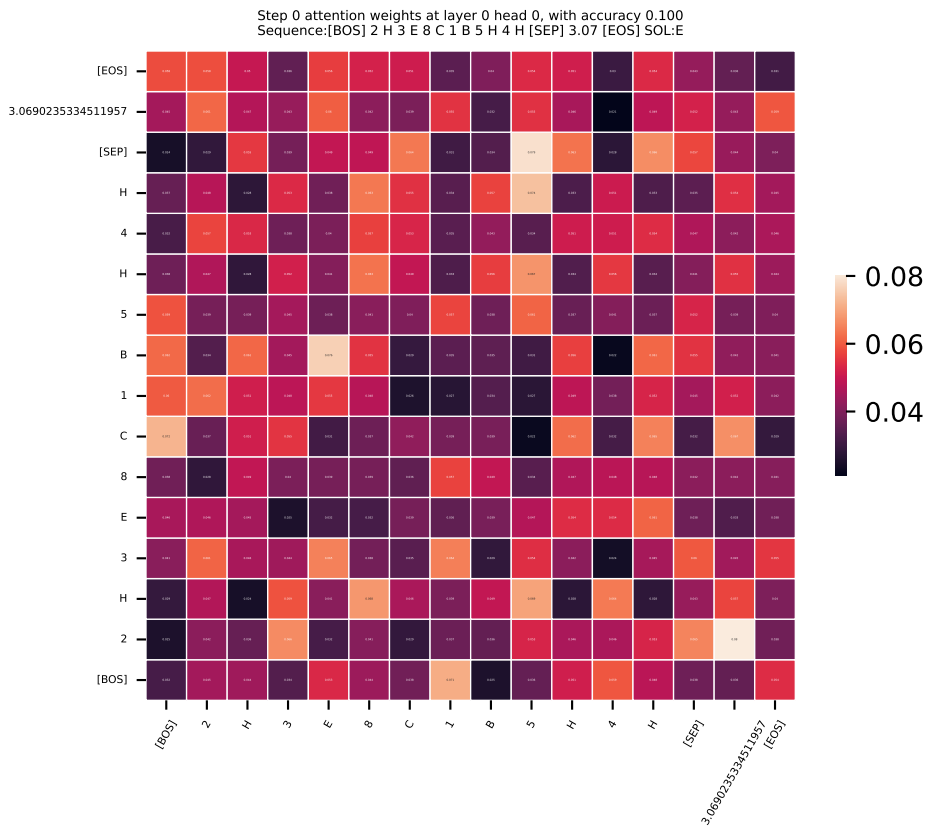
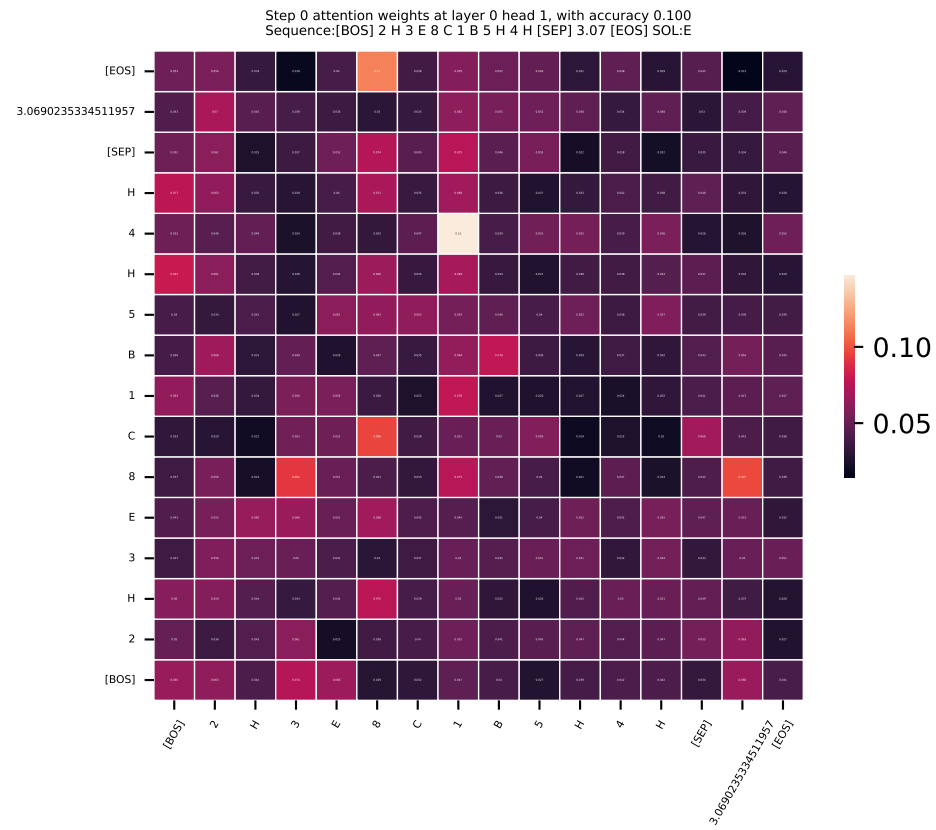
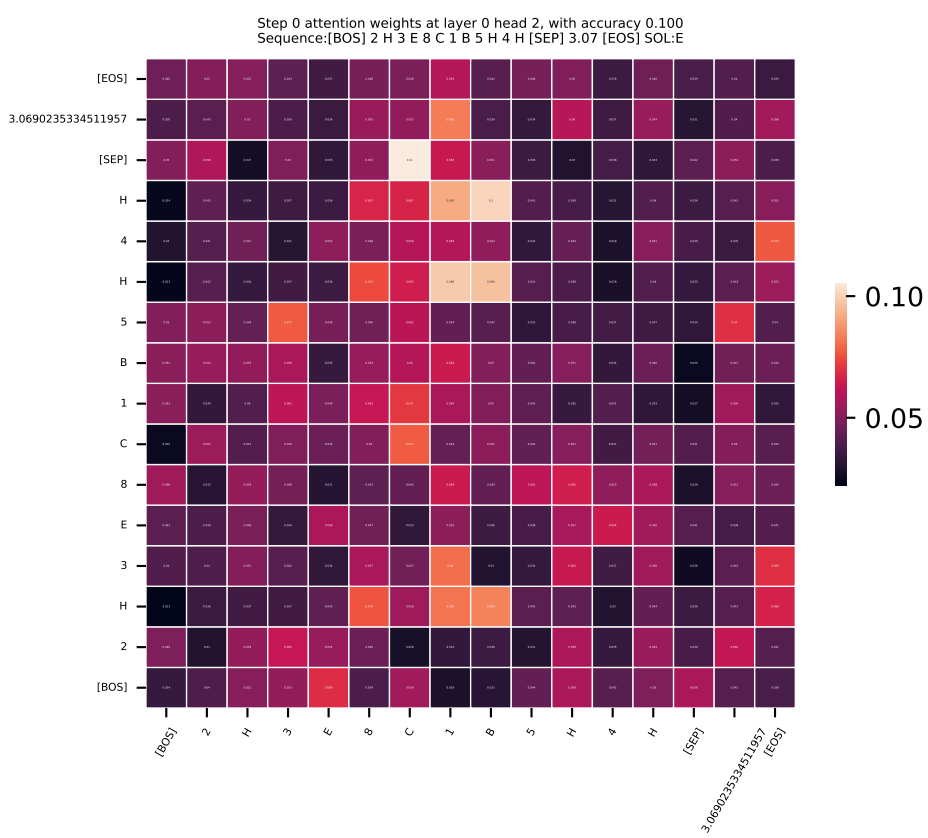
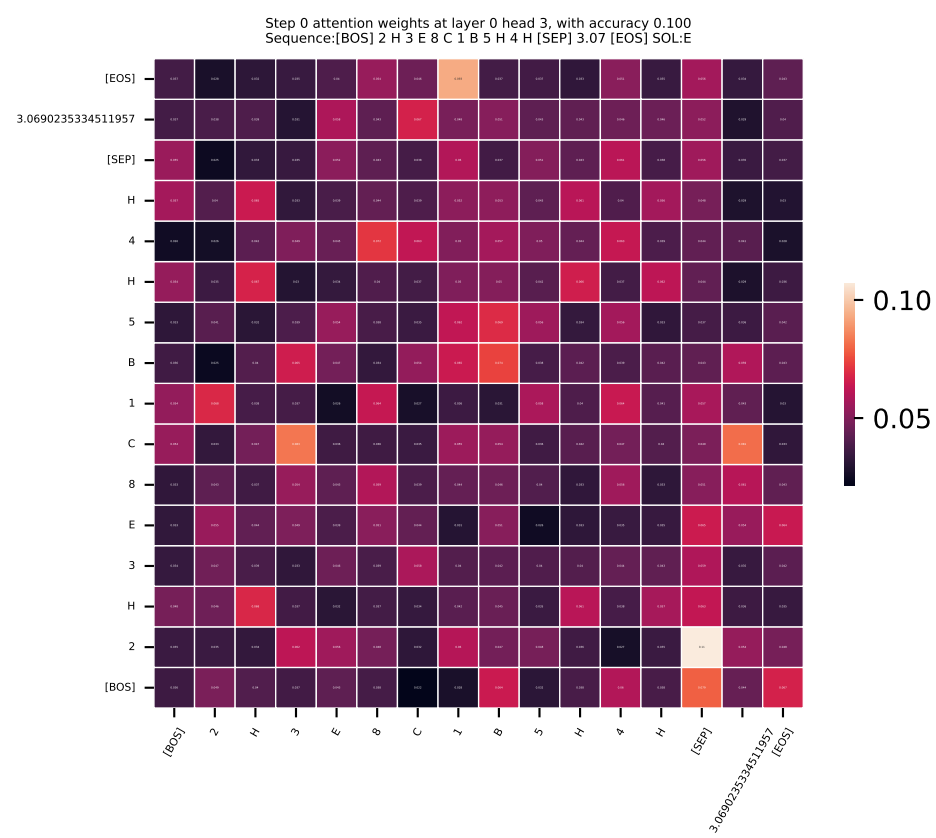
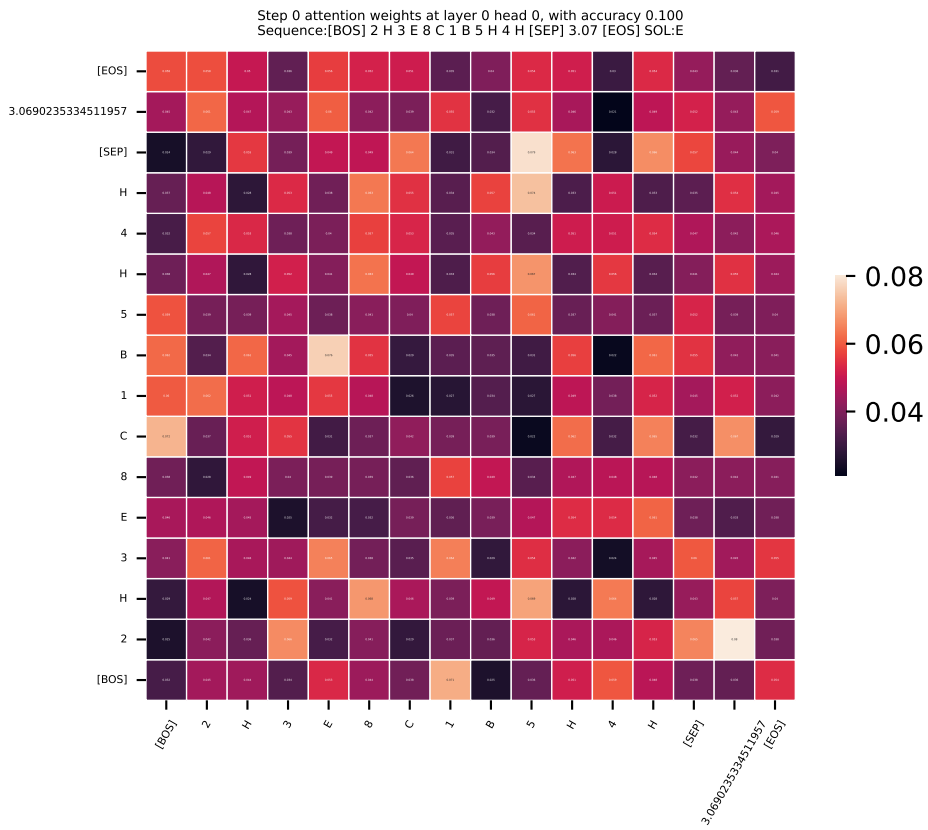
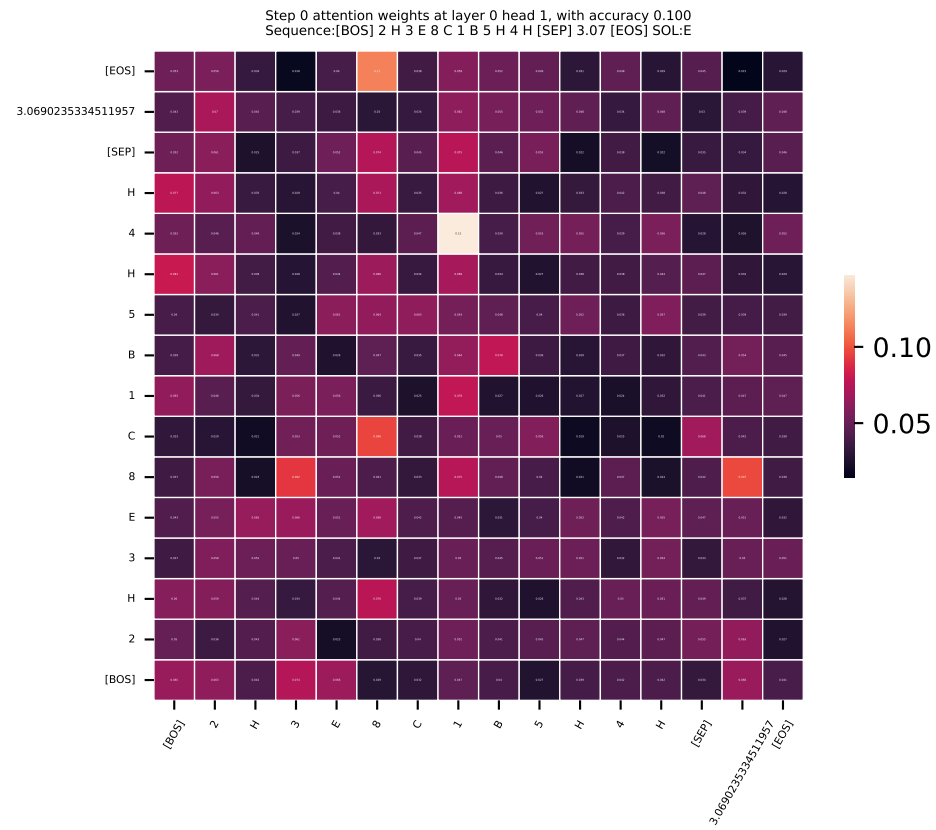

Step 0 attention weights at layer 0 head 0, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.08
0.06
0.04

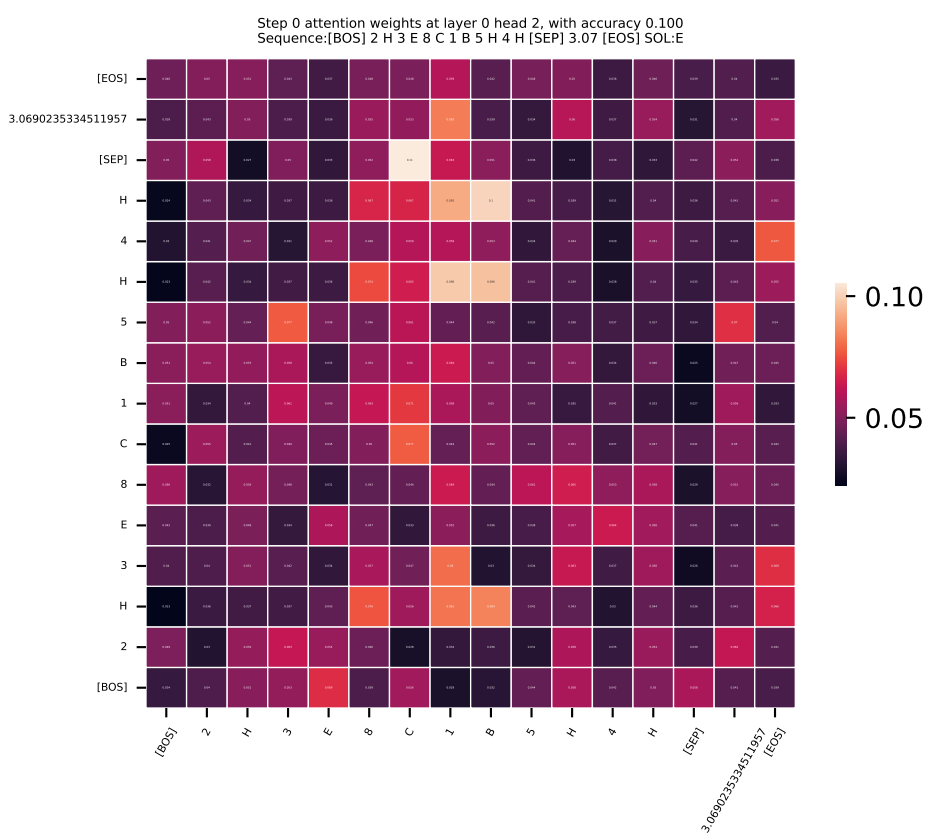

Step 0 attention weights at layer 0 head 1, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.10
0.05

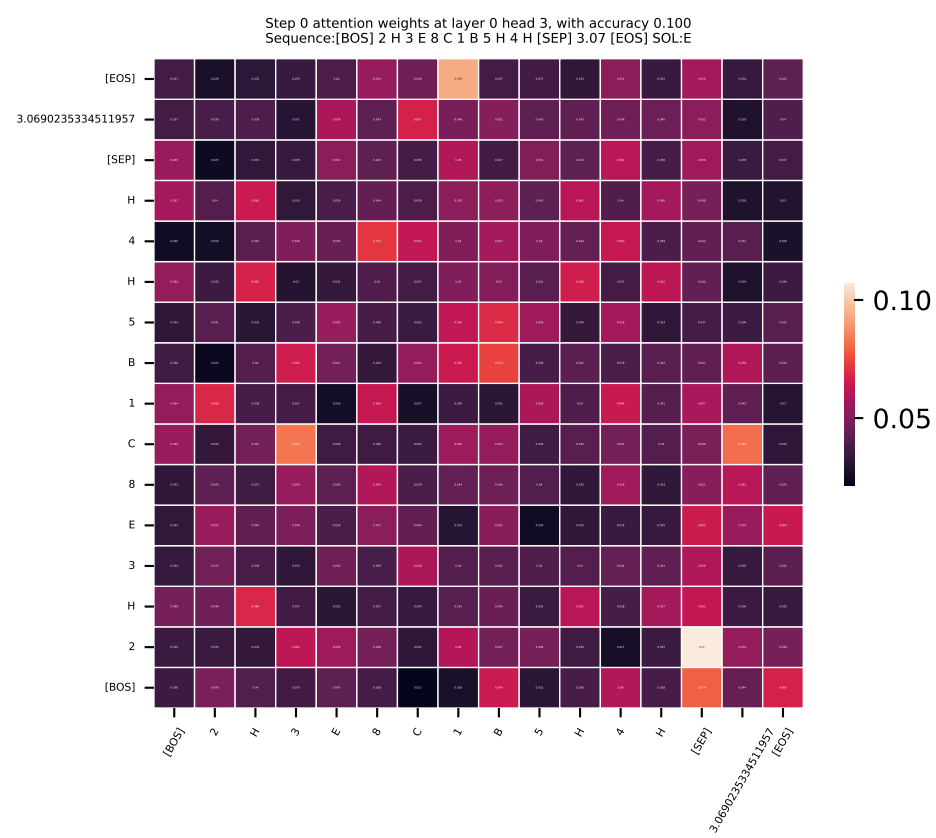

Step 0 attention weights at layer 0 head 2, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.10
0.05

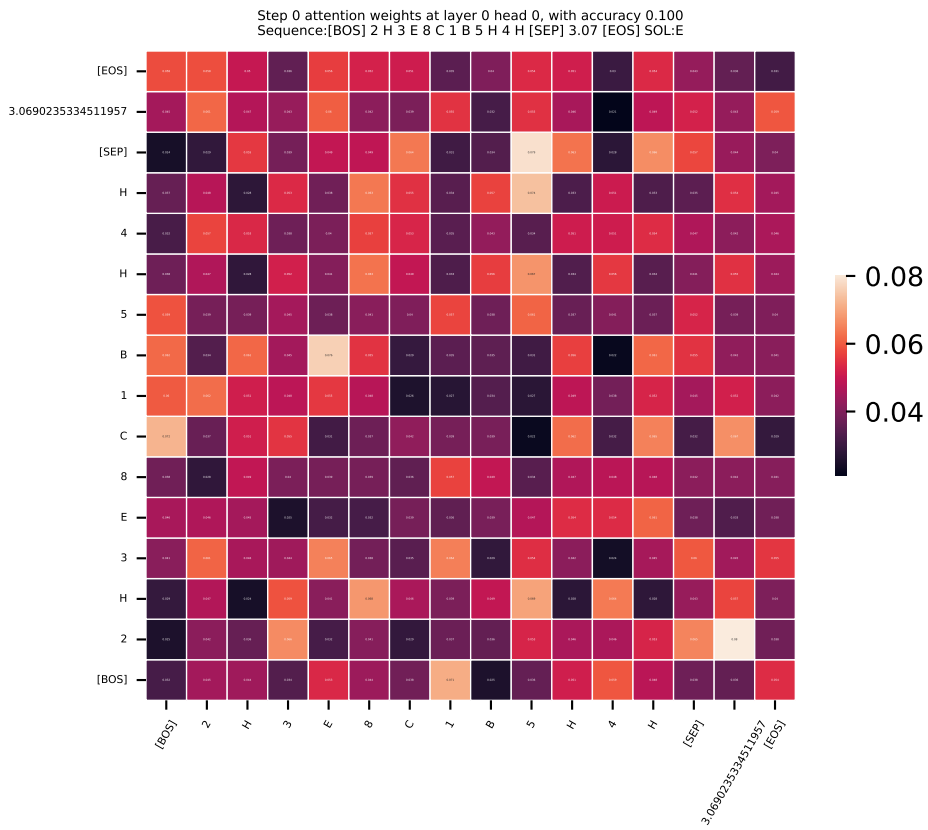

Step 0 attention weights at layer 0 head 3, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.10
0.05

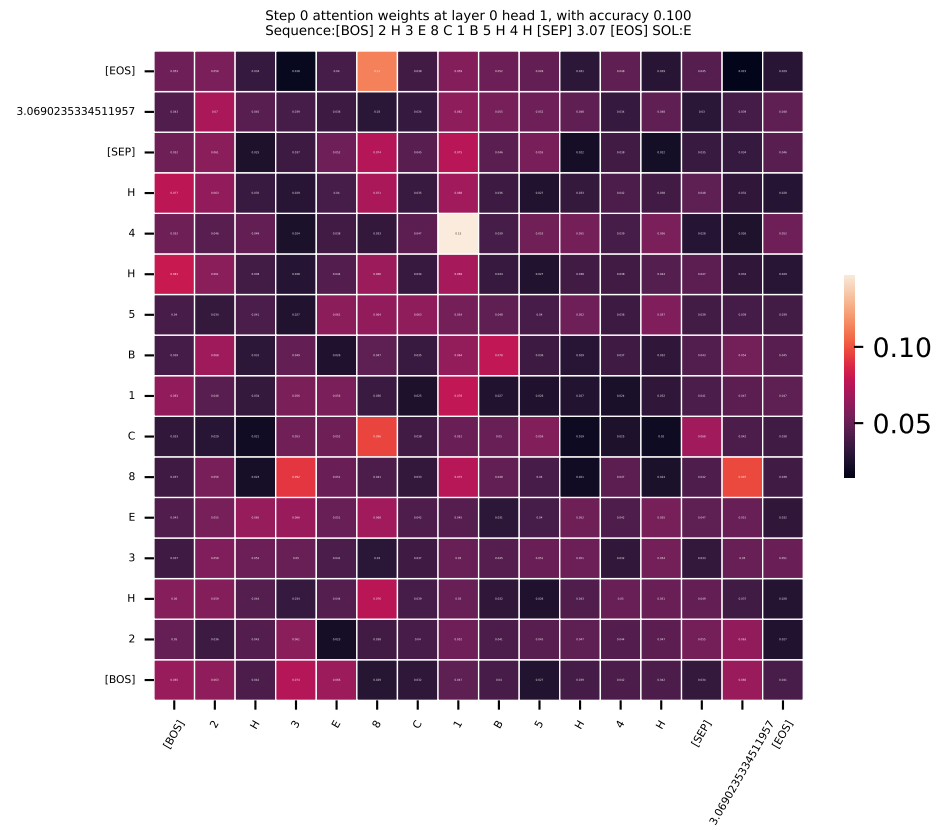

Step 0 attention weights at layer 0 head 0, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.08
0.06
0.04

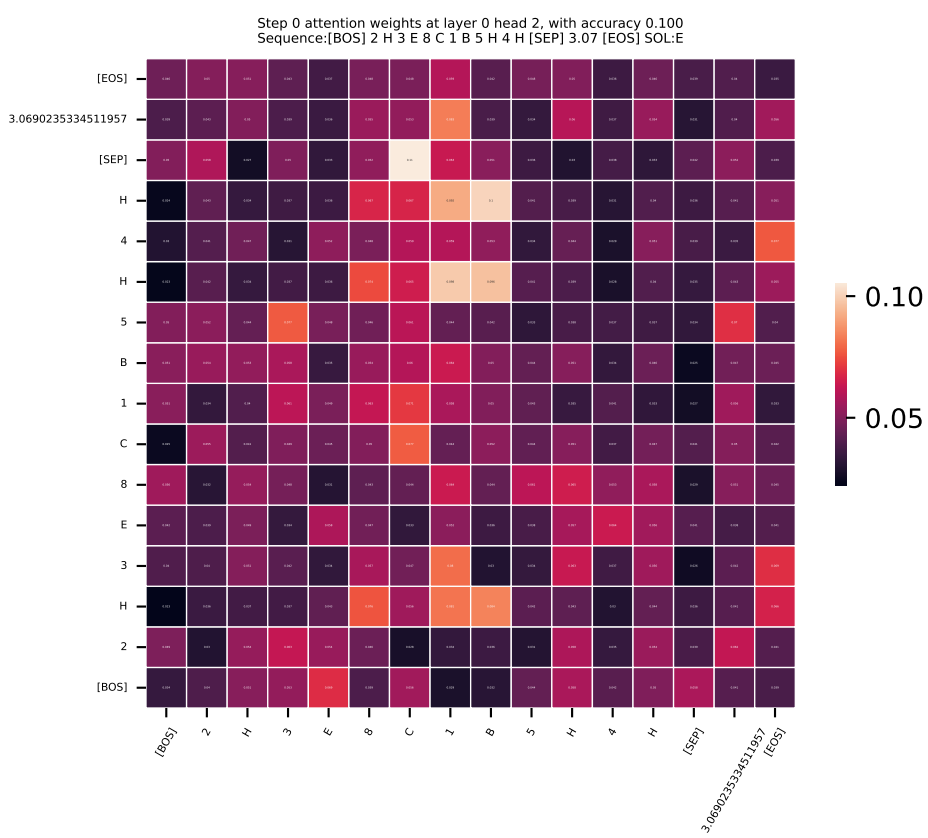

Step 0 attention weights at layer 0 head 1, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.10
0.05

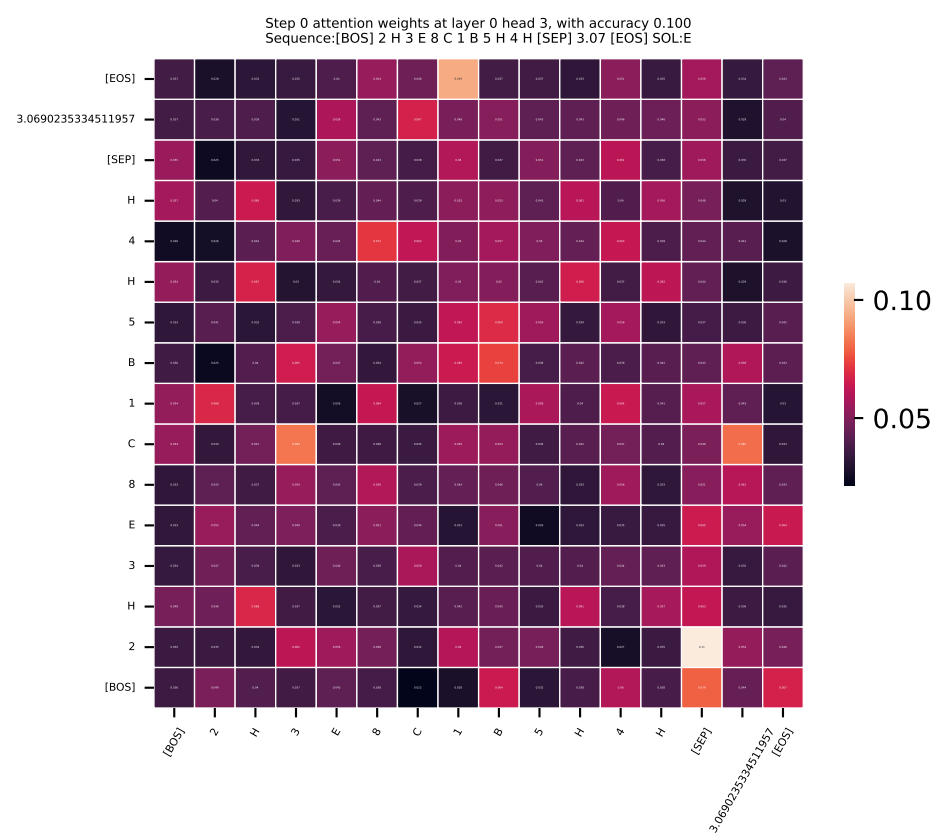

Step 0 attention weights at layer 0 head 2, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.10
0.05

