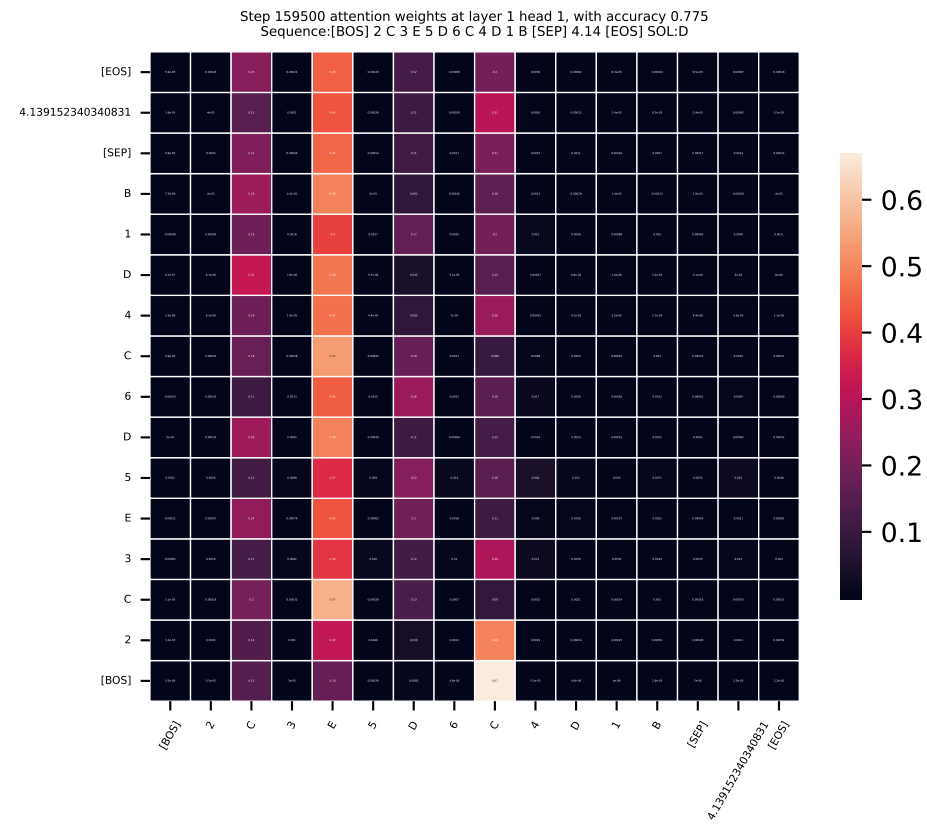

Step 150000 attention weights at layer 1 head 0, with accuracy 0.785
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.8
0.6
0.4
0.2

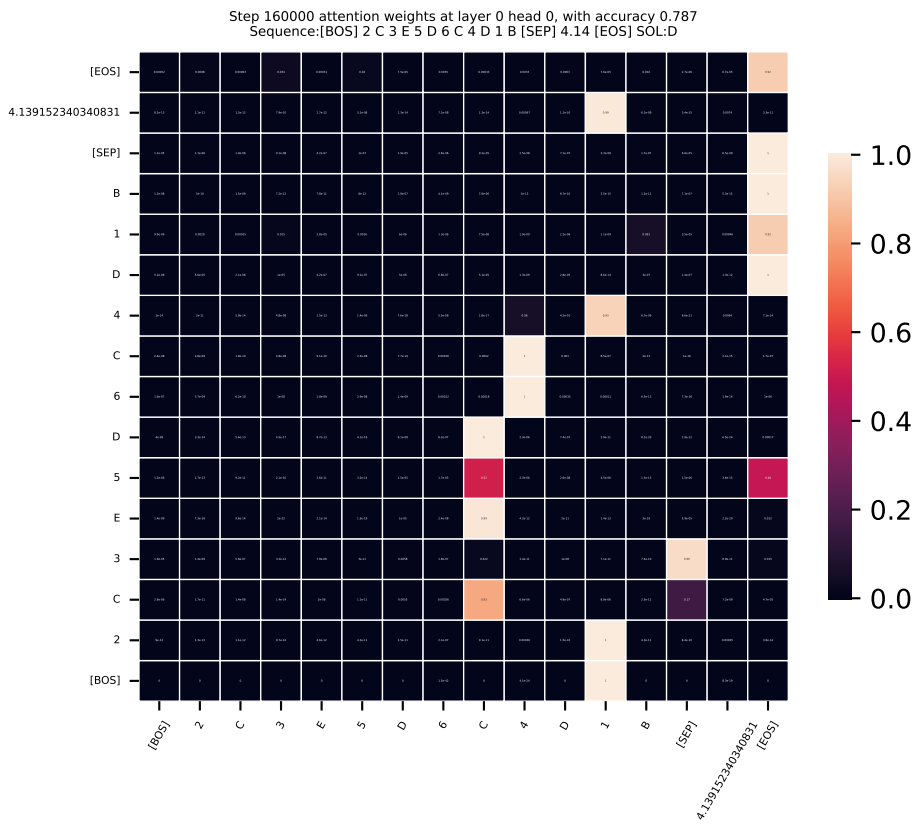

Step 150000 attention weights at layer 1 head 1, with accuracy 0.785
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.6
0.5
0.4
0.3
0.2
0.1

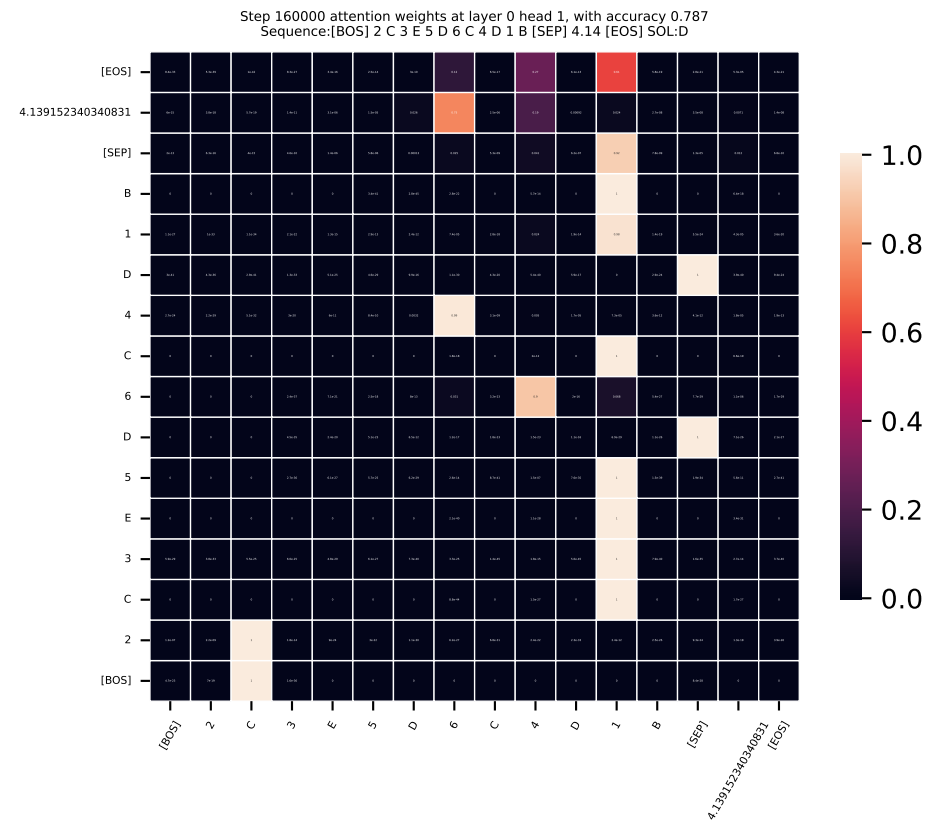

Step 159500 attention weights at layer 0 head 0, with accuracy 0.775
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
1.0
0.8
0.6
0.4
0.2
0.0

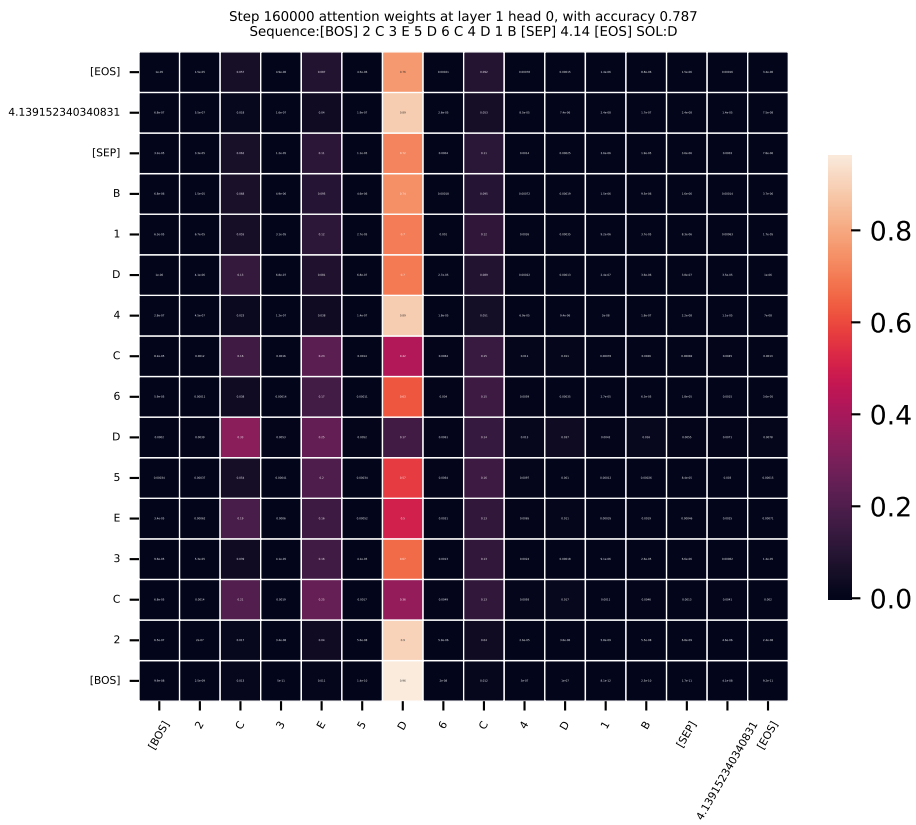

Step 159500 attention weights at layer 0 head 1, with accuracy 0.775
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
1.0
0.8
0.6
0.4
0.2
0.0

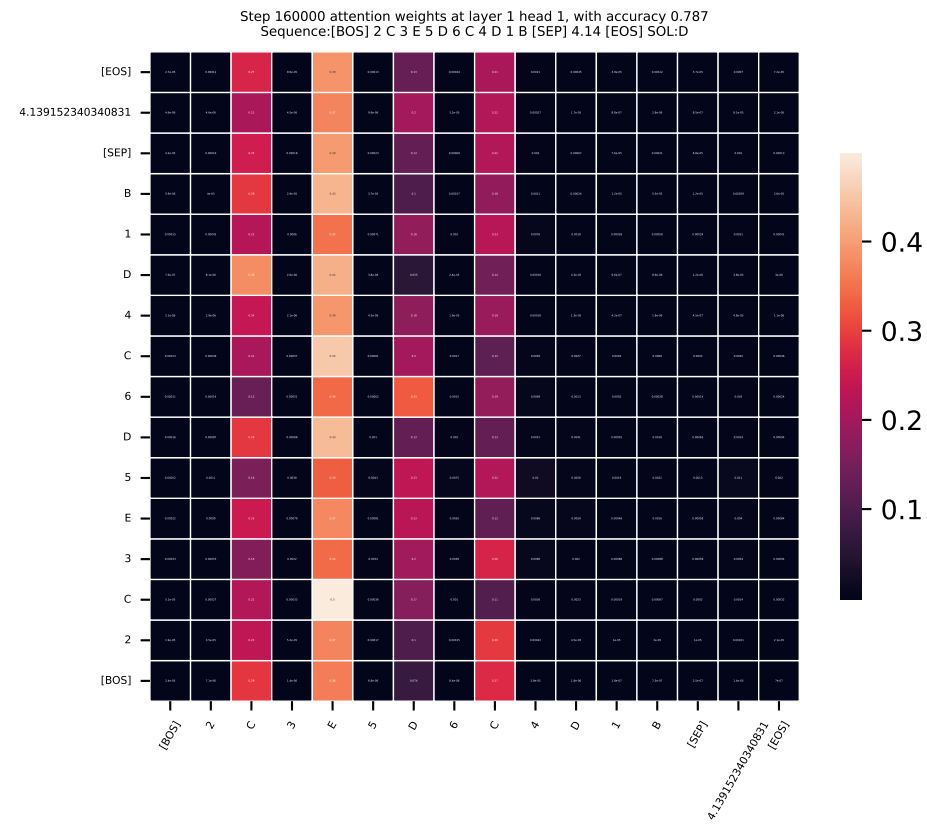

Step 159500 attention weights at layer 1 head 0, with accuracy 0.775
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2
0.0

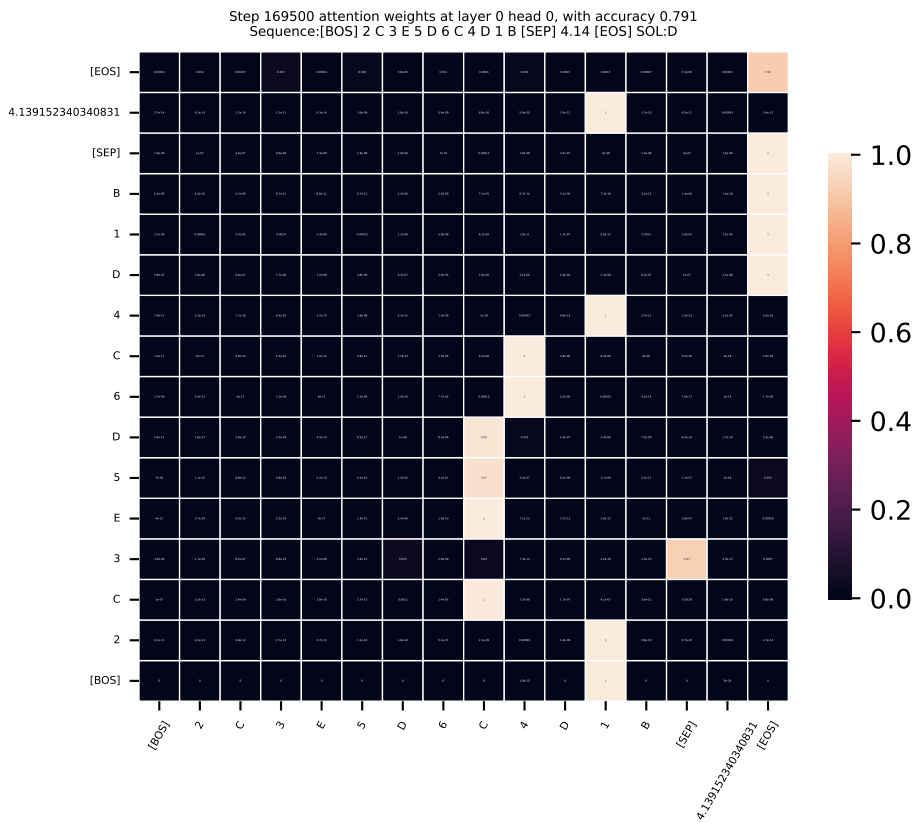

Step 159500 attention weights at layer 1 head 1, with accuracy 0.775
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.6
0.5
0.4
0.3
0.2
0.1

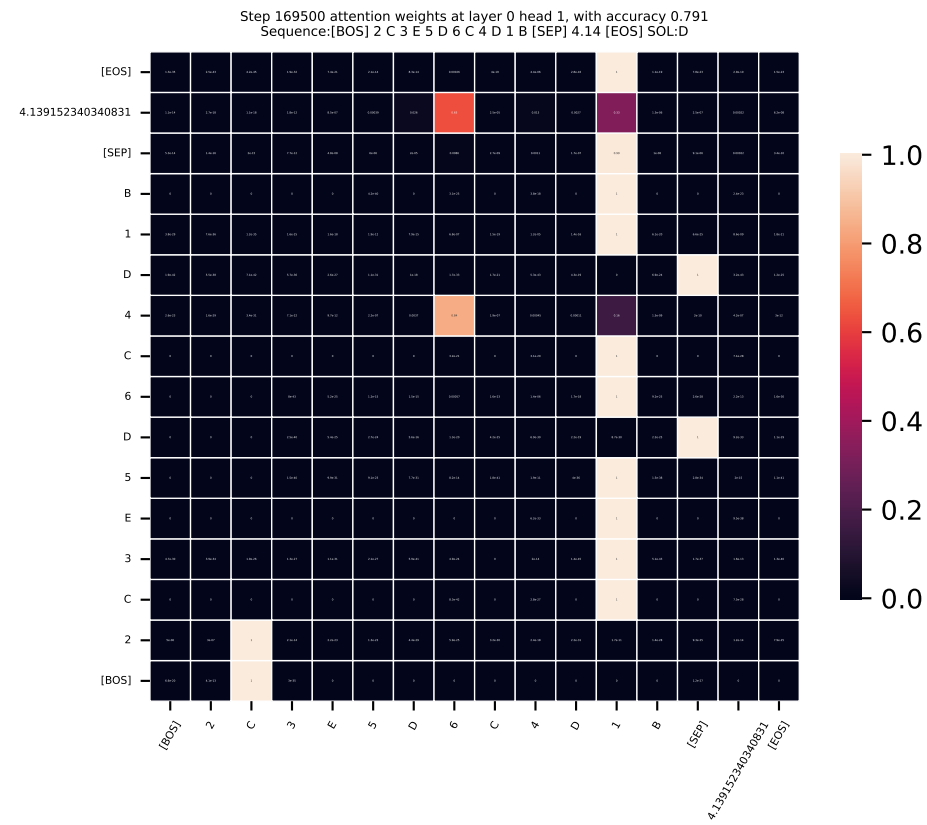

Step 160000 attention weights at layer 0 head 0, with accuracy 0.787
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

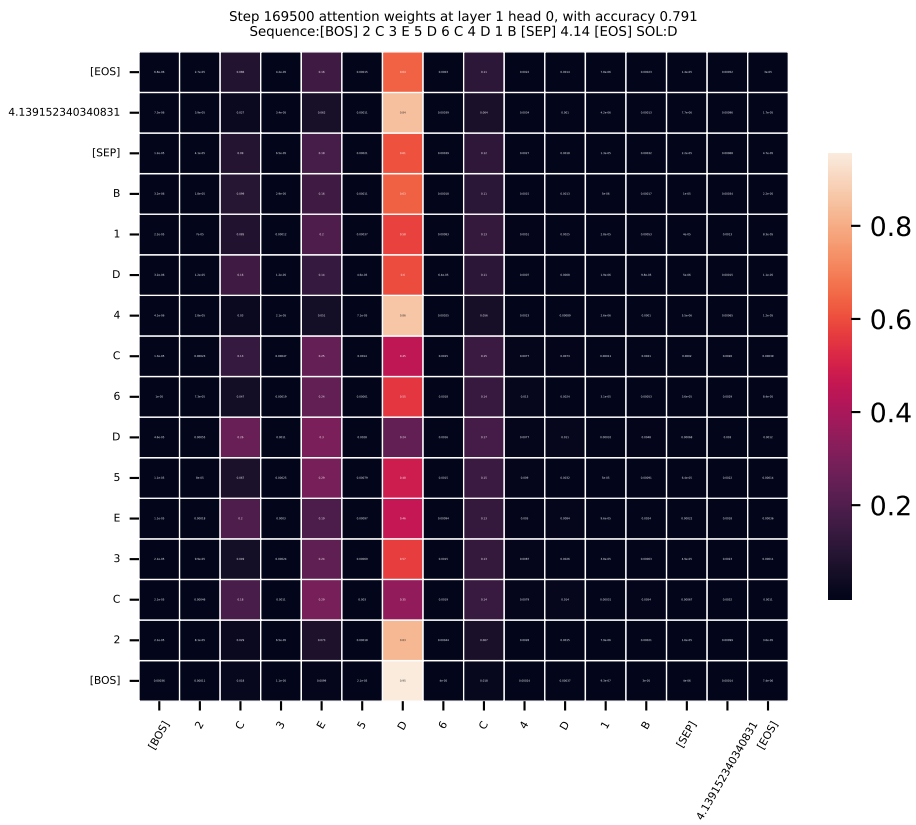

Step 160000 attention weights at layer 0 head 1, with accuracy 0.787
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

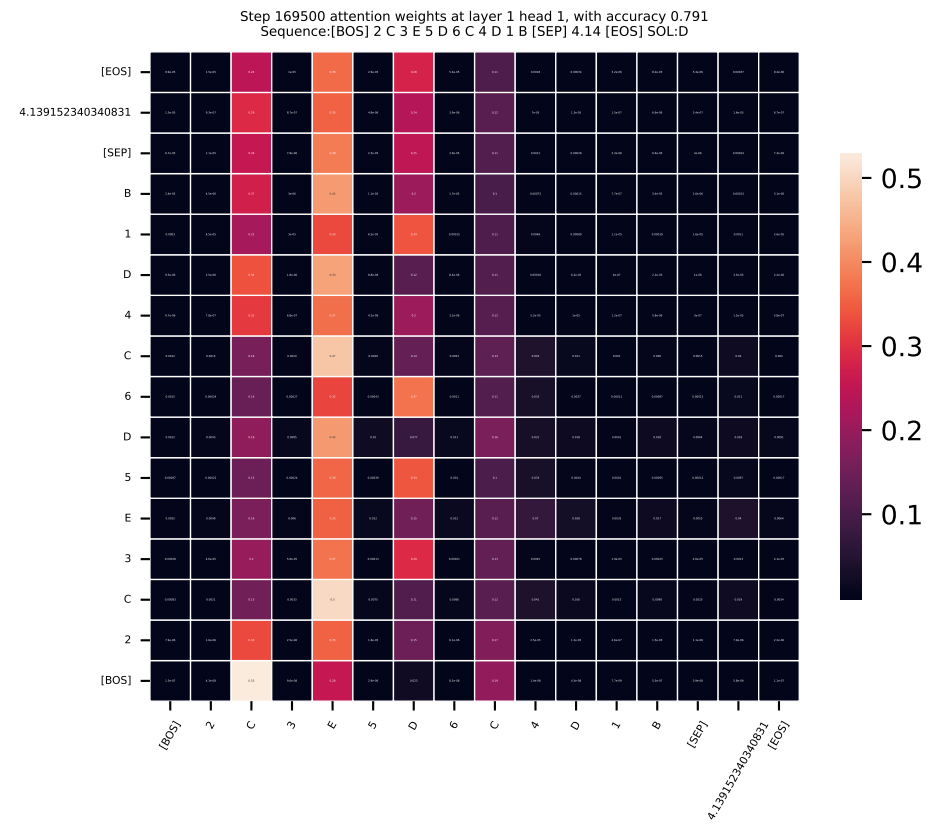

Step 160000 attention weights at layer 1 head 0, with accuracy 0.787
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.8
0.6
0.4
0.2
0.0

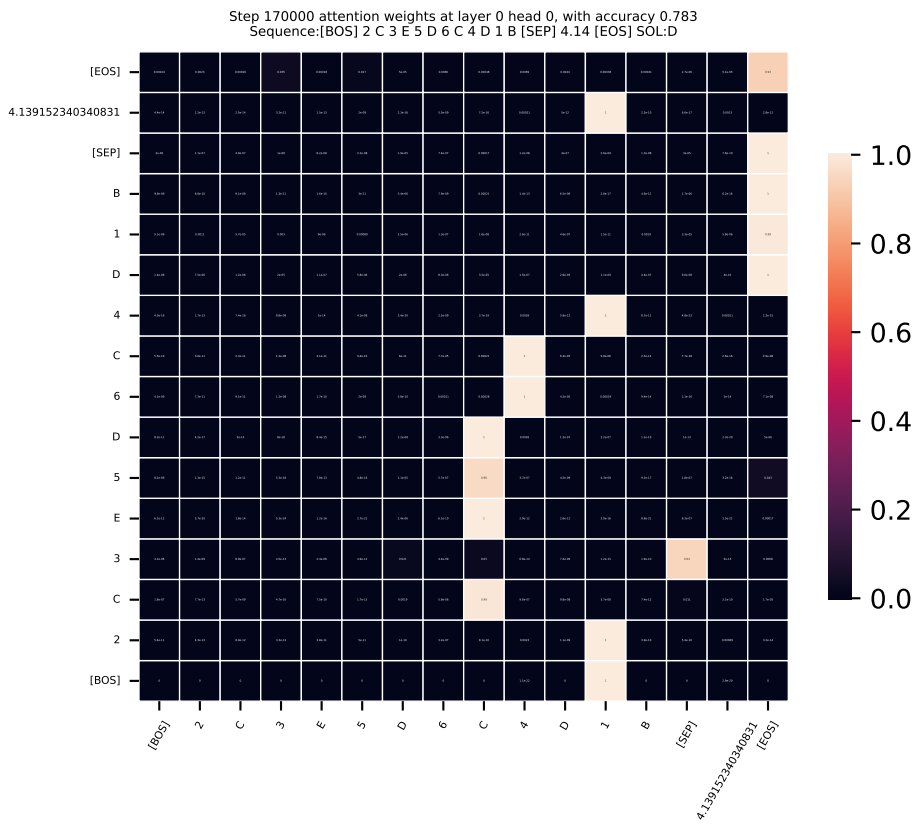

Step 160000 attention weights at layer 1 head 1, with accuracy 0.787
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.4
0.3
0.2
0.1

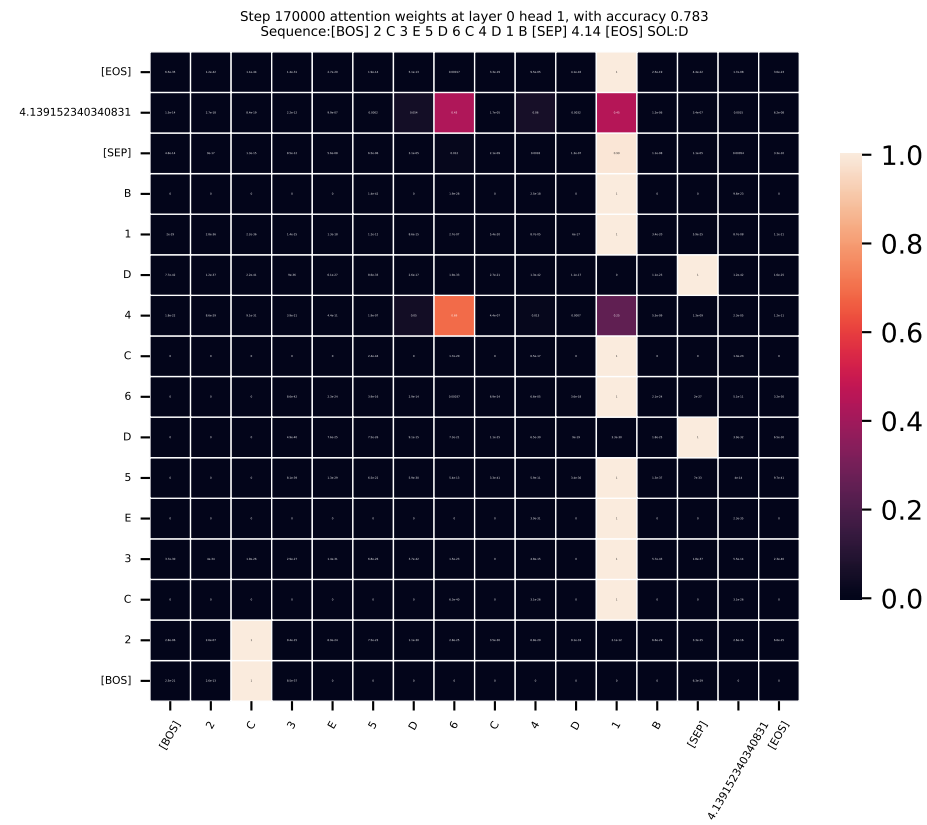

Step 169500 attention weights at layer 0 head 0, with accuracy 0.791
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
1.0
0.8
0.6
0.4
0.2
0.0

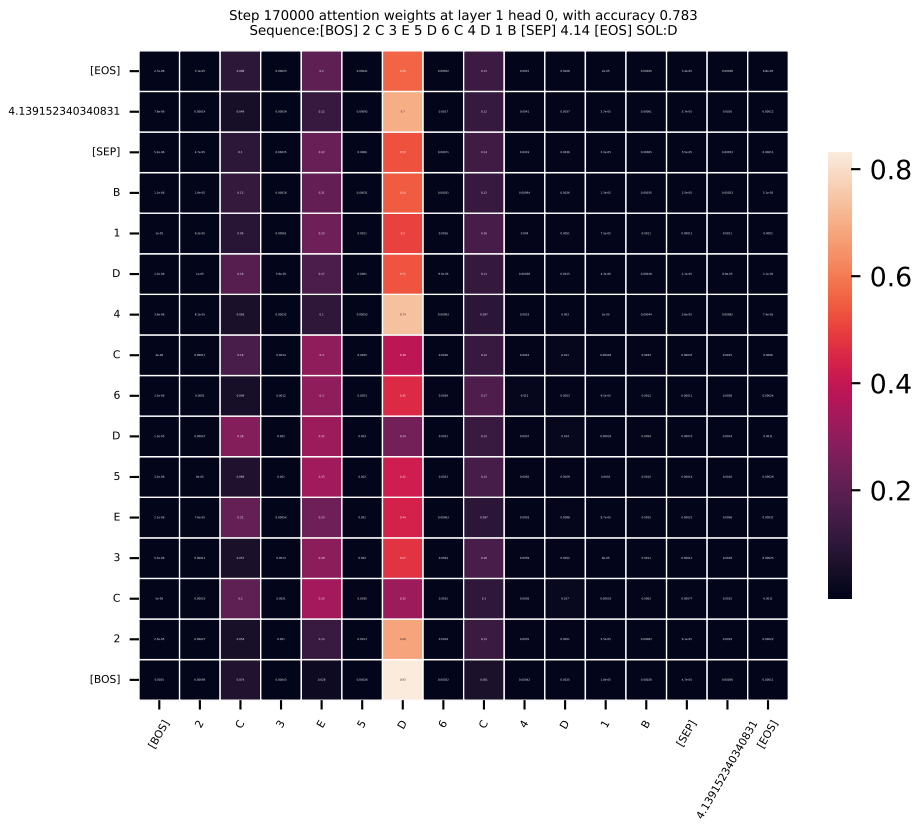

Step 169500 attention weights at layer 0 head 1, with accuracy 0.791
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
1.0
0.8
0.6
0.4
0.2
0.0

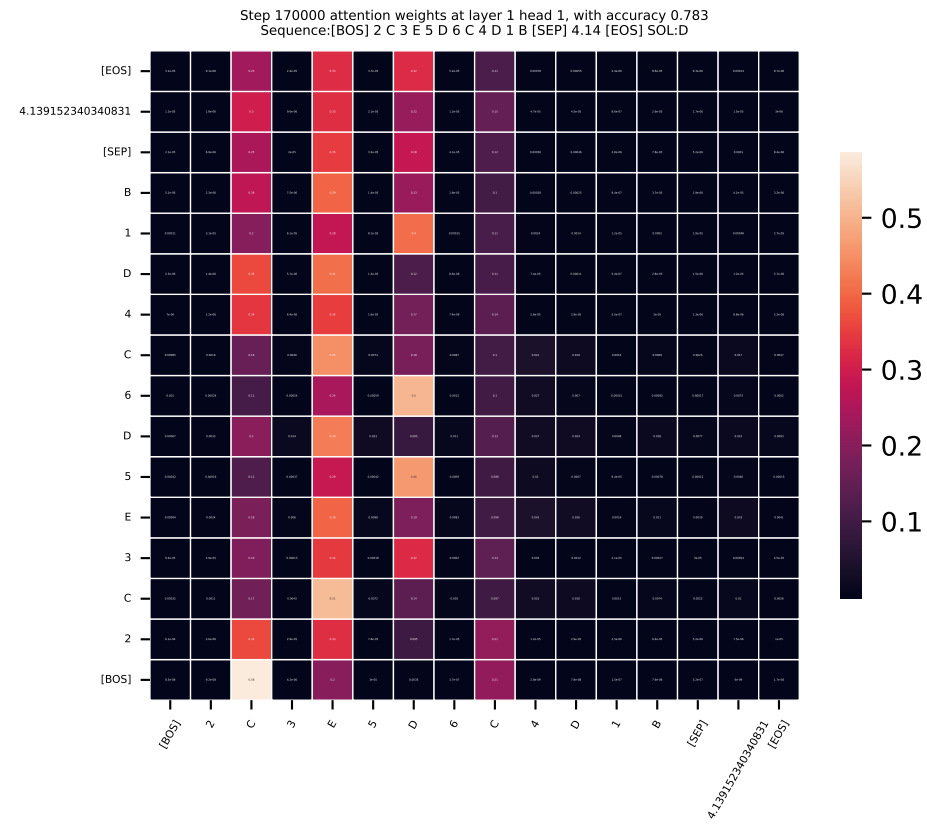

Step 169500 attention weights at layer 1 head 0, with accuracy 0.791
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2

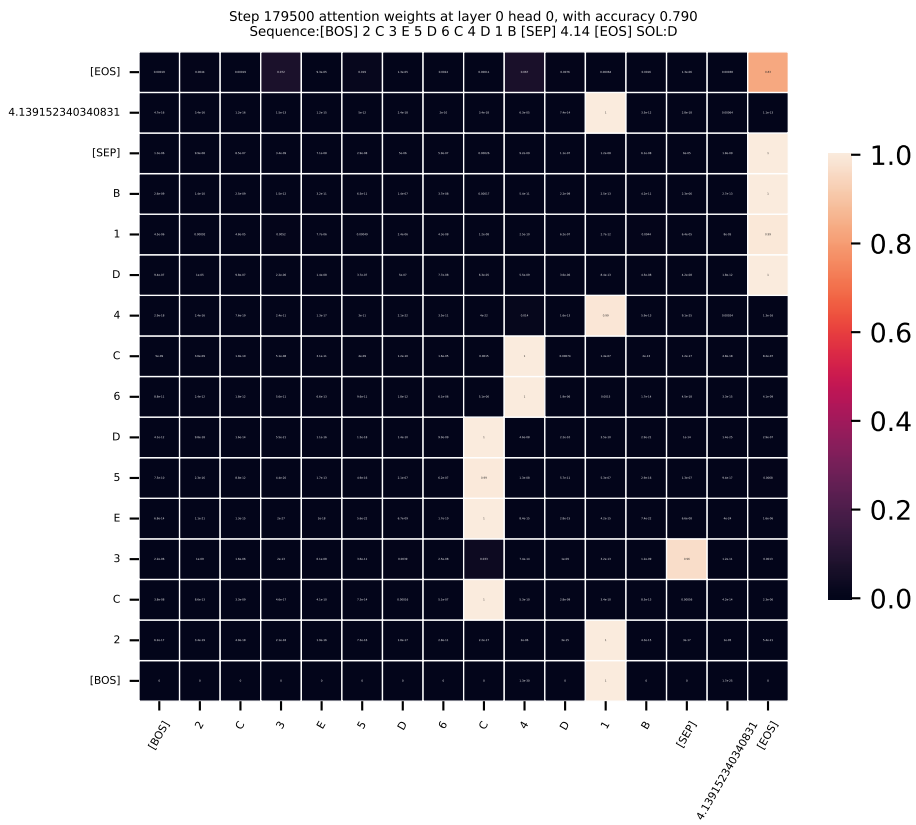

Step 169500 attention weights at layer 1 head 1, with accuracy 0.791
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.5
0.4
0.3
0.2
0.1

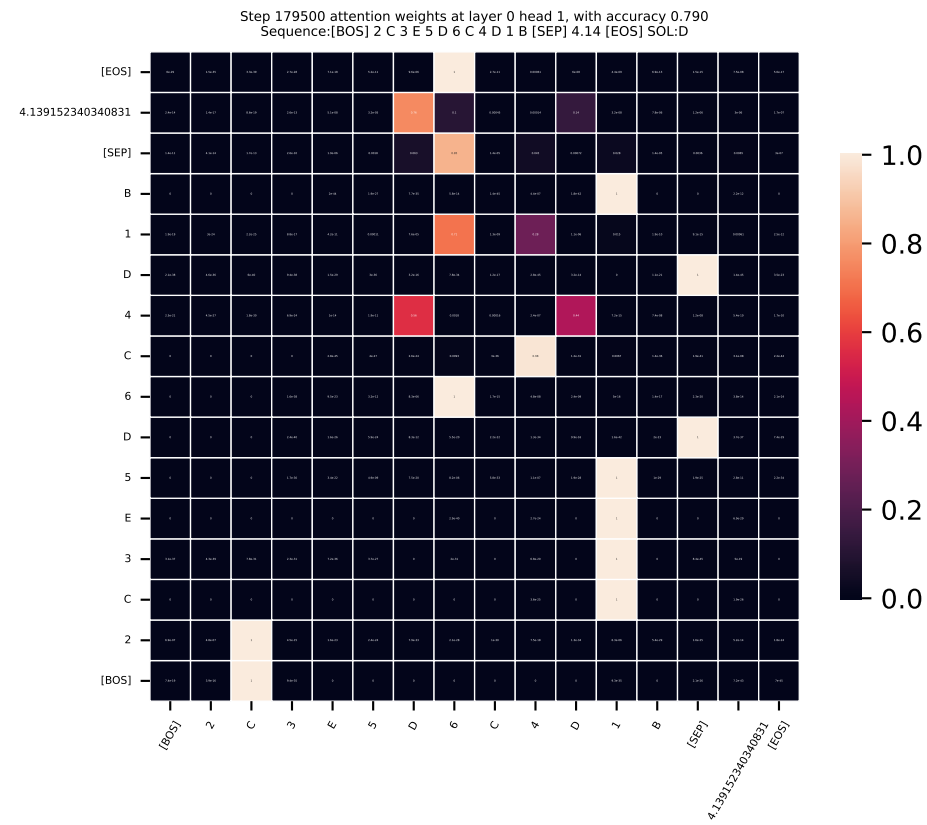

Step 170000 attention weights at layer 0 head 0, with accuracy 0.783
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
1.0
0.8
0.6
0.4
0.2
0.0

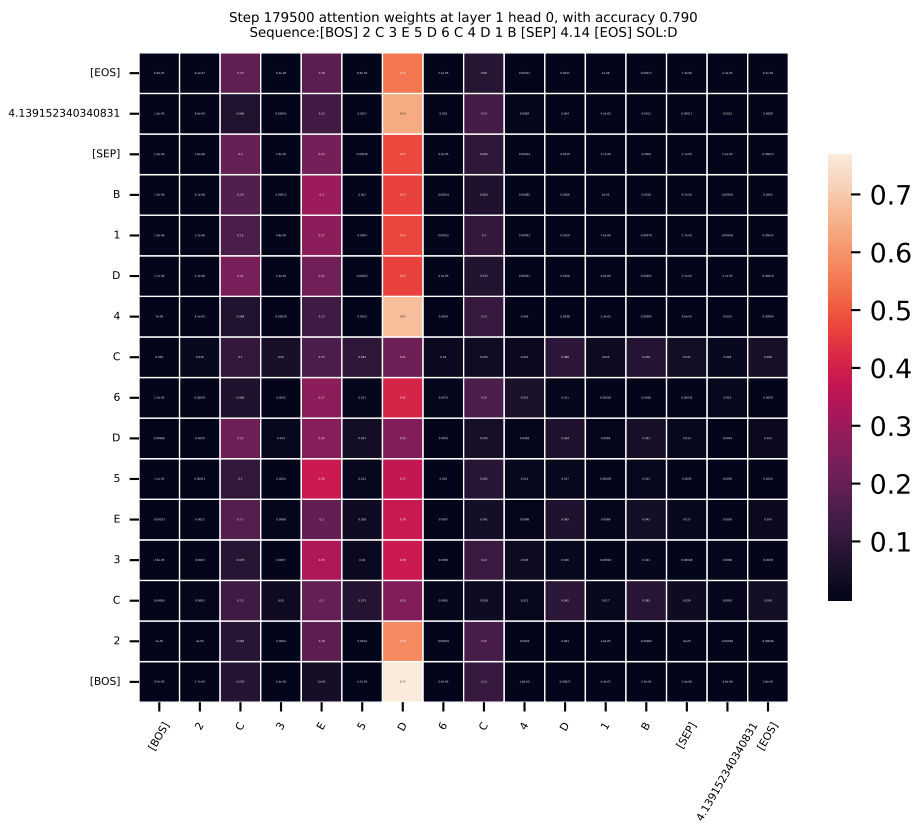

Step 170000 attention weights at layer 0 head 1, with accuracy 0.783
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
1.0
0.8
0.6
0.4
0.2
0.0

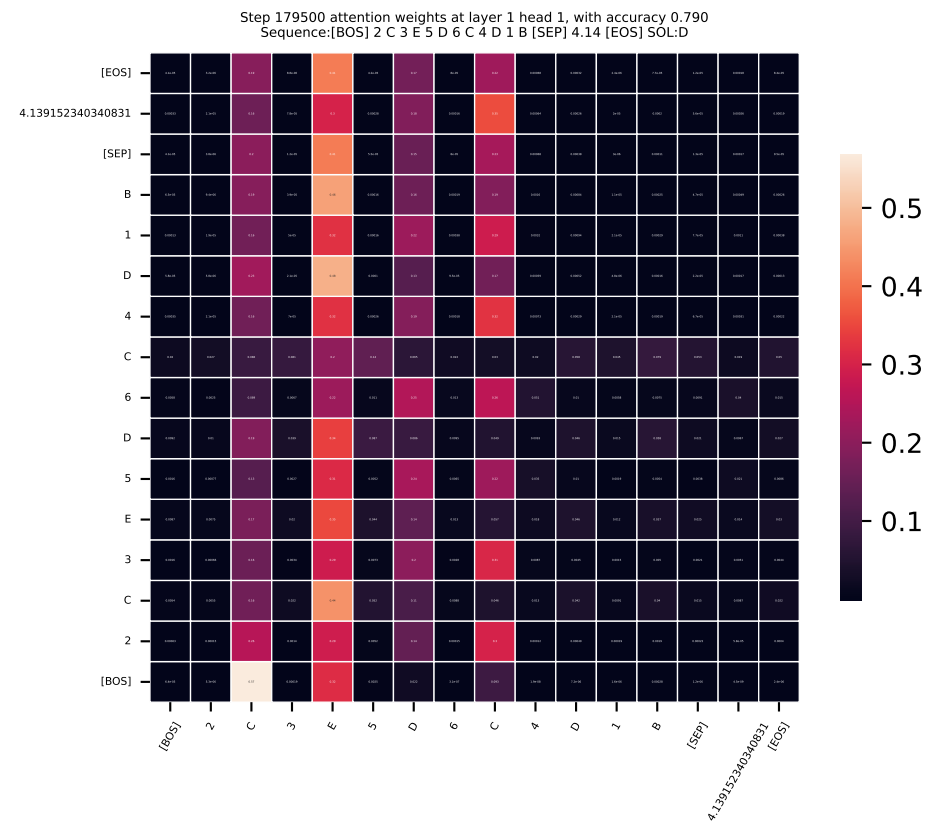

Step 170000 attention weights at layer 1 head 0, with accuracy 0.783
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2

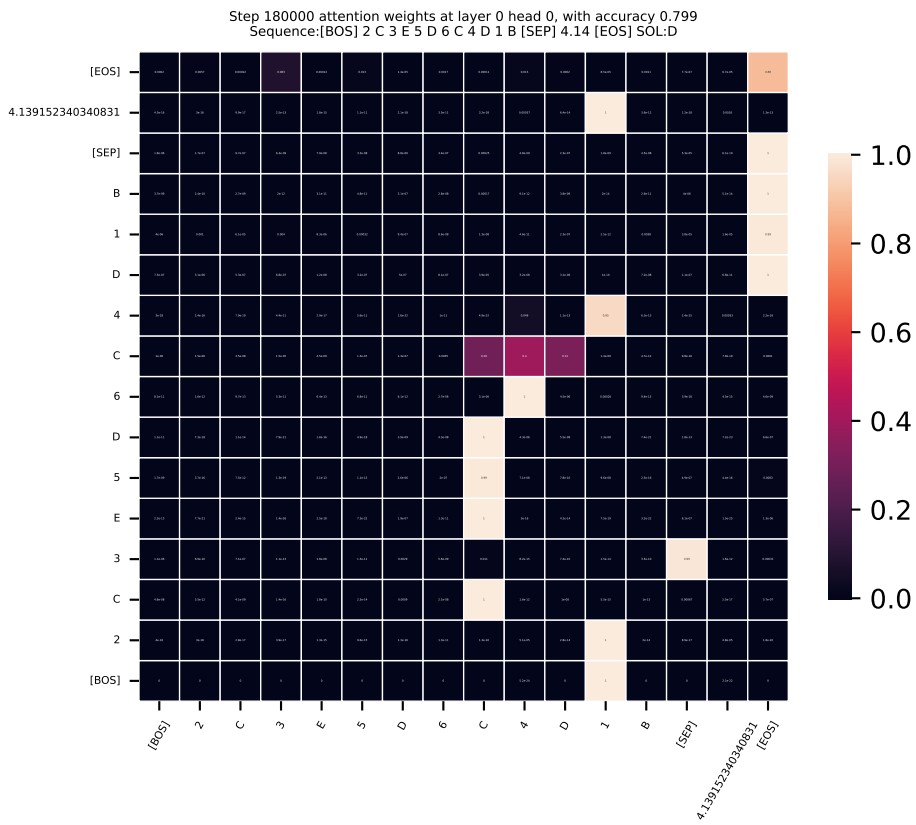

Step 170000 attention weights at layer 1 head 1, with accuracy 0.783
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.5
0.4
0.3
0.2
0.1

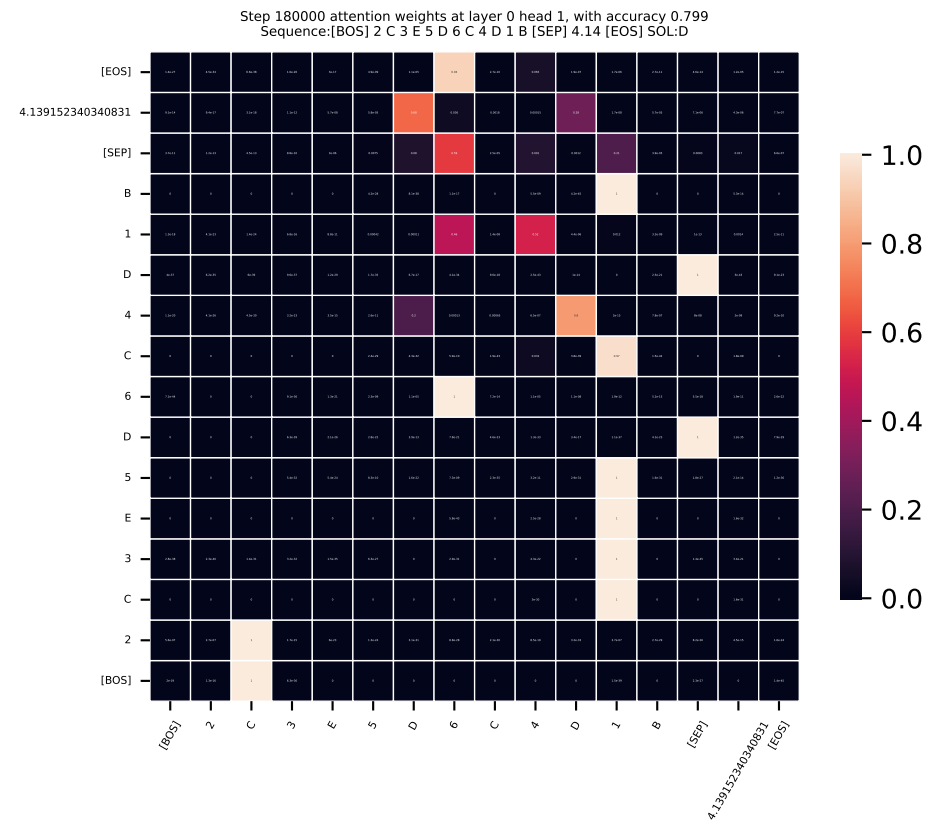

Step 179500 attention weights at layer 0 head 0, with accuracy 0.790
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
1.0
0.8
0.6
0.4
0.2
0.0