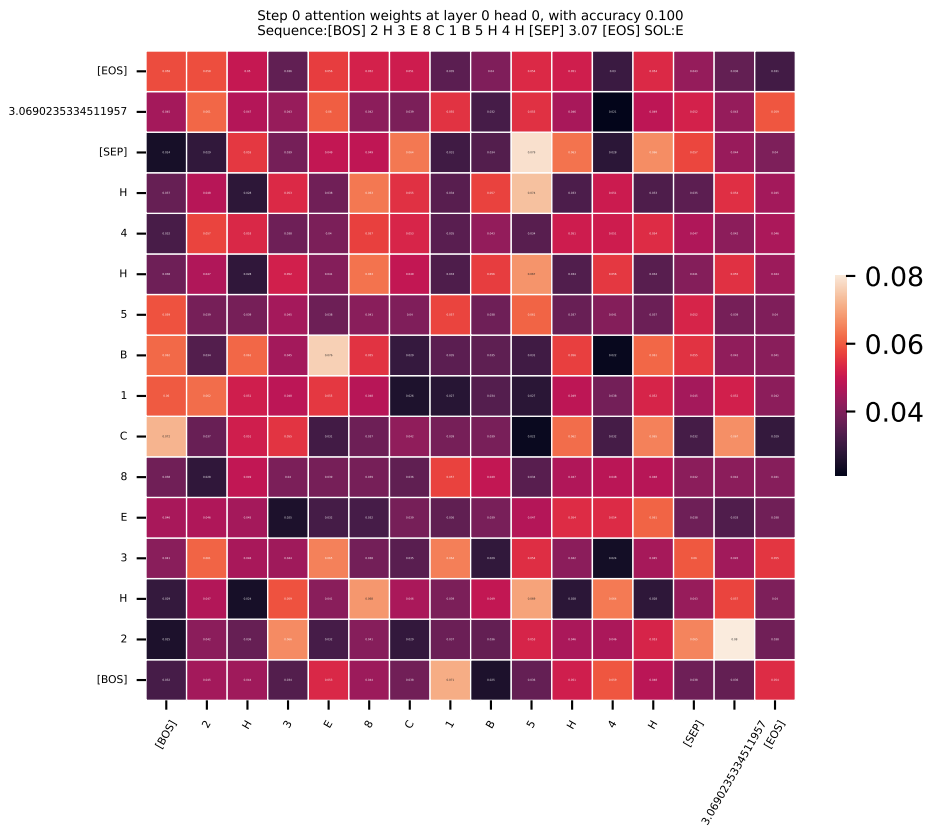

Step 0 attention weights at layer 0 head 3, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.10
0.05

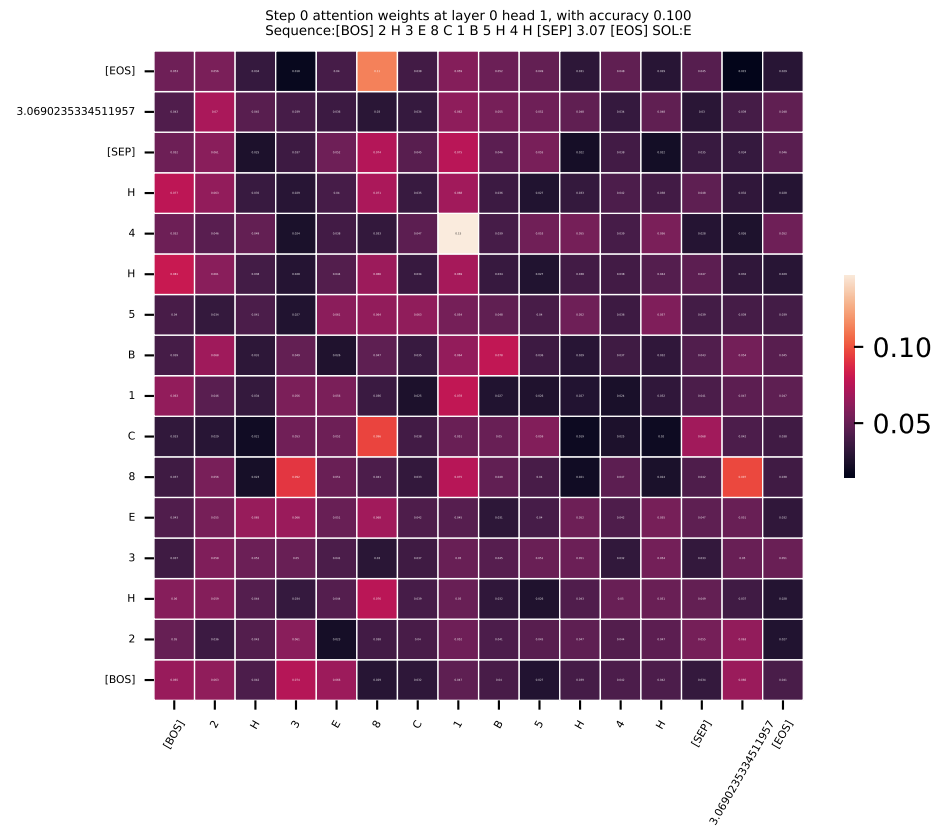

Step 0 attention weights at layer 0 head 0, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.08
0.06
0.04

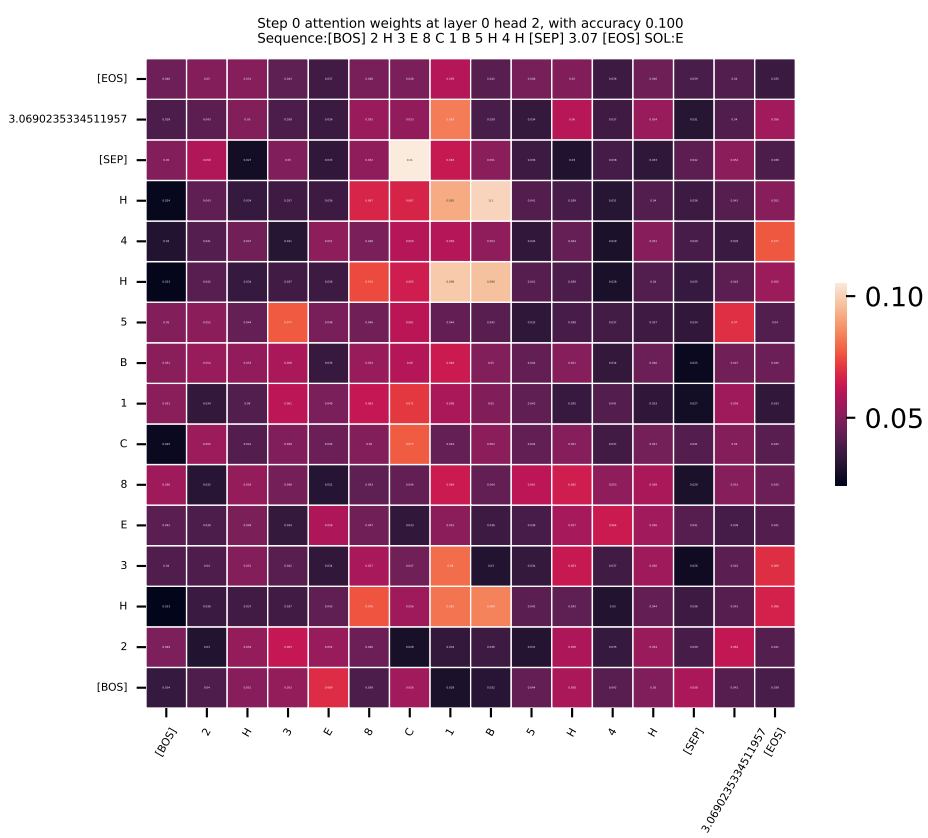

Step 0 attention weights at layer 0 head 1, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.10
0.05

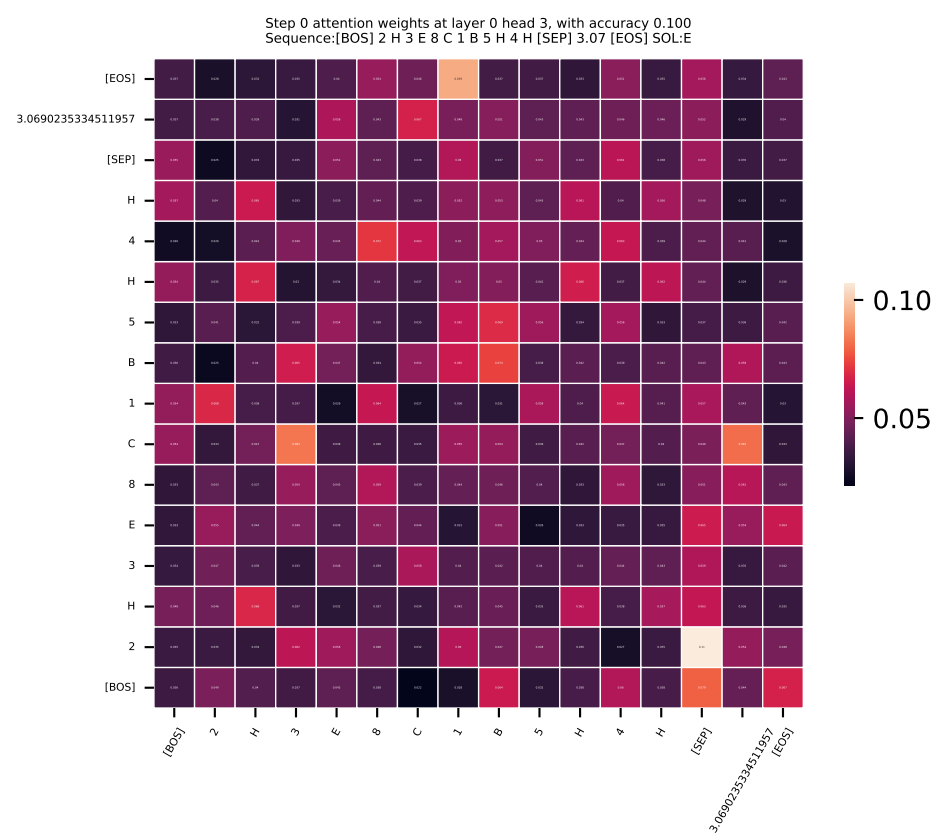

Step 0 attention weights at layer 0 head 2, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.10
0.05

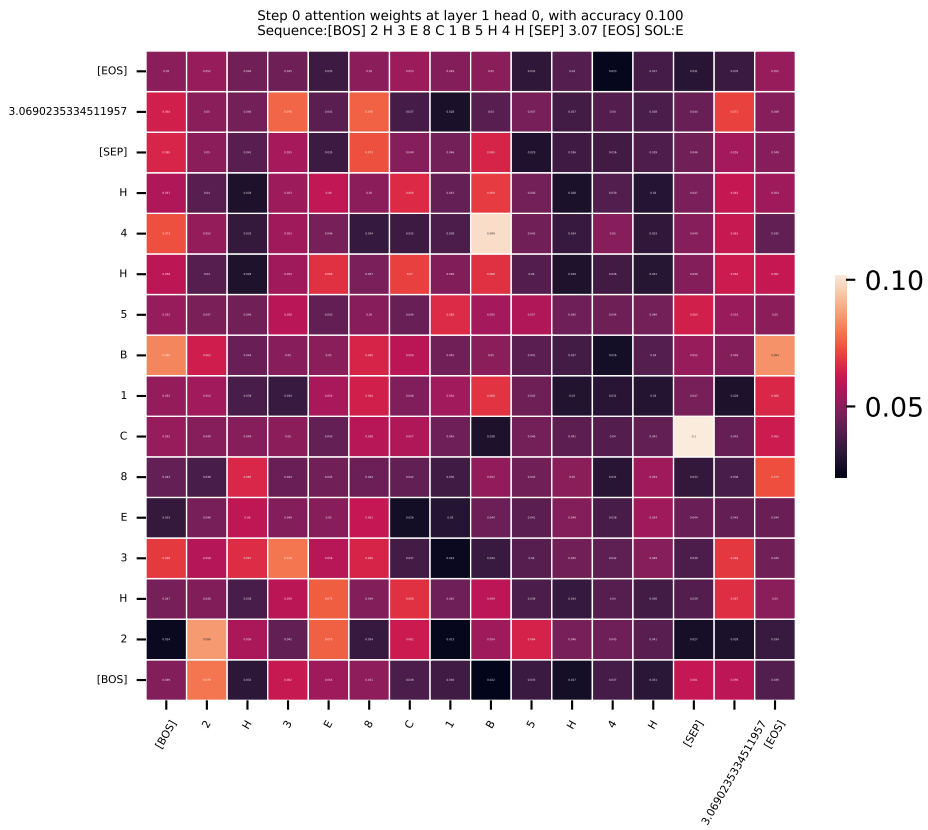

Step 0 attention weights at layer 0 head 3, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.10
0.05

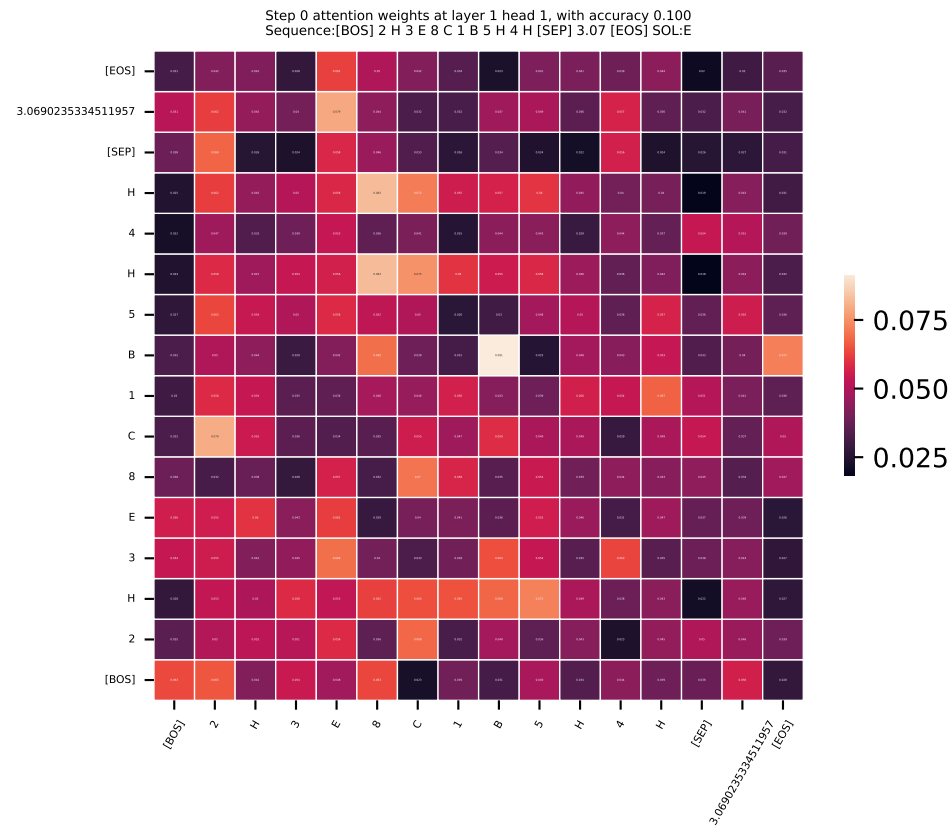

Step 0 attention weights at layer 0 head 0, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.08
0.06
0.04

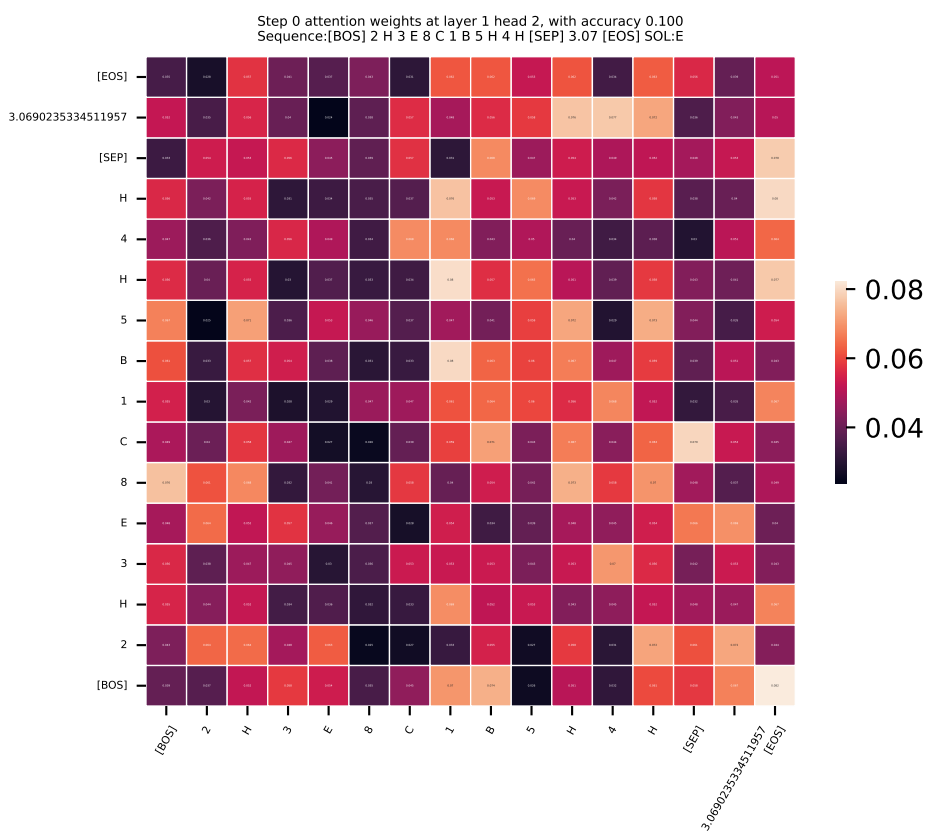

Step 0 attention weights at layer 0 head 1, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.10
0.05

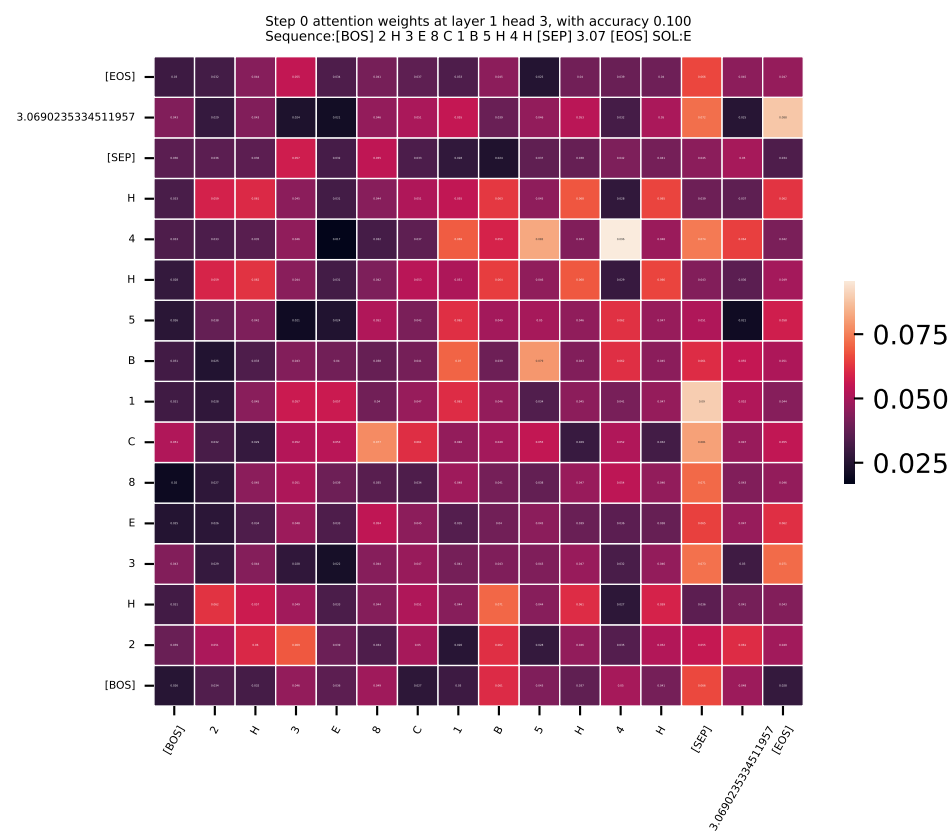

Step 0 attention weights at layer 0 head 2, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.10
0.05

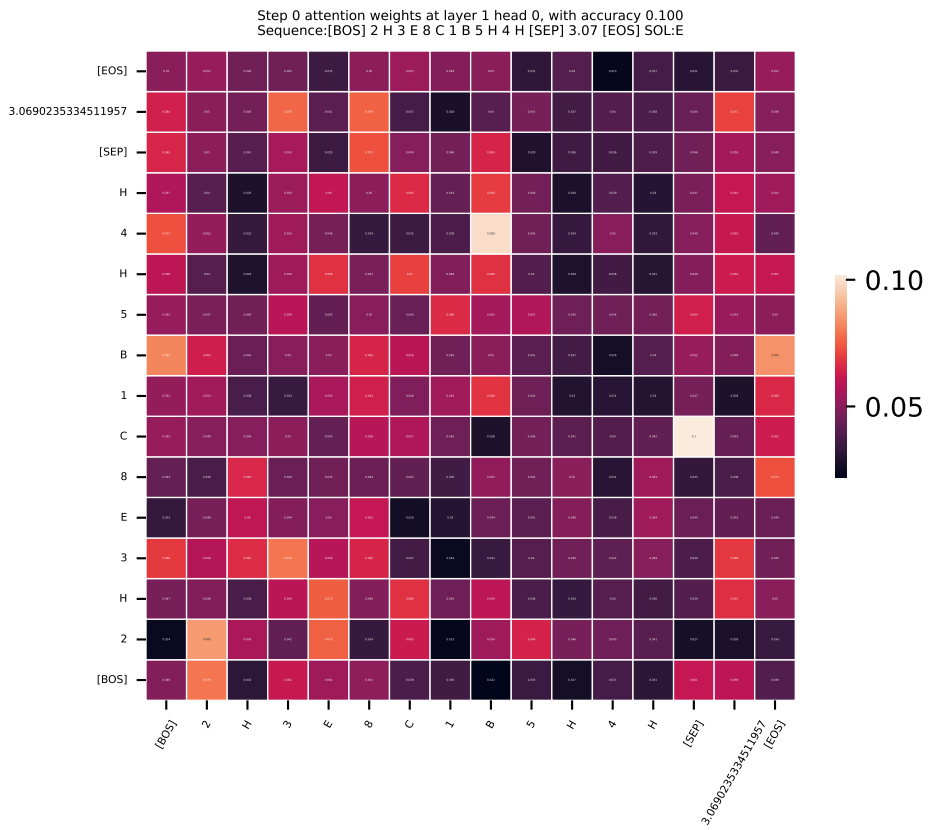

Step 0 attention weights at layer 0 head 3, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.10
0.05

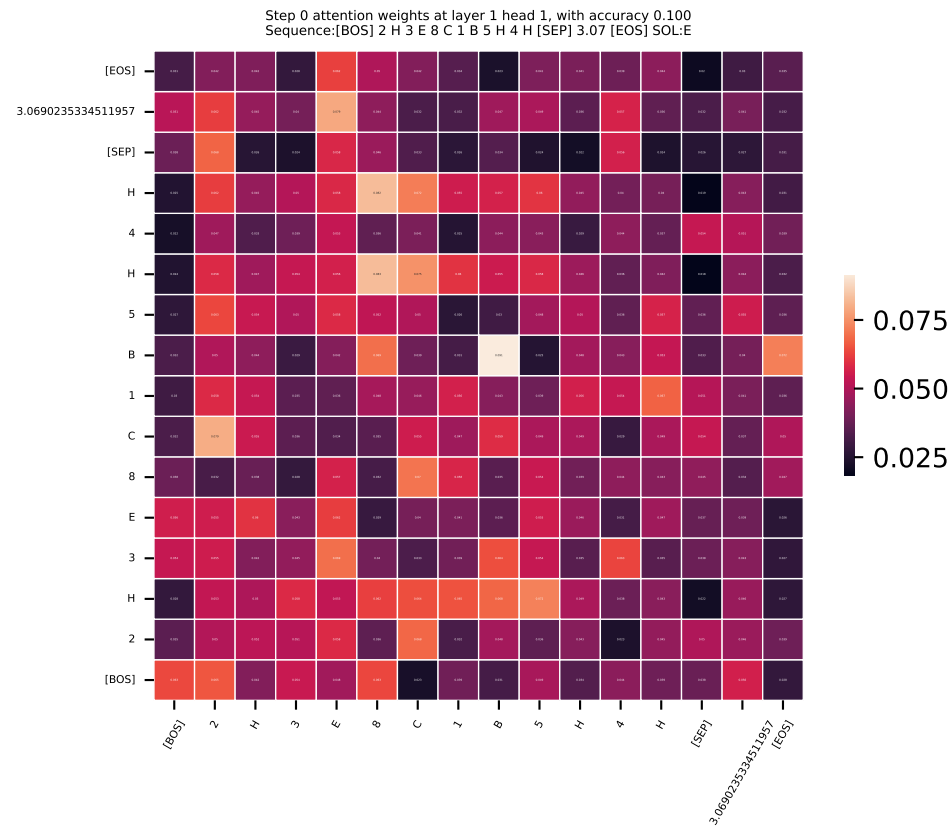

Step 0 attention weights at layer 1 head 0, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.10
0.05

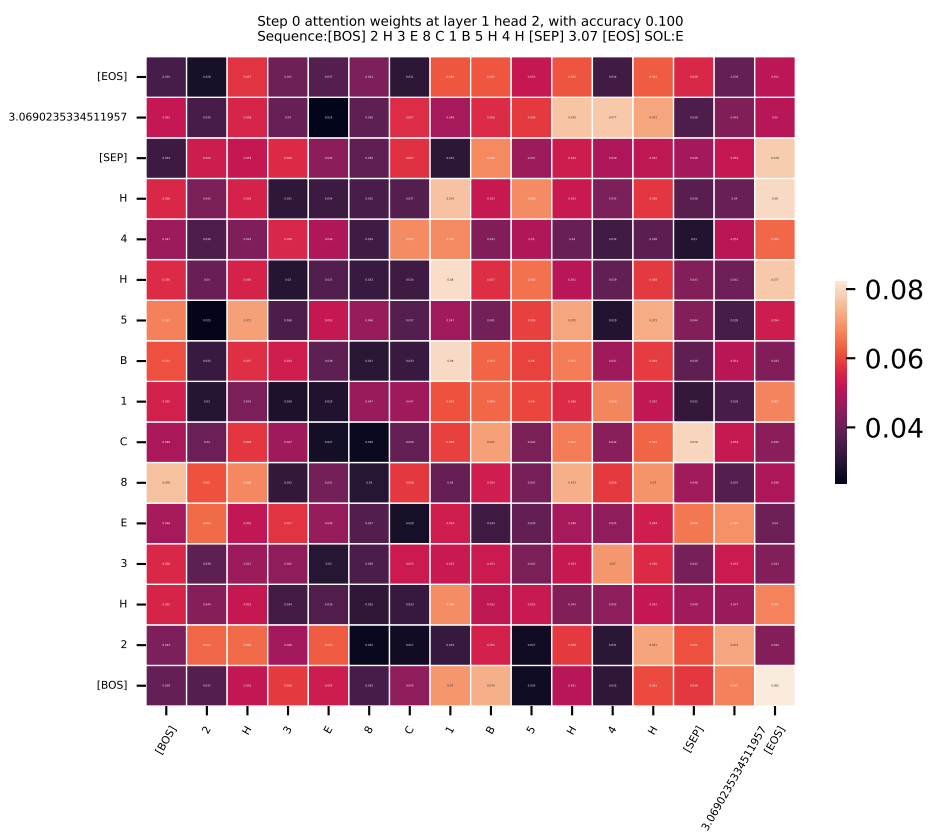

Step 0 attention weights at layer 1 head 1, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.075
0.050
0.025

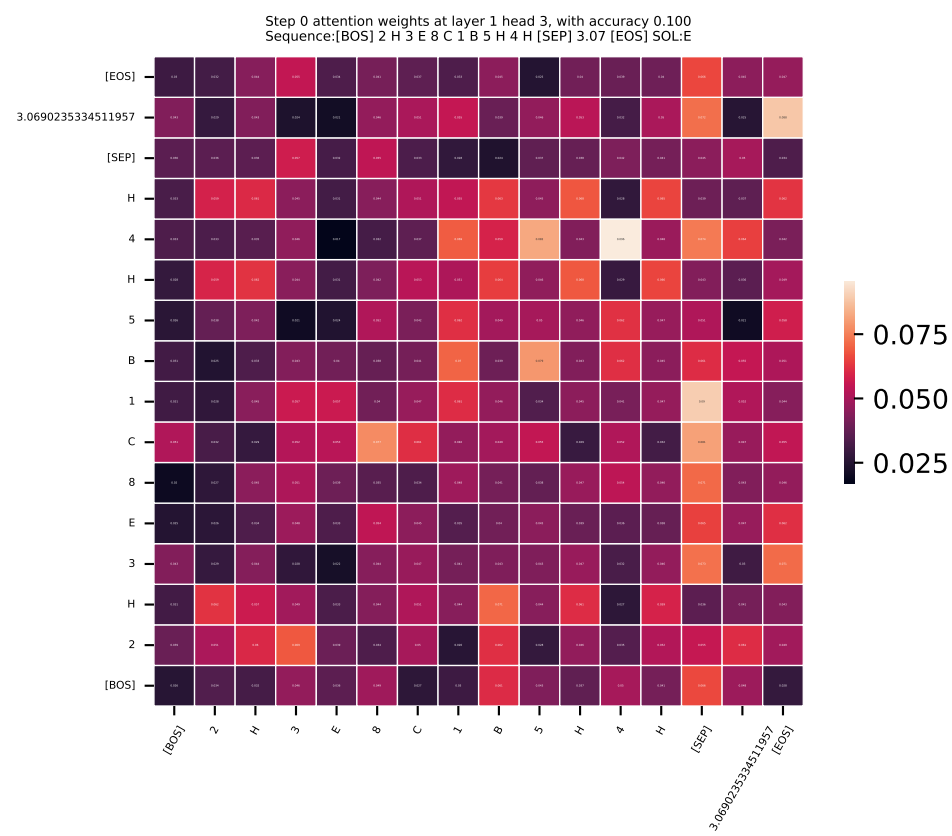

Step 0 attention weights at layer 1 head 2, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.08
0.06
0.04

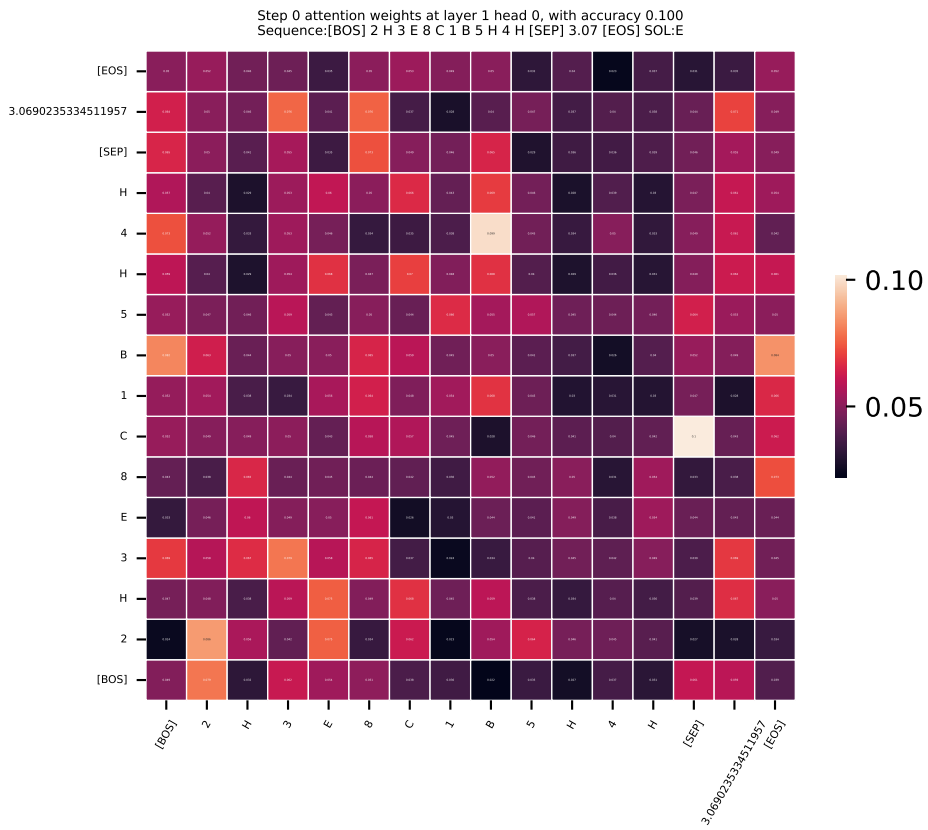

Step 0 attention weights at layer 1 head 3, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.075
0.050
0.025

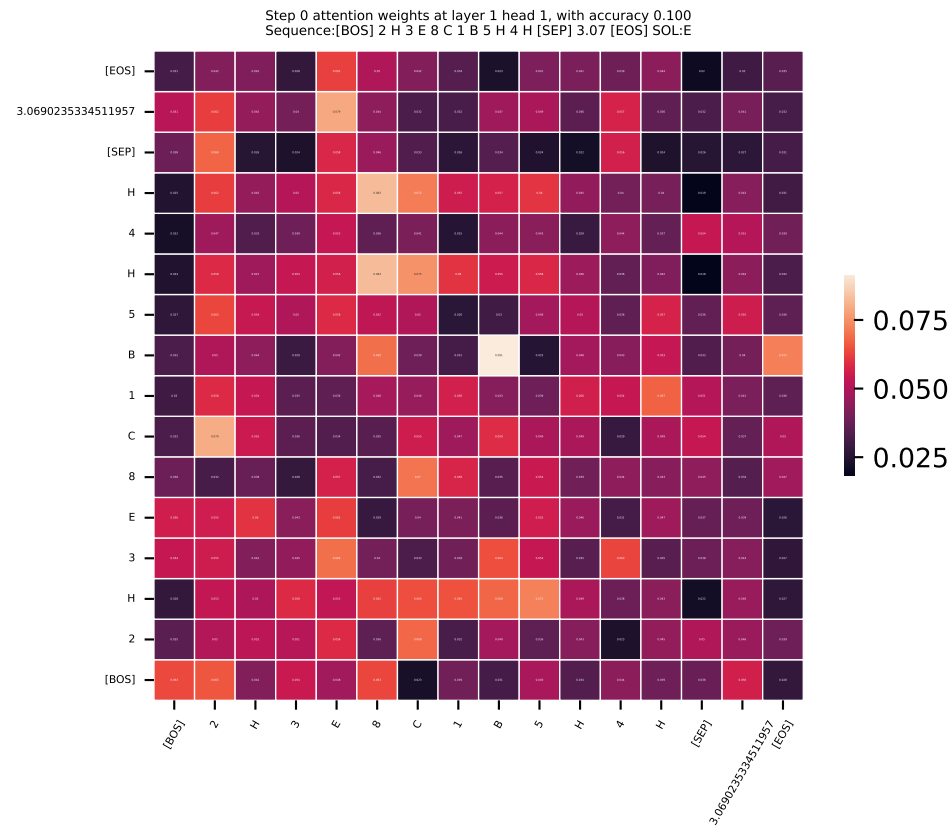

Step 0 attention weights at layer 1 head 0, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.10
0.05

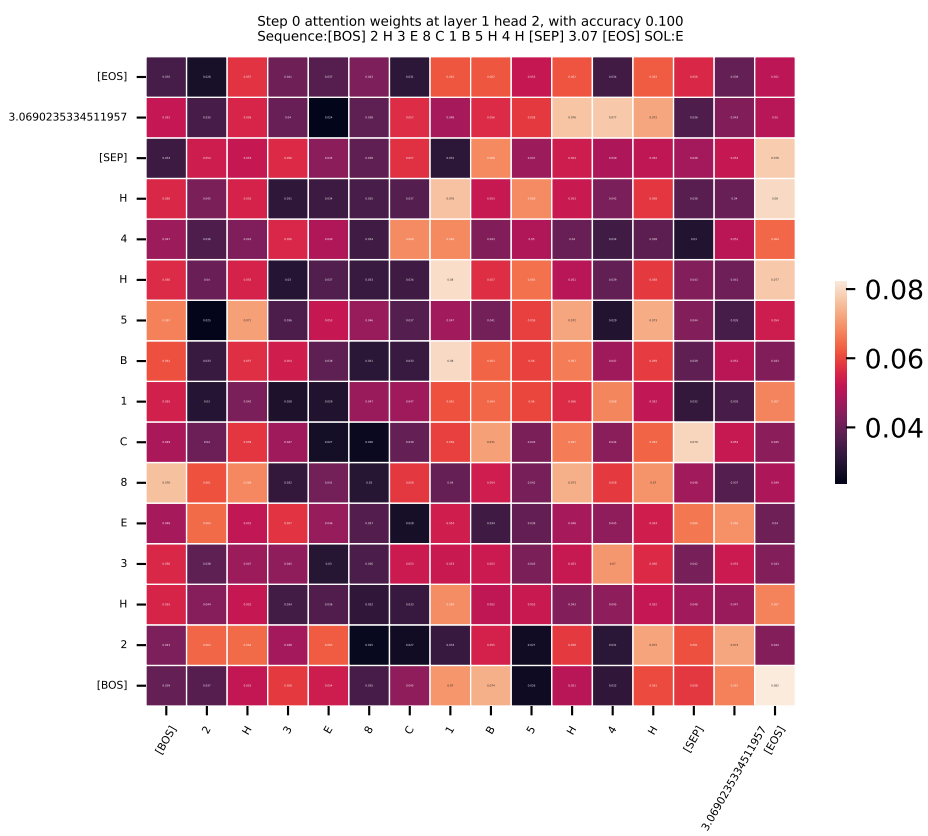

Step 0 attention weights at layer 1 head 1, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.075
0.050
0.025

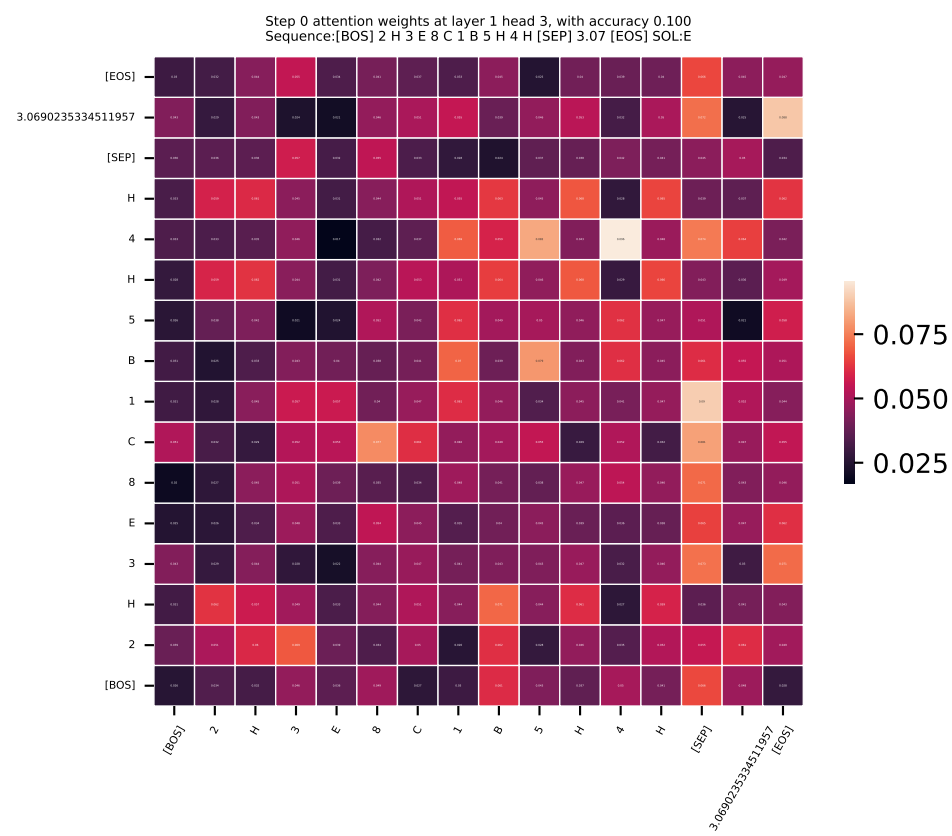

Step 0 attention weights at layer 1 head 2, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.08
0.06
0.04

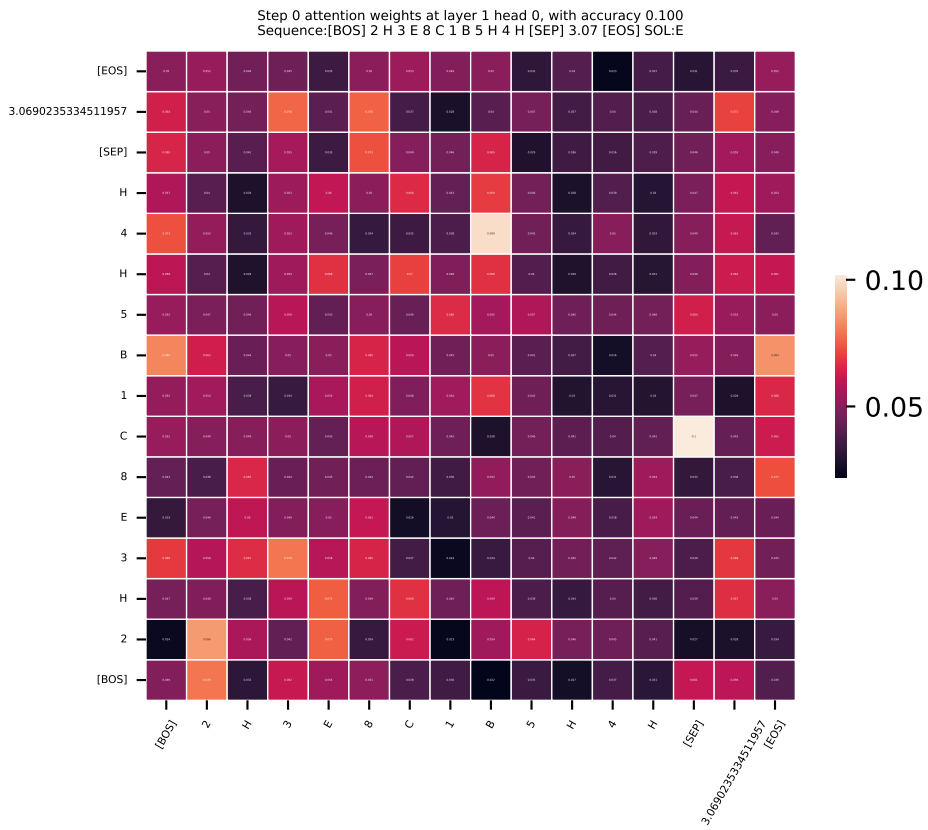

Step 0 attention weights at layer 1 head 3, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.075
0.050
0.025

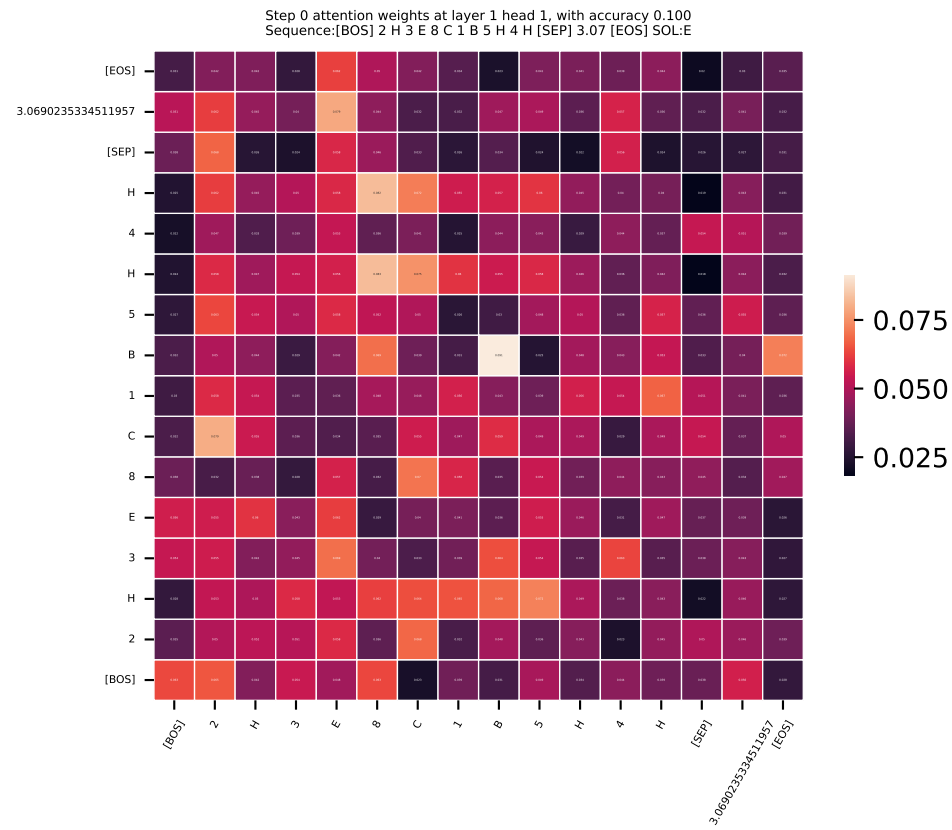

Step 0 attention weights at layer 1 head 0, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.10
0.05

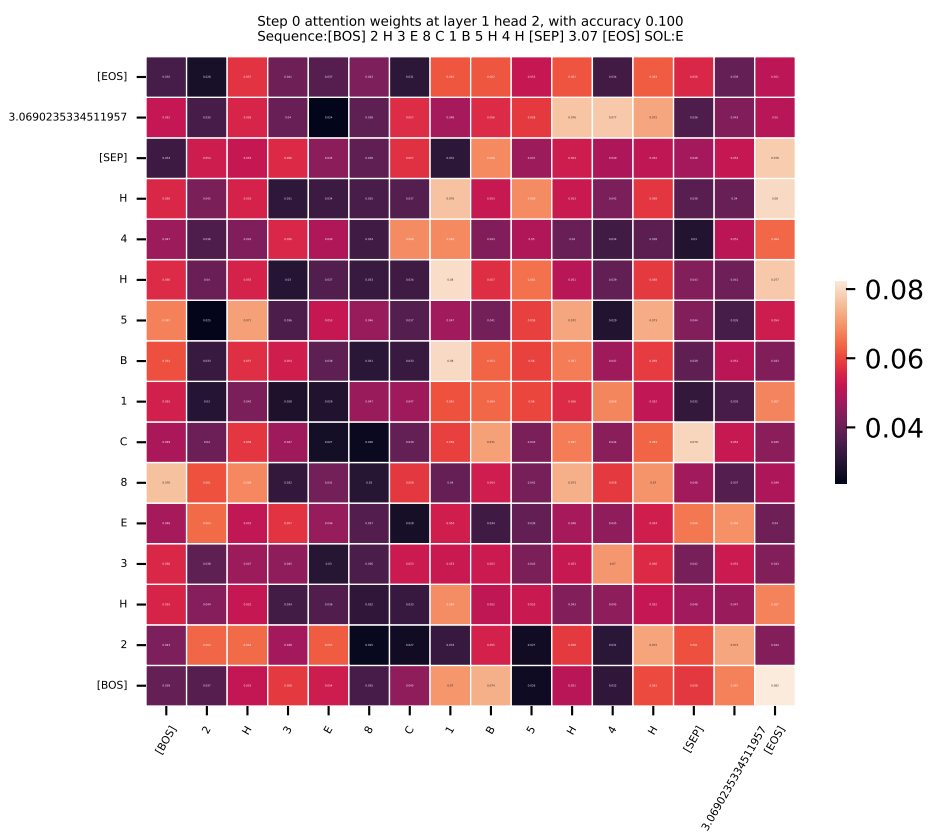

Step 0 attention weights at layer 1 head 1, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.075
0.050
0.025

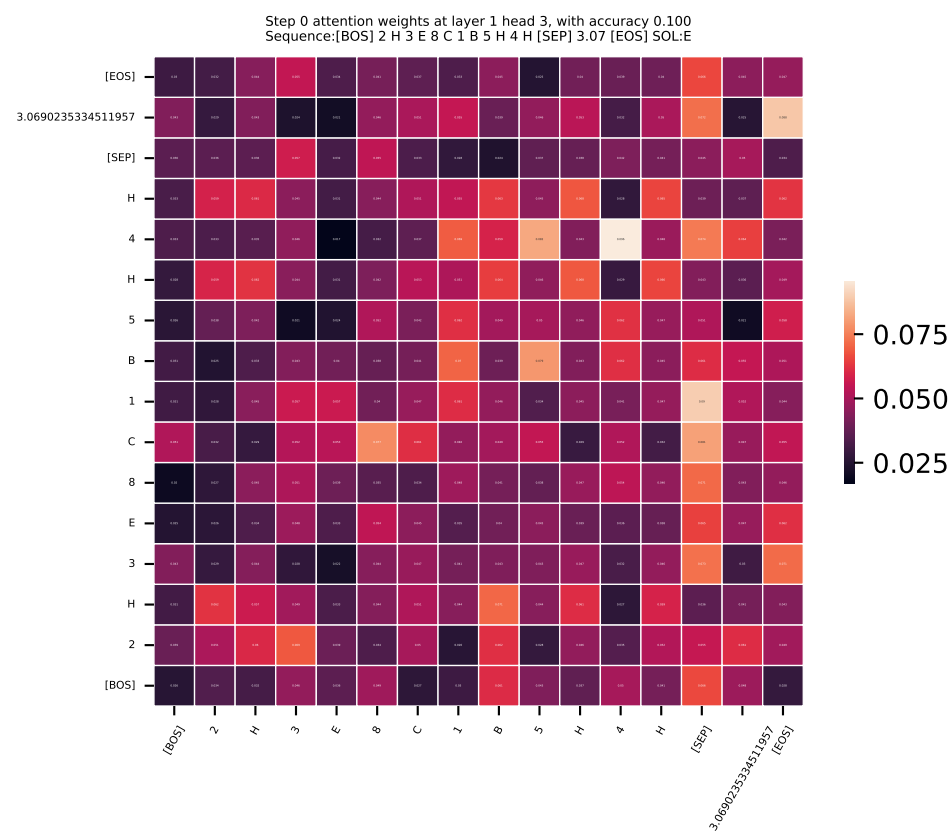

Step 0 attention weights at layer 1 head 2, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.08
0.06
0.04

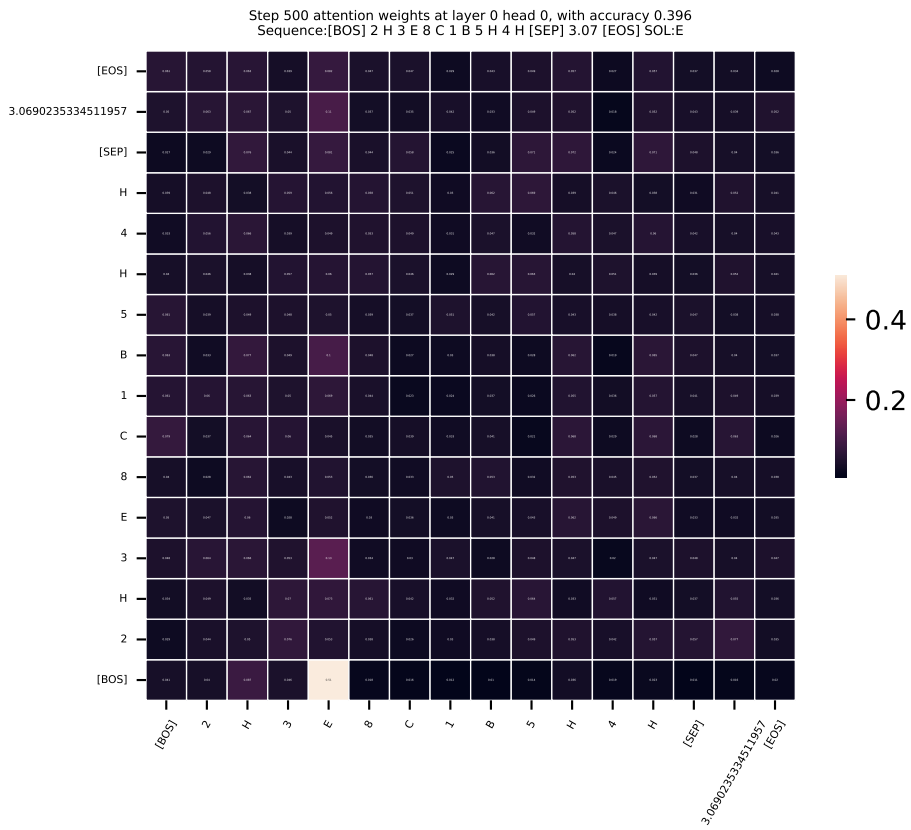

Step 0 attention weights at layer 1 head 3, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.075
0.050
0.025

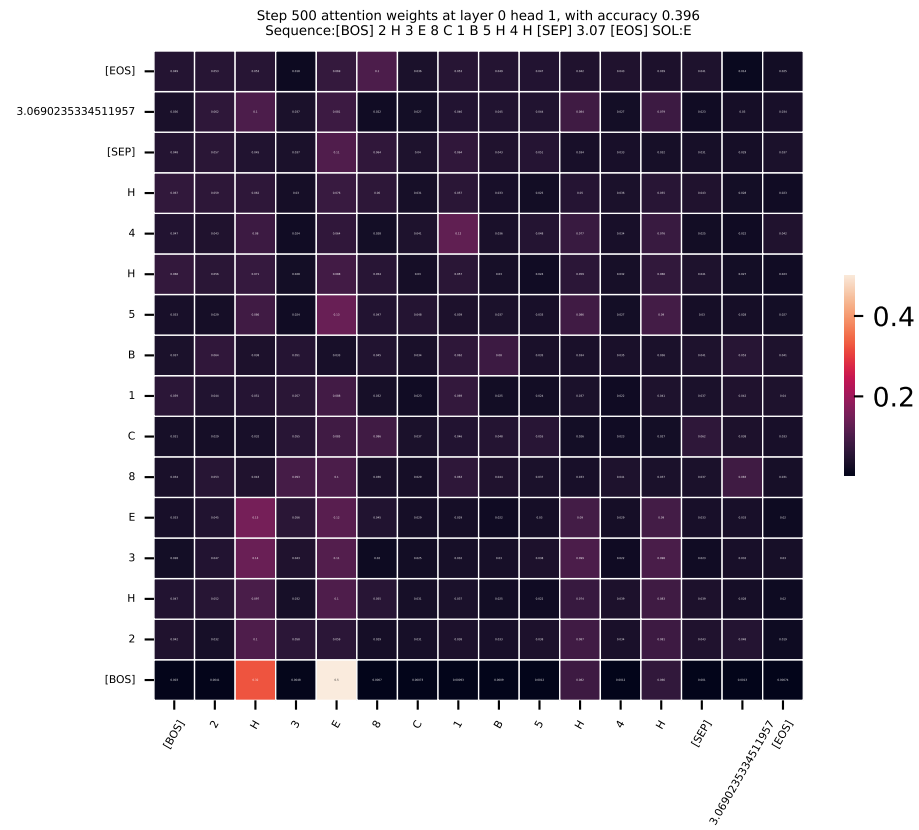

Step 0 attention weights at layer 1 head 0, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.10
0.05

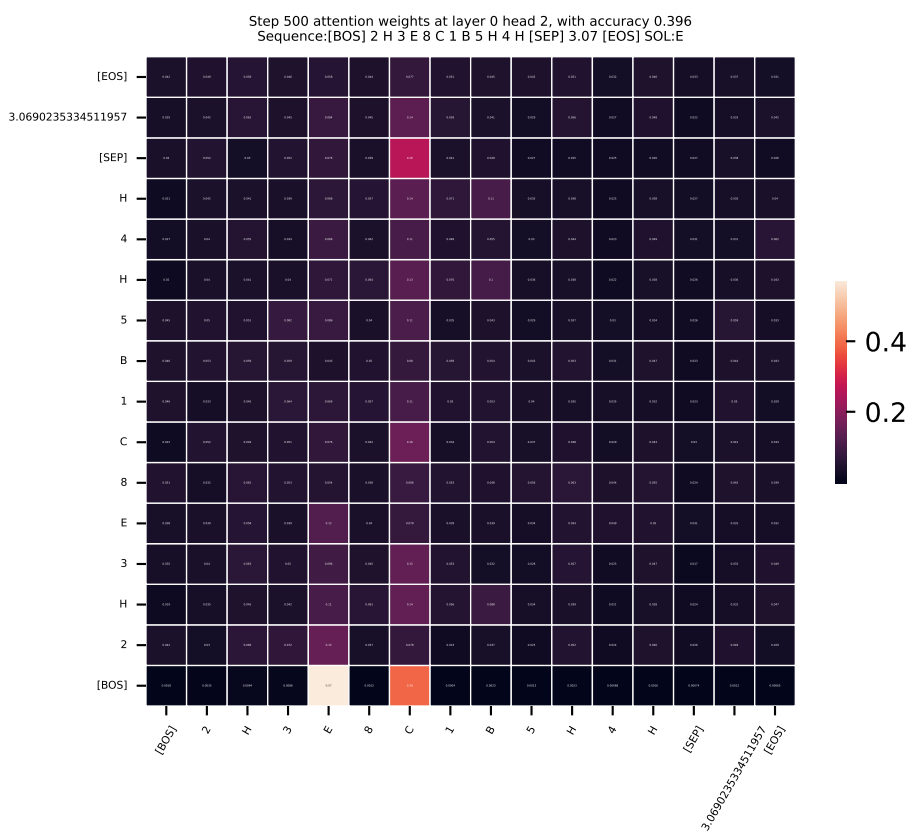

Step 0 attention weights at layer 1 head 1, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.075
0.050
0.025

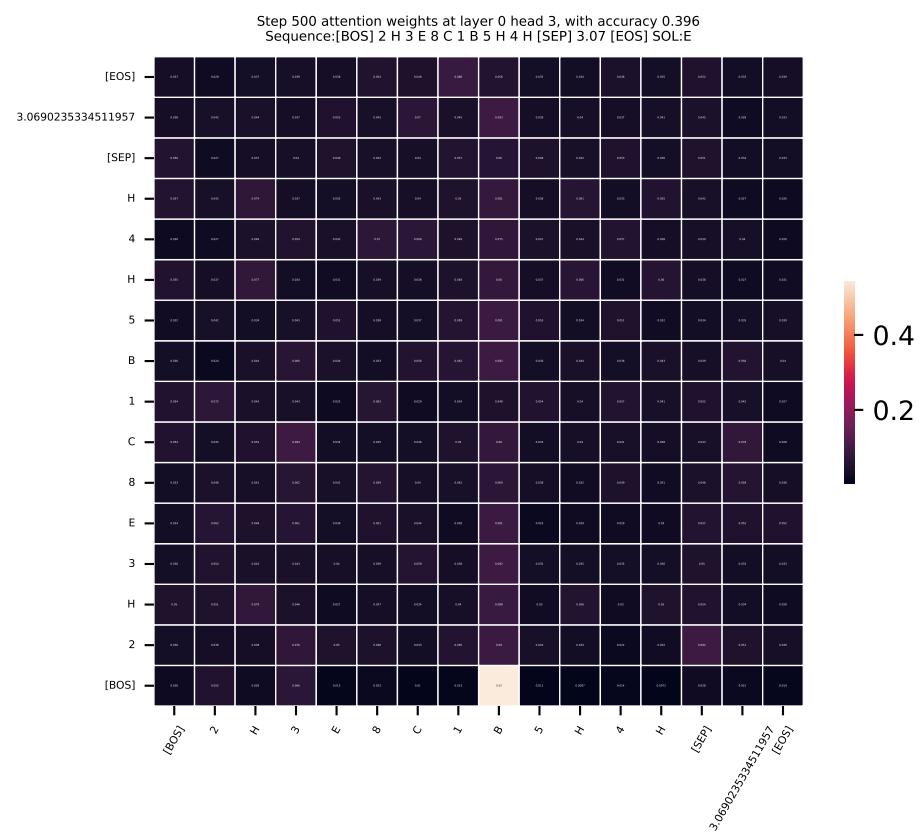

Step 0 attention weights at layer 1 head 2, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.08
0.06
0.04