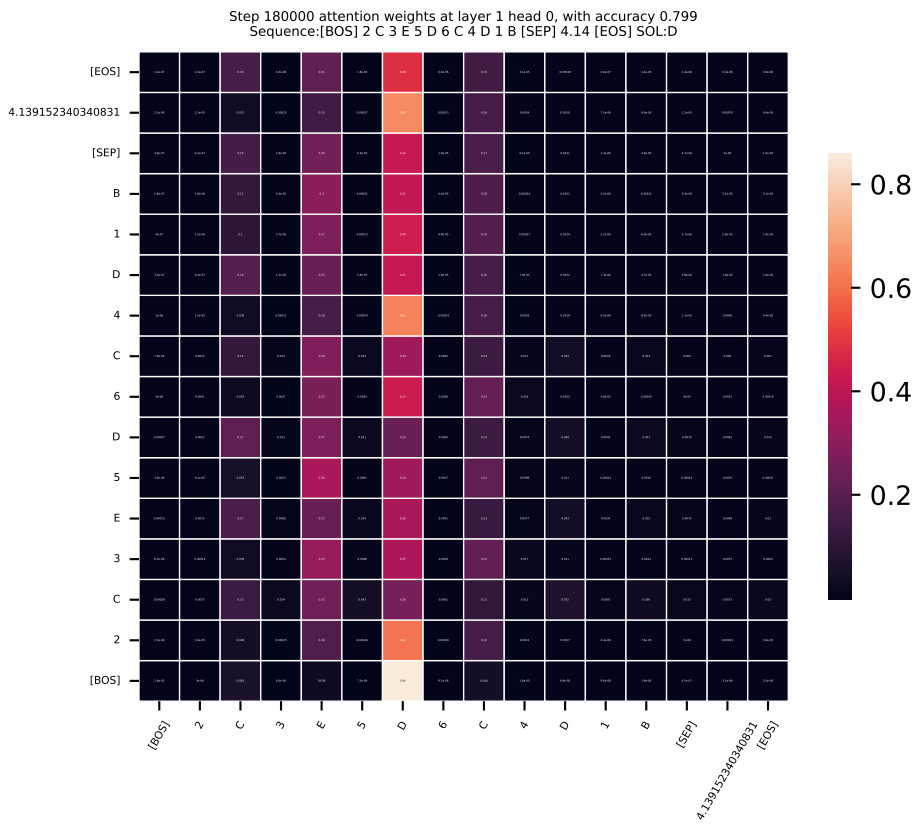

Step 179500 attention weights at layer 0 head 1, with accuracy 0.790
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
1.0
0.8
0.6
0.4
0.2
0.0

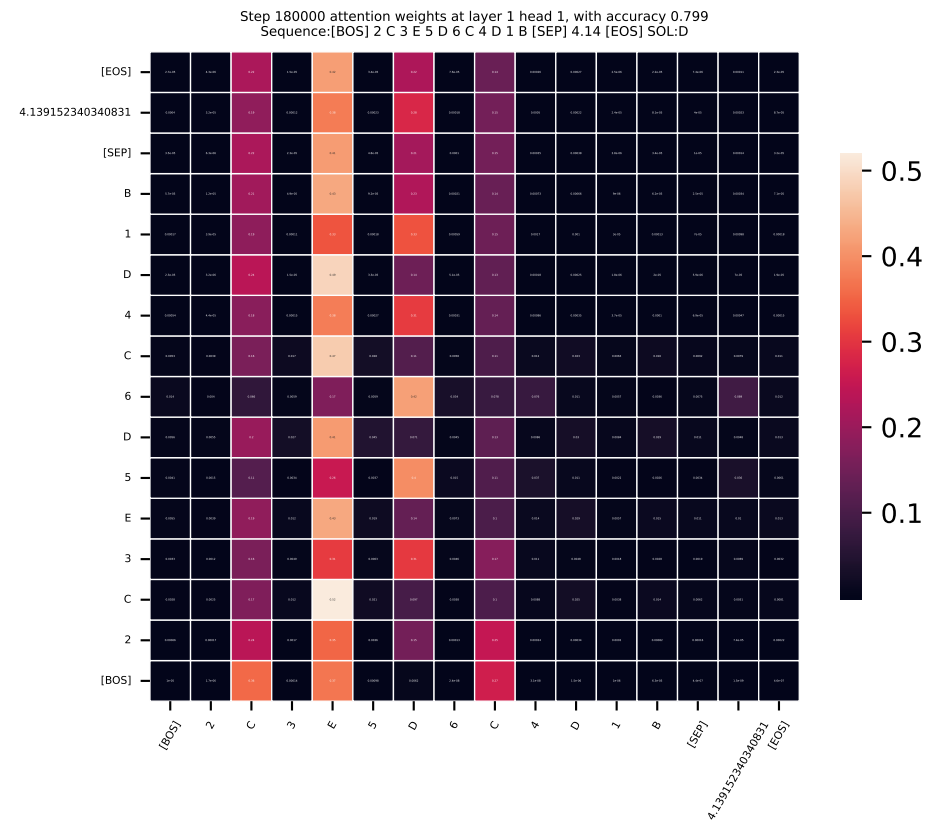

Step 179500 attention weights at layer 1 head 0, with accuracy 0.790
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.7
0.6
0.5
0.4
0.3
0.2
0.1

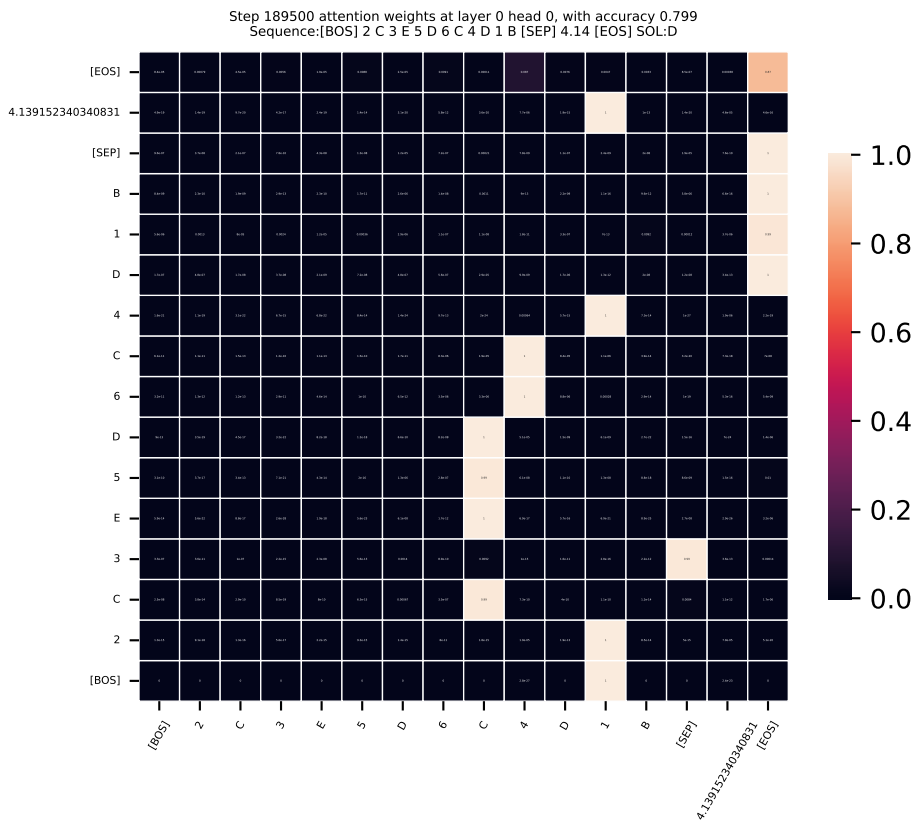

Step 179500 attention weights at layer 1 head 1, with accuracy 0.790
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.5
0.4
0.3
0.2
0.1

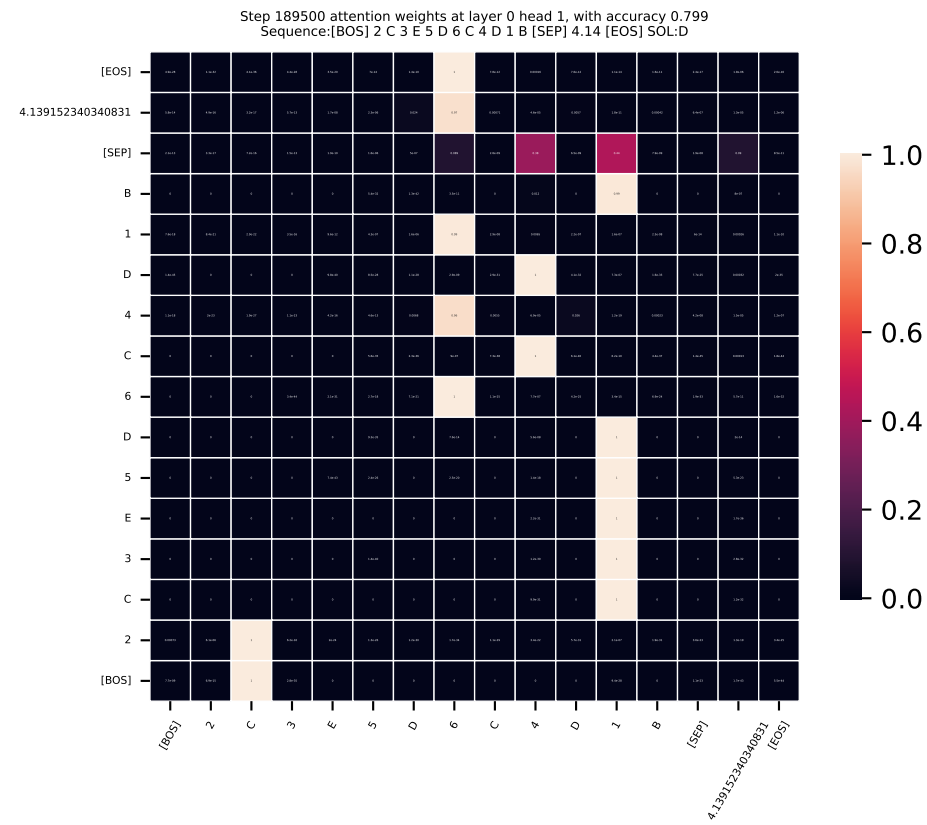

Step 180000 attention weights at layer 0 head 0, with accuracy 0.799
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

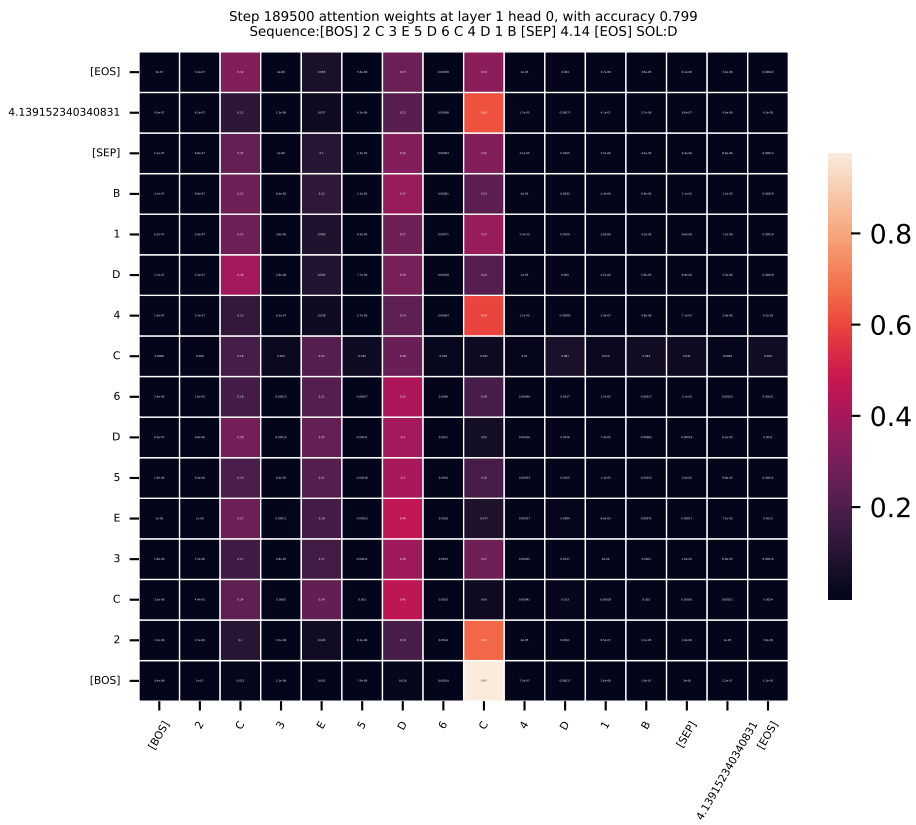

Step 180000 attention weights at layer 0 head 1, with accuracy 0.799
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

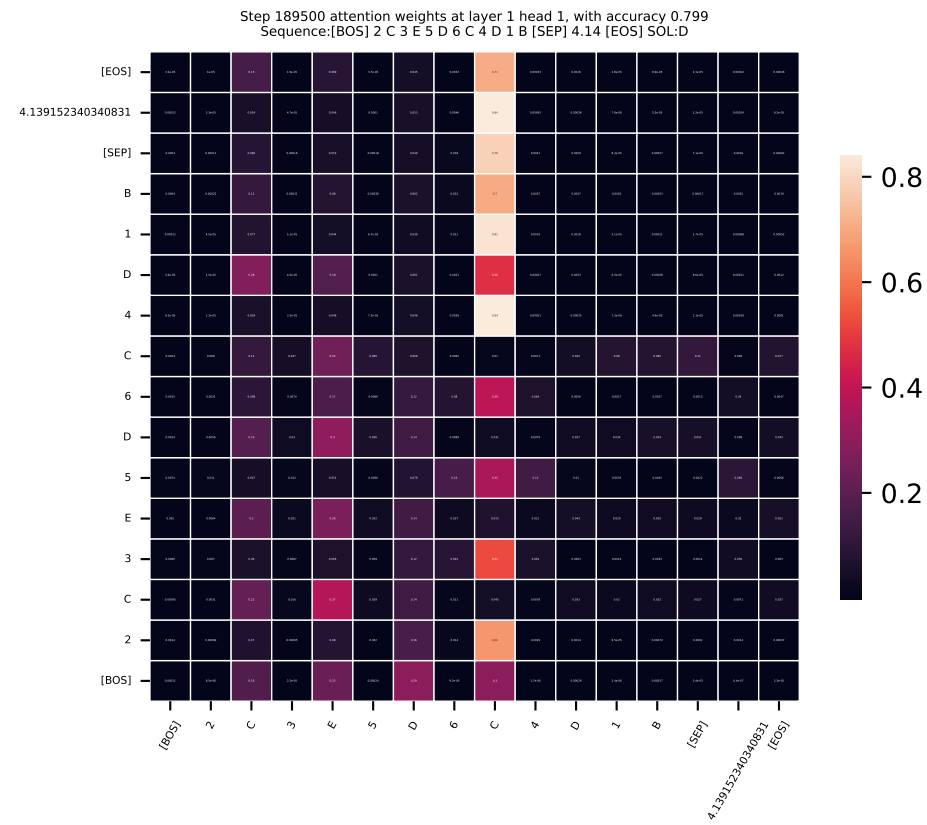

Step 180000 attention weights at layer 1 head 0, with accuracy 0.799
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2

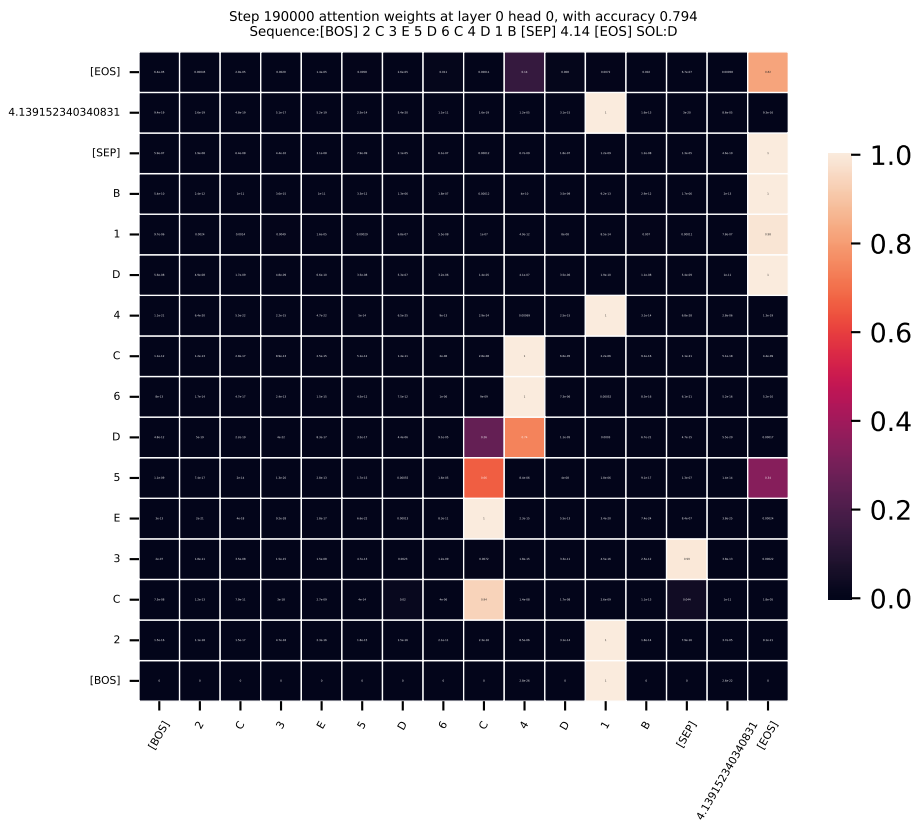

Step 180000 attention weights at layer 1 head 1, with accuracy 0.799
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.5
0.4
0.3
0.2
0.1

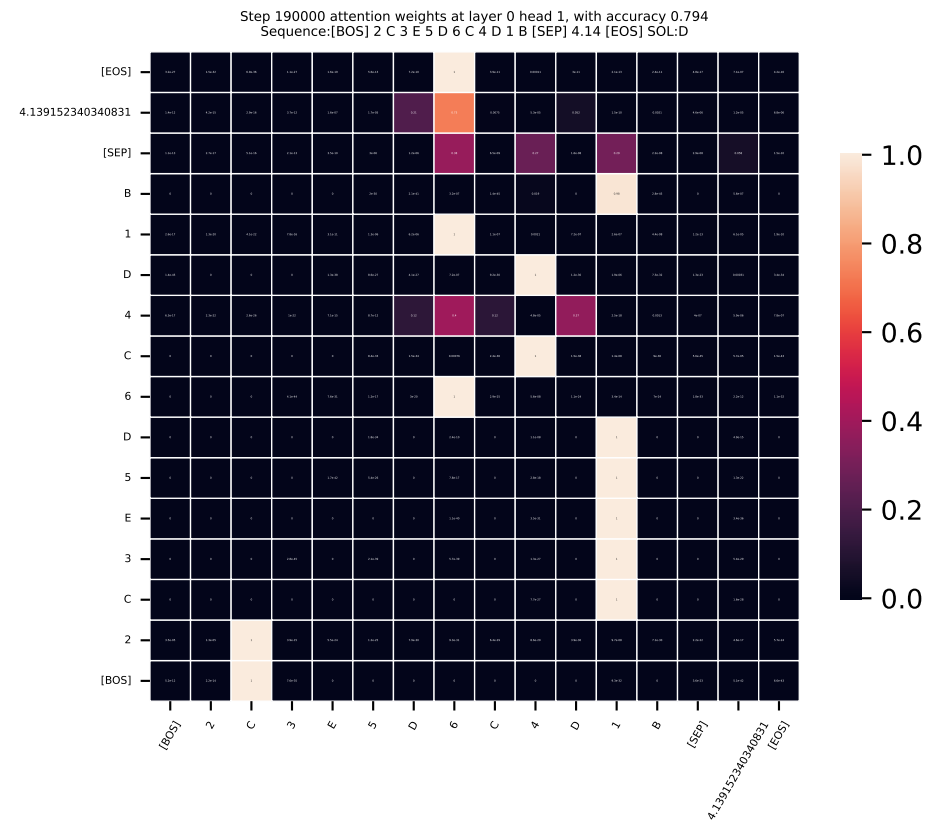

Step 189500 attention weights at layer 0 head 0, with accuracy 0.799
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
1.0
0.8
0.6
0.4
0.2
0.0

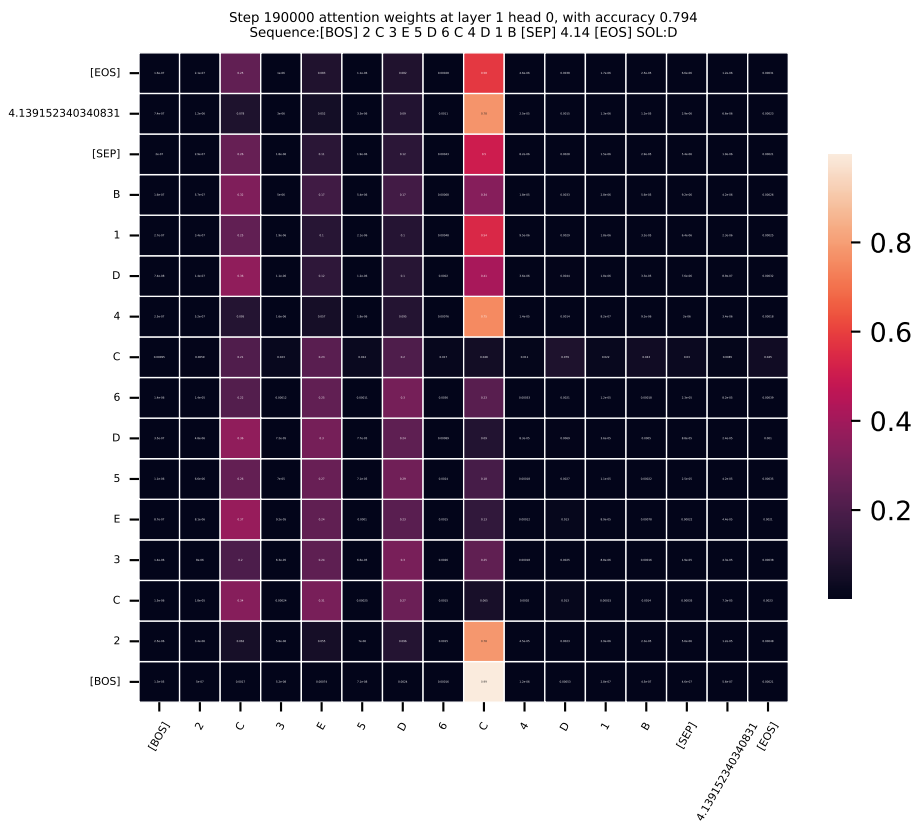

Step 189500 attention weights at layer 0 head 1, with accuracy 0.799
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
1.0
0.8
0.6
0.4
0.2
0.0

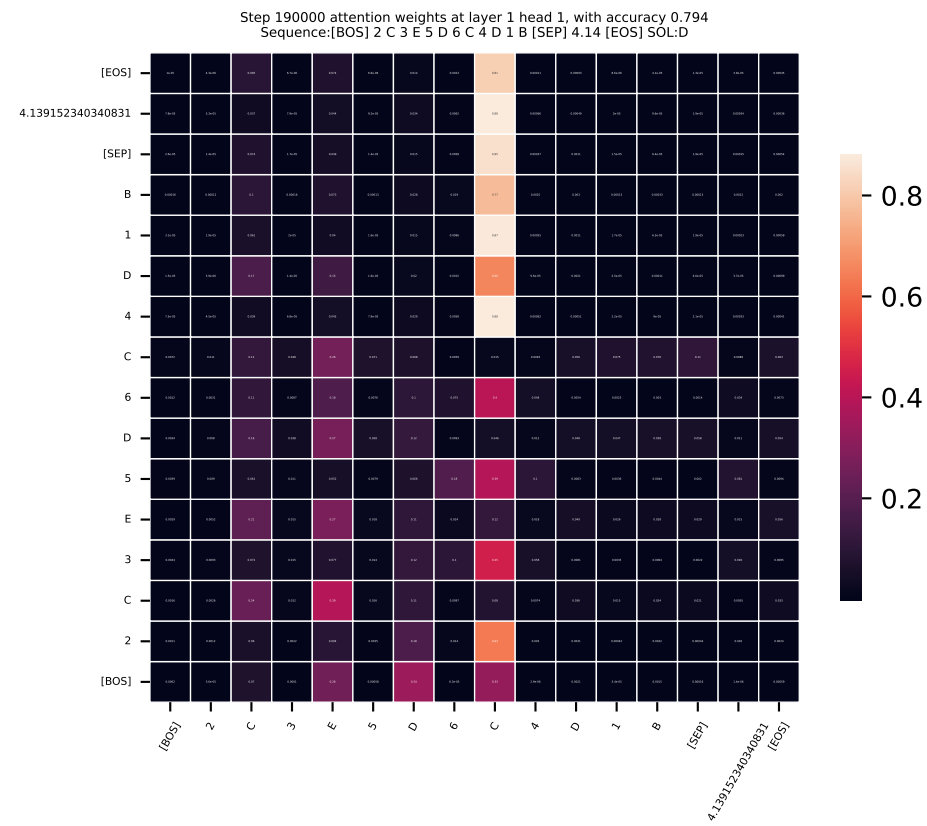

Step 189500 attention weights at layer 1 head 0, with accuracy 0.799
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2

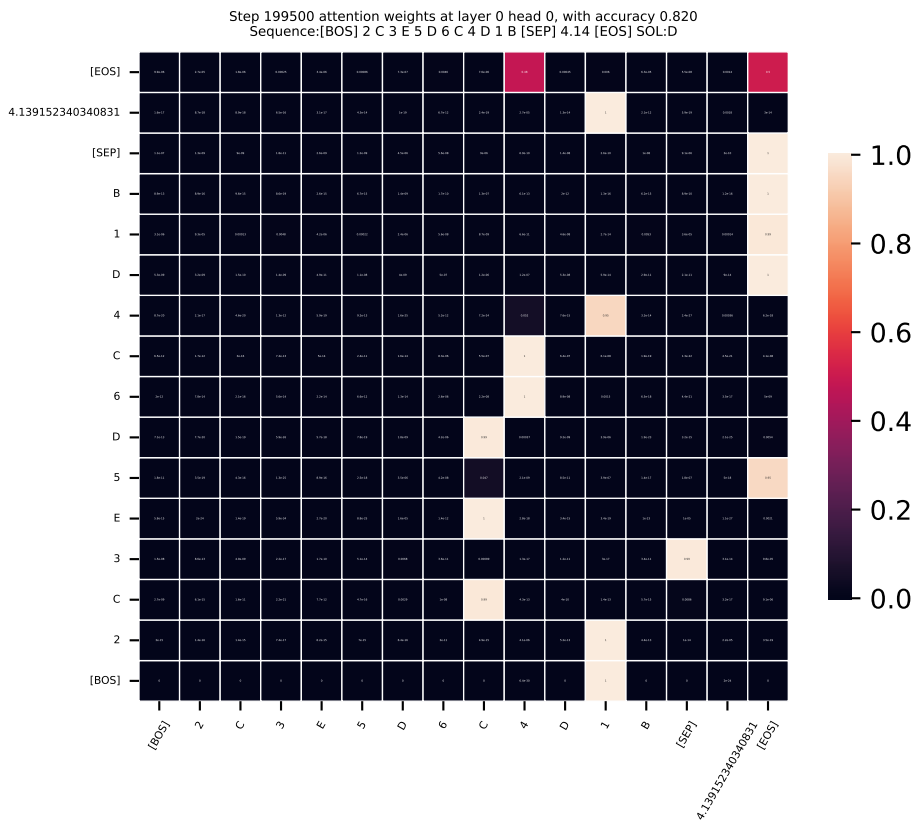

Step 189500 attention weights at layer 1 head 1, with accuracy 0.799
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2

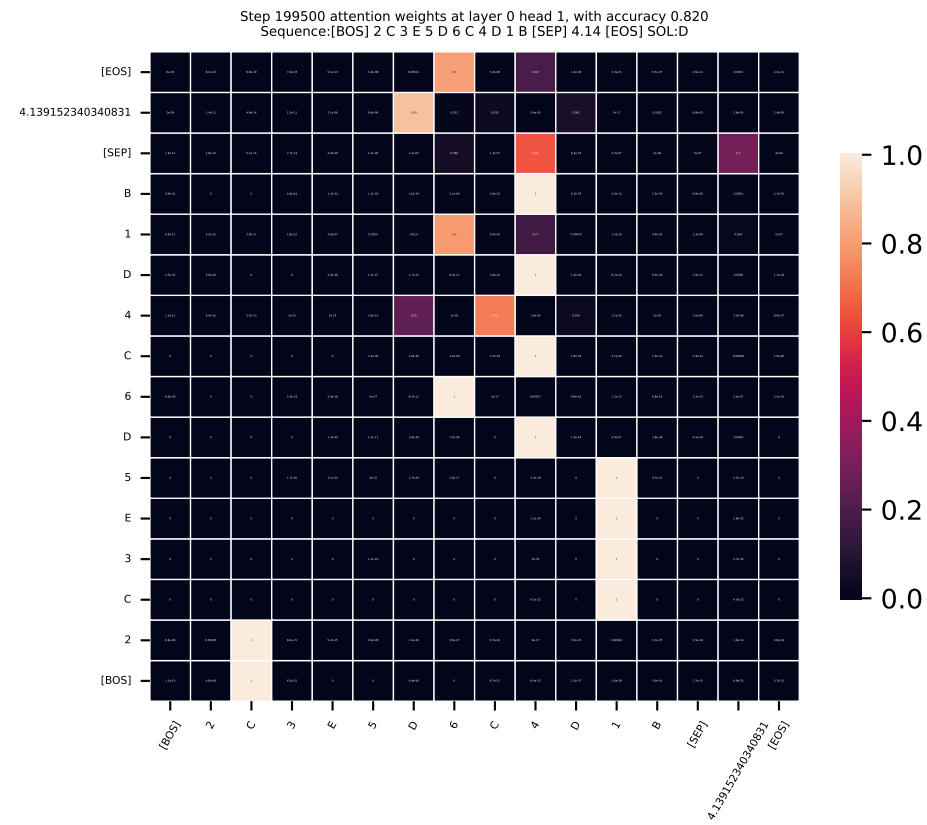

Step 190000 attention weights at layer 0 head 0, with accuracy 0.794
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
1.0
0.8
0.6
0.4
0.2
0.0

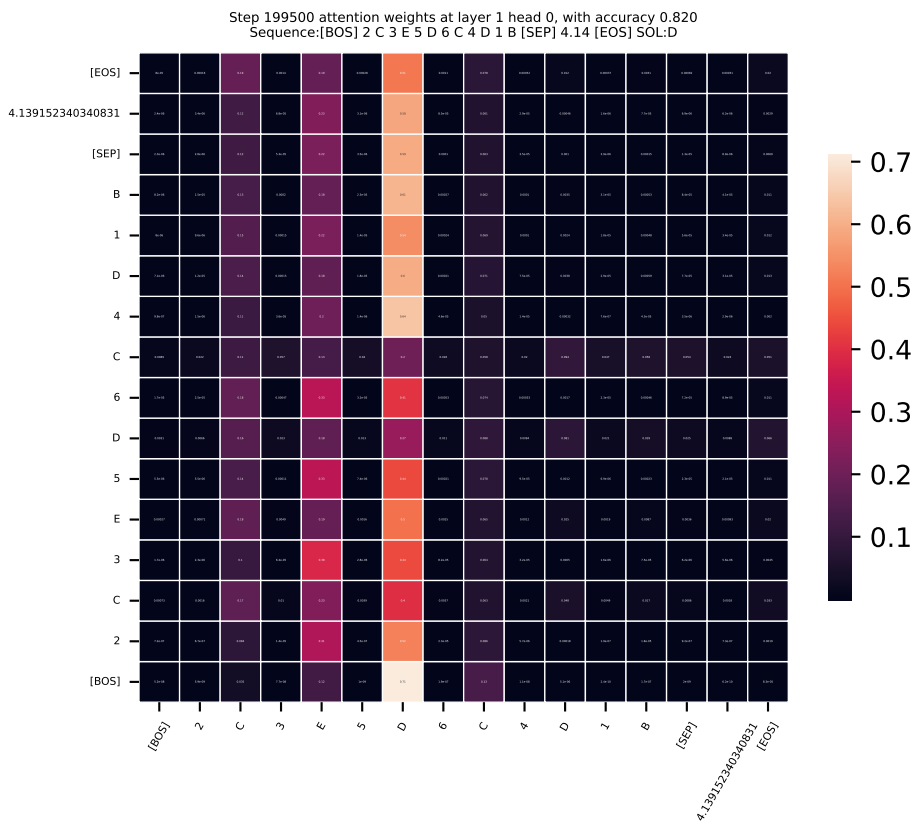

Step 190000 attention weights at layer 0 head 1, with accuracy 0.794
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
1.0
0.8
0.6
0.4
0.2
0.0

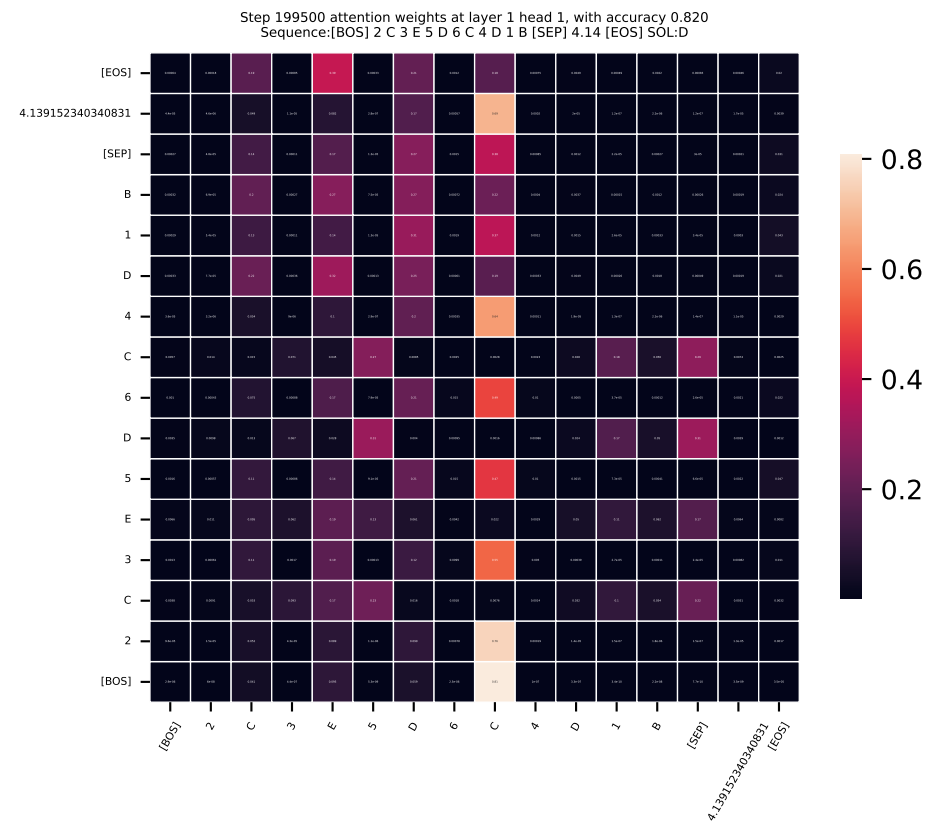

Step 190000 attention weights at layer 1 head 0, with accuracy 0.794
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2

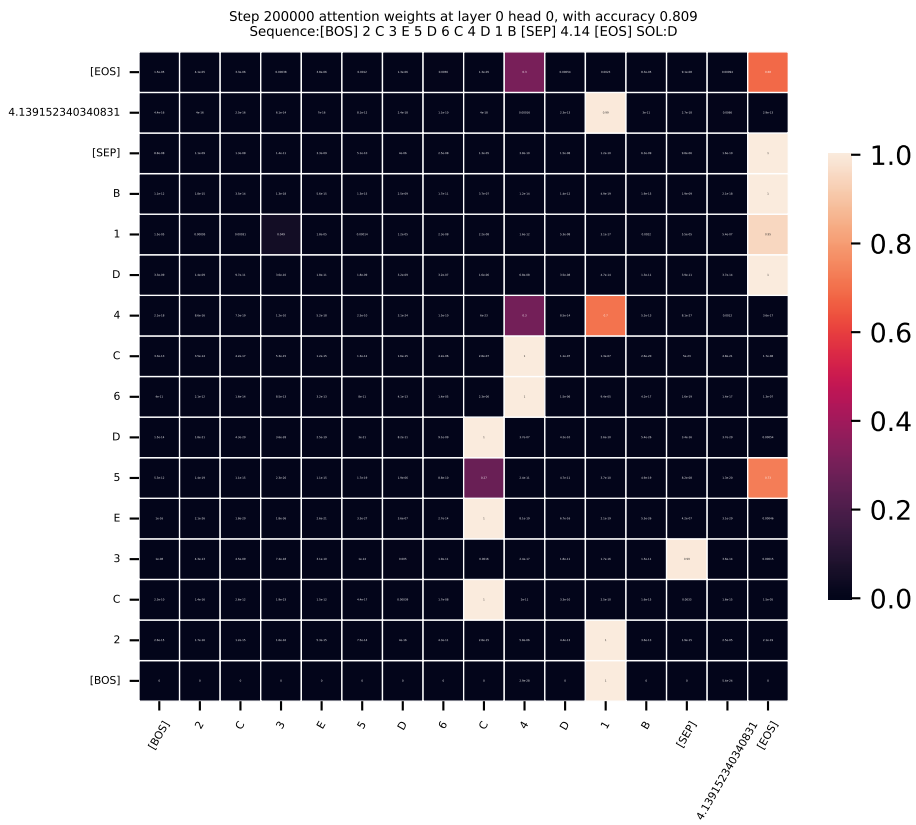

Step 190000 attention weights at layer 1 head 1, with accuracy 0.794
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2

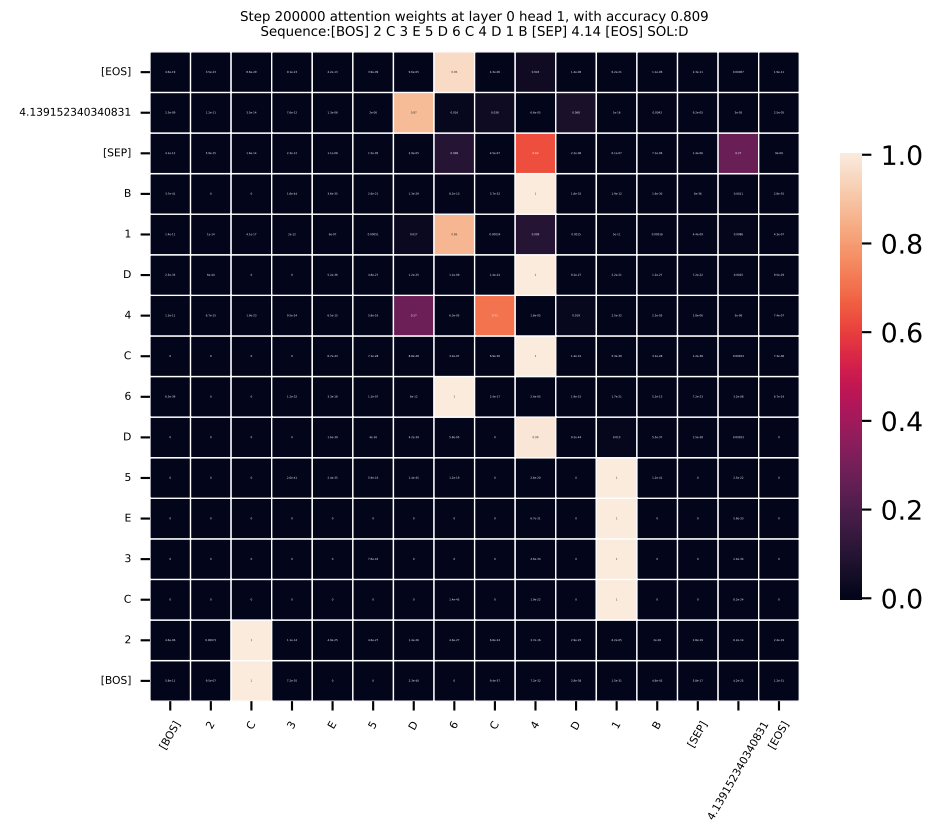

Step 199500 attention weights at layer 0 head 0, with accuracy 0.820
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

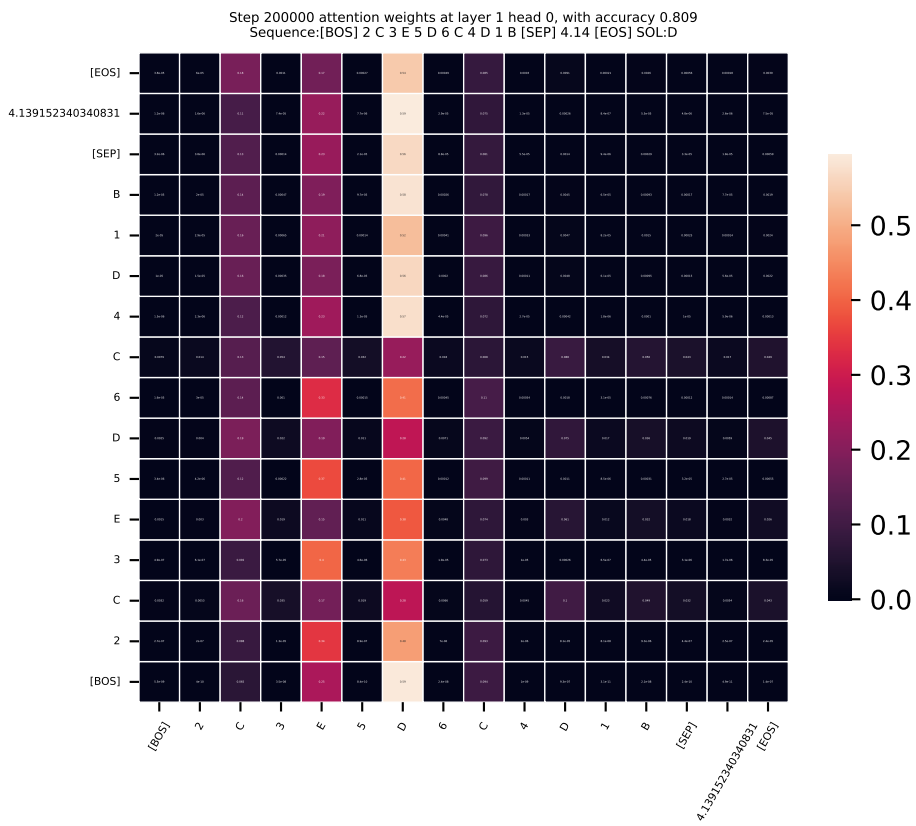

Step 199500 attention weights at layer 0 head 1, with accuracy 0.820
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

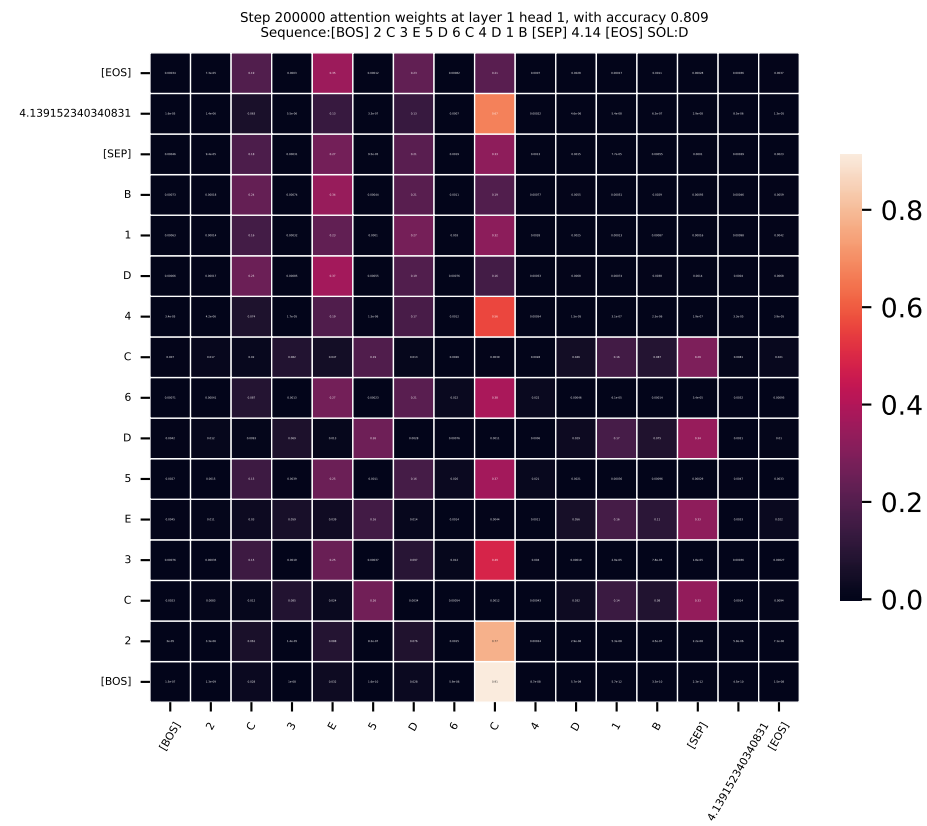

Step 199500 attention weights at layer 1 head 0, with accuracy 0.820
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.7
0.6
0.5
0.4
0.3
0.2
0.1

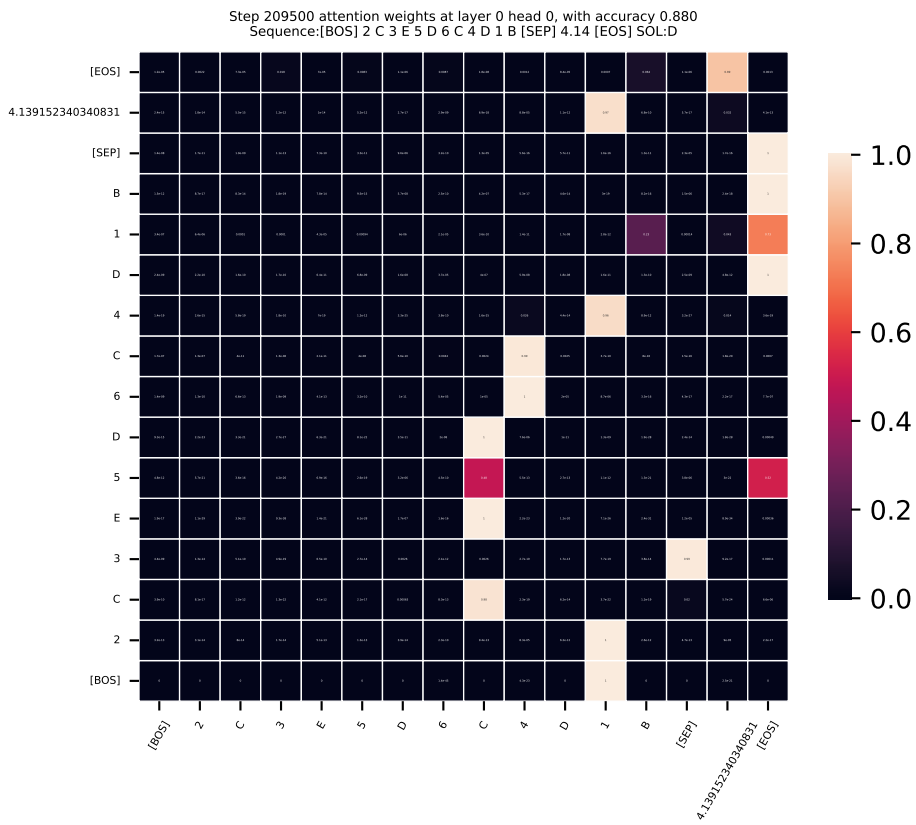

Step 199500 attention weights at layer 1 head 1, with accuracy 0.820
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2

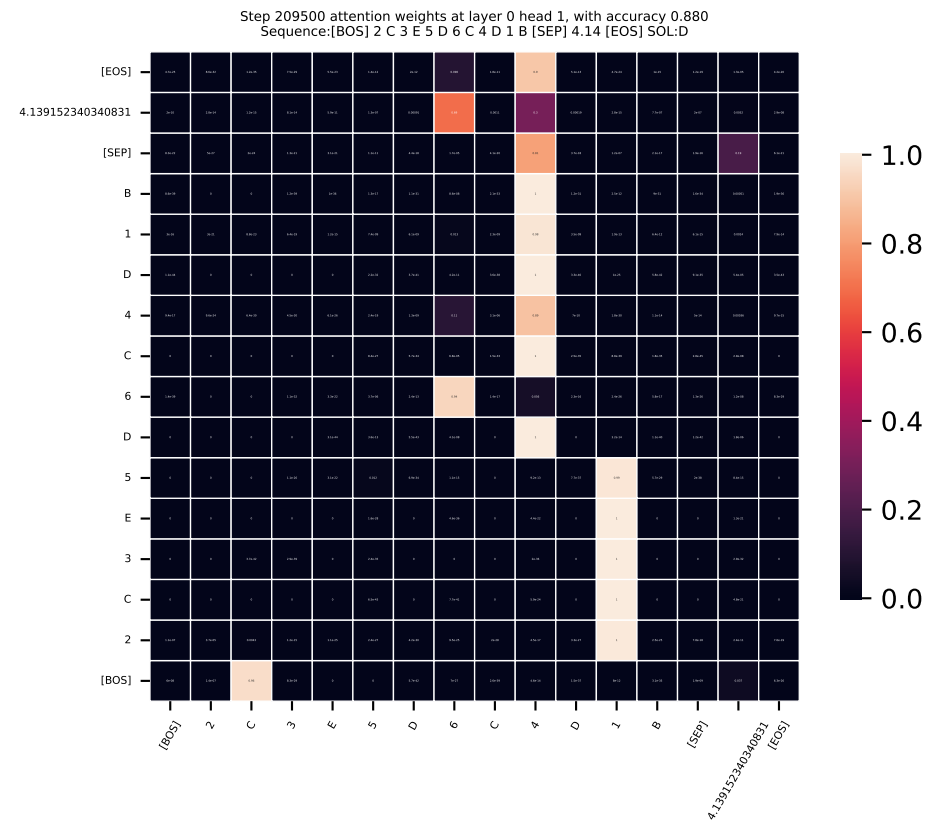

Step 200000 attention weights at layer 0 head 0, with accuracy 0.809
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