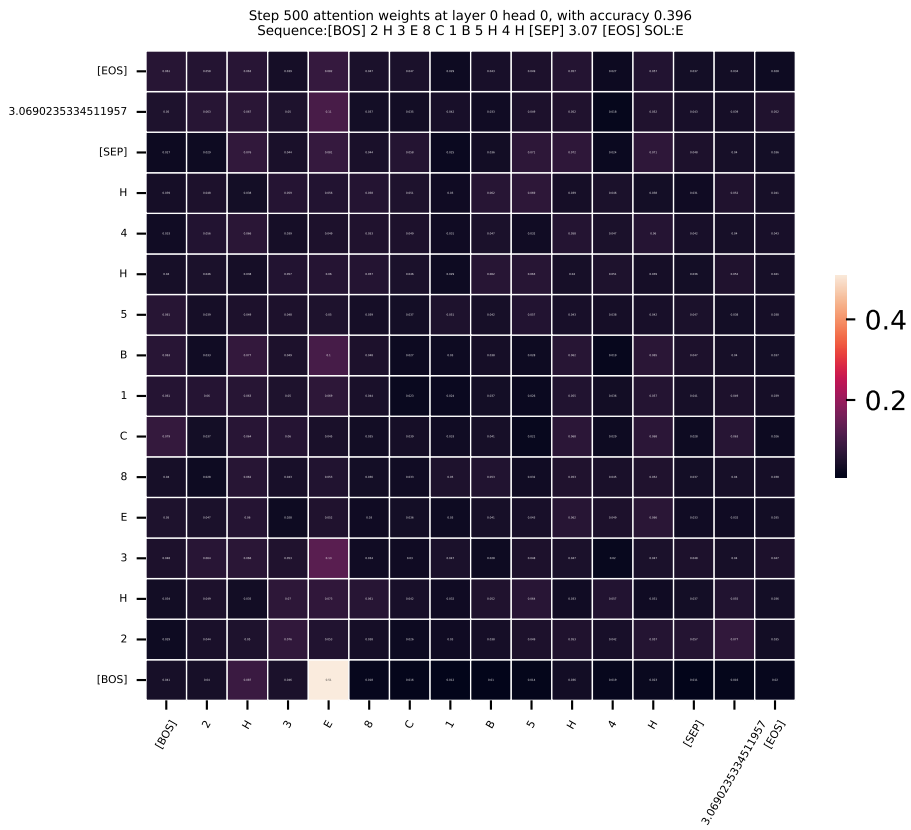

Step 0 attention weights at layer 1 head 3, with accuracy 0.100
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.075
0.050
0.025

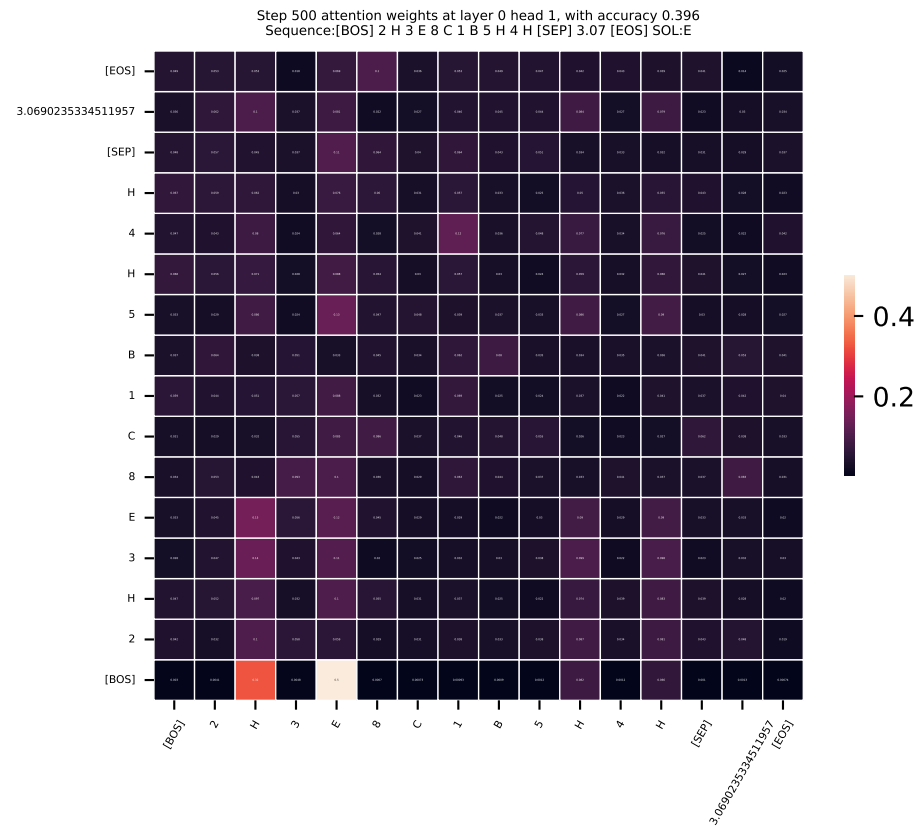

Step 500 attention weights at layer 0 head 0, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.4
0.2

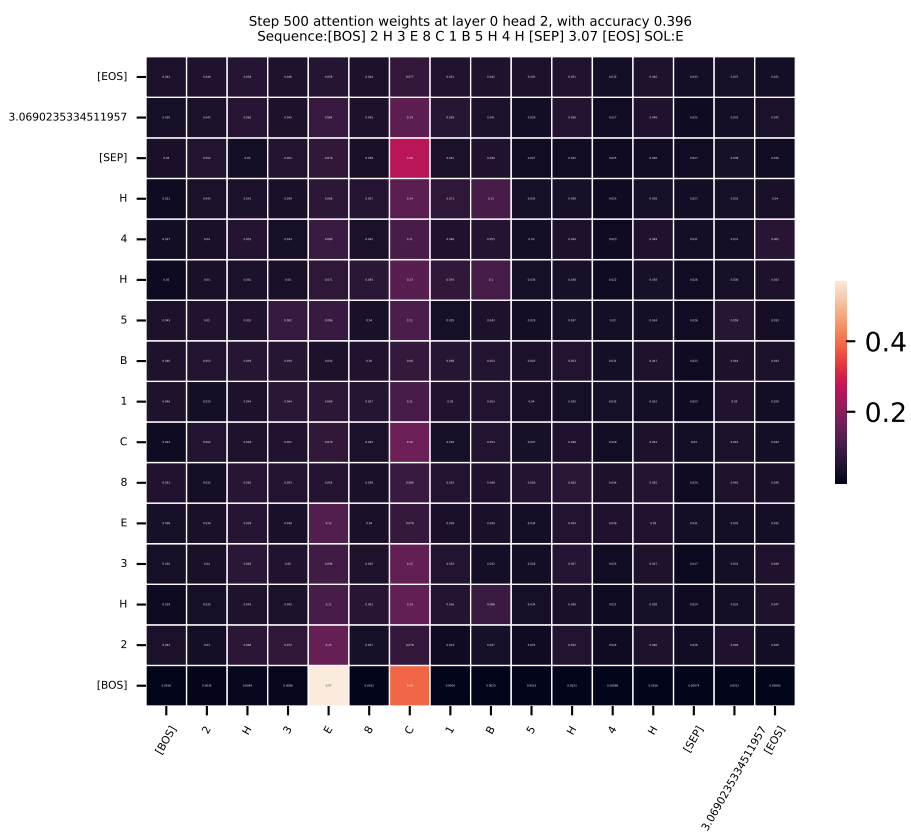

Step 500 attention weights at layer 0 head 1, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.4
0.2

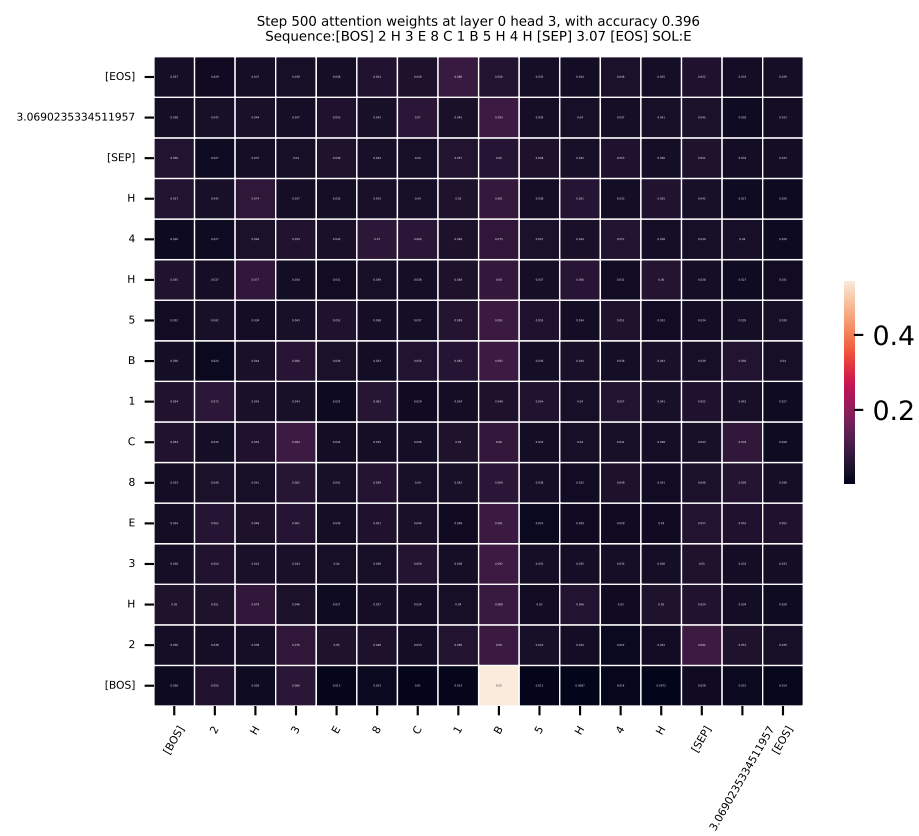

Step 500 attention weights at layer 0 head 2, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.4
0.2

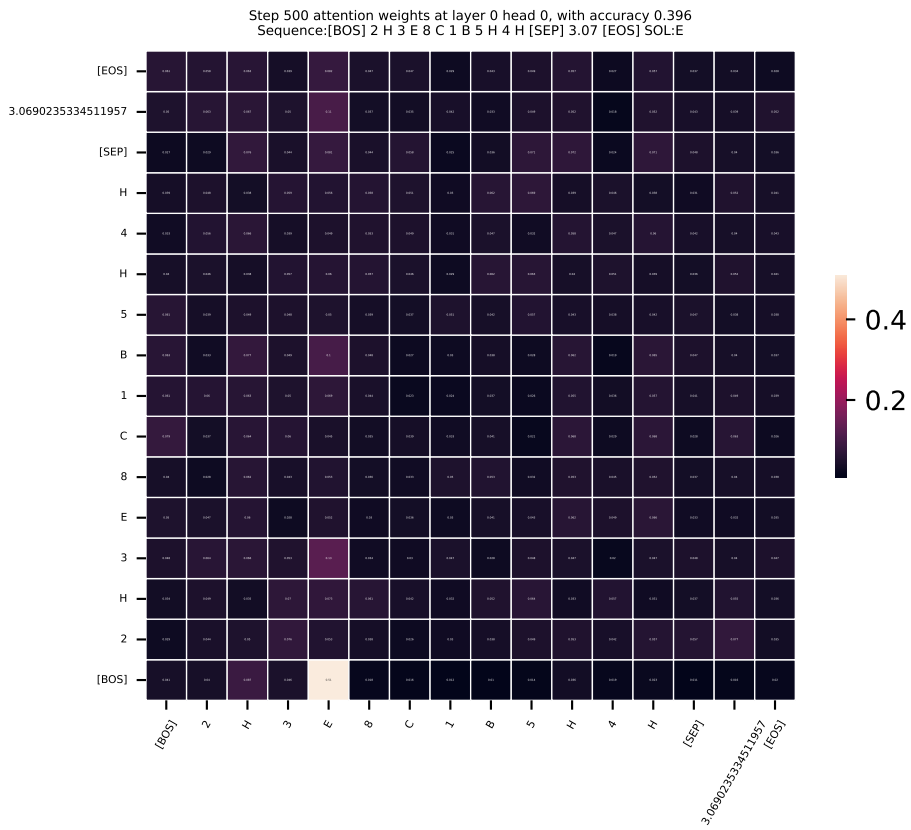

Step 500 attention weights at layer 0 head 3, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.4
0.2

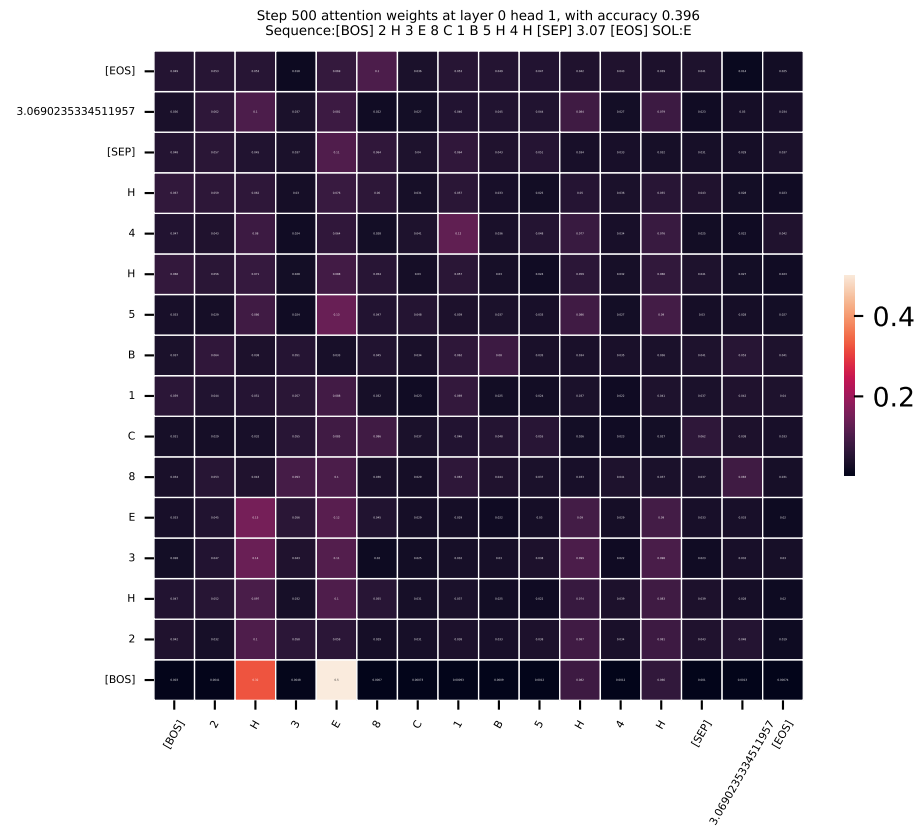

Step 500 attention weights at layer 0 head 0, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.4
0.2

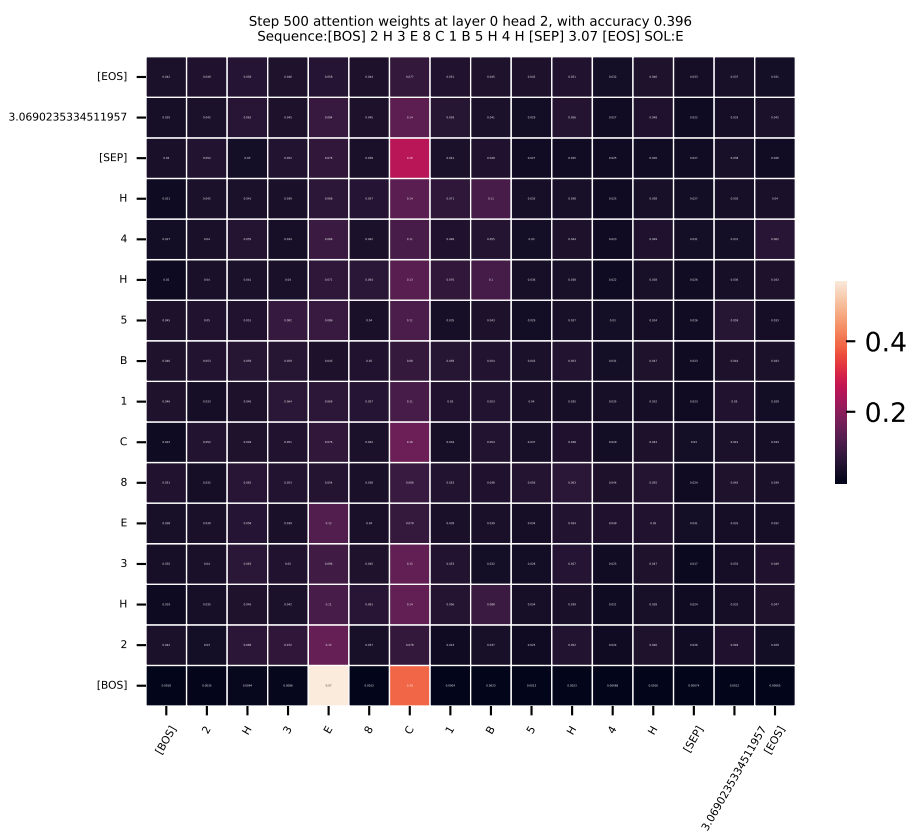

Step 500 attention weights at layer 0 head 1, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.4
0.2

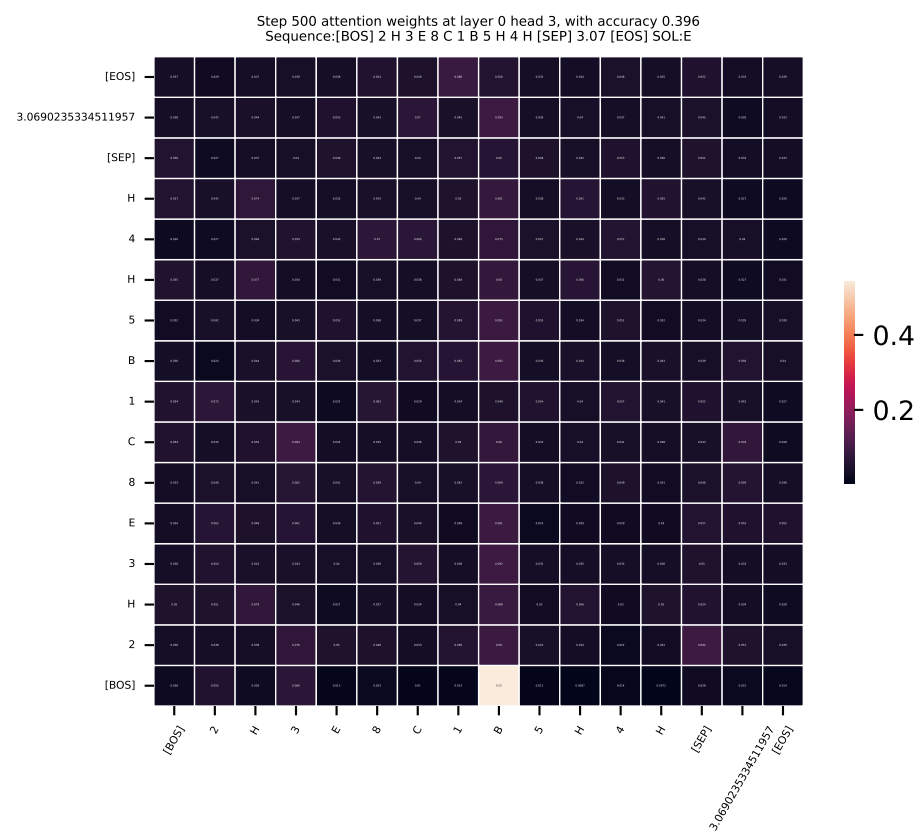

Step 500 attention weights at layer 0 head 2, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.4
0.2

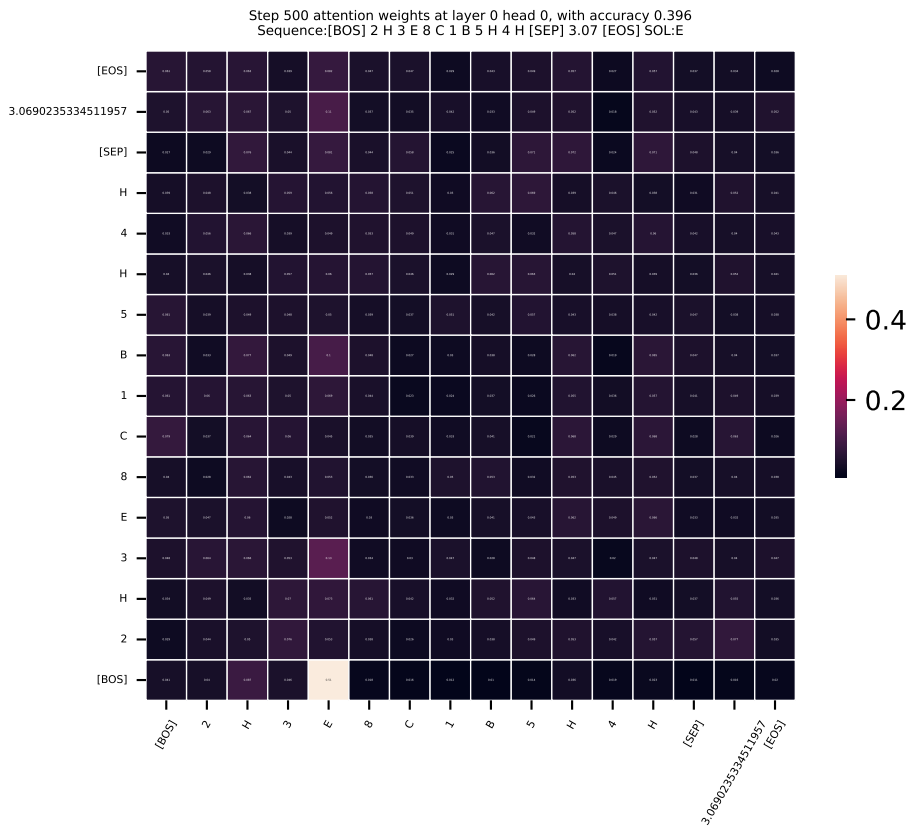

Step 500 attention weights at layer 0 head 3, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.4
0.2

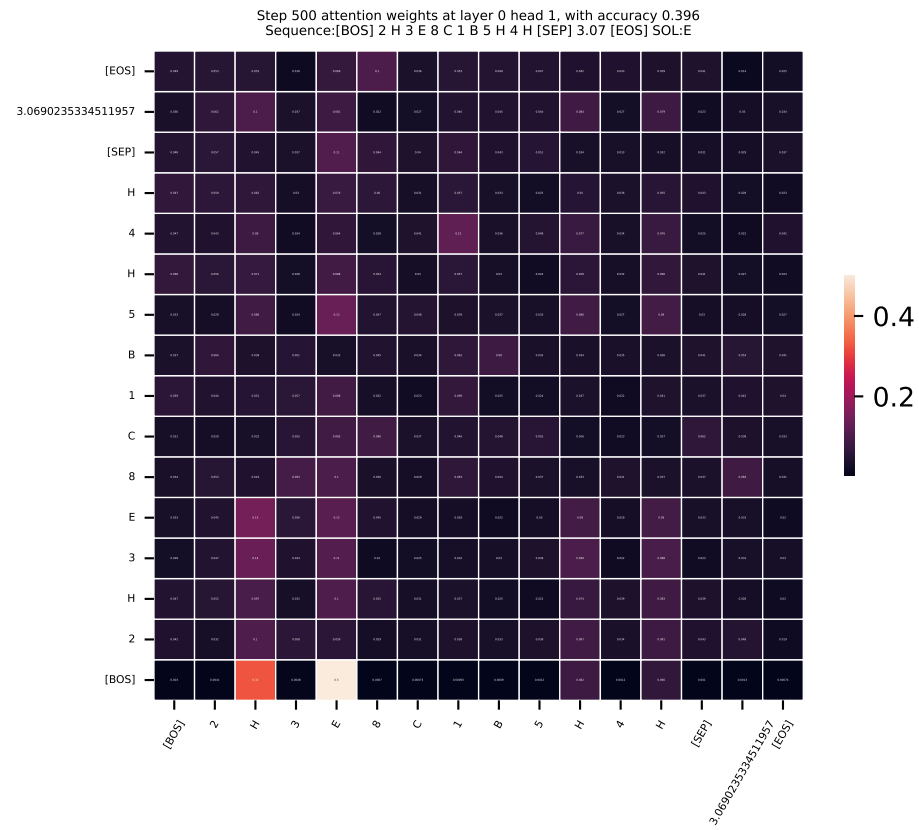

Step 500 attention weights at layer 0 head 0, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.4
0.2

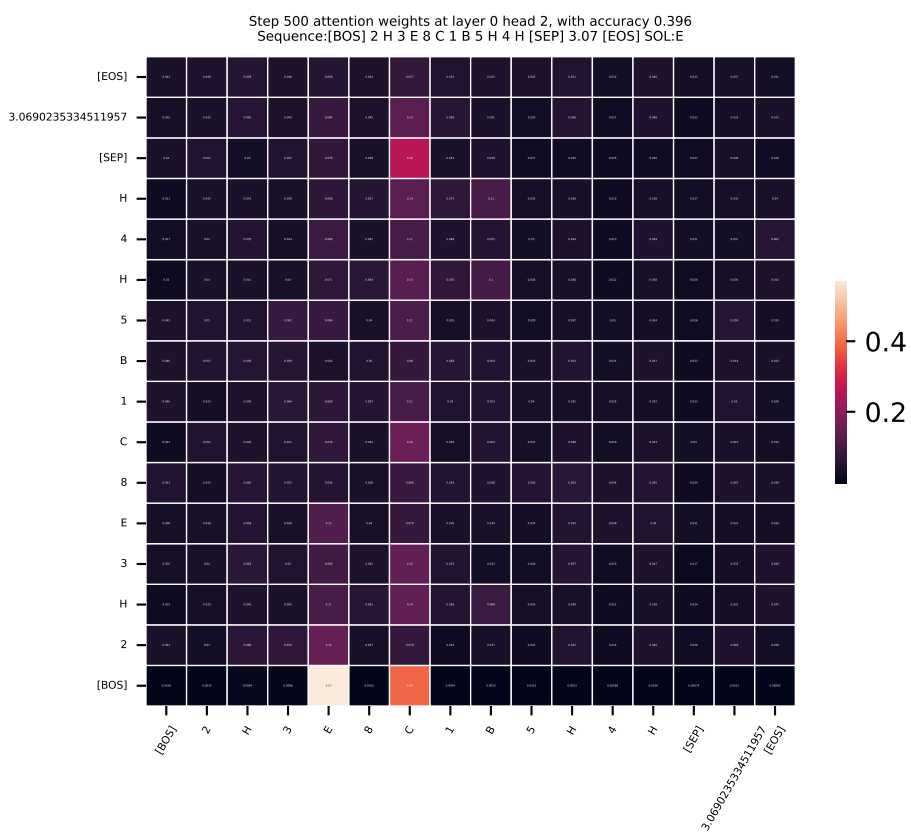

Step 500 attention weights at layer 0 head 1, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.4
0.2

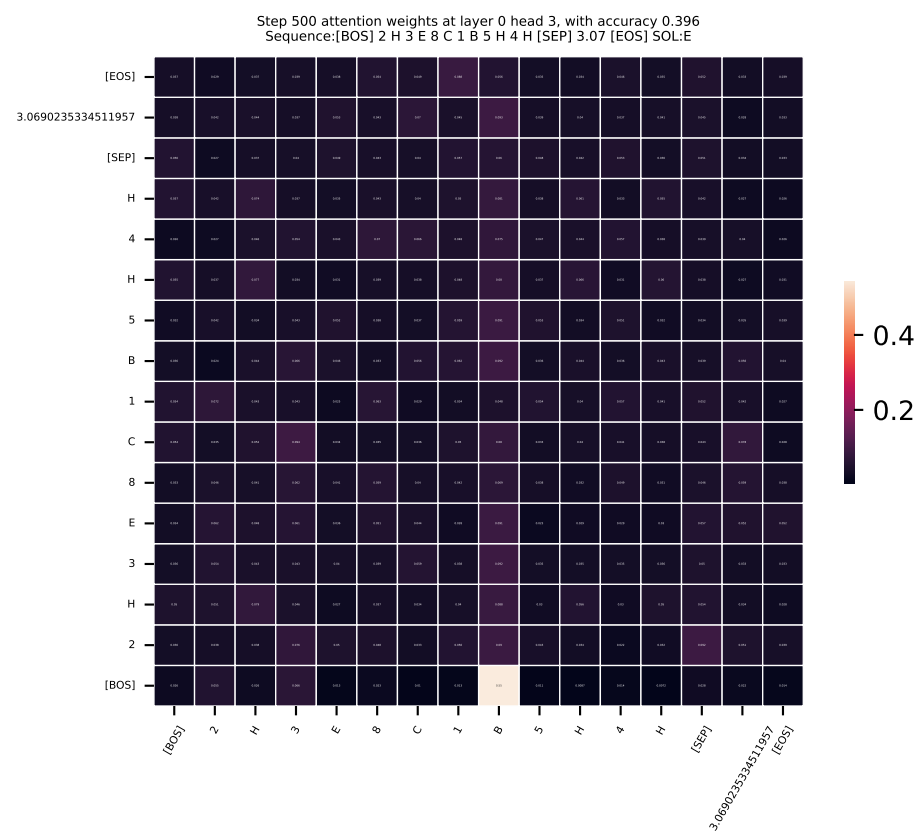

Step 500 attention weights at layer 0 head 2, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.4
0.2