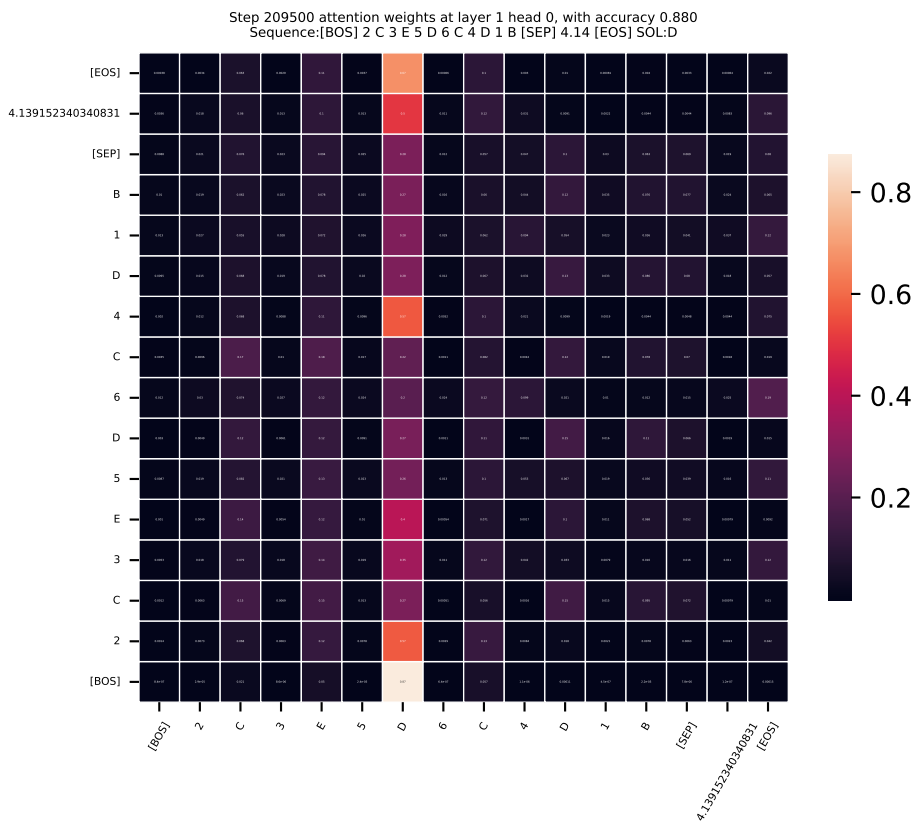

Step 200000 attention weights at layer 0 head 1, with accuracy 0.809
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

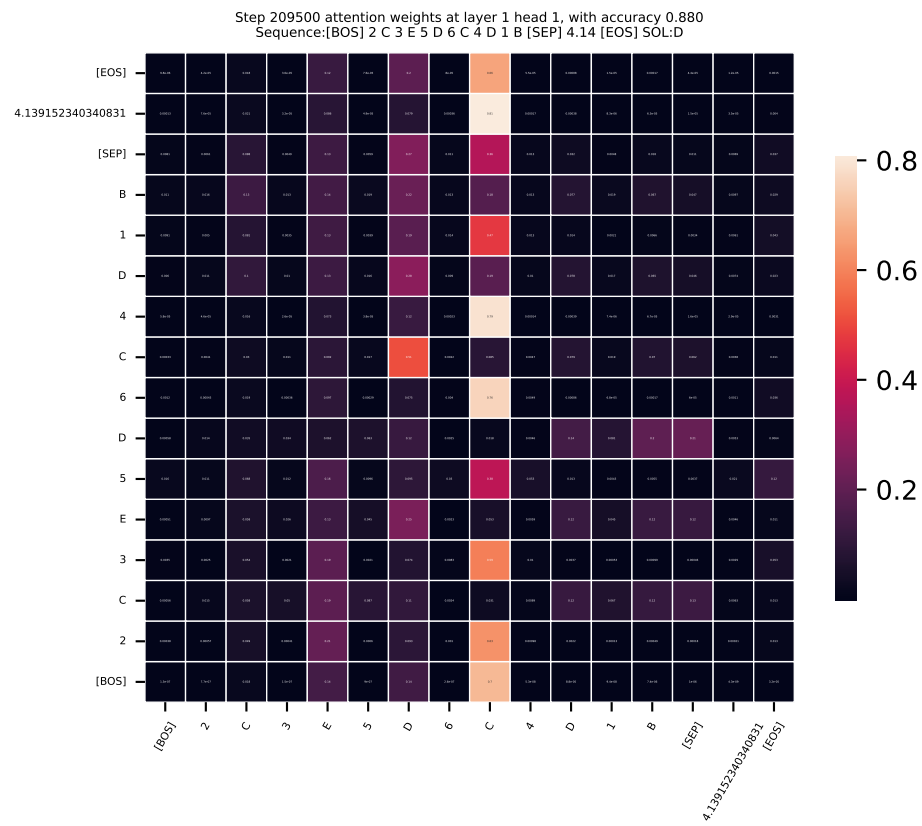

Step 200000 attention weights at layer 1 head 0, with accuracy 0.809
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.5
0.4
0.3
0.2
0.1
0.0

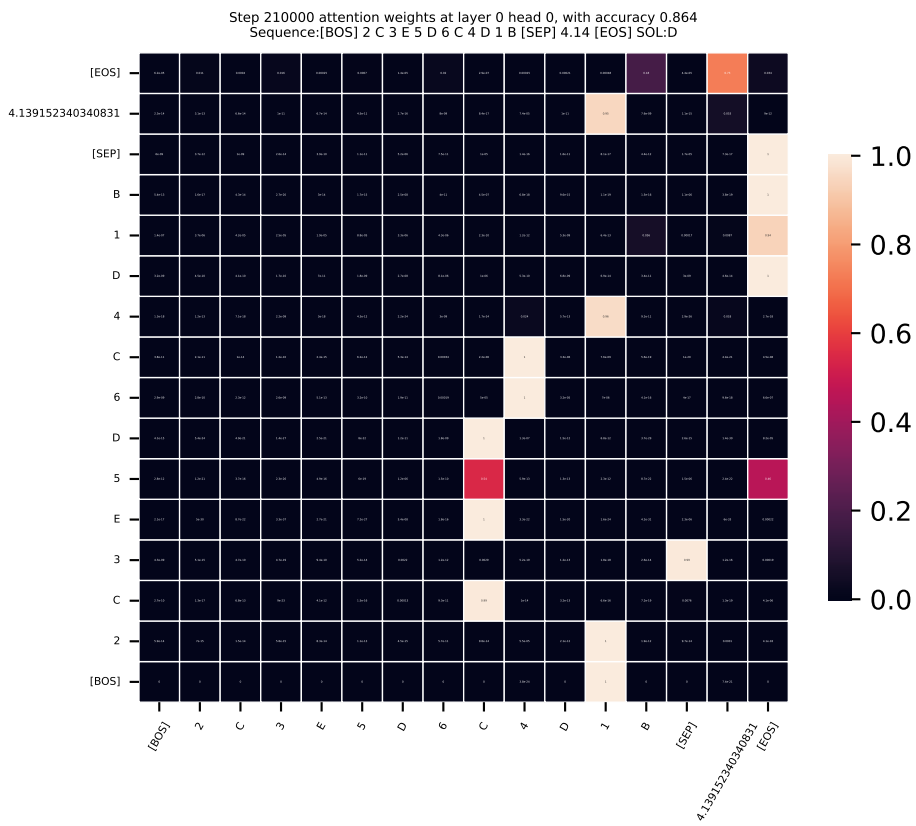

Step 200000 attention weights at layer 1 head 1, with accuracy 0.809
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2
0.0

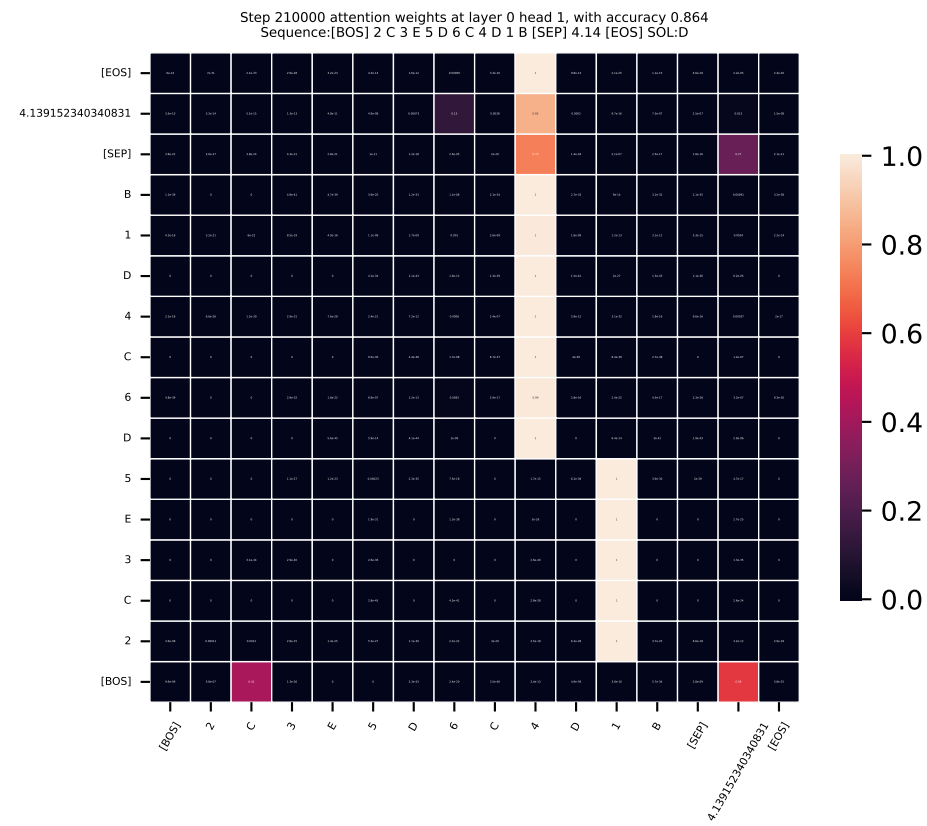

Step 209500 attention weights at layer 0 head 0, with accuracy 0.880
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
1.0
0.8
0.6
0.4
0.2
0.0

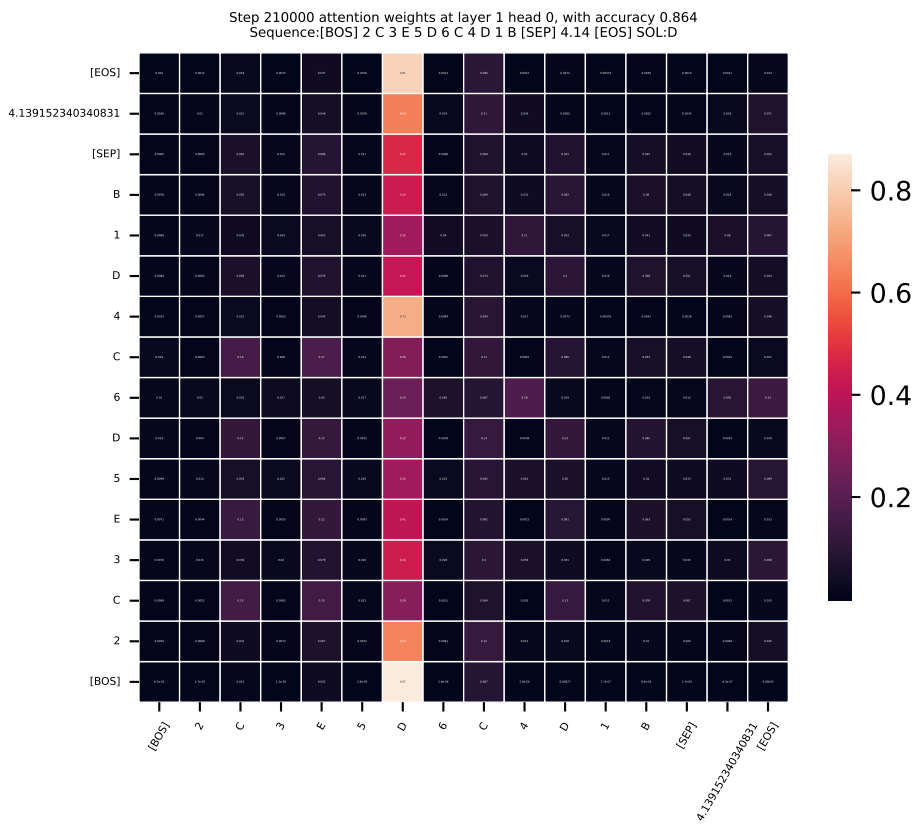

Step 209500 attention weights at layer 0 head 1, with accuracy 0.880
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
1.0
0.8
0.6
0.4
0.2
0.0

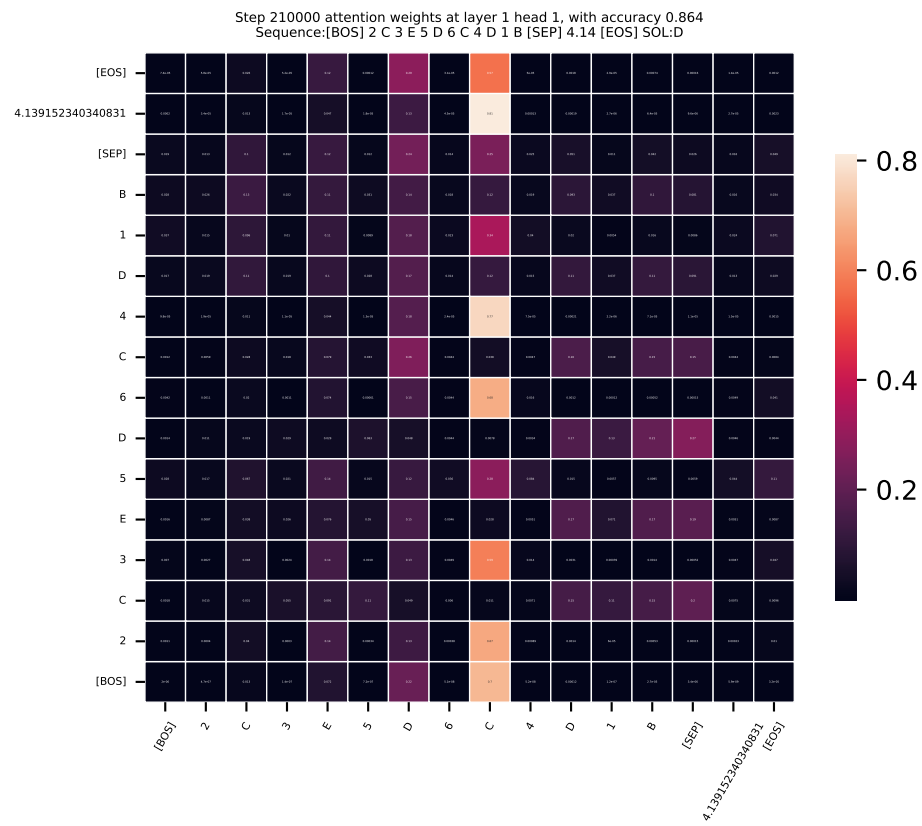

Step 209500 attention weights at layer 1 head 0, with accuracy 0.880
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2

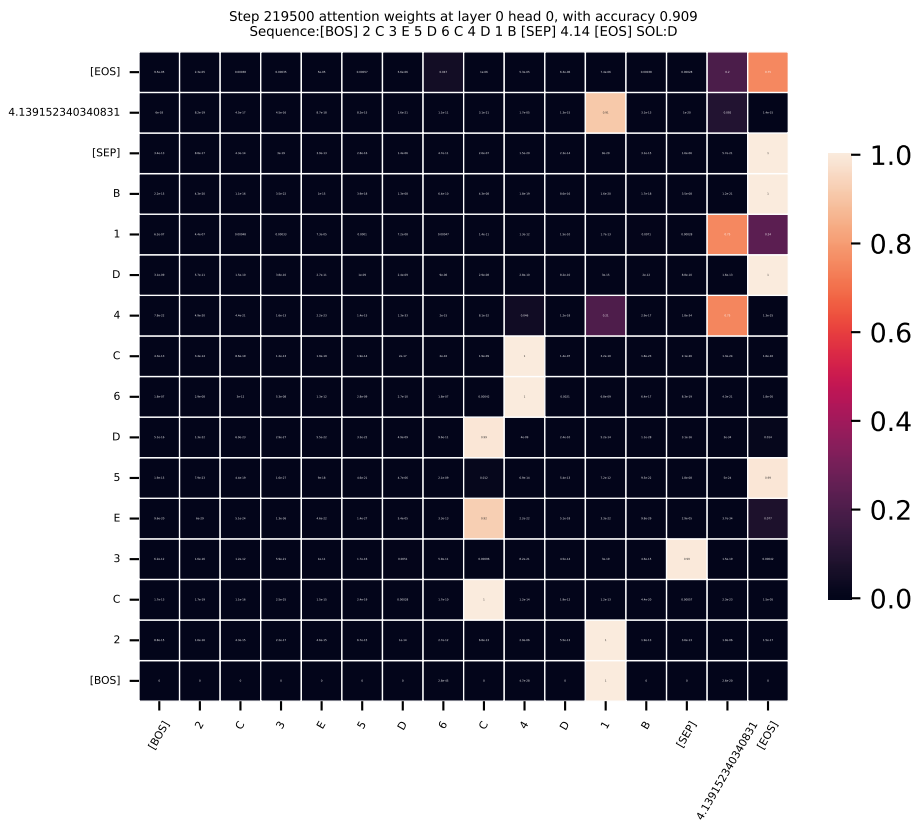

Step 209500 attention weights at layer 1 head 1, with accuracy 0.880
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2

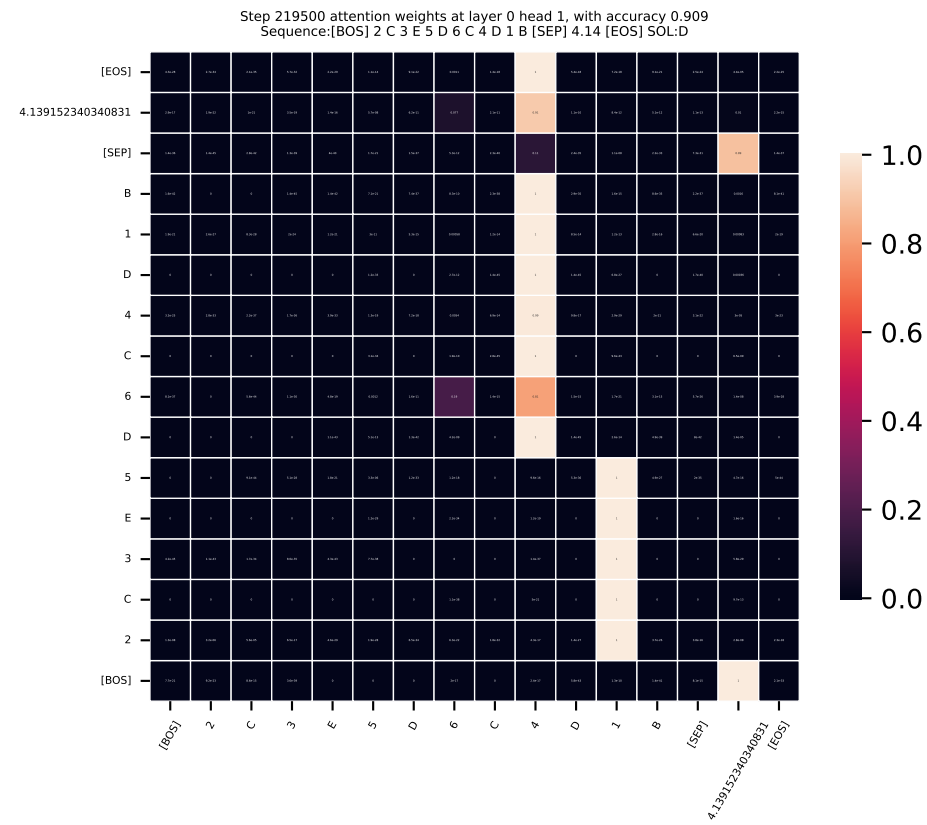

Step 210000 attention weights at layer 0 head 0, with accuracy 0.864
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

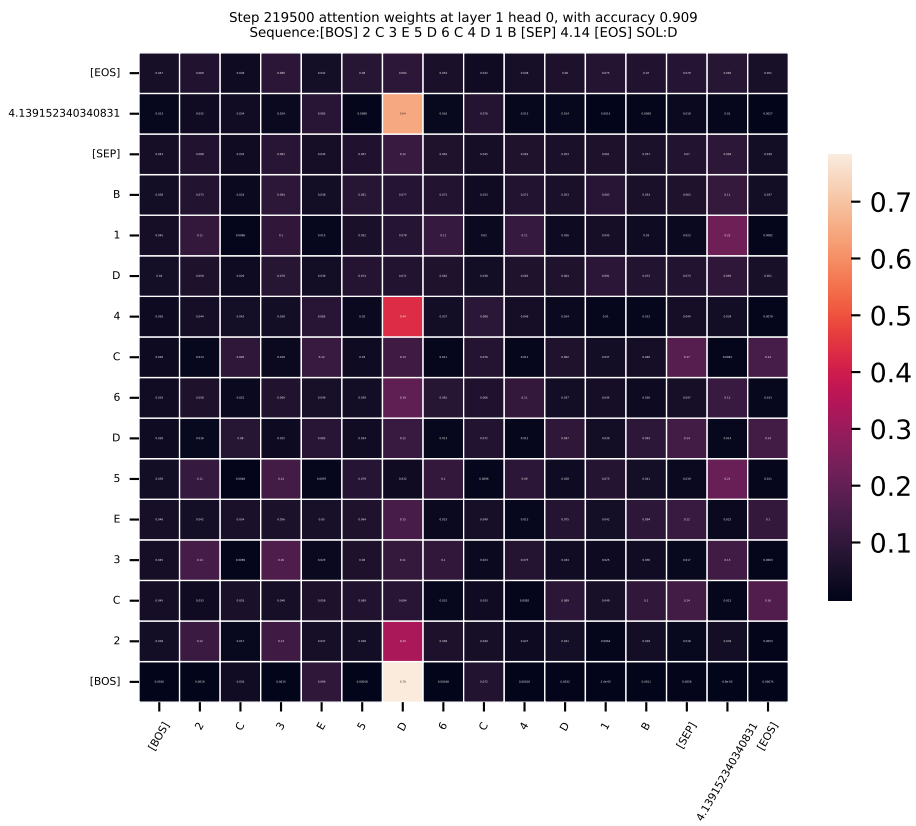

Step 210000 attention weights at layer 0 head 1, with accuracy 0.864
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

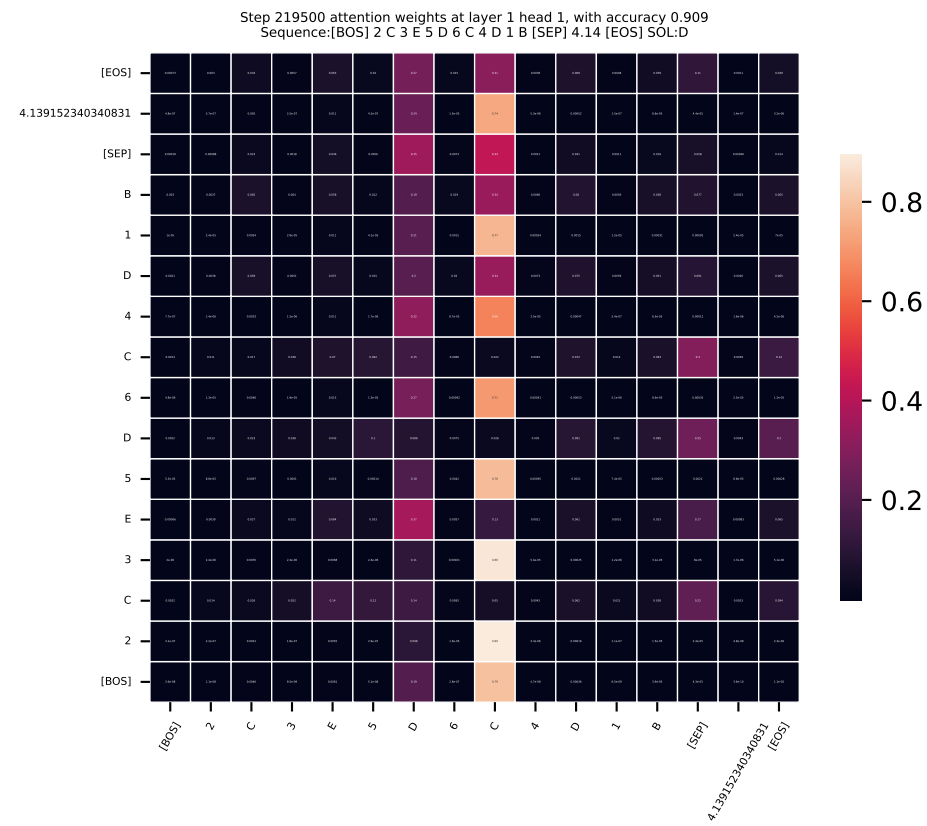

Step 210000 attention weights at layer 1 head 0, with accuracy 0.864
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2

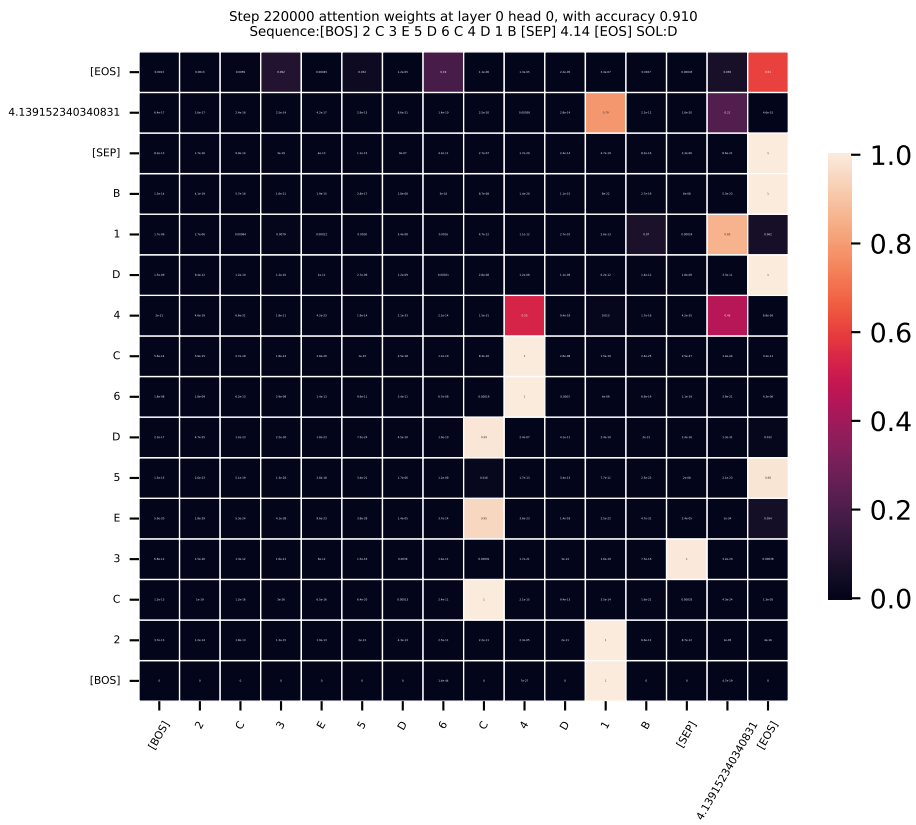

Step 210000 attention weights at layer 1 head 1, with accuracy 0.864
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2

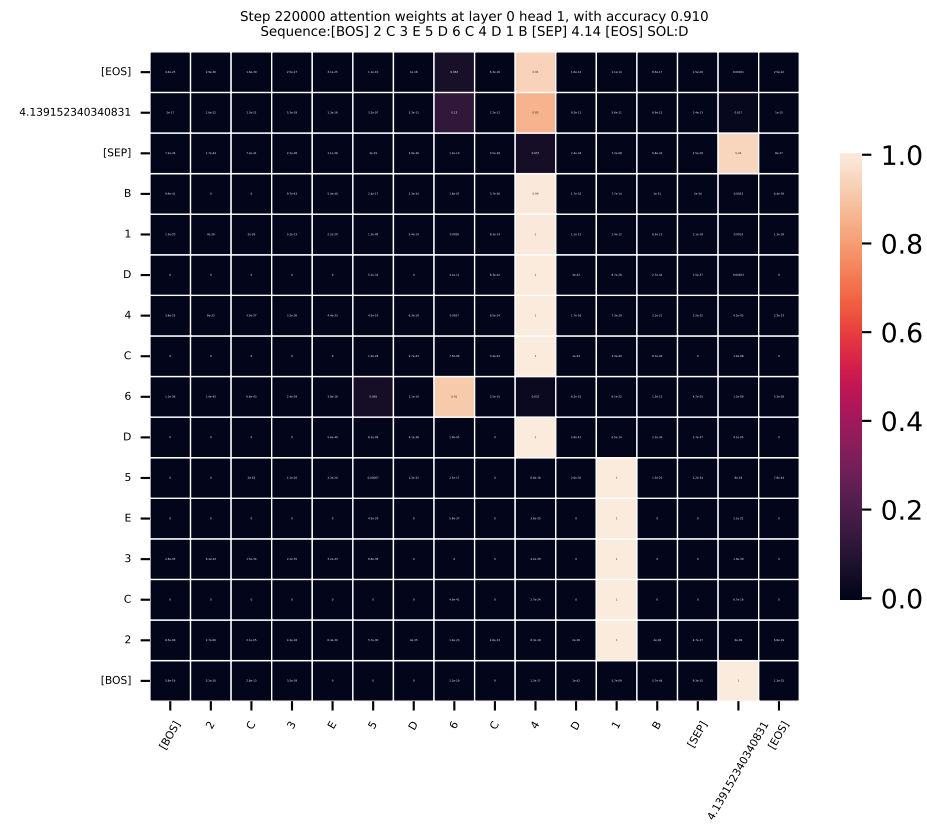

Step 219500 attention weights at layer 0 head 0, with accuracy 0.909
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
1.0
0.8
0.6
0.4
0.2
0.0

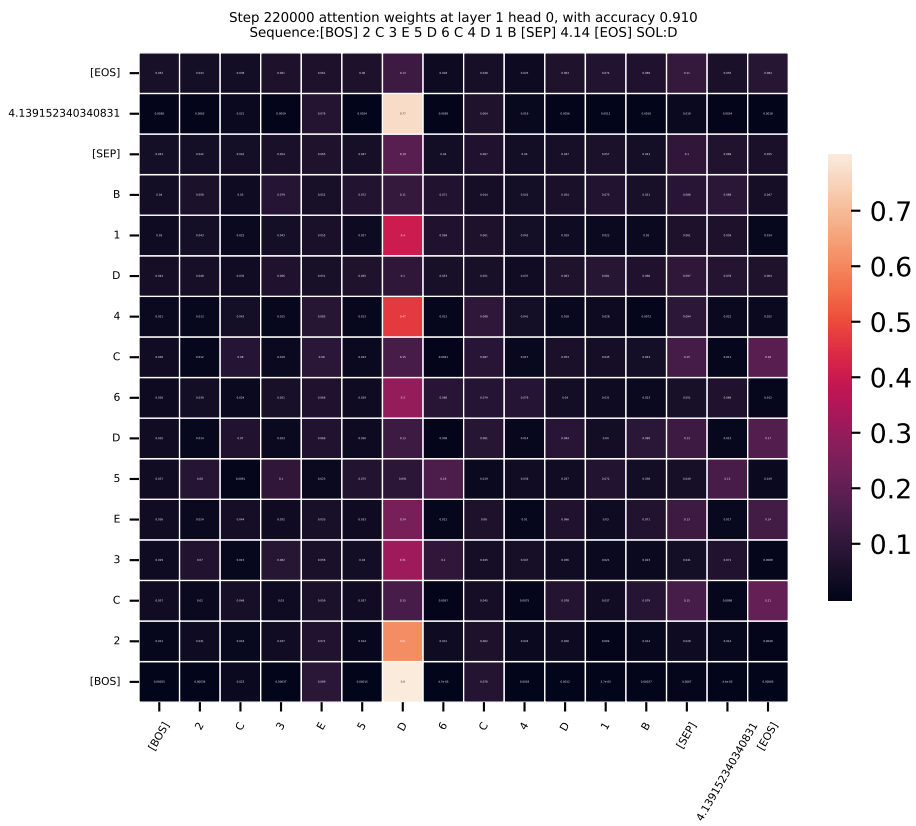

Step 219500 attention weights at layer 0 head 1, with accuracy 0.909
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
1.0
0.8
0.6
0.4
0.2
0.0

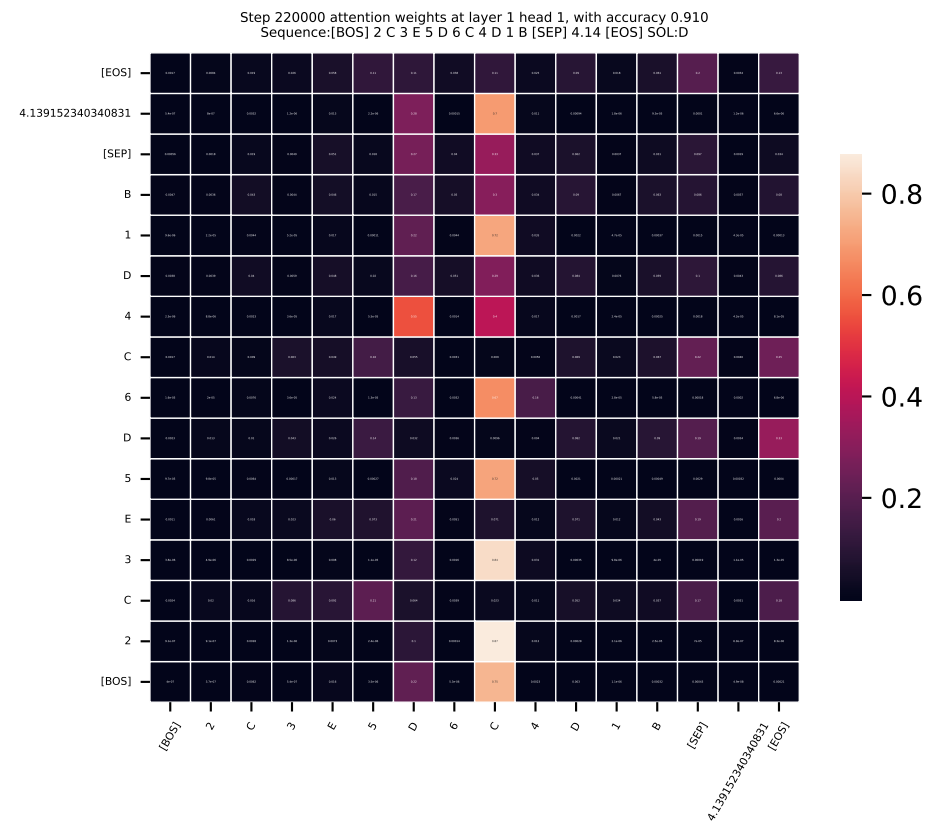

Step 219500 attention weights at layer 1 head 0, with accuracy 0.909
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.7
0.6
0.5
0.4
0.3
0.2
0.1

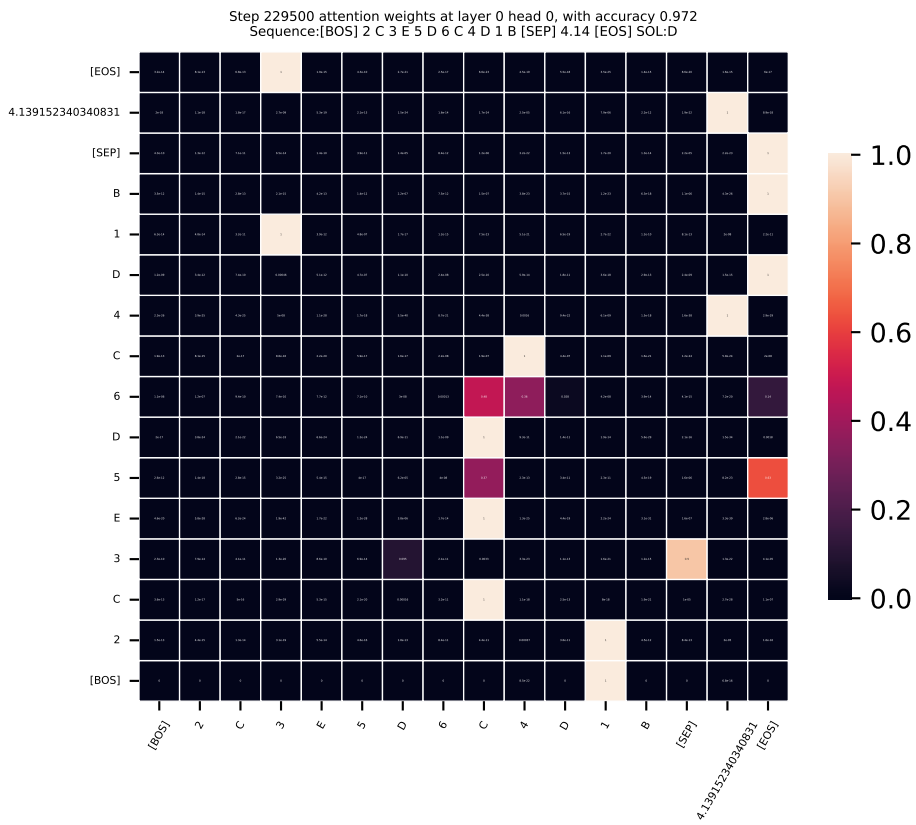

Step 219500 attention weights at layer 1 head 1, with accuracy 0.909
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2

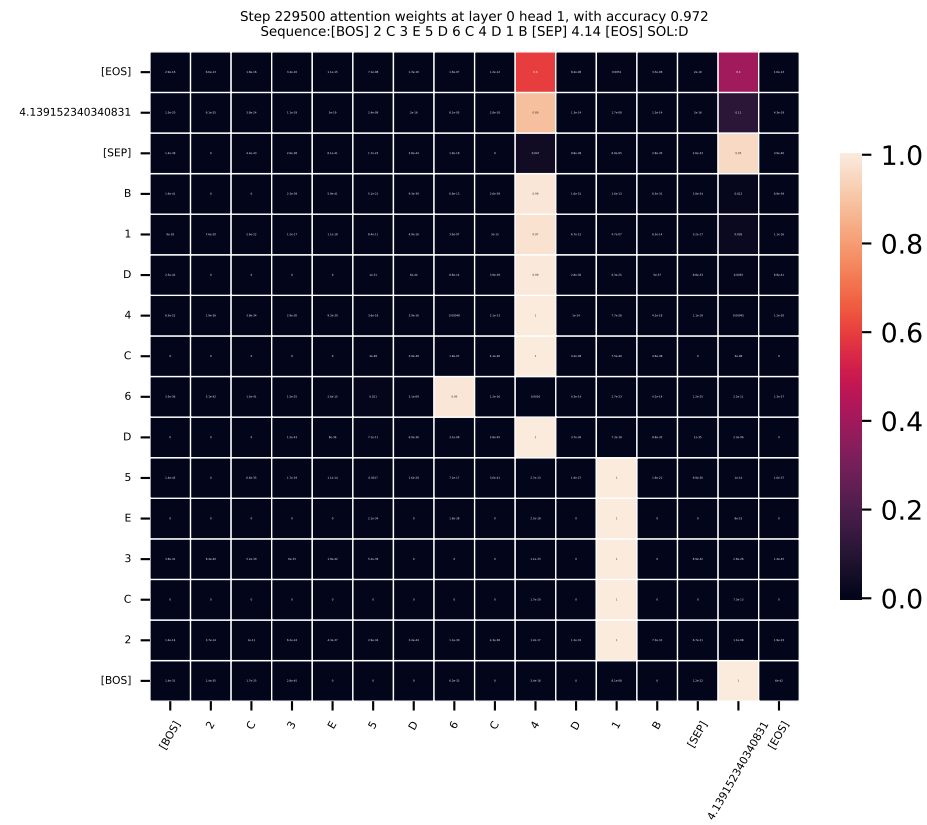

Step 220000 attention weights at layer 0 head 0, with accuracy 0.910
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

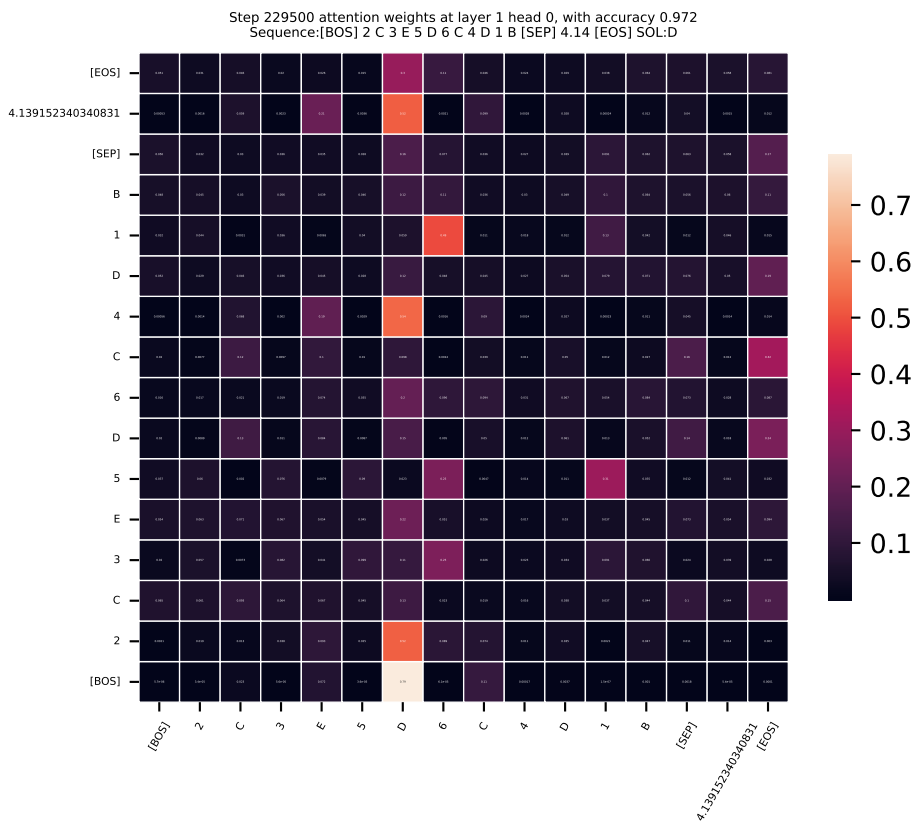

Step 220000 attention weights at layer 0 head 1, with accuracy 0.910
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

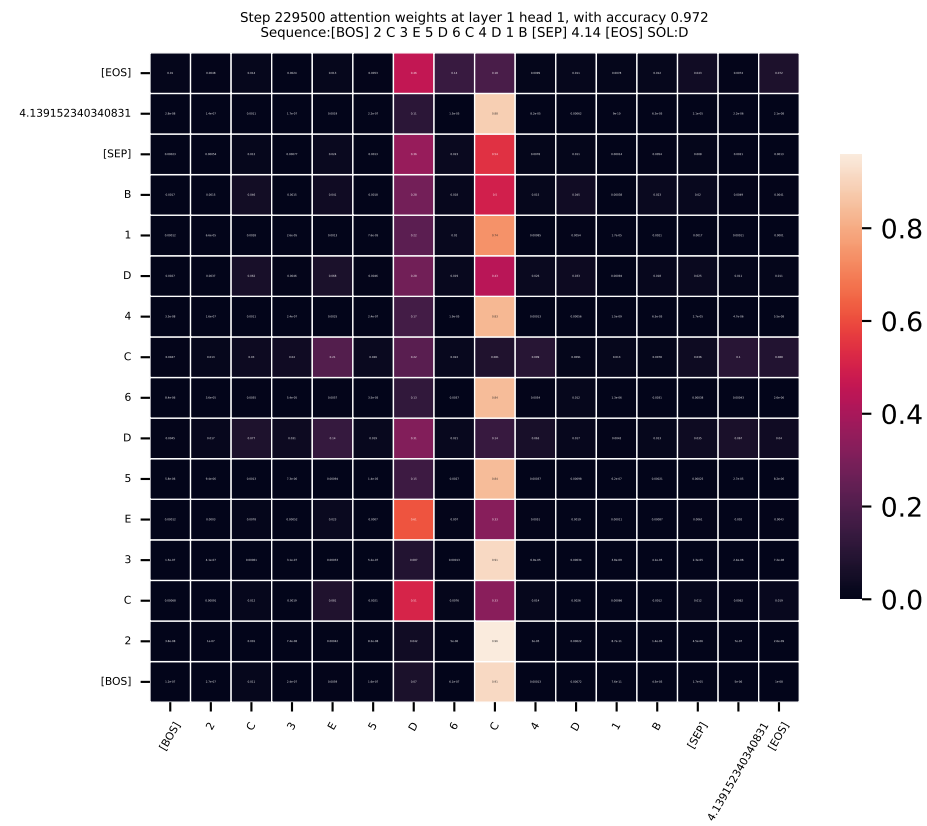

Step 220000 attention weights at layer 1 head 0, with accuracy 0.910
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.7
0.6
0.5
0.4
0.3
0.2
0.1

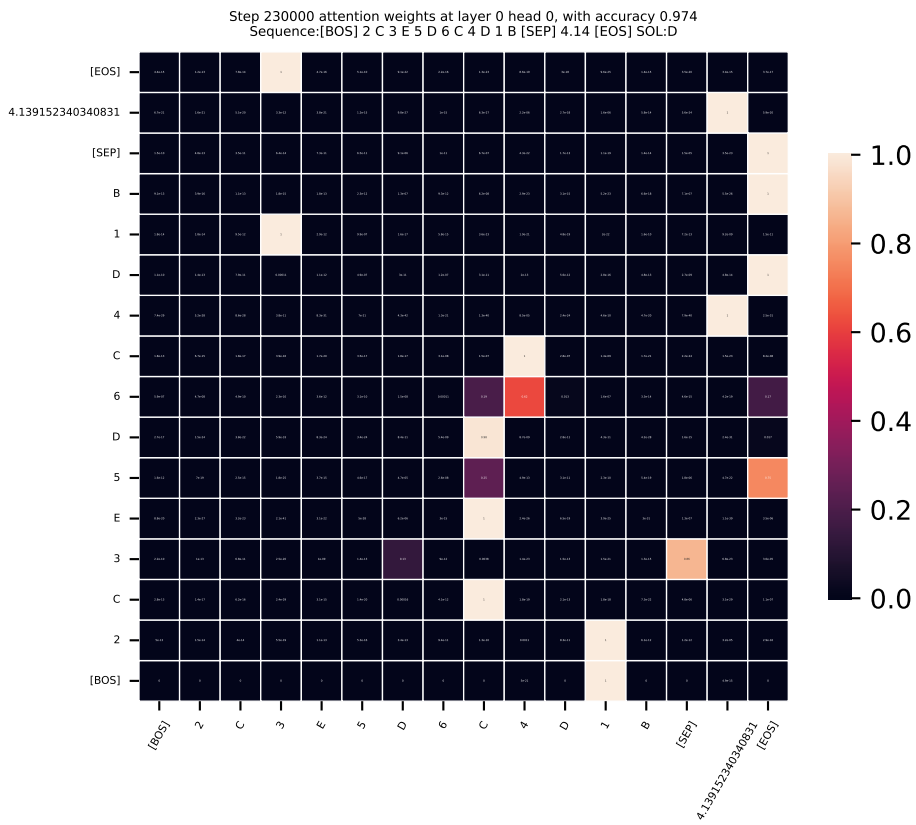

Step 220000 attention weights at layer 1 head 1, with accuracy 0.910
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.8
0.6
0.4
0.2