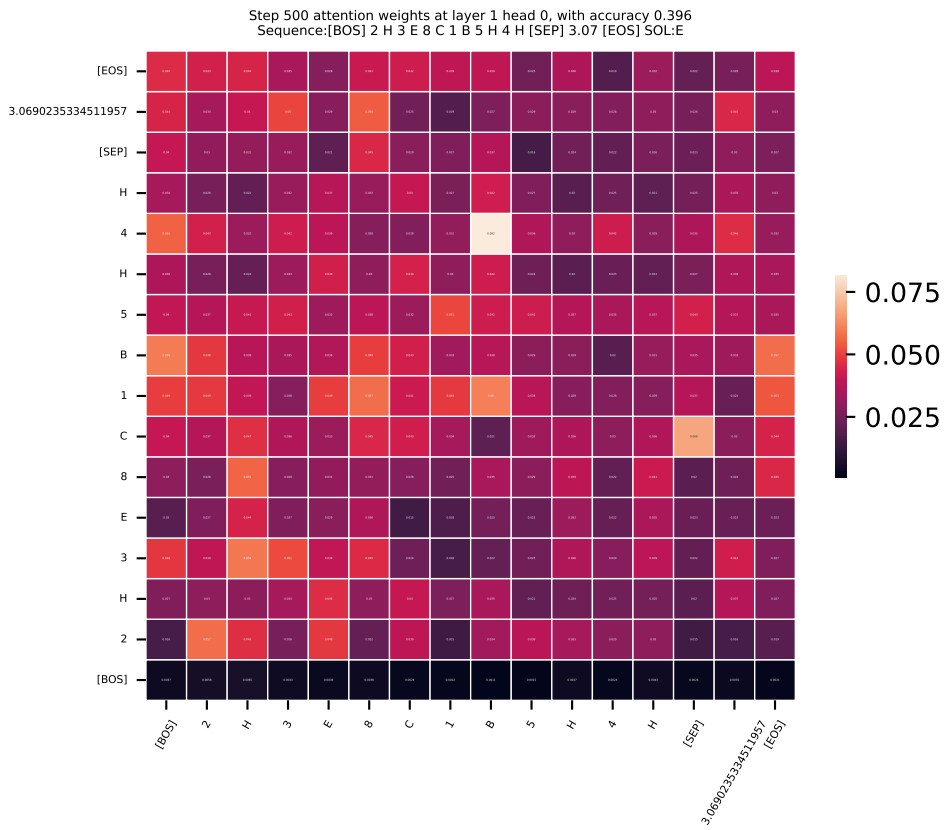

Step 500 attention weights at layer 0 head 3, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.4
0.2

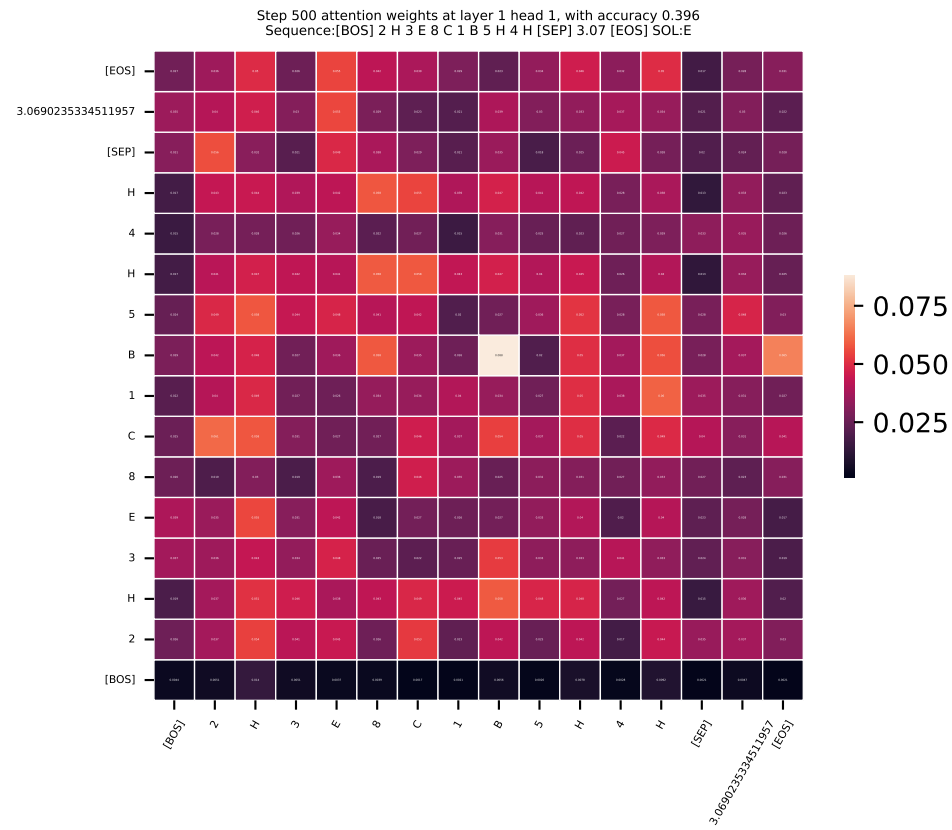

Step 500 attention weights at layer 0 head 0, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.4
0.2

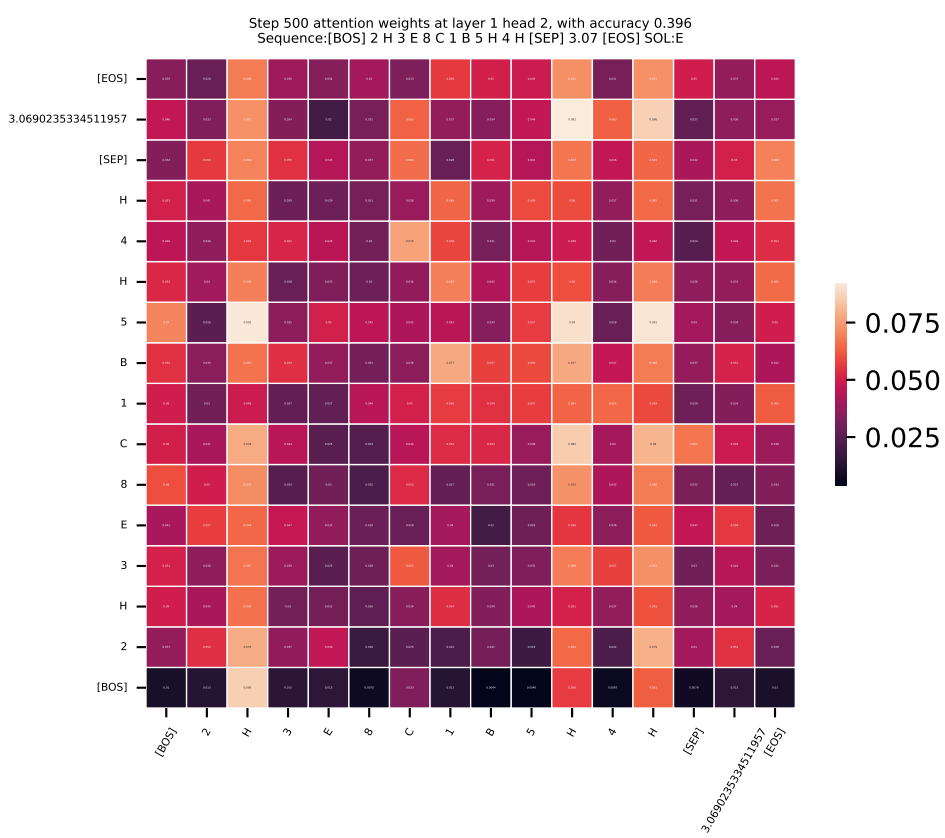

Step 500 attention weights at layer 0 head 1, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.4
0.2

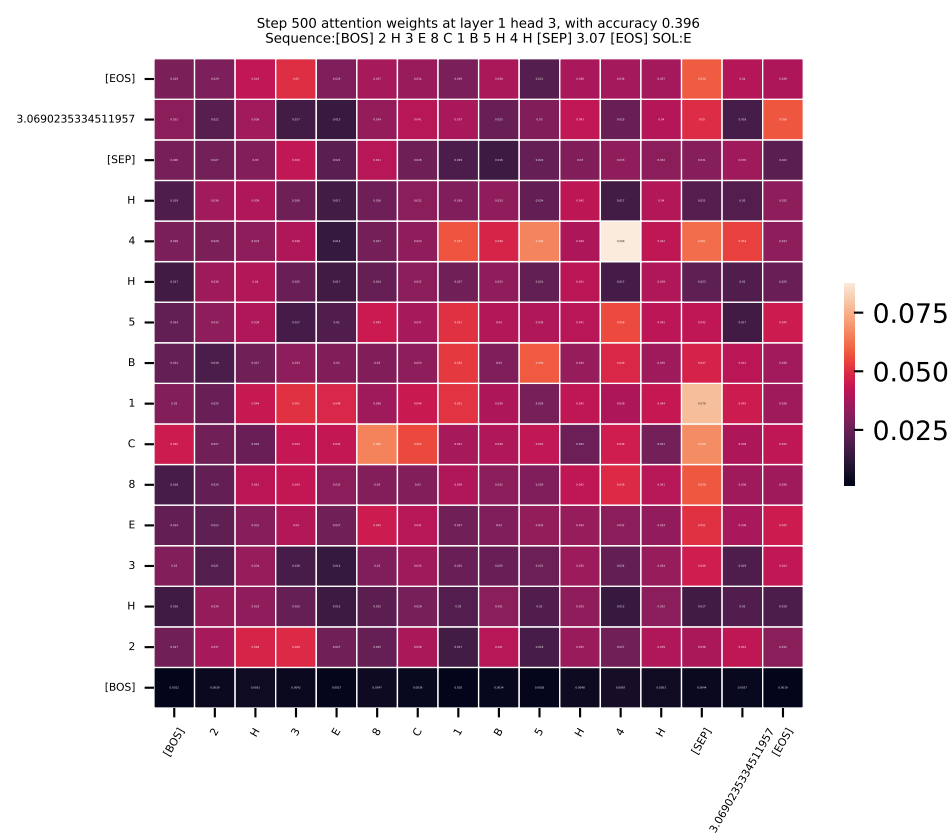

Step 500 attention weights at layer 0 head 2, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.4
0.2

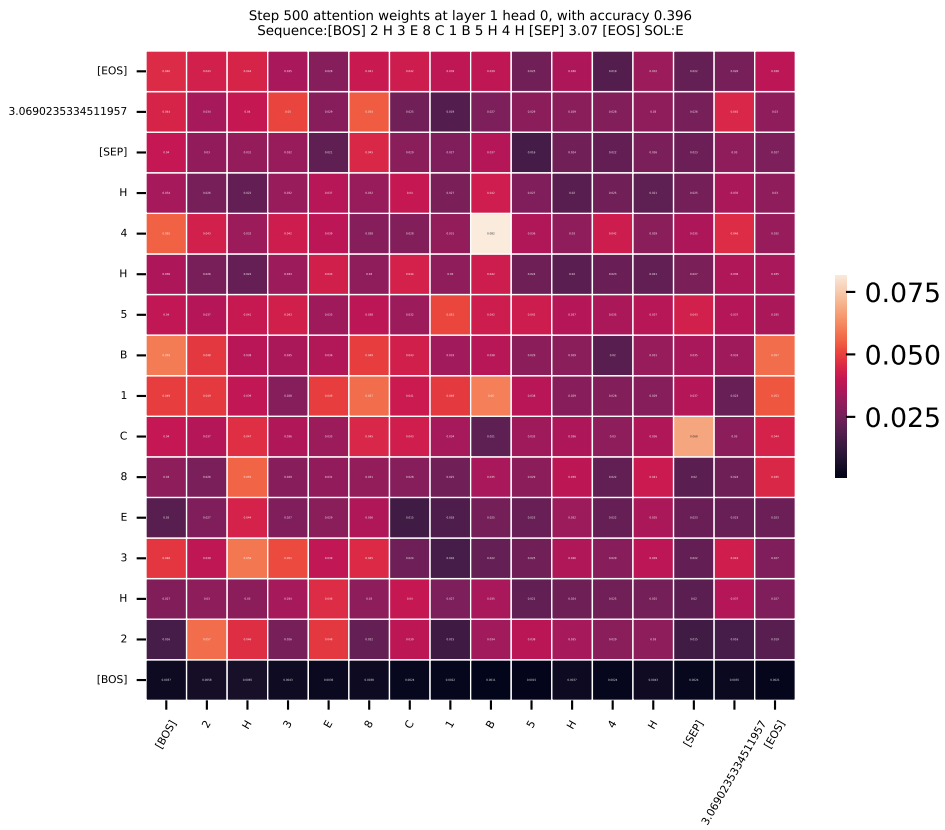

Step 500 attention weights at layer 0 head 3, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.4
0.2

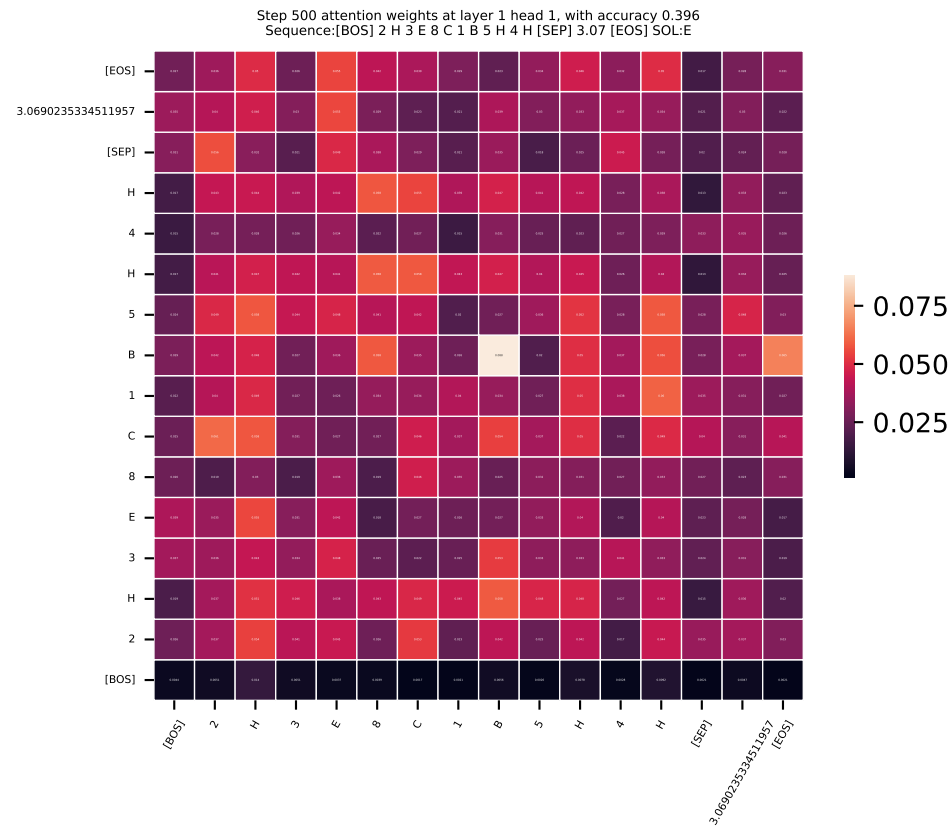

Step 500 attention weights at layer 1 head 0, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.075
0.050
0.025

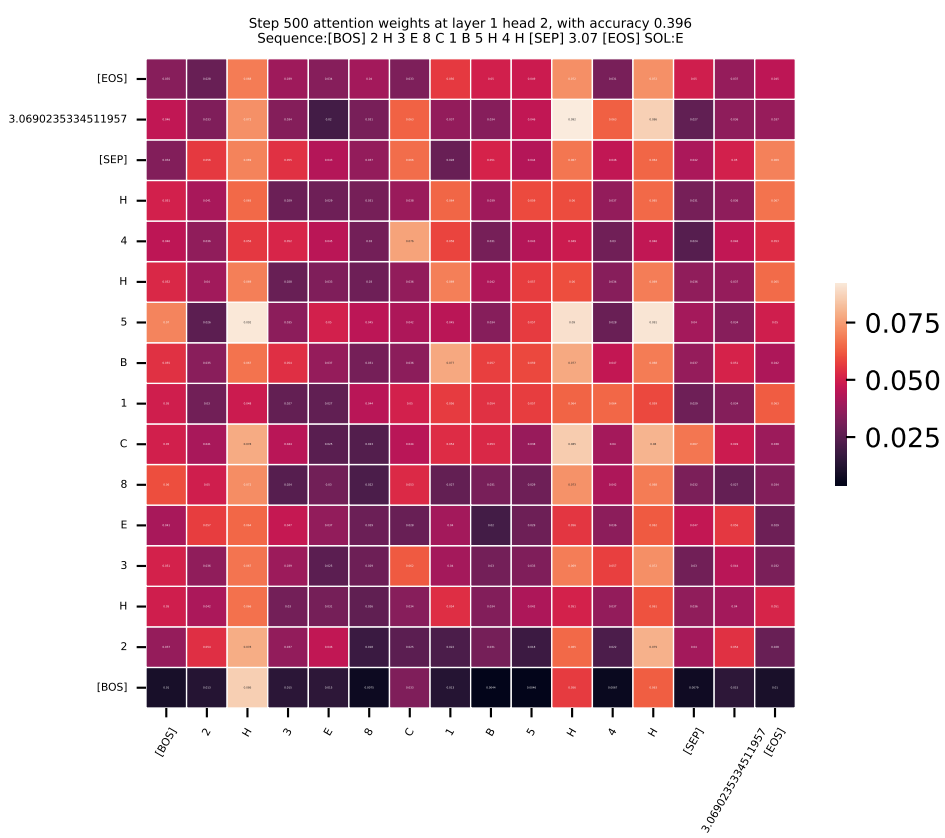

Step 500 attention weights at layer 1 head 1, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.075
0.050
0.025

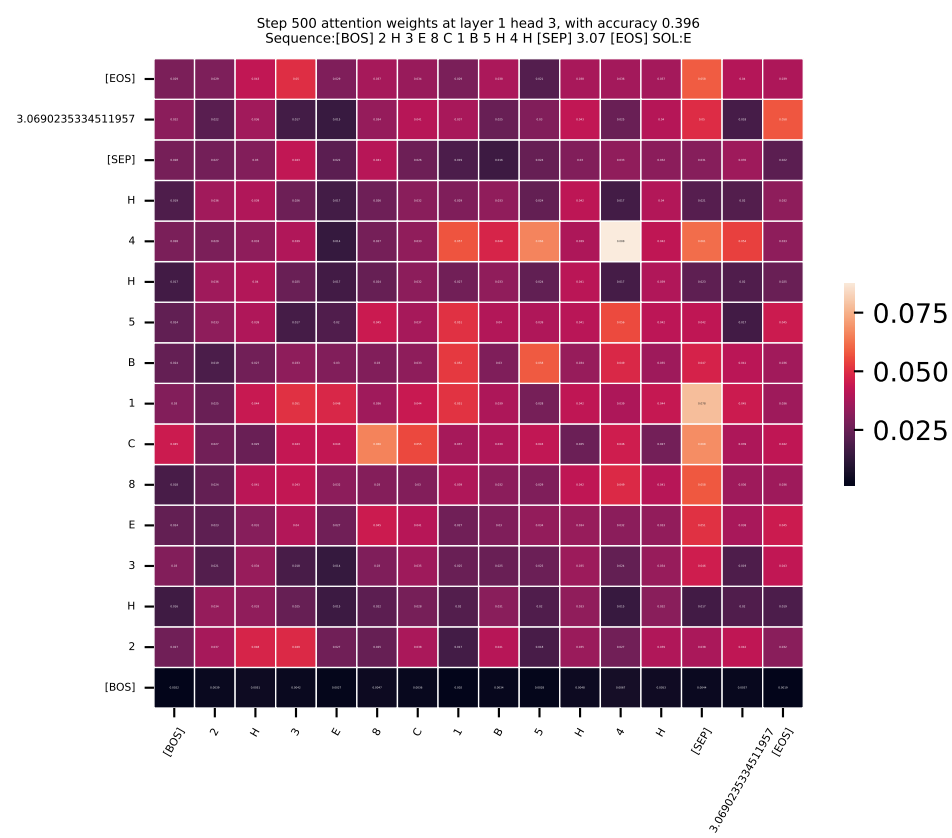

Step 500 attention weights at layer 1 head 2, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.075
0.050
0.025

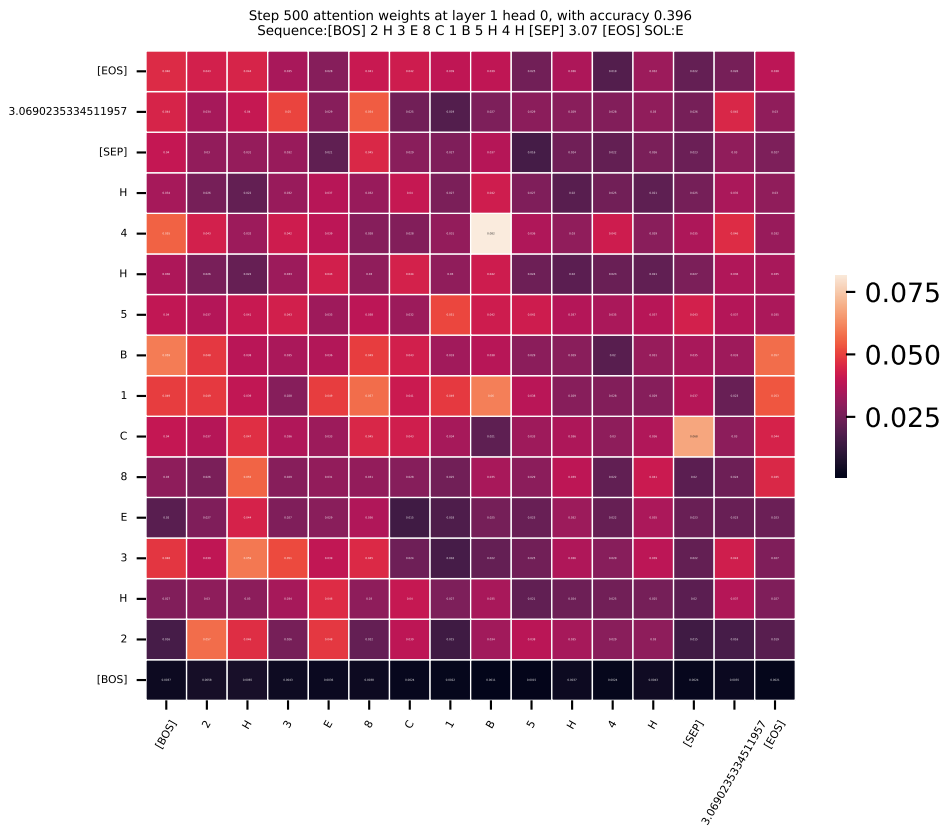

Step 500 attention weights at layer 1 head 3, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.075
0.050
0.025

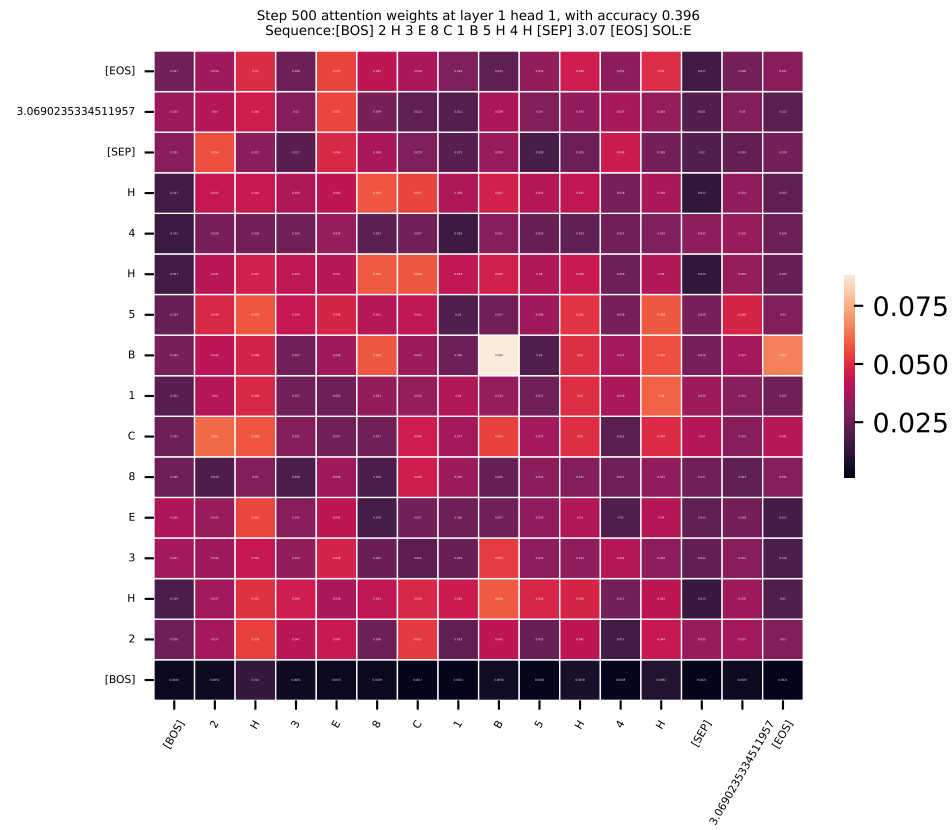

Step 500 attention weights at layer 1 head 0, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.075
0.050
0.025

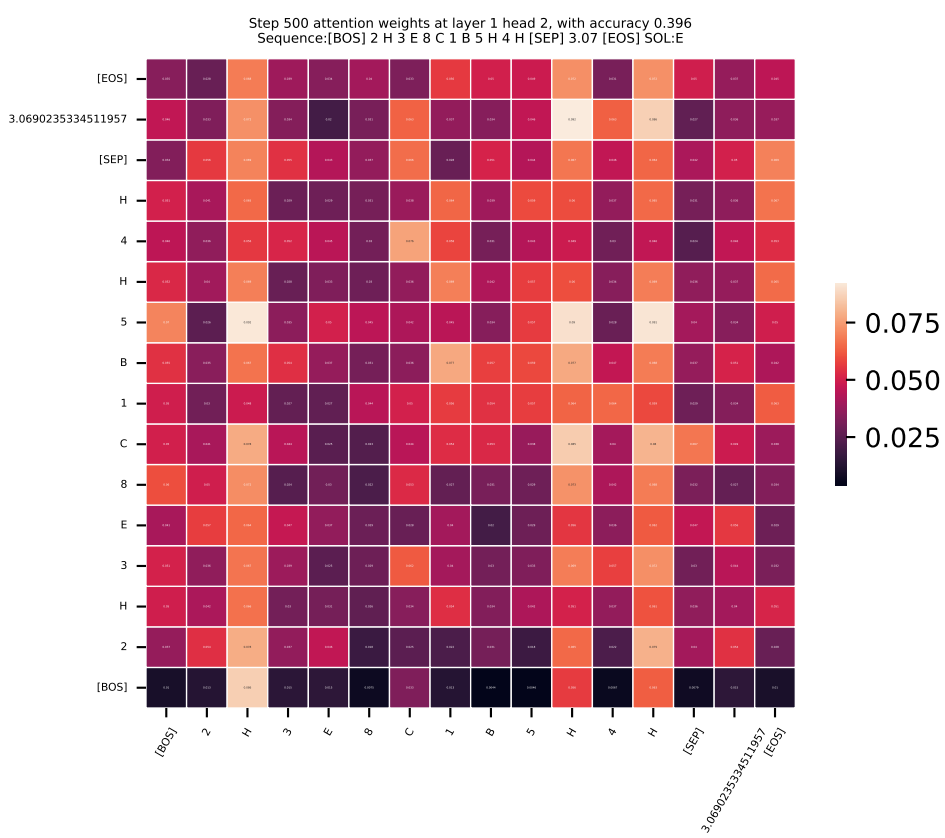

Step 500 attention weights at layer 1 head 1, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.075
0.050
0.025

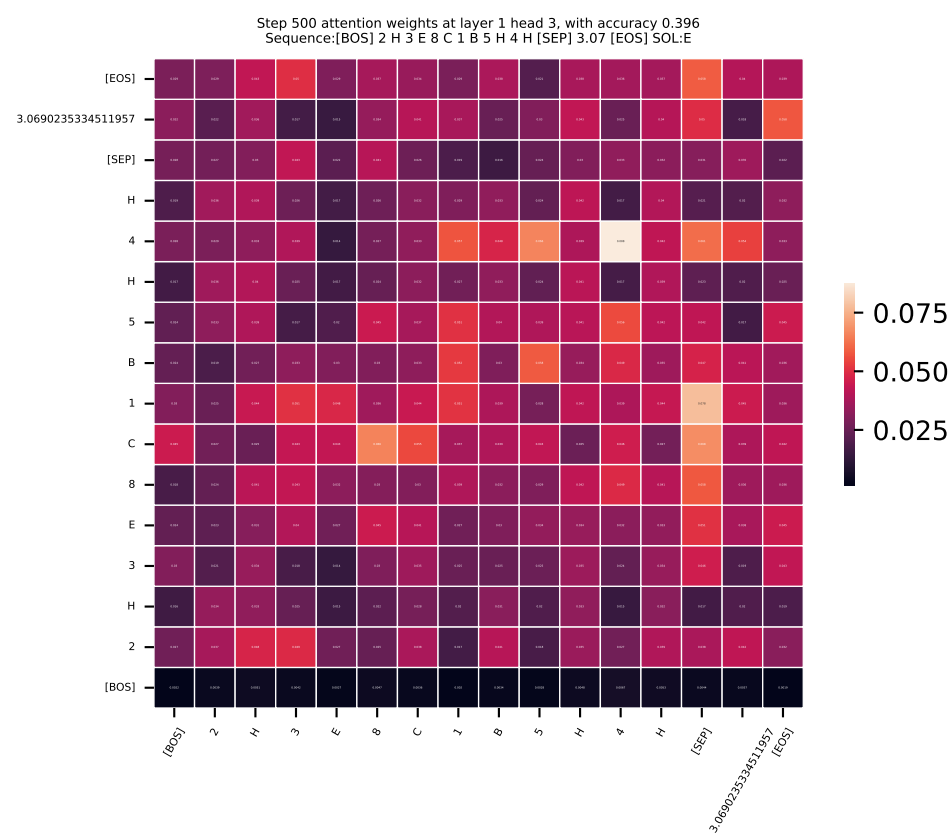

Step 500 attention weights at layer 1 head 2, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.075
0.050
0.025

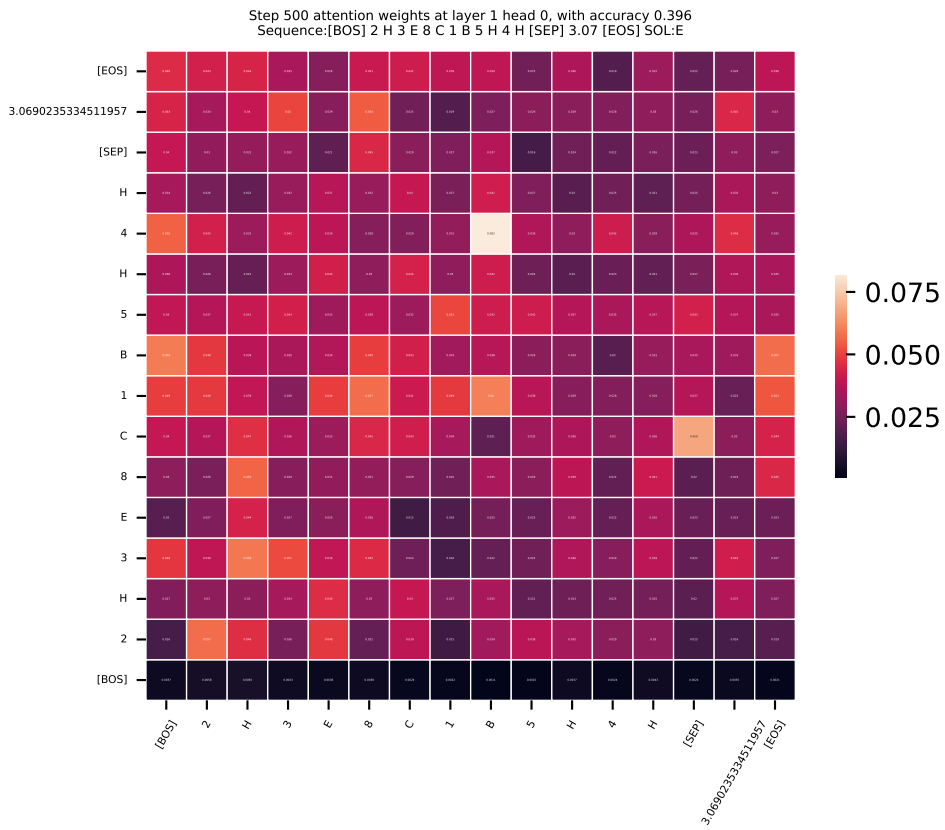

Step 500 attention weights at layer 1 head 3, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.075
0.050
0.025

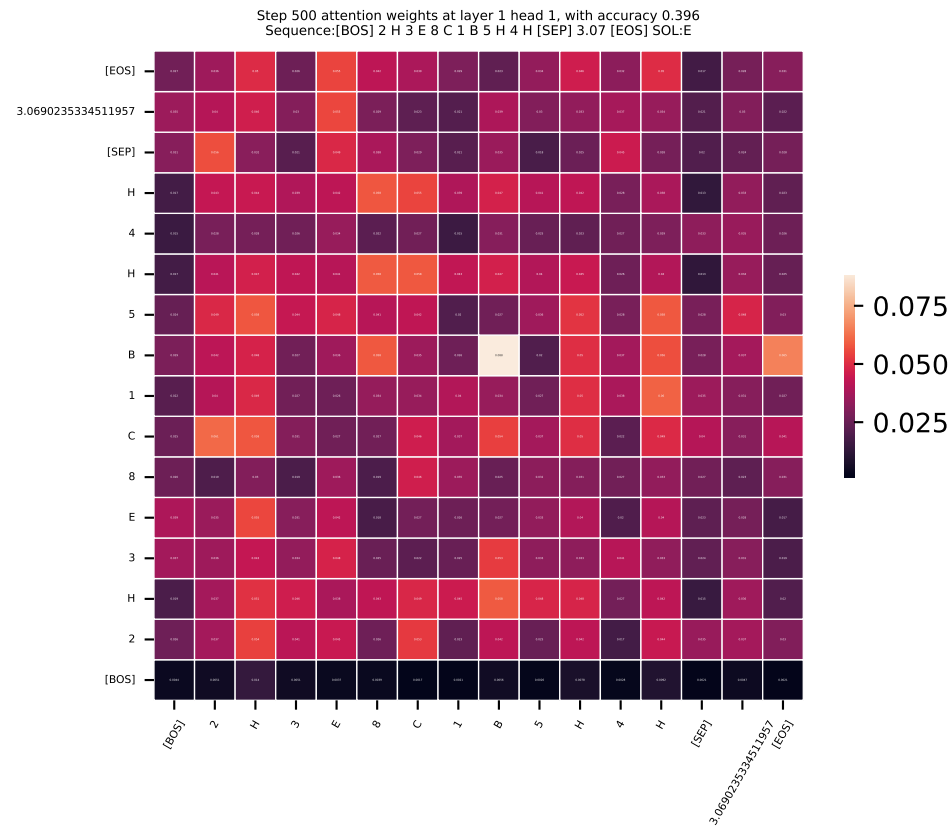

Step 500 attention weights at layer 1 head 0, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.075
0.050
0.025

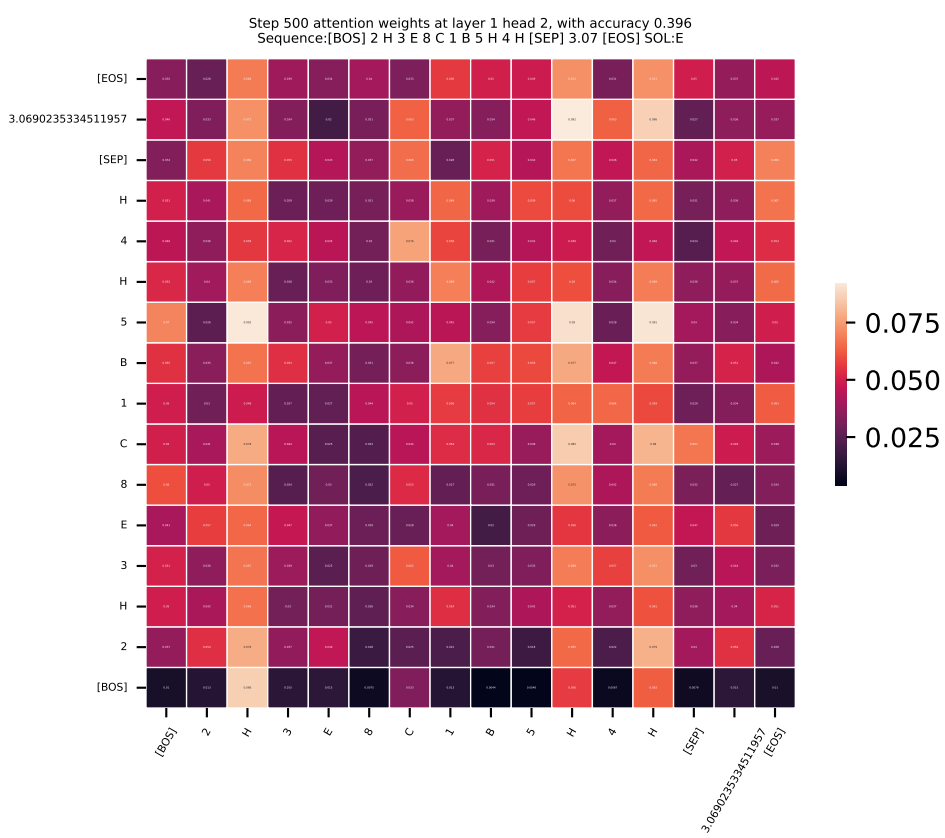

Step 500 attention weights at layer 1 head 1, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.075
0.050
0.025

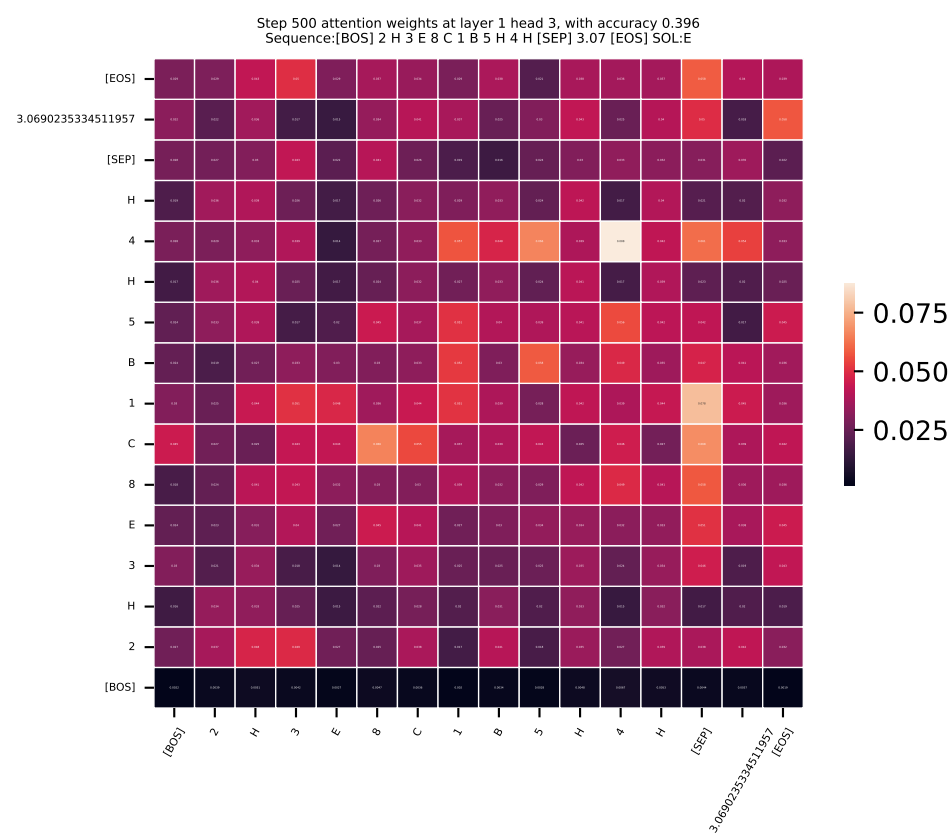

Step 500 attention weights at layer 1 head 2, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.075
0.050
0.025

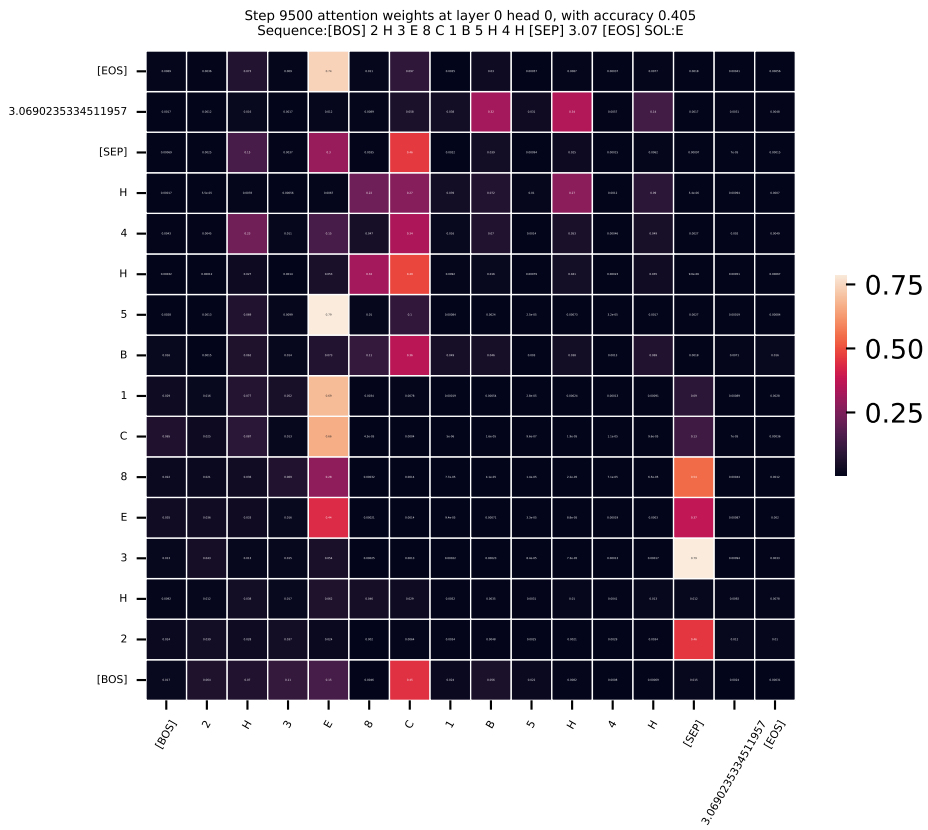

Step 500 attention weights at layer 1 head 3, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.075
0.050
0.025

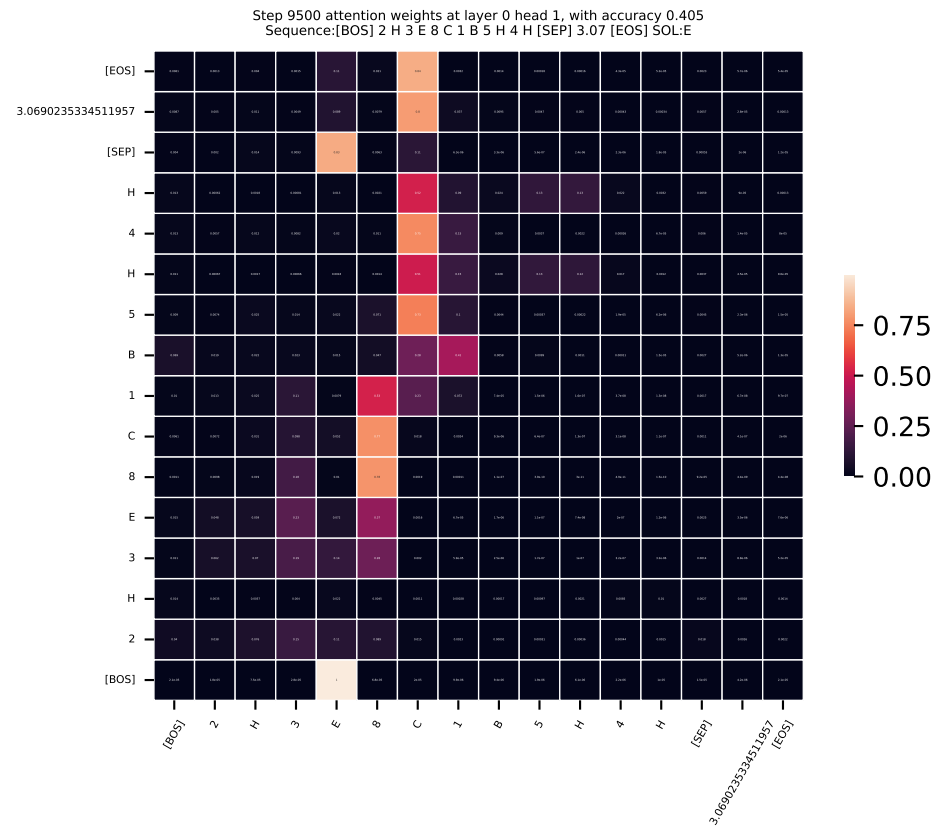

Step 500 attention weights at layer 1 head 0, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.075
0.050
0.025

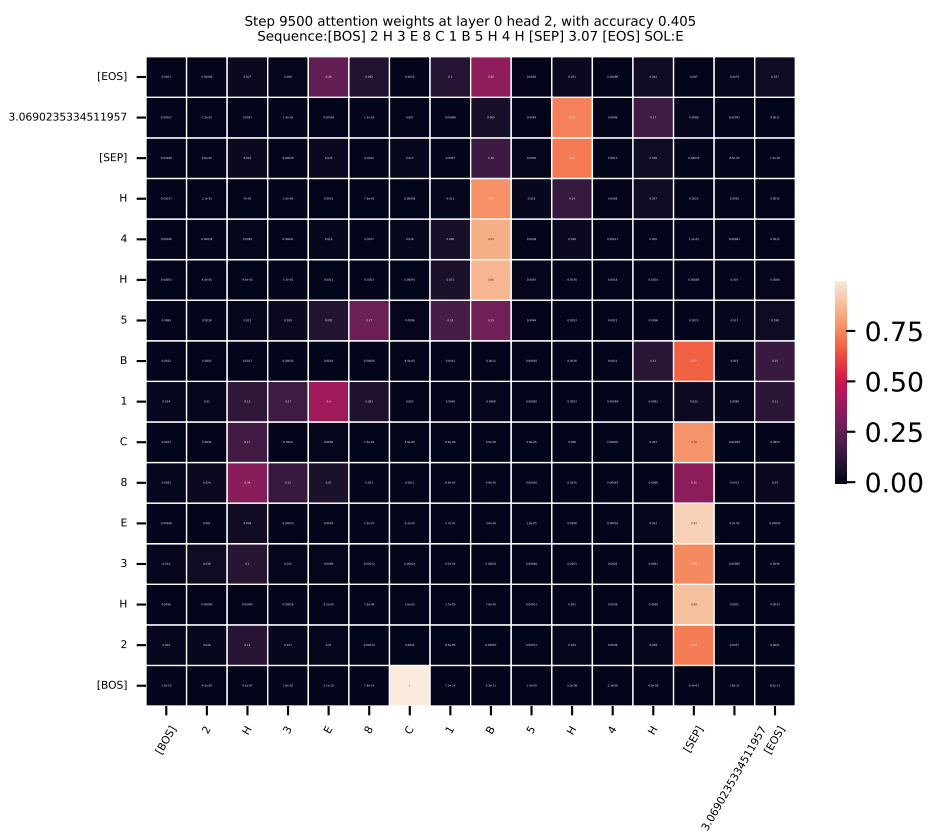

Step 500 attention weights at layer 1 head 1, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.075
0.050
0.025

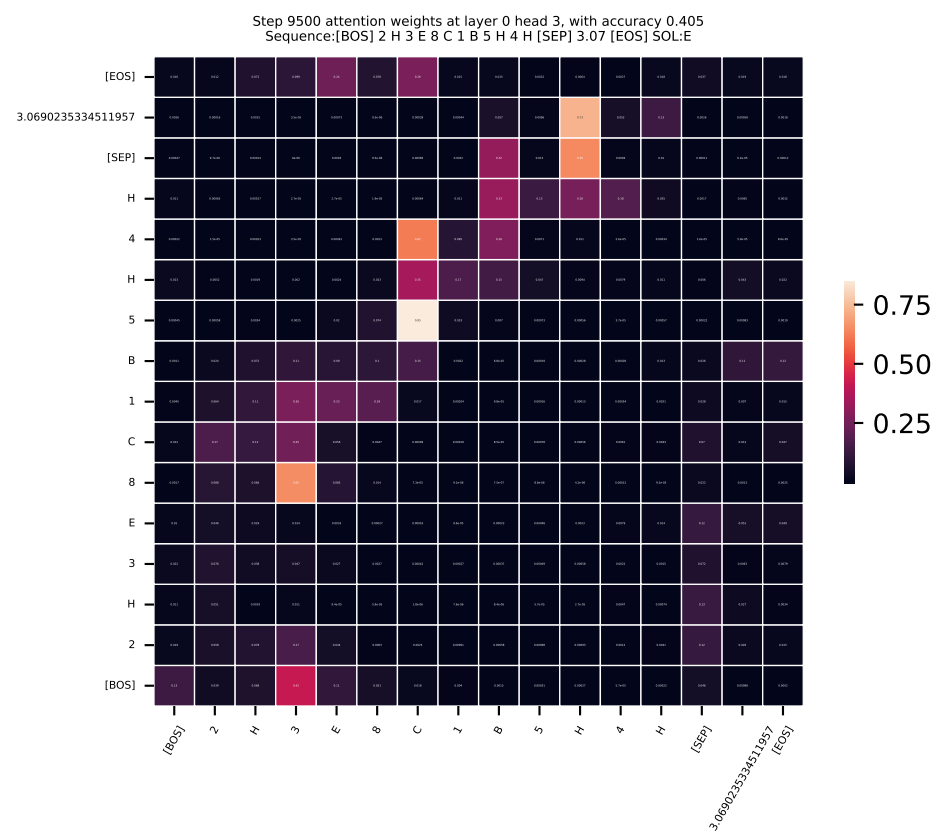

Step 500 attention weights at layer 1 head 2, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.075
0.050
0.025

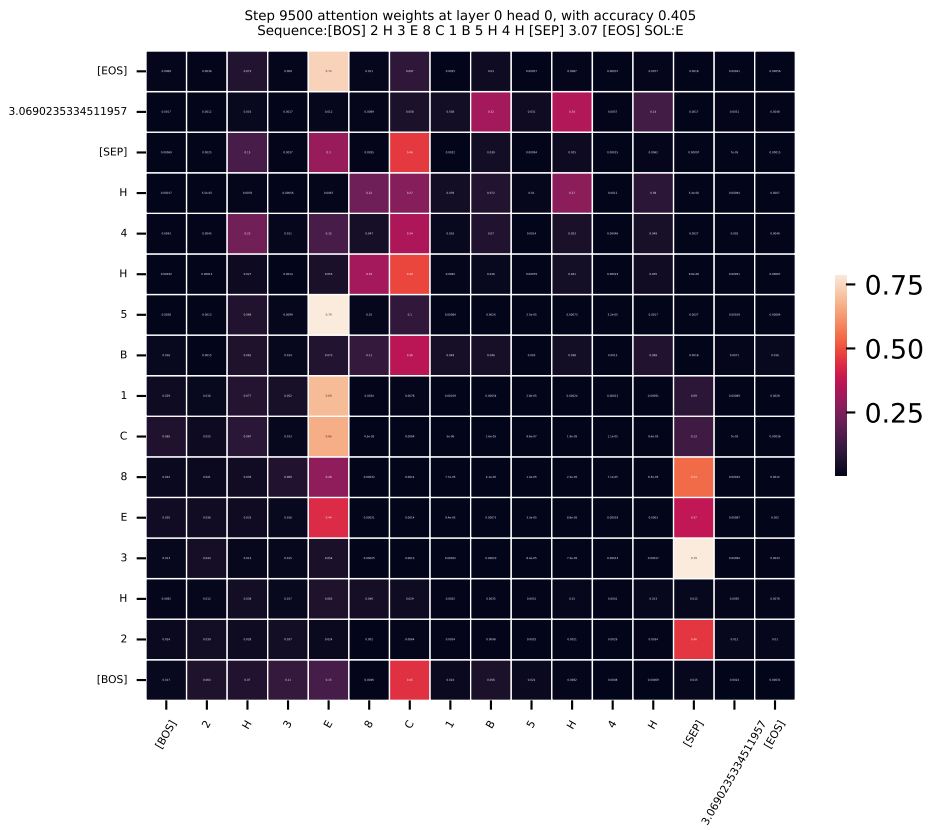

Step 500 attention weights at layer 1 head 3, with accuracy 0.396
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.075
0.050
0.025

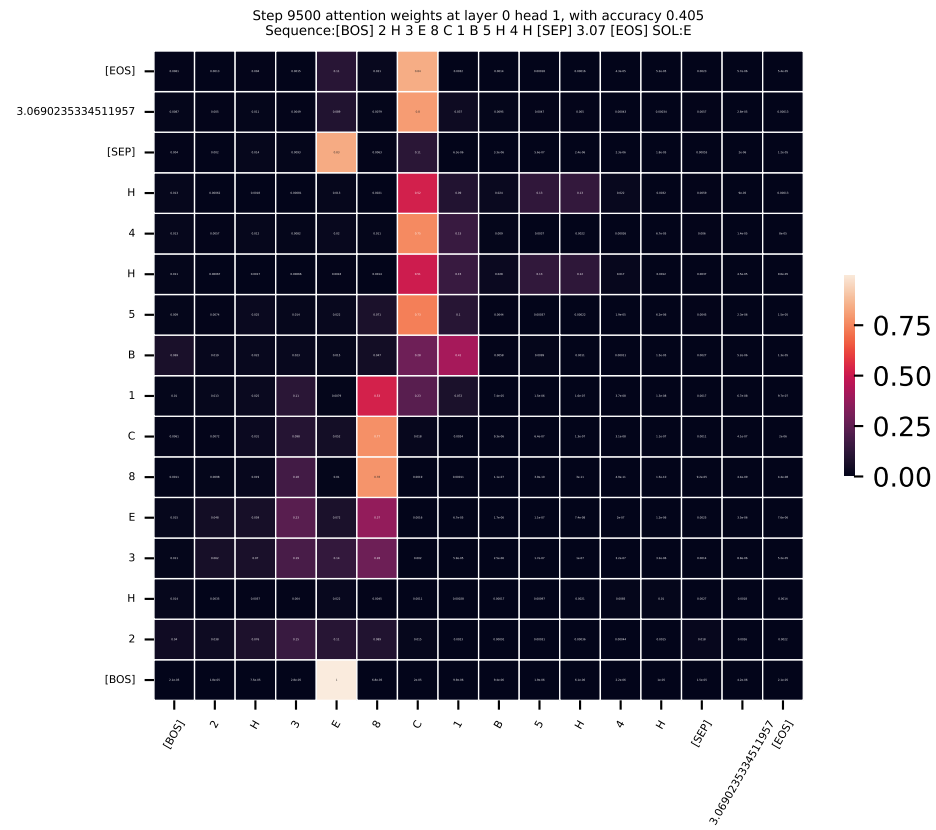

Step 9500 attention weights at layer 0 head 0, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

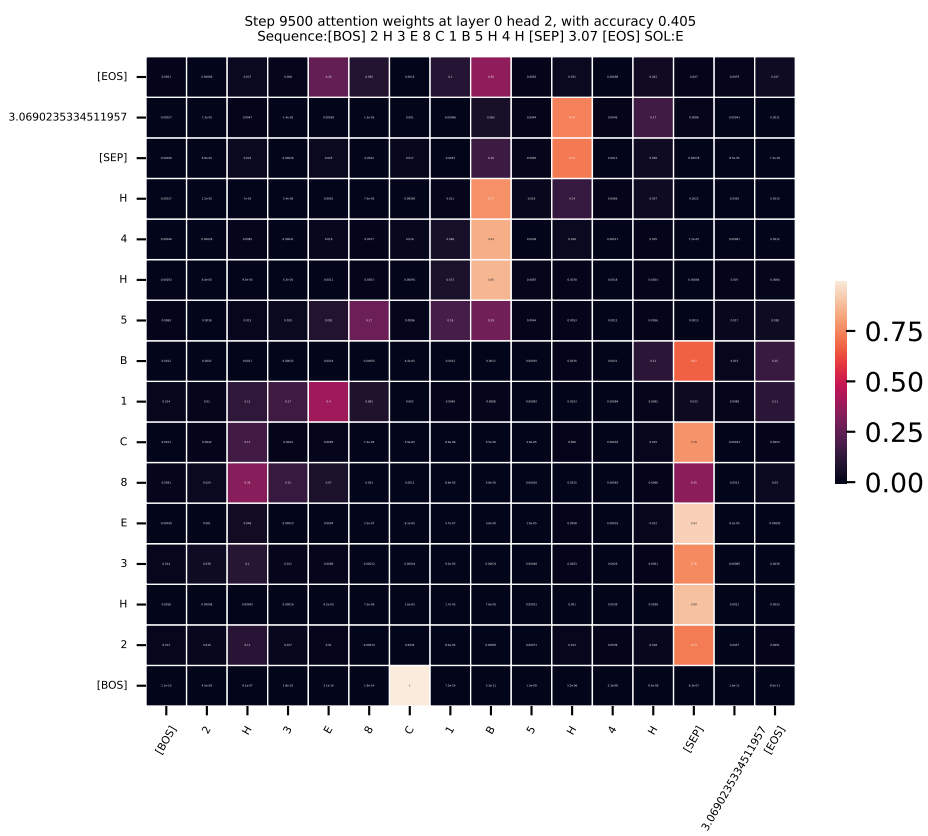

Step 9500 attention weights at layer 0 head 1, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