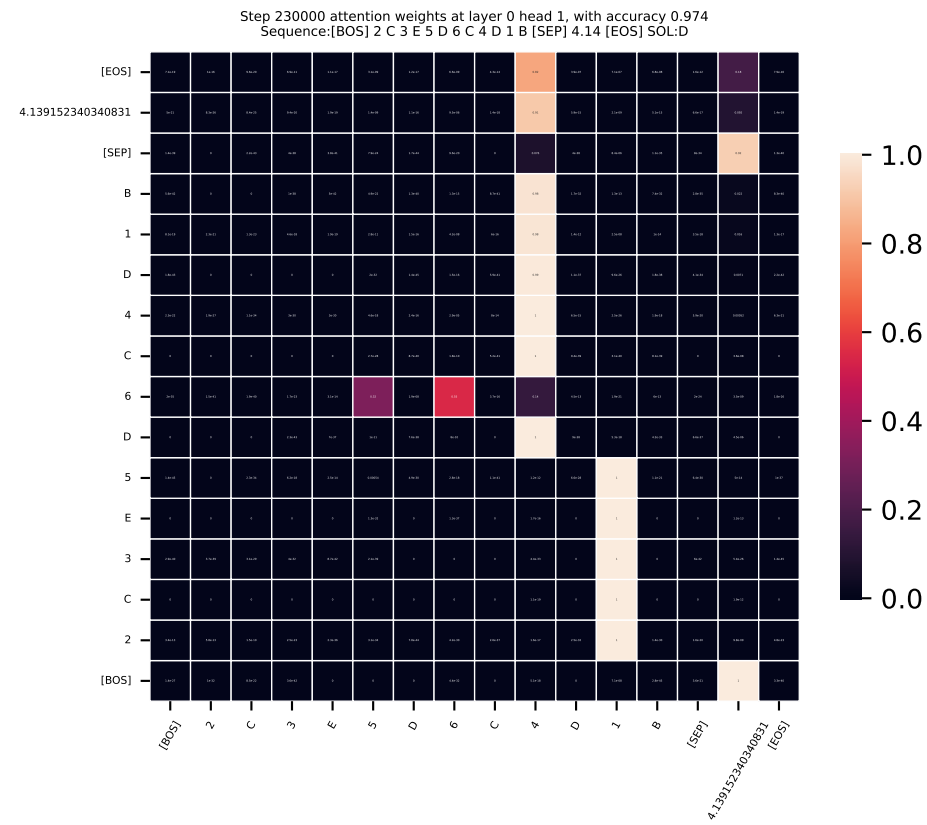

Step 229500 attention weights at layer 0 head 0, with accuracy 0.972
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
1.0
0.8
0.6
0.4
0.2
0.0

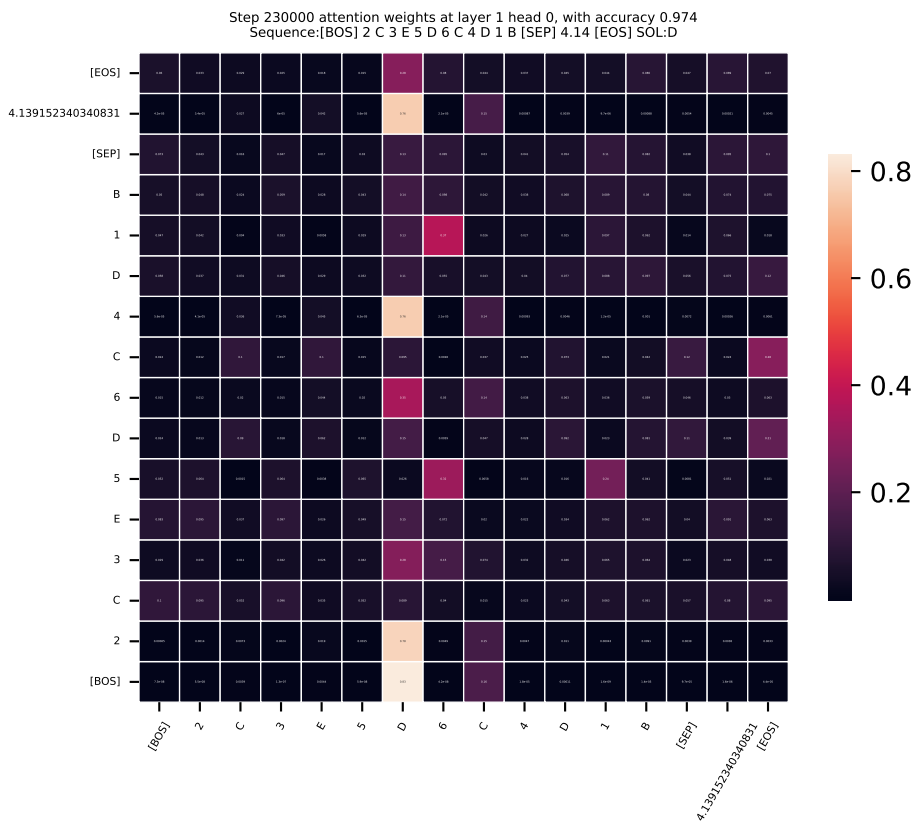

Step 229500 attention weights at layer 0 head 1, with accuracy 0.972
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
1.0
0.8
0.6
0.4
0.2
0.0

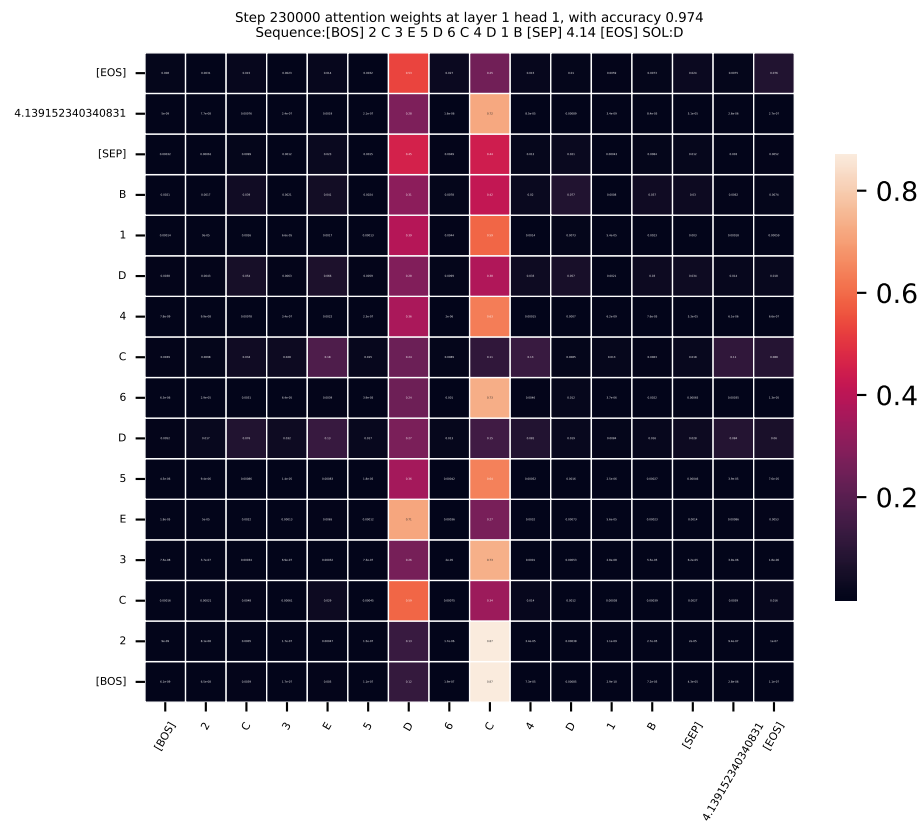

Step 229500 attention weights at layer 1 head 0, with accuracy 0.972
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.7
0.6
0.5
0.4
0.3
0.2
0.1

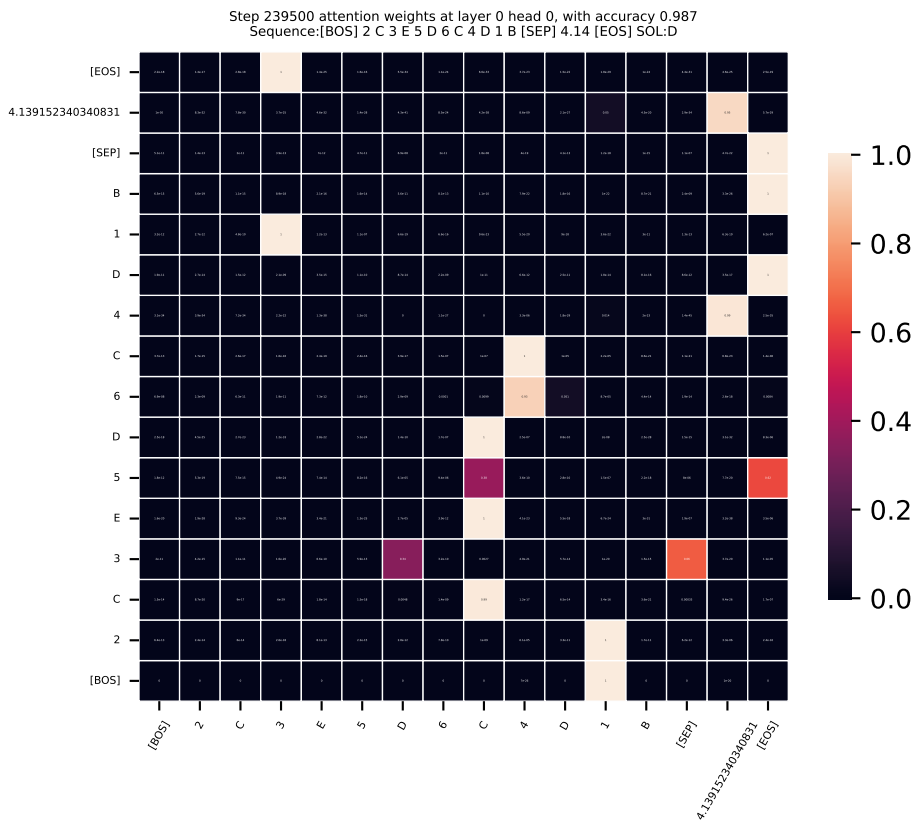

Step 229500 attention weights at layer 1 head 1, with accuracy 0.972
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.8
0.6
0.4
0.2
0.0

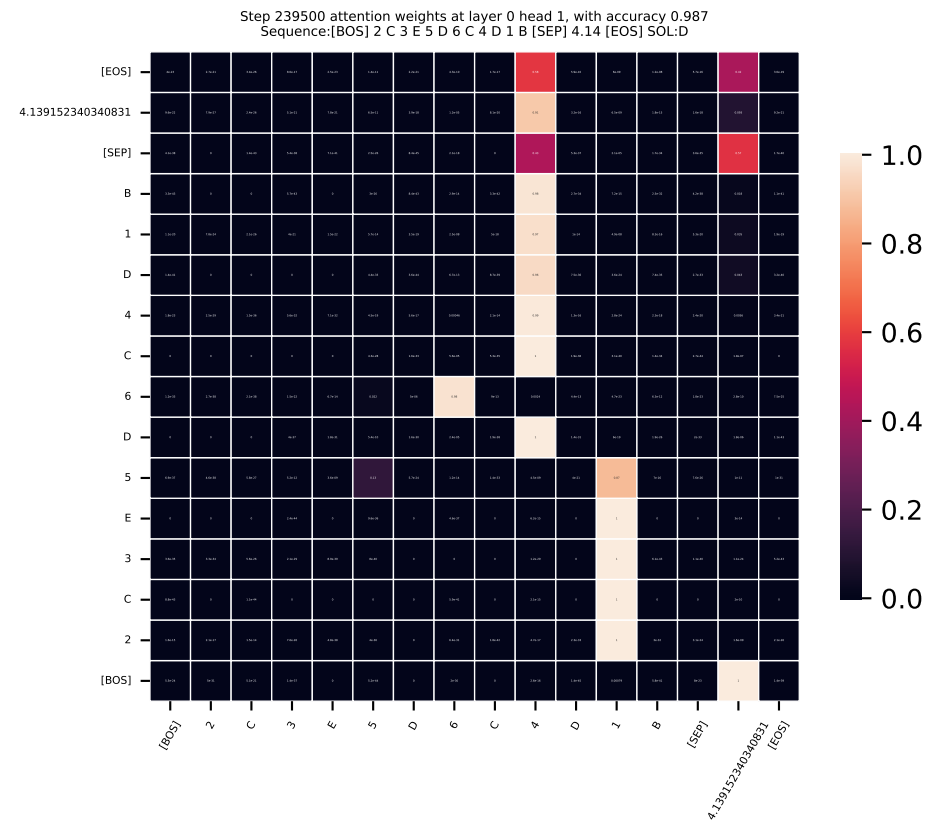

Step 230000 attention weights at layer 0 head 0, with accuracy 0.974
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

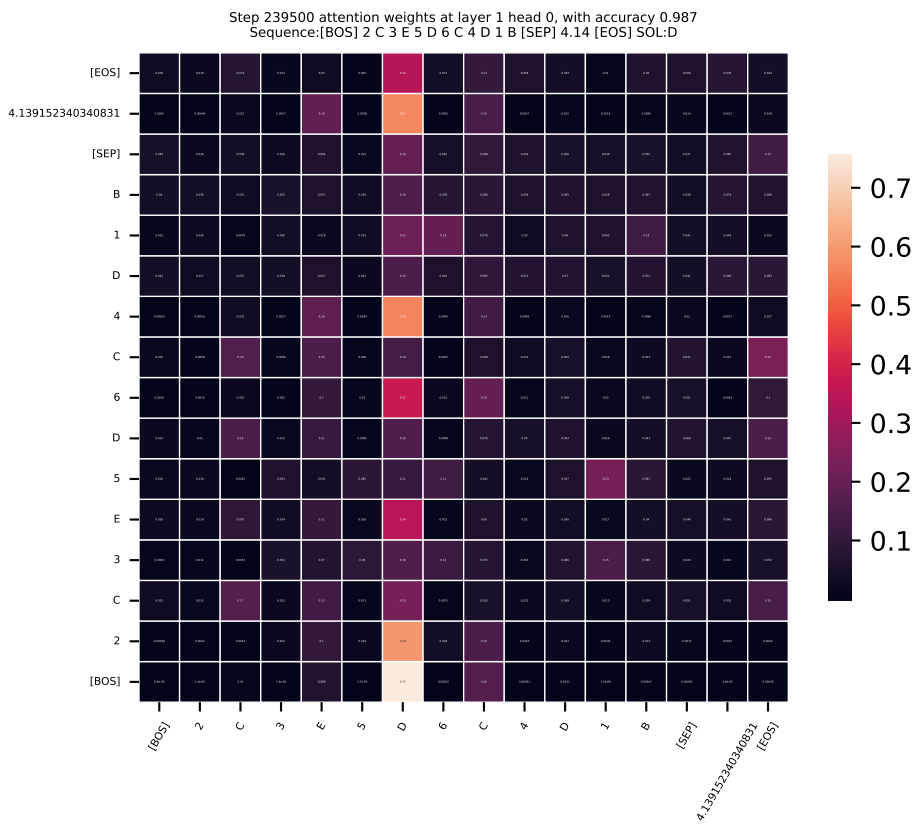

Step 230000 attention weights at layer 0 head 1, with accuracy 0.974
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

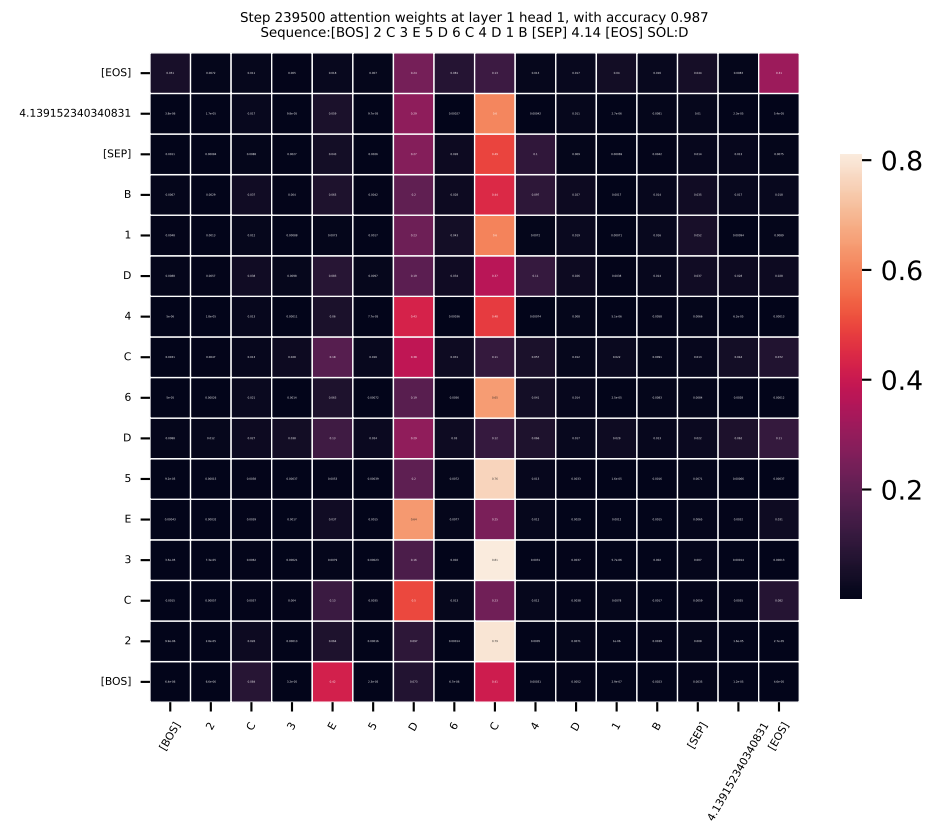

Step 230000 attention weights at layer 1 head 0, with accuracy 0.974
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2

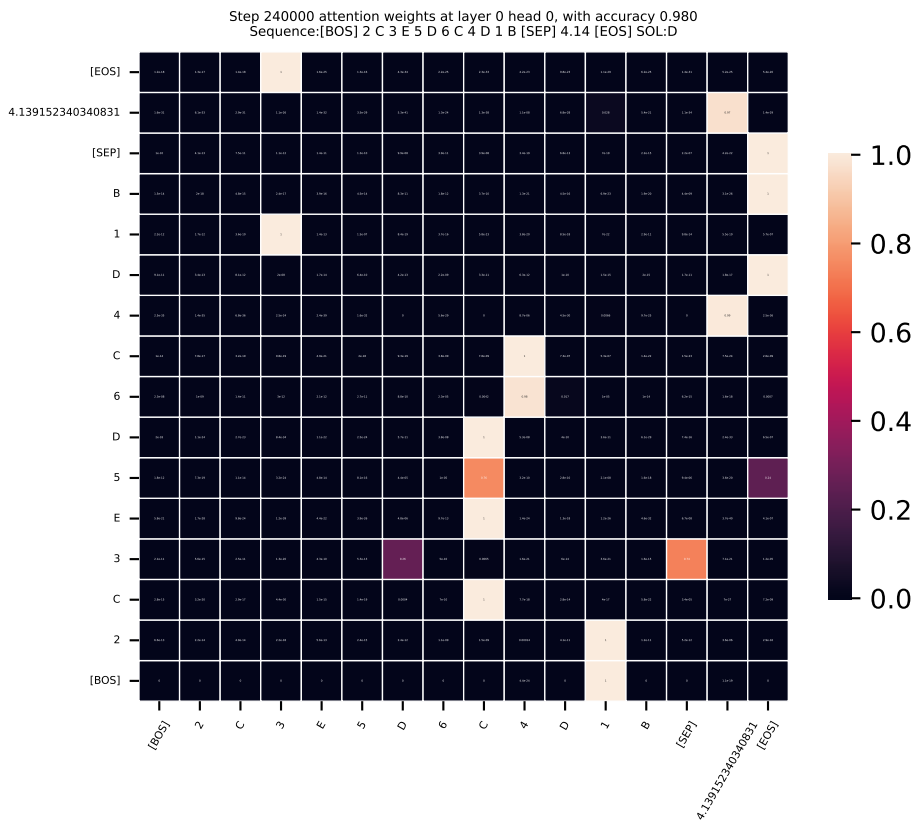

Step 230000 attention weights at layer 1 head 1, with accuracy 0.974
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2

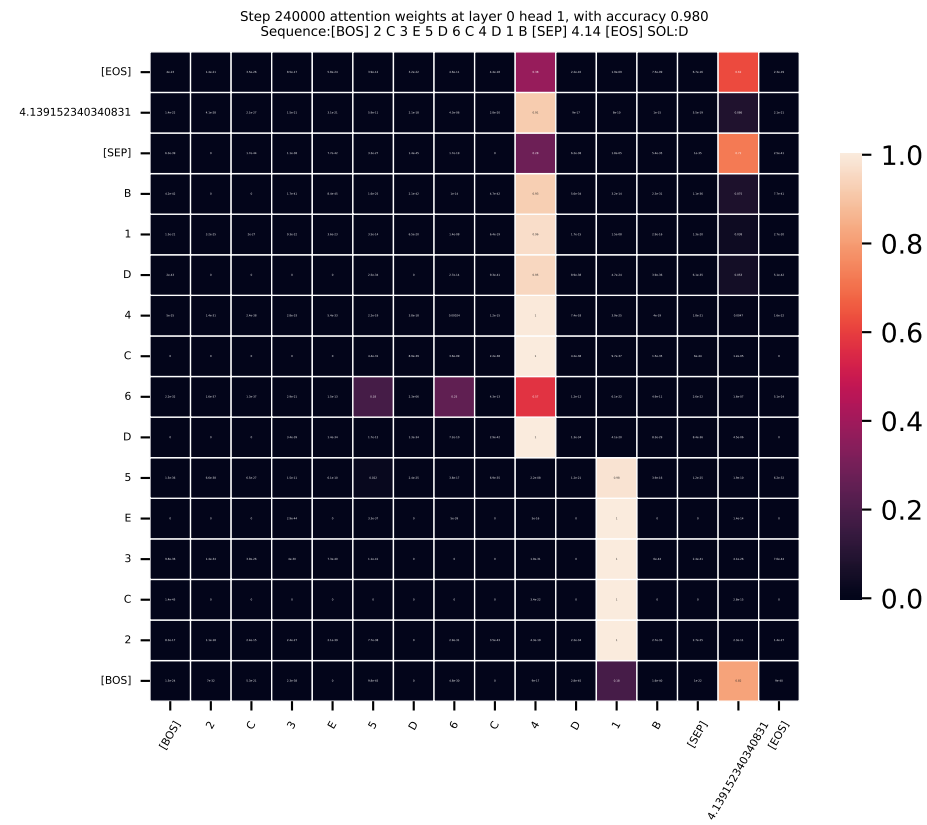

Step 239500 attention weights at layer 0 head 0, with accuracy 0.987
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

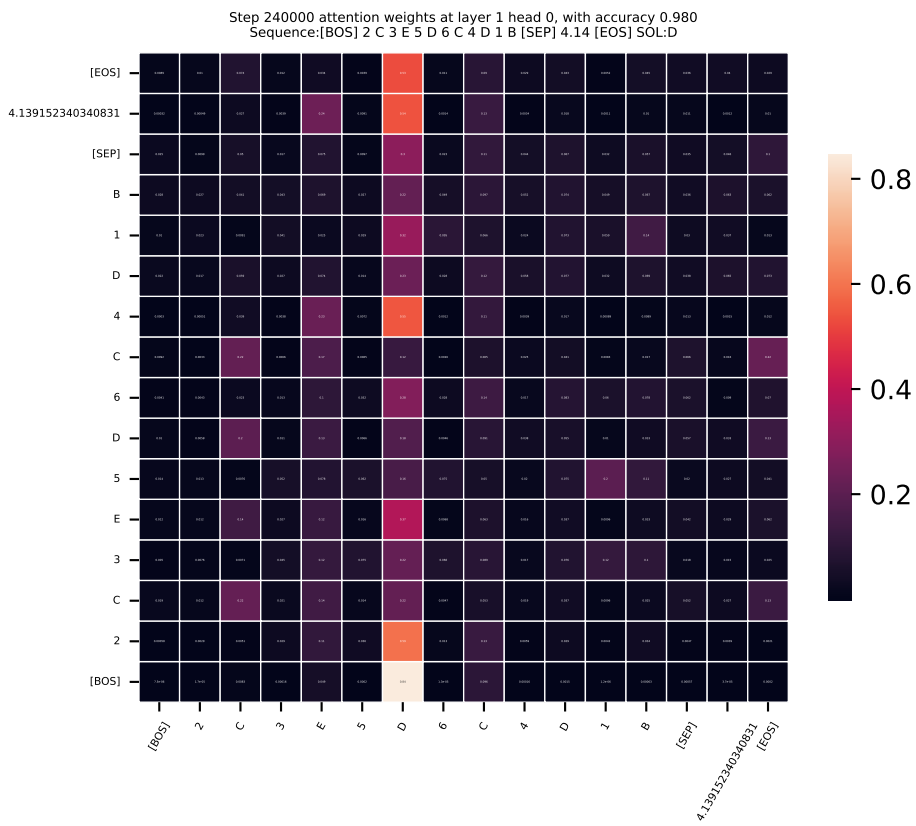

Step 239500 attention weights at layer 0 head 1, with accuracy 0.987
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