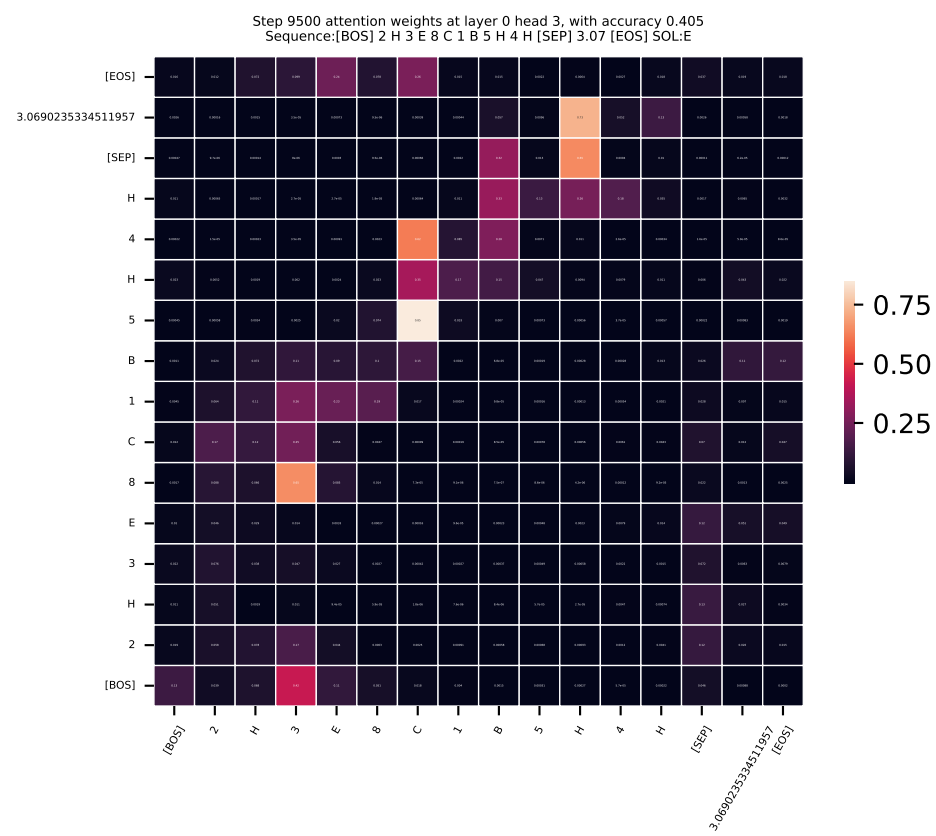

Step 9500 attention weights at layer 0 head 2, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

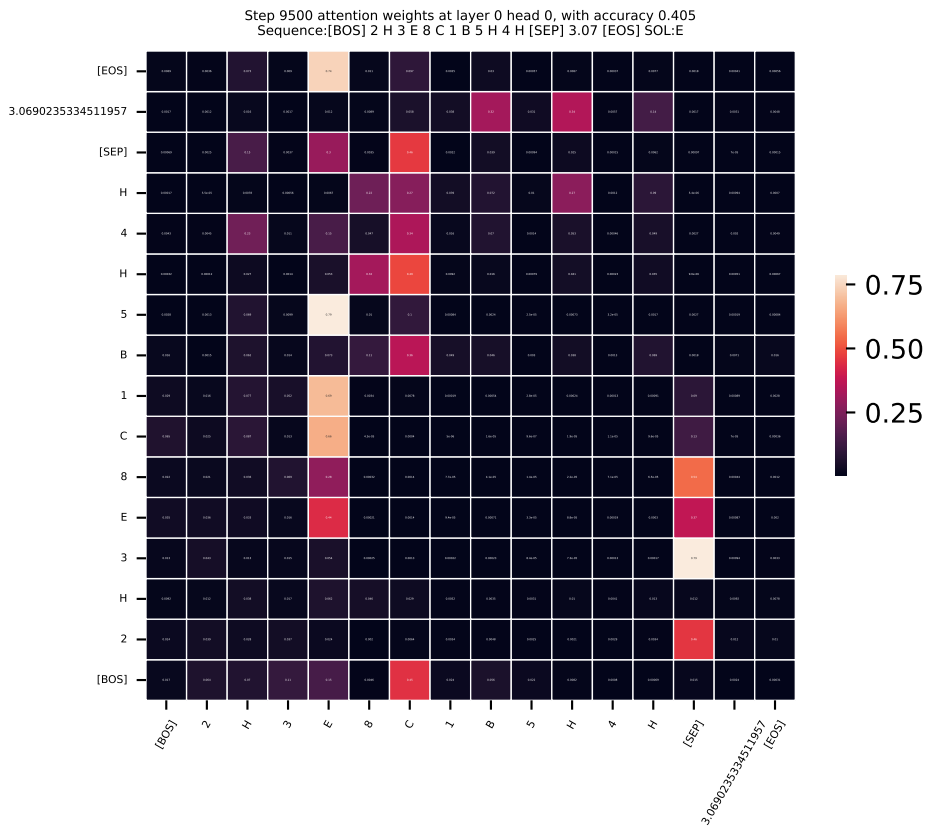

Step 9500 attention weights at layer 0 head 3, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

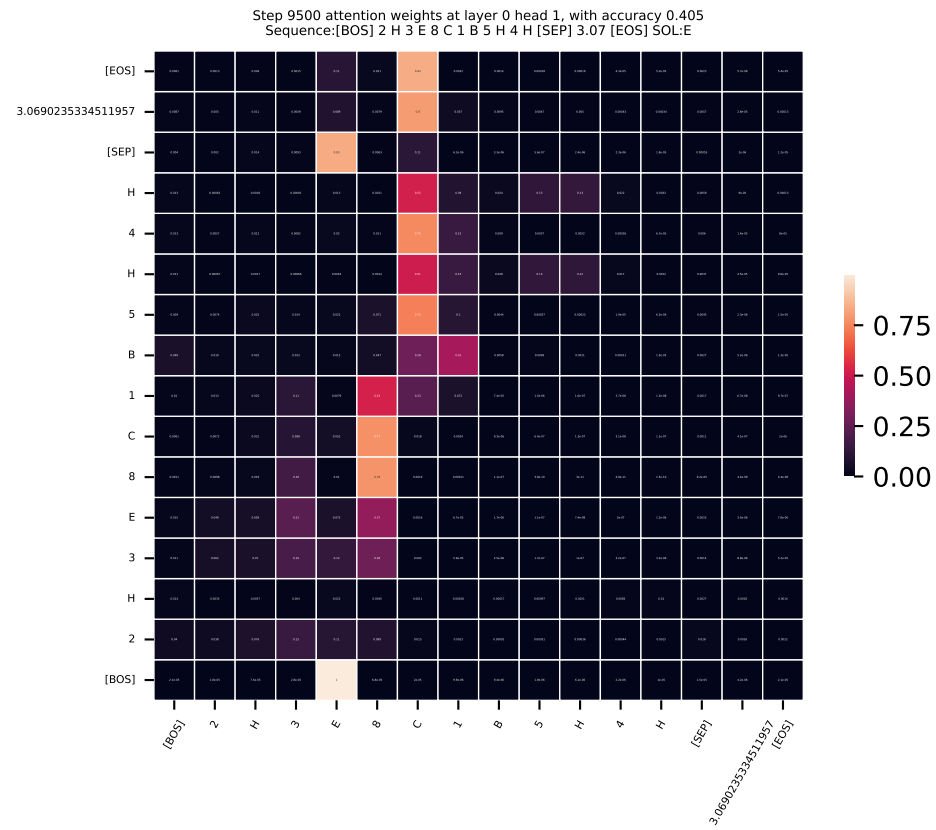

Step 9500 attention weights at layer 0 head 0, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

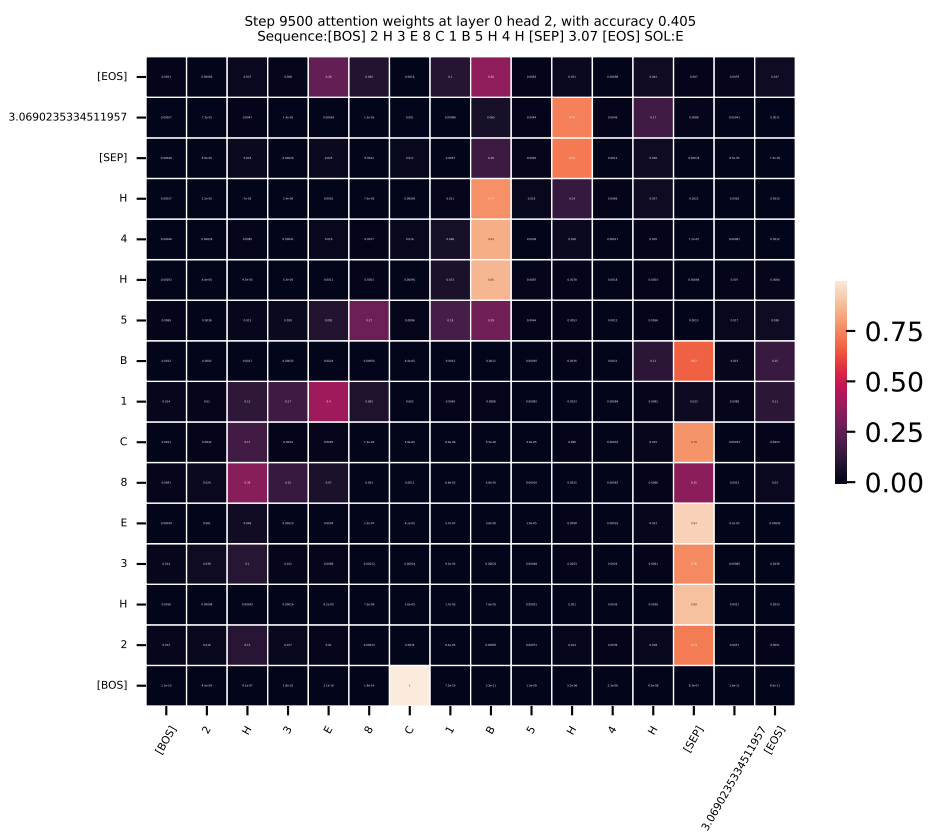

Step 9500 attention weights at layer 0 head 1, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

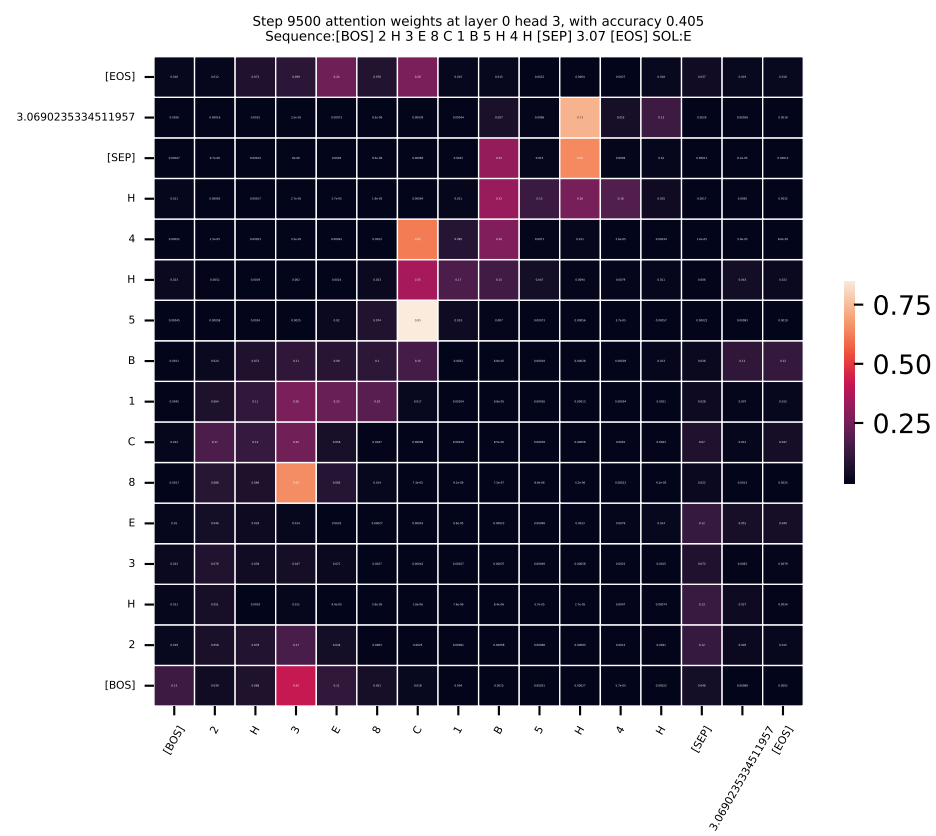

Step 9500 attention weights at layer 0 head 2, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

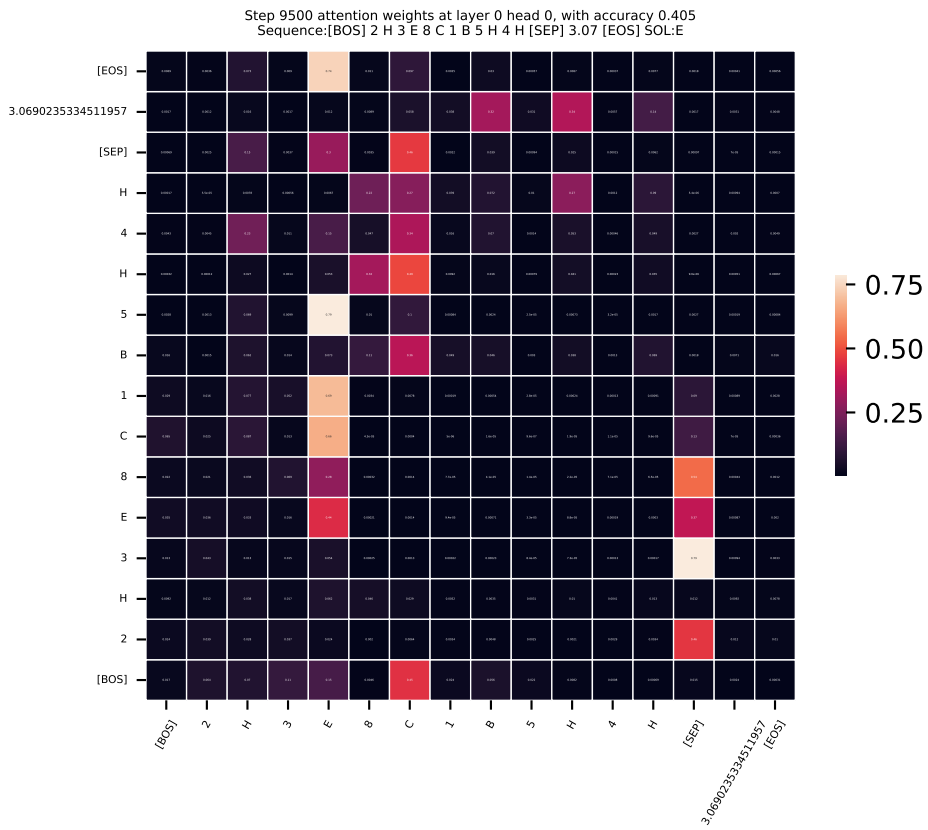

Step 9500 attention weights at layer 0 head 3, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

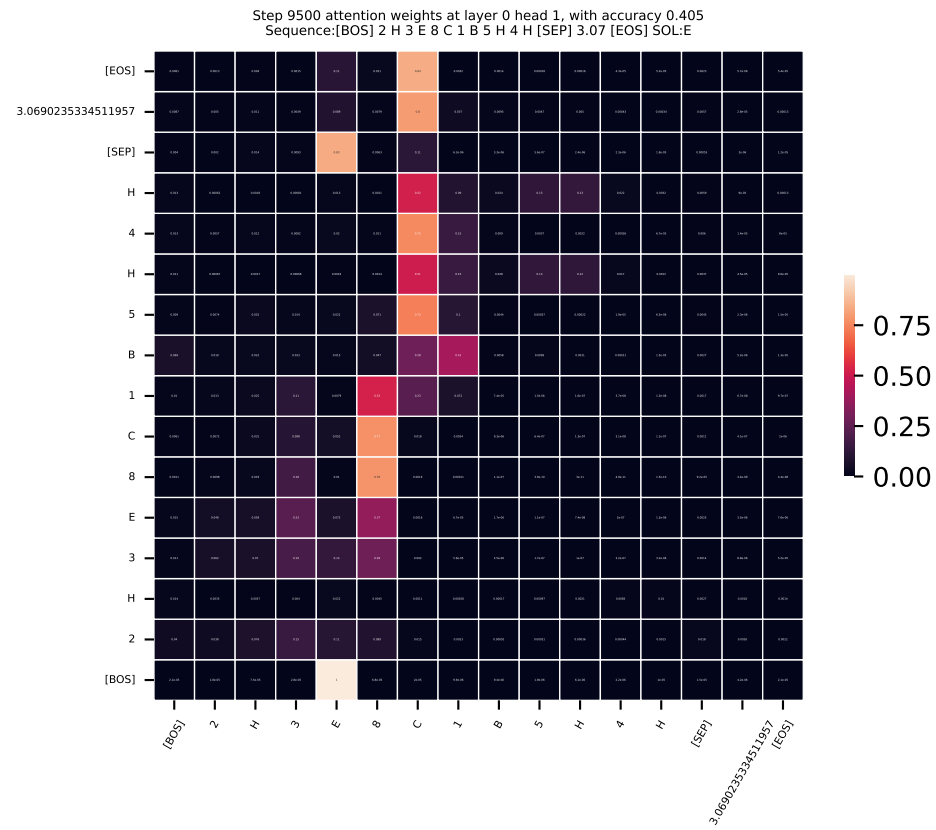

Step 9500 attention weights at layer 0 head 0, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

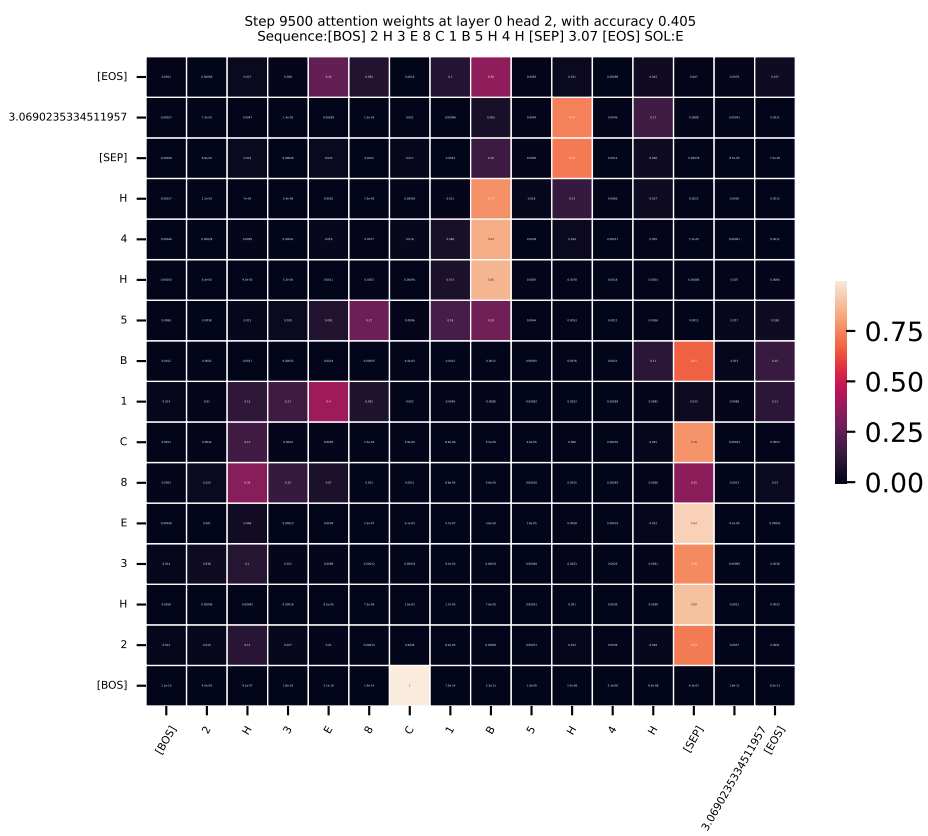

Step 9500 attention weights at layer 0 head 1, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

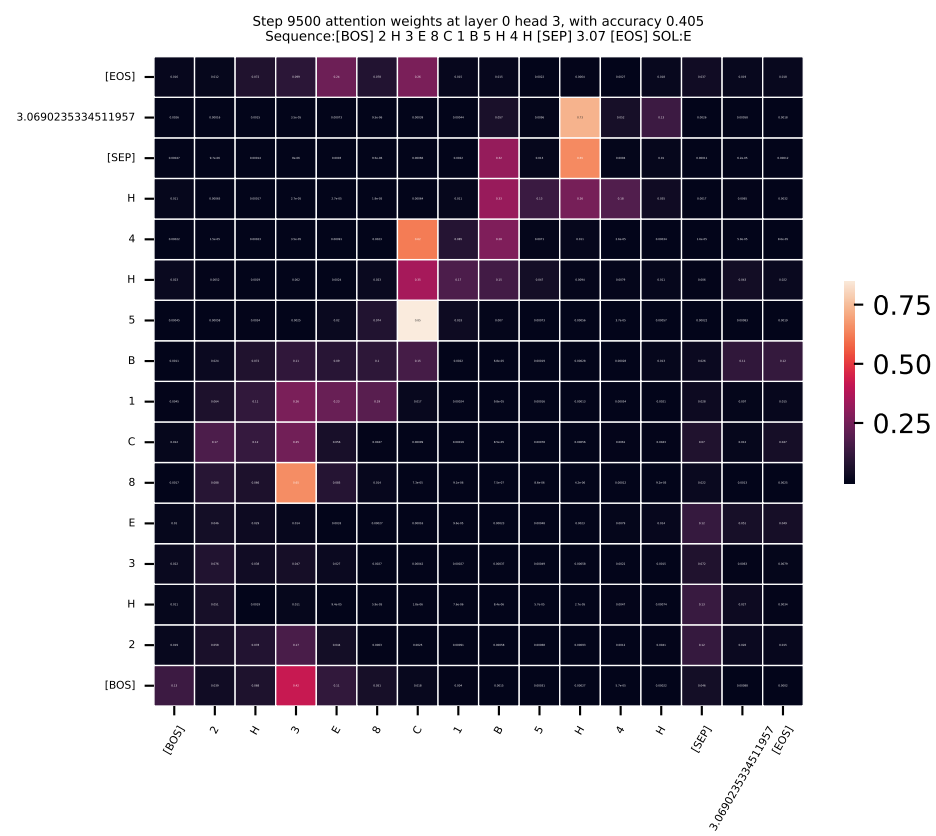

Step 9500 attention weights at layer 0 head 2, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

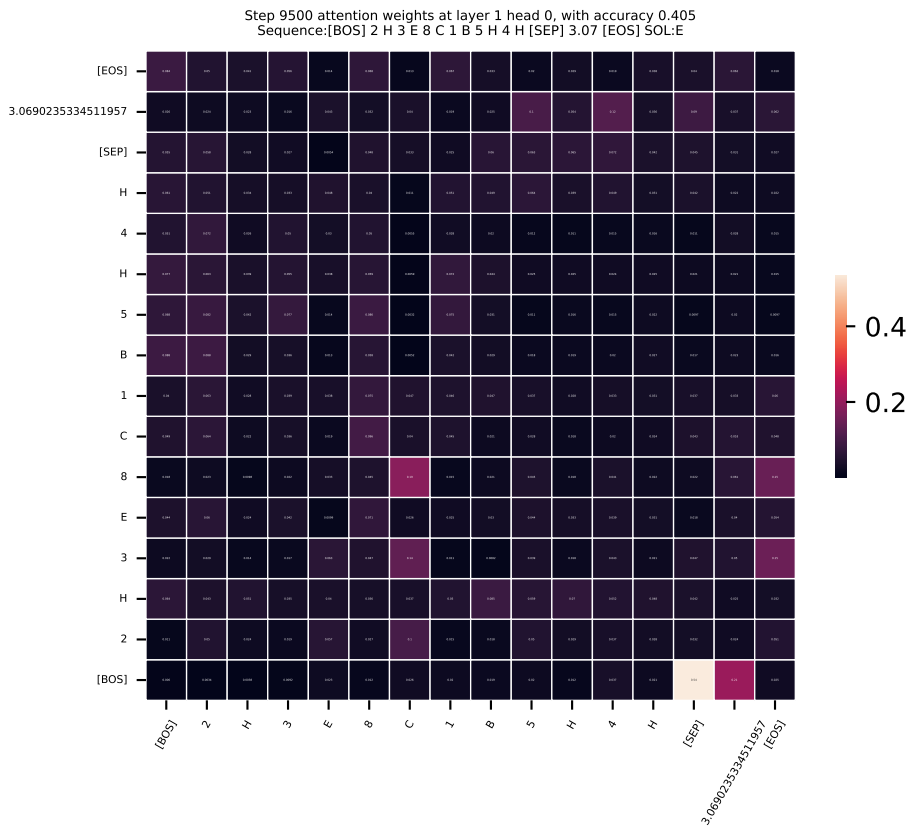

Step 9500 attention weights at layer 0 head 3, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

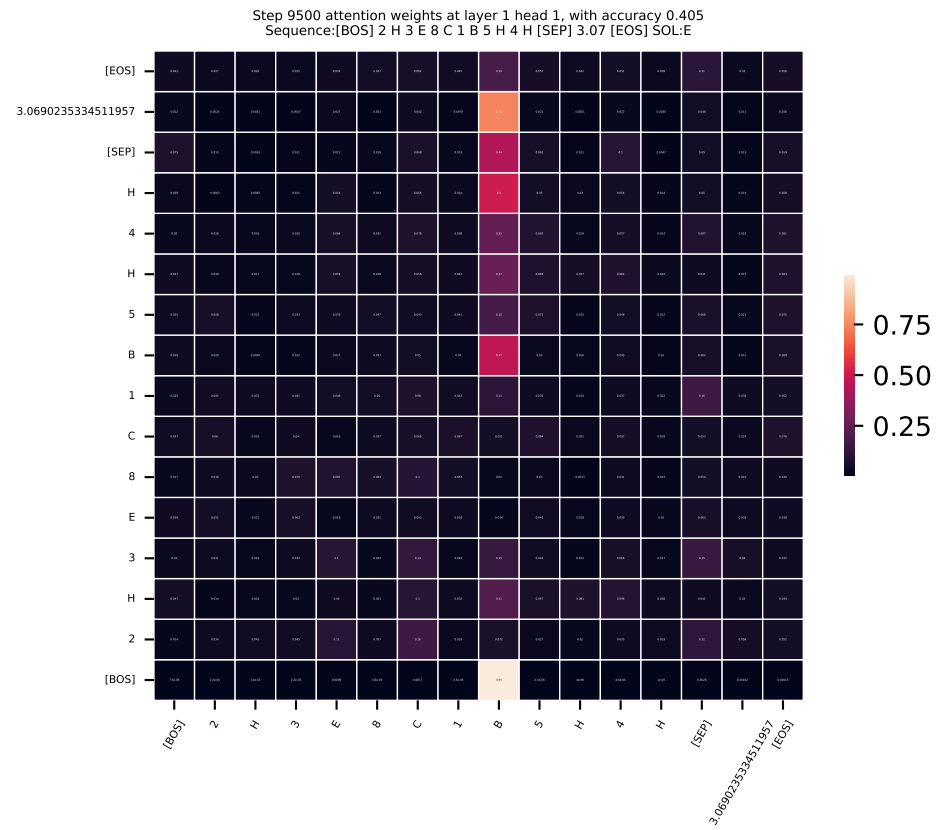

Step 9500 attention weights at layer 0 head 0, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

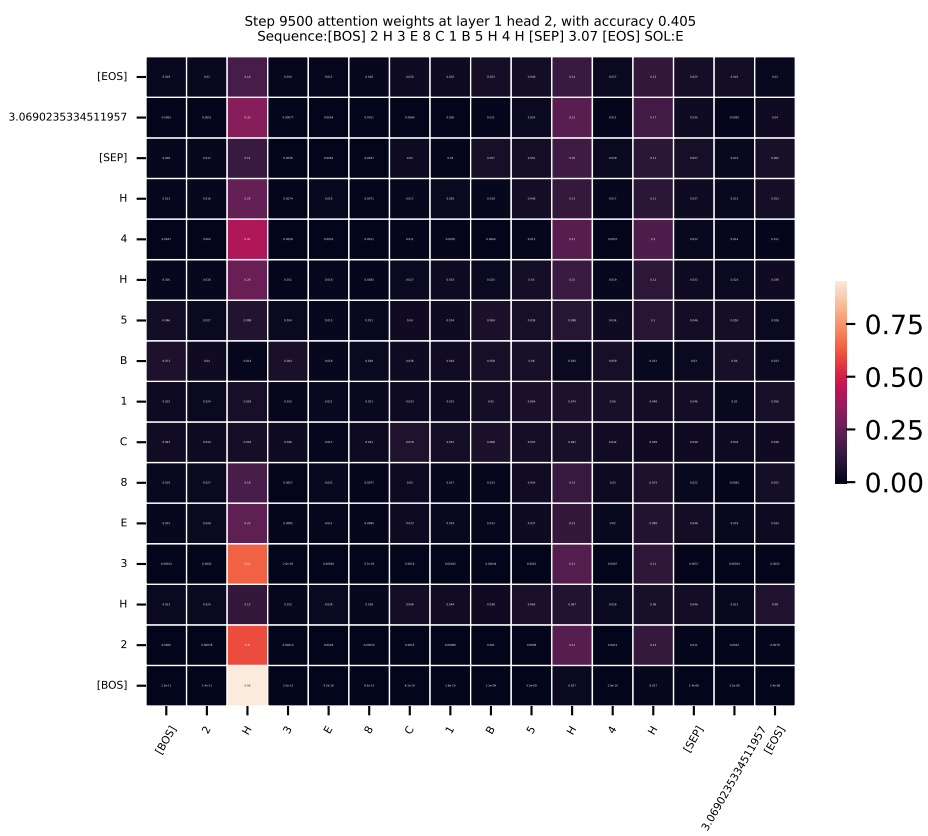

Step 9500 attention weights at layer 0 head 1, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