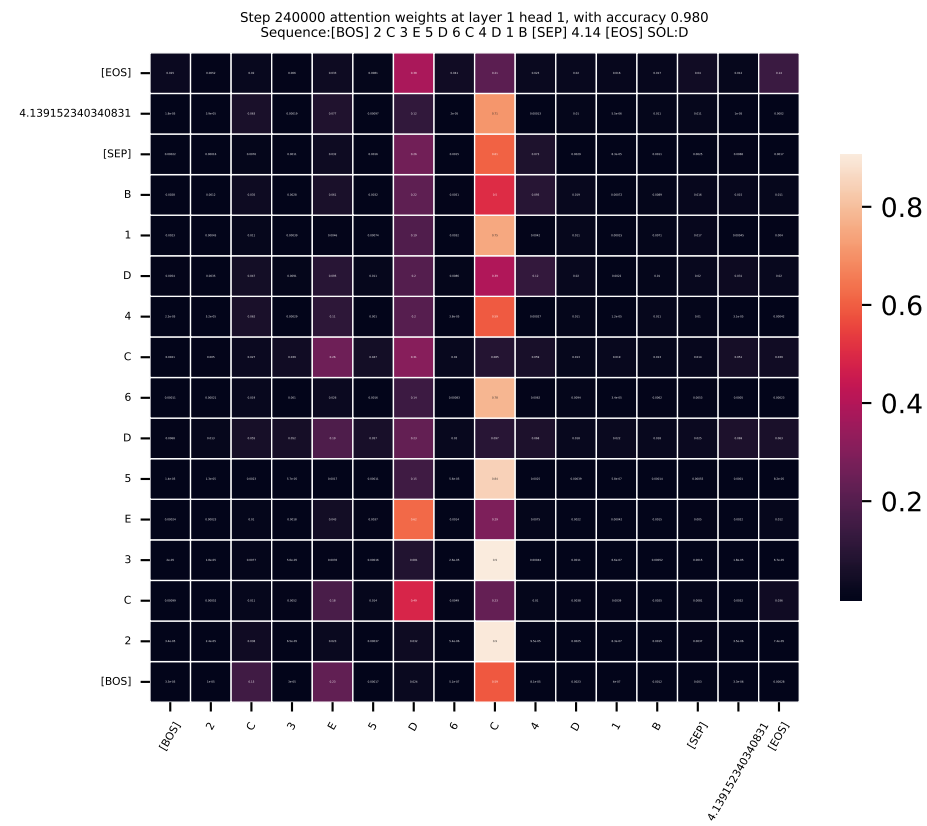

Step 239500 attention weights at layer 1 head 0, with accuracy 0.987
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.7
0.6
0.5
0.4
0.3
0.2
0.1

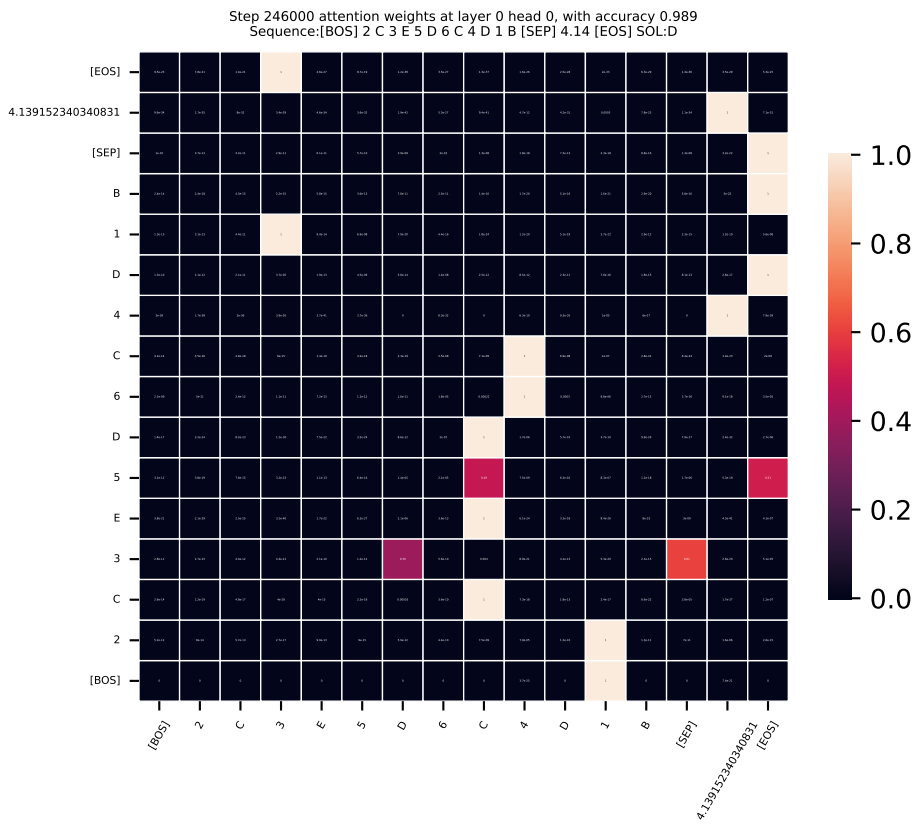

Step 239500 attention weights at layer 1 head 1, with accuracy 0.987
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2

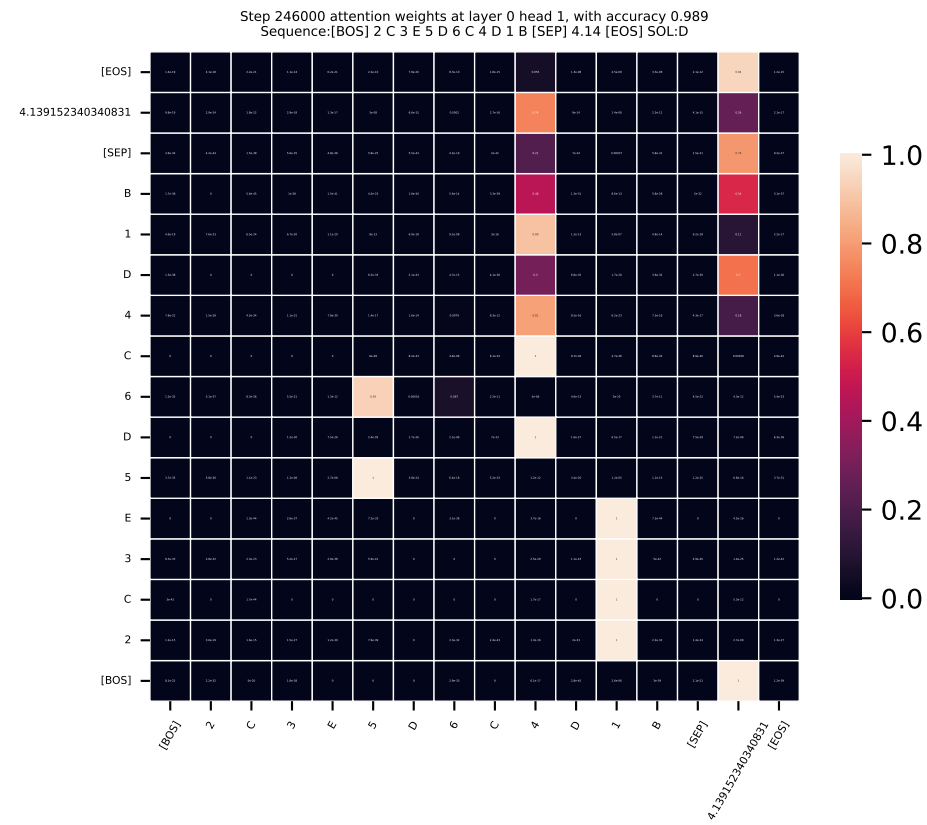

Step 240000 attention weights at layer 0 head 0, with accuracy 0.980
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

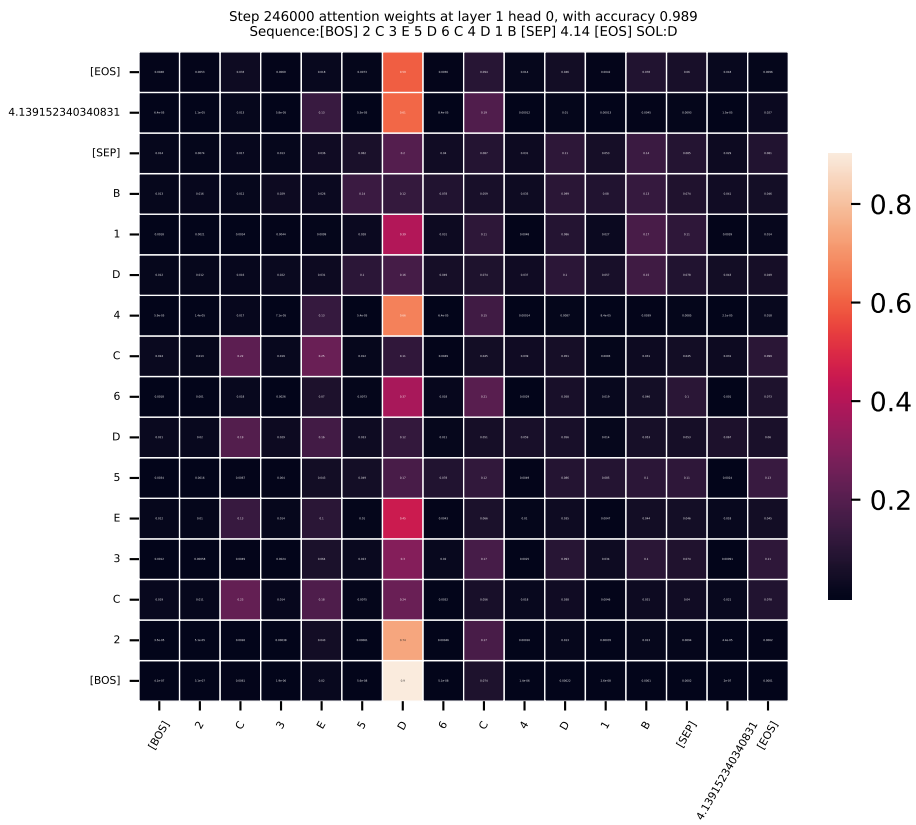

Step 240000 attention weights at layer 0 head 1, with accuracy 0.980
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

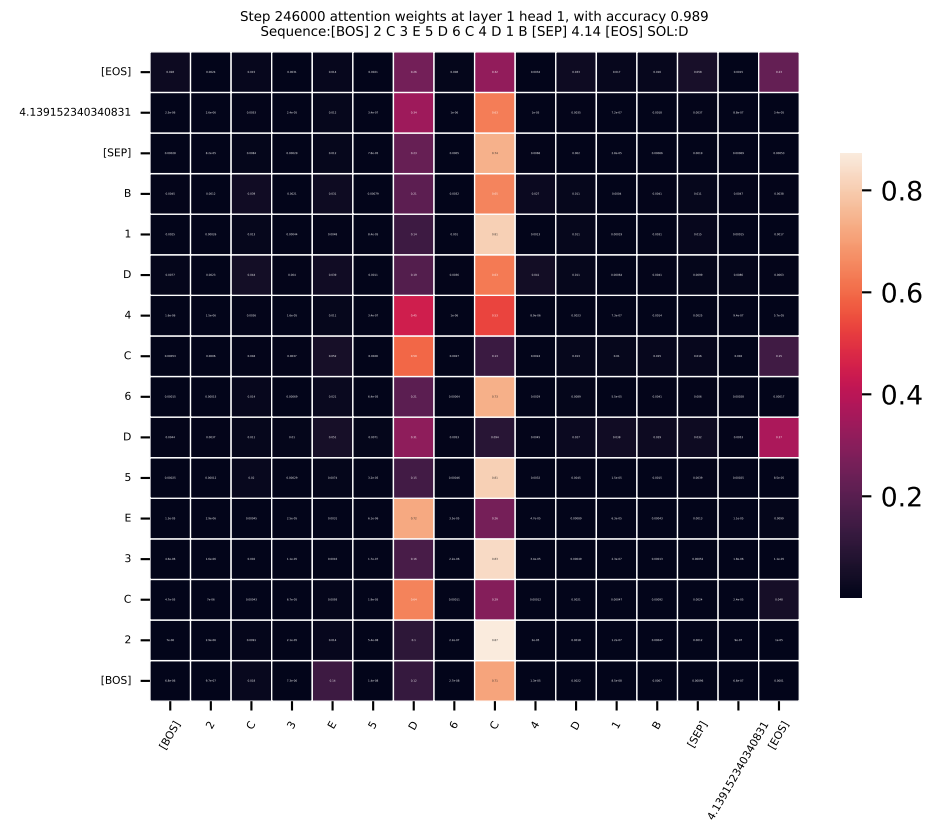

Step 240000 attention weights at layer 1 head 0, with accuracy 0.980
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.8
0.6
0.4
0.2

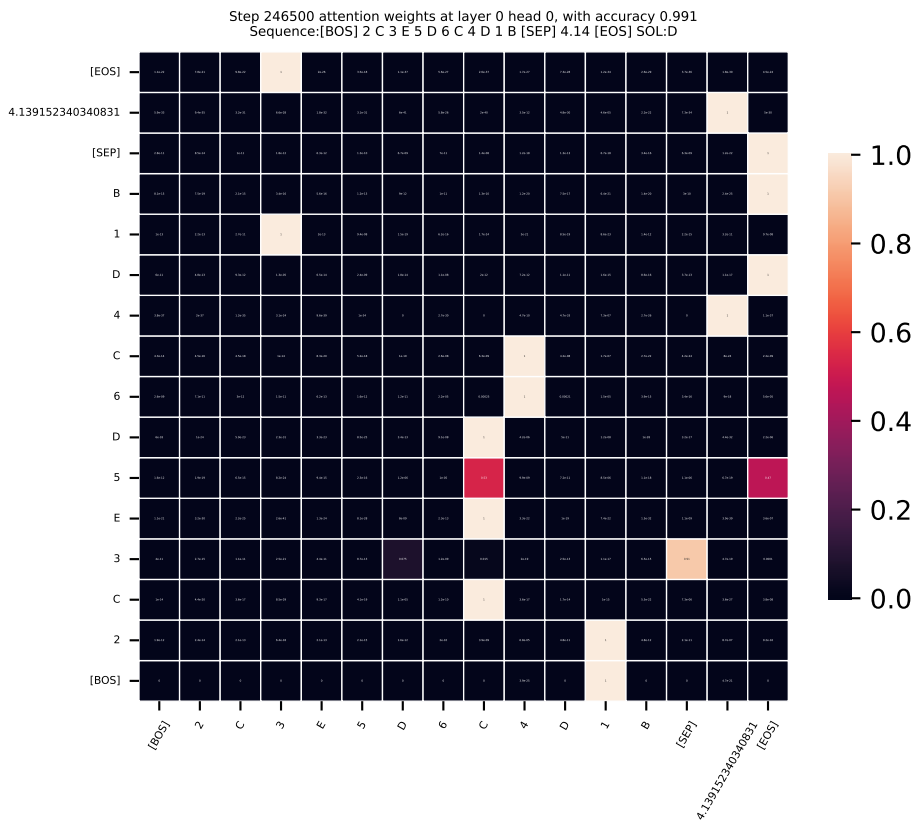

Step 240000 attention weights at layer 1 head 1, with accuracy 0.980
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.139152340340831 [EOS]
0.8
0.6
0.4
0.2

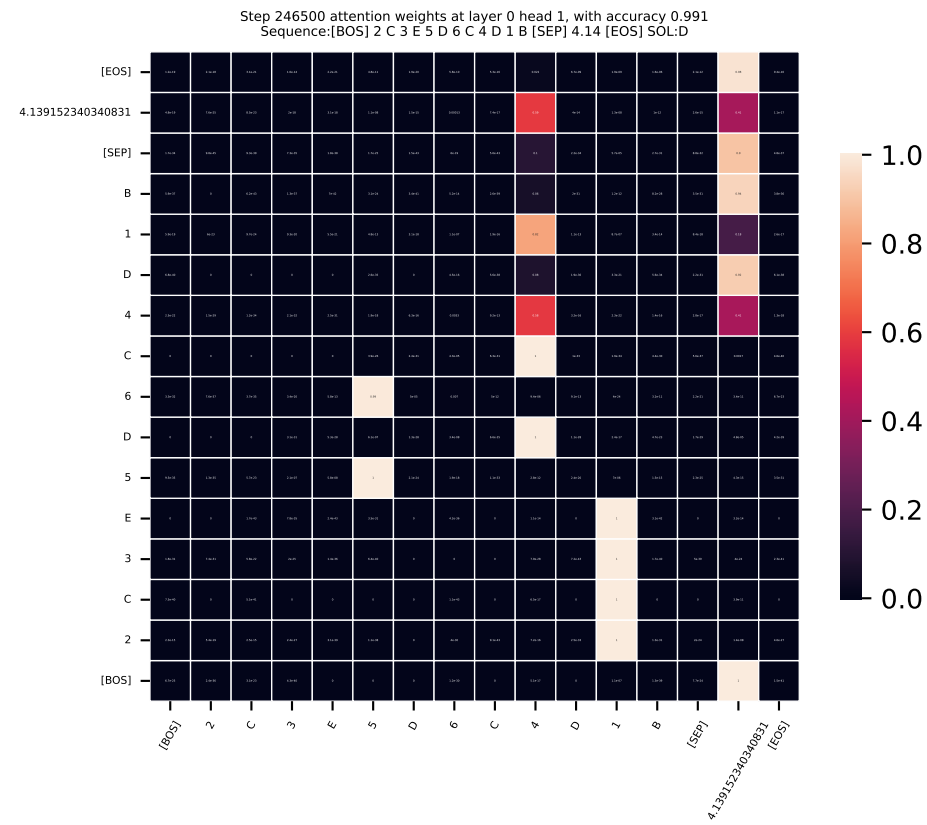

Step 246000 attention weights at layer 0 head 0, with accuracy 0.989
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

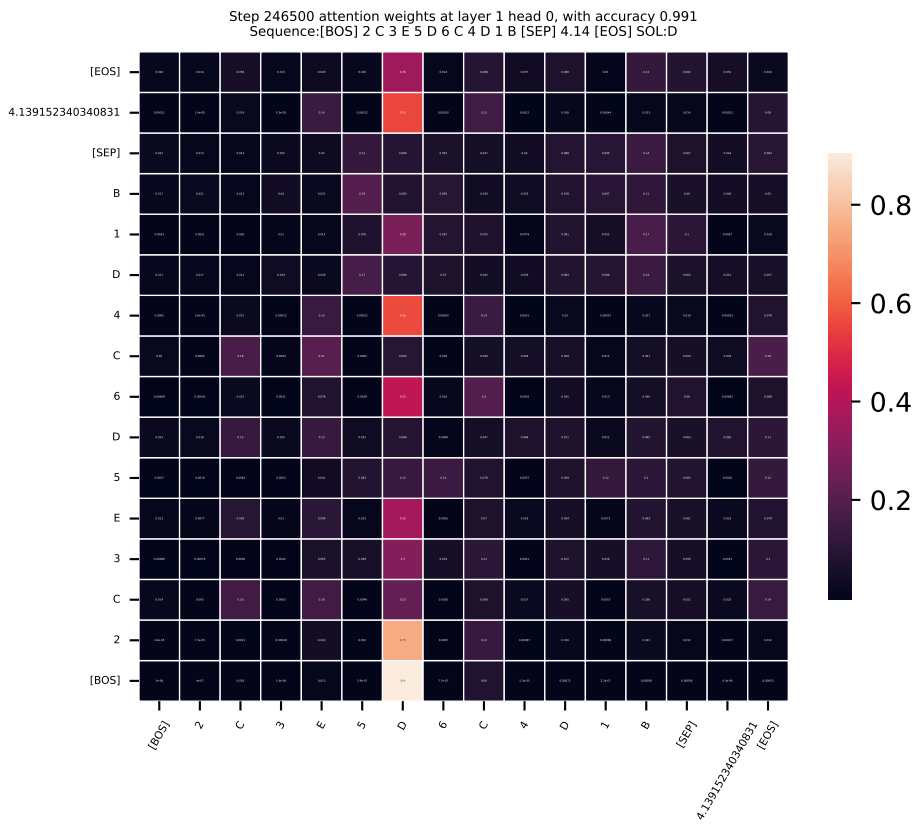

Step 246000 attention weights at layer 0 head 1, with accuracy 0.989
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0

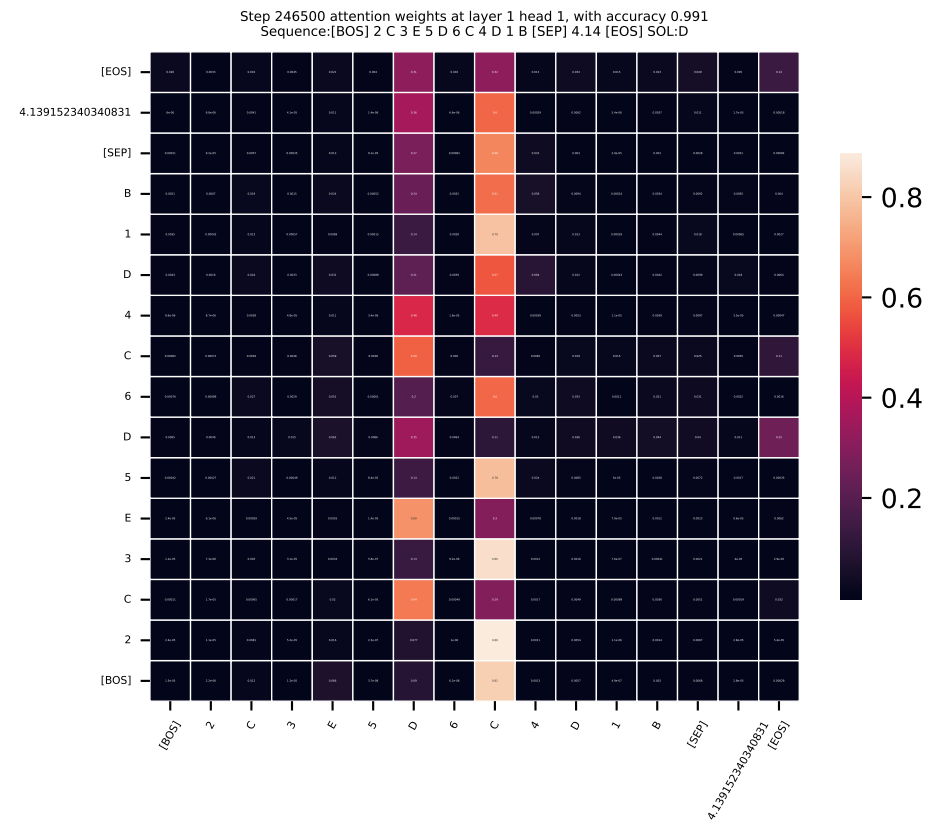

Step 246000 attention weights at layer 1 head 0, with accuracy 0.989
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2


Step 246000 attention weights at layer 1 head 1, with accuracy 0.989
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2


Step 246500 attention weights at layer 0 head 0, with accuracy 0.991
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0


Step 246500 attention weights at layer 0 head 1, with accuracy 0.991
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
1.0
0.8
0.6
0.4
0.2
0.0


Step 246500 attention weights at layer 1 head 0, with accuracy 0.991
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2


Step 246500 attention weights at layer 1 head 1, with accuracy 0.991
Sequence:[BOS] 2 C 3 E 5 D 6 C 4 D 1 B [SEP] 4.14 [EOS] SOL:D
[EOS]
4.139152340340831
[SEP]
B
1
D
4
C
6
D
5
E
3
C
2
[BOS]
[BOS]
2
C
3
E
5
D
6
C
4
D
1
B
[SEP]
4.139152340340831
[EOS]
0.8
0.6
0.4
0.2