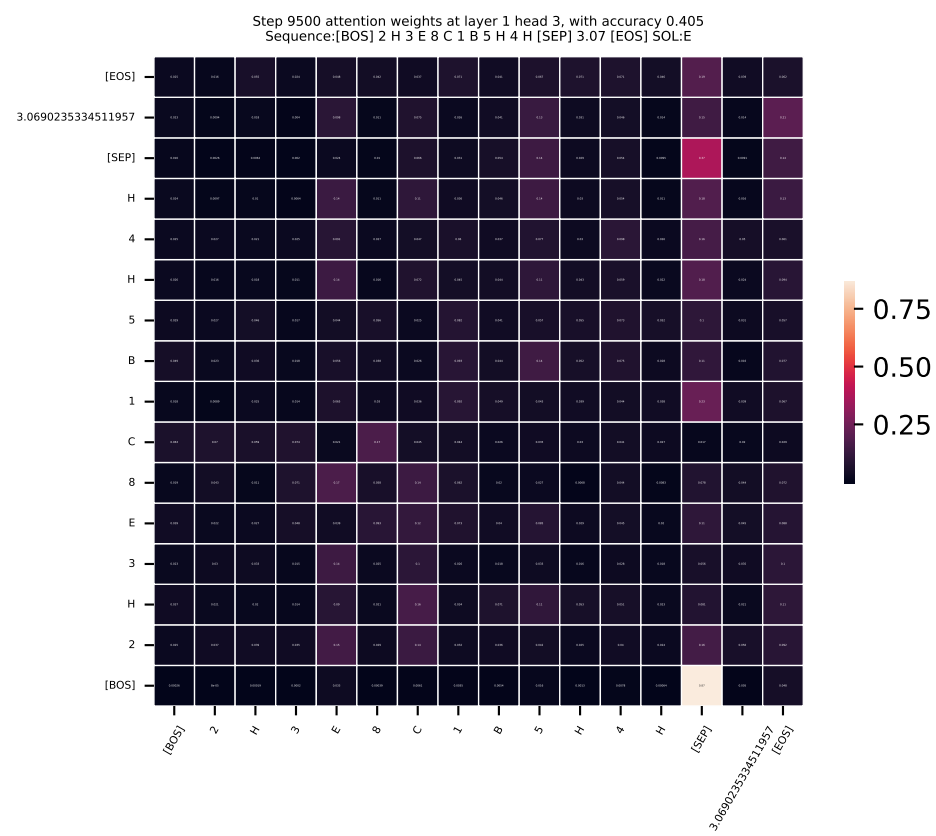

Step 9500 attention weights at layer 0 head 2, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

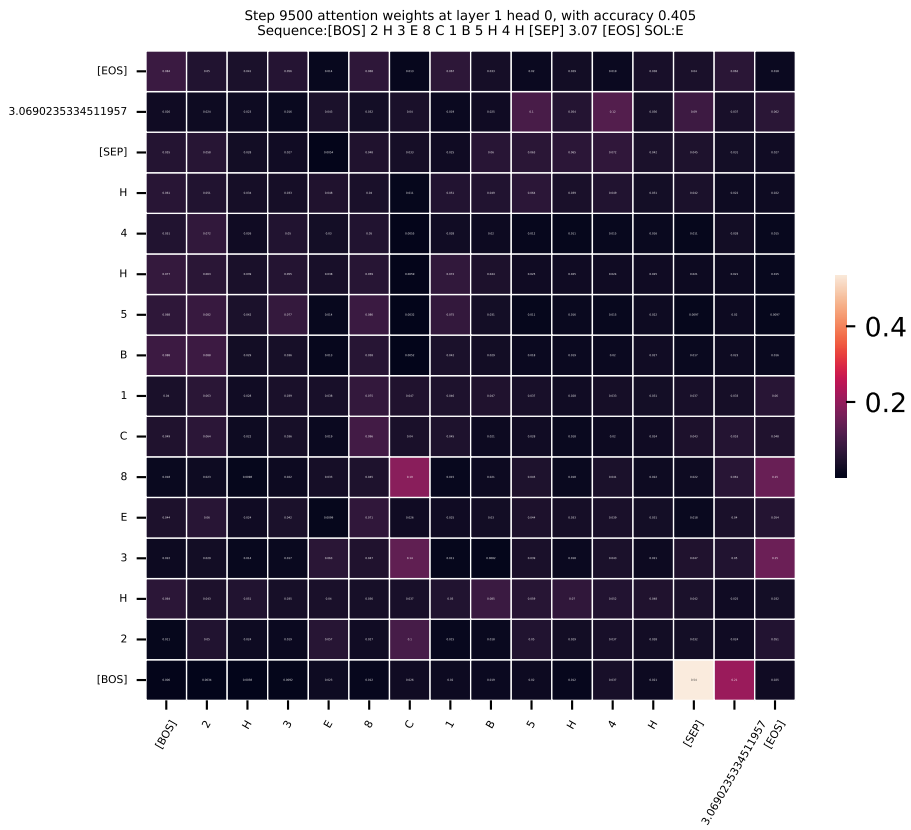

Step 9500 attention weights at layer 0 head 3, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

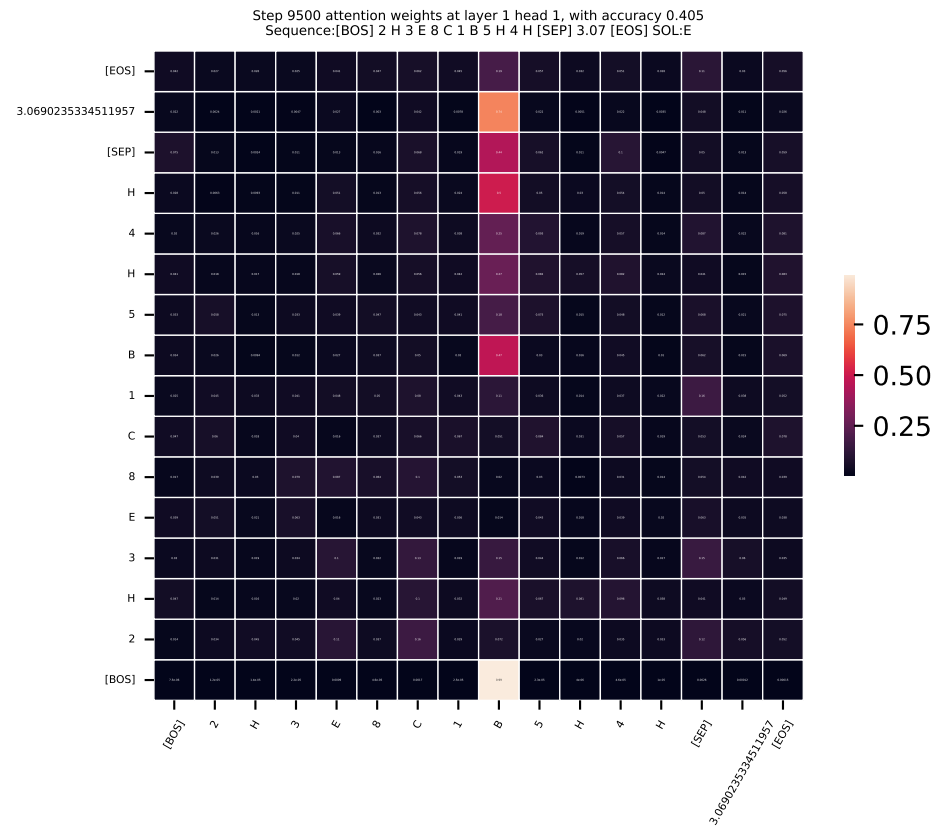

Step 9500 attention weights at layer 1 head 0, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.4
0.2

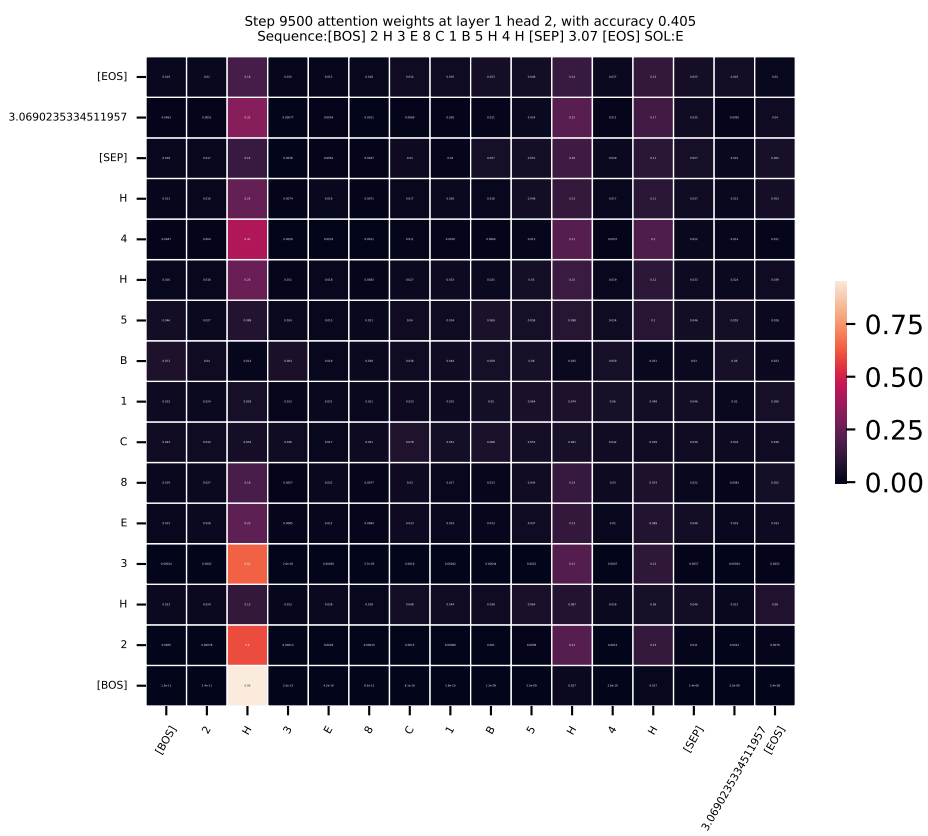

Step 9500 attention weights at layer 1 head 1, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

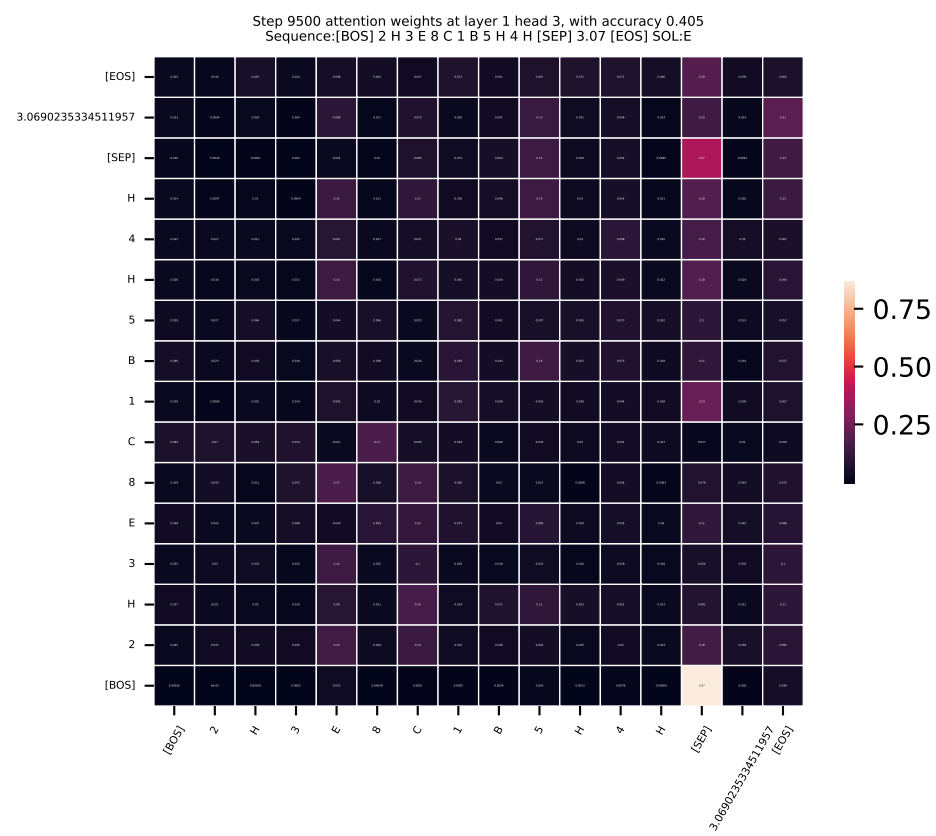

Step 9500 attention weights at layer 1 head 2, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

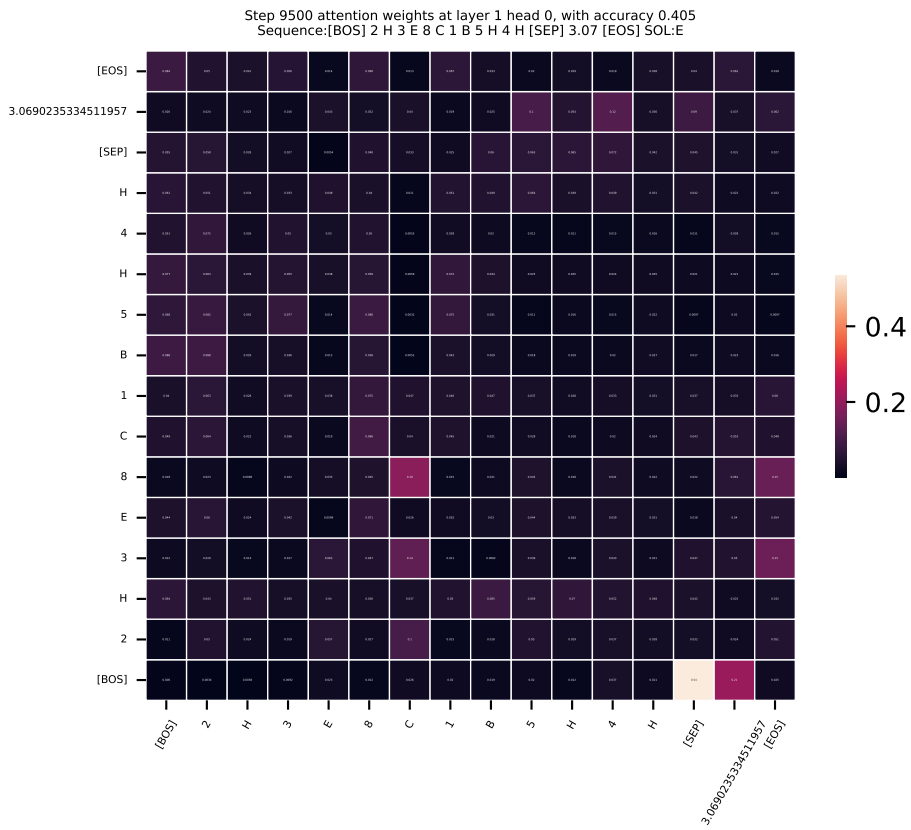

Step 9500 attention weights at layer 1 head 3, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

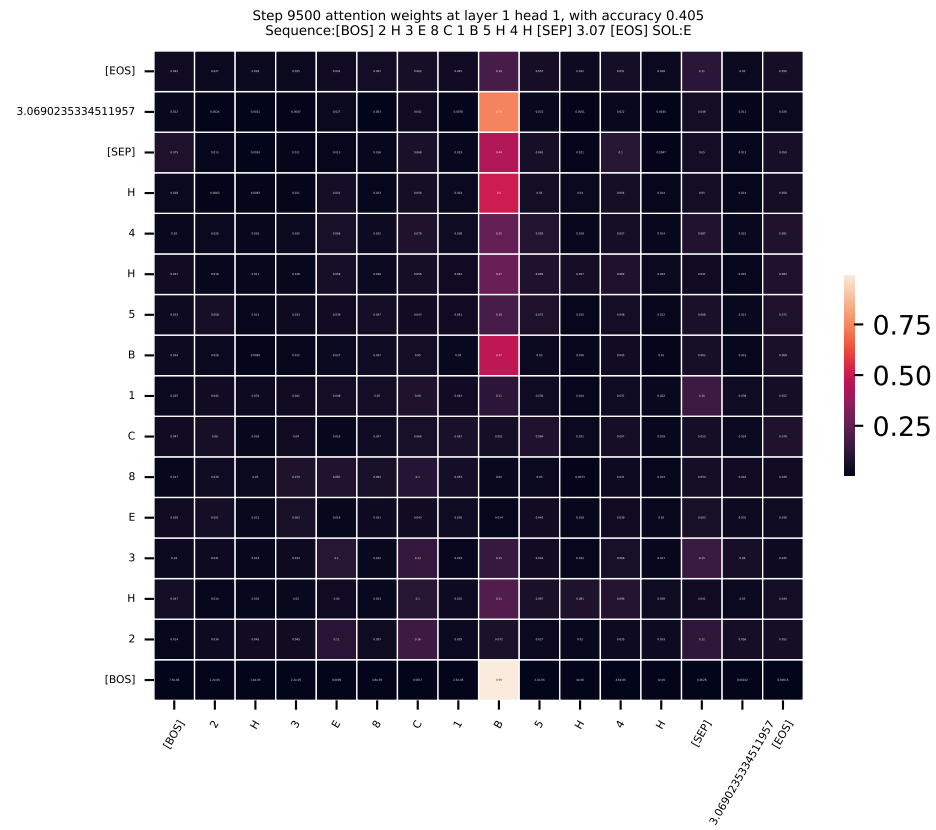

Step 9500 attention weights at layer 1 head 0, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.4
0.2

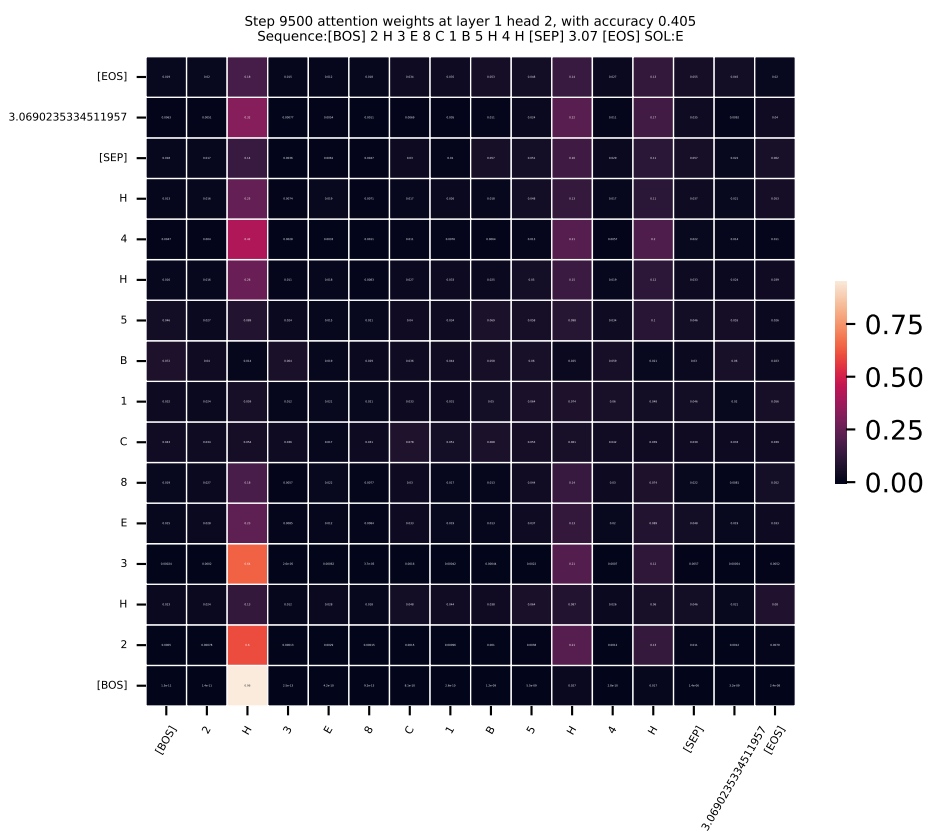

Step 9500 attention weights at layer 1 head 1, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

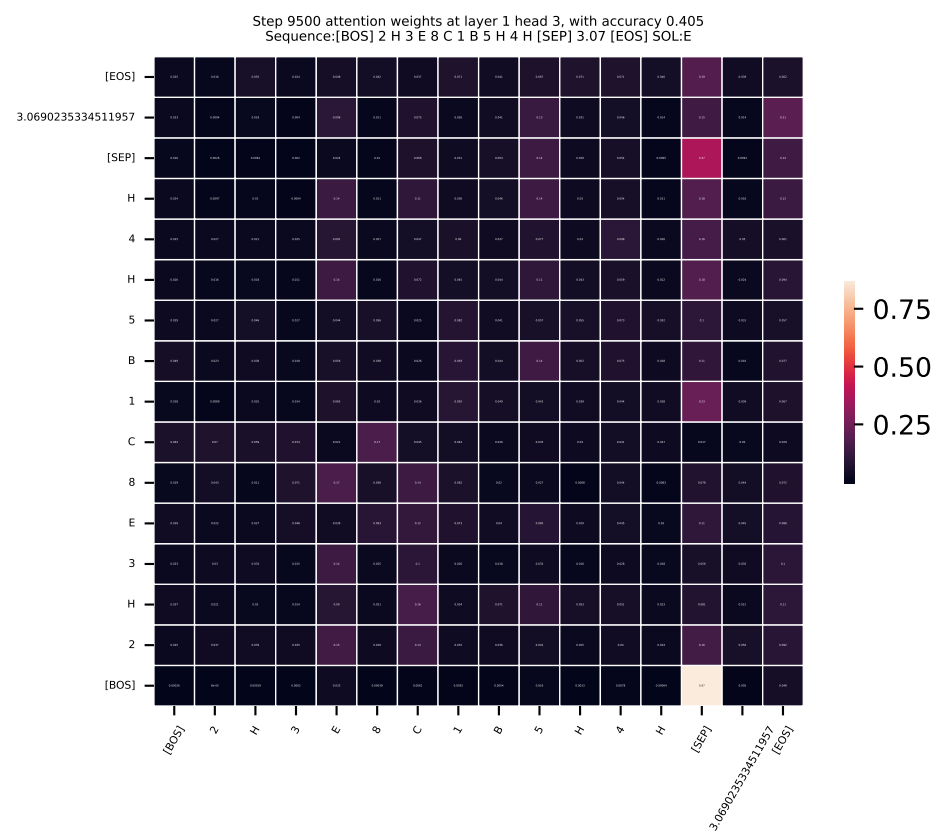

Step 9500 attention weights at layer 1 head 2, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

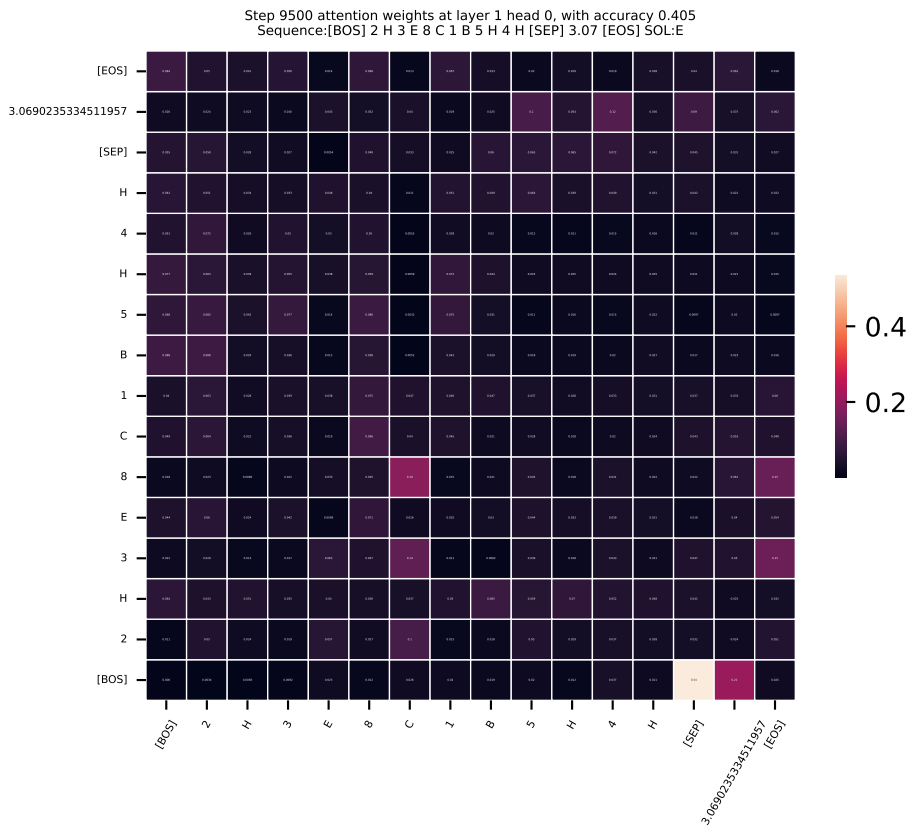

Step 9500 attention weights at layer 1 head 3, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

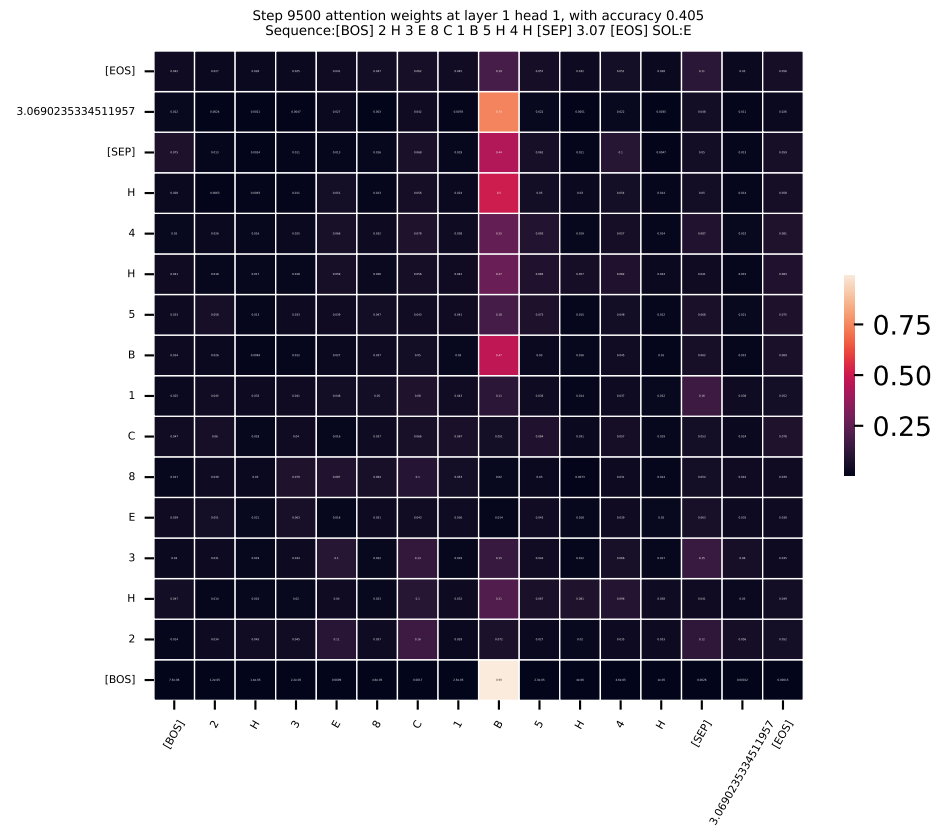

Step 9500 attention weights at layer 1 head 0, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.4
0.2

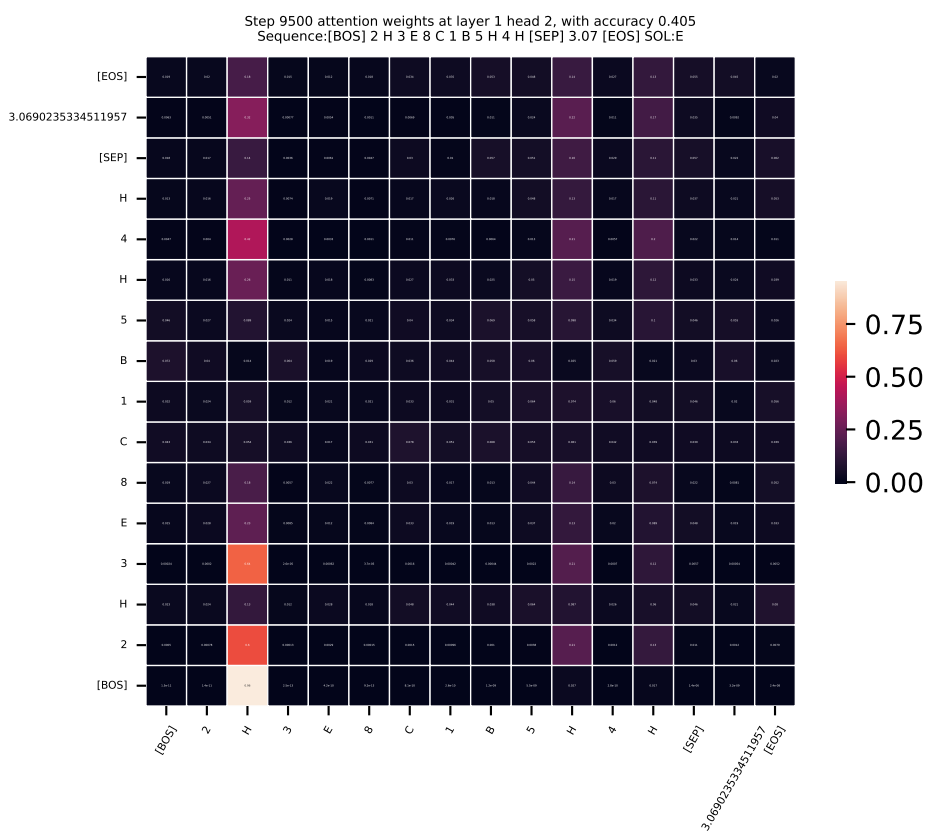

Step 9500 attention weights at layer 1 head 1, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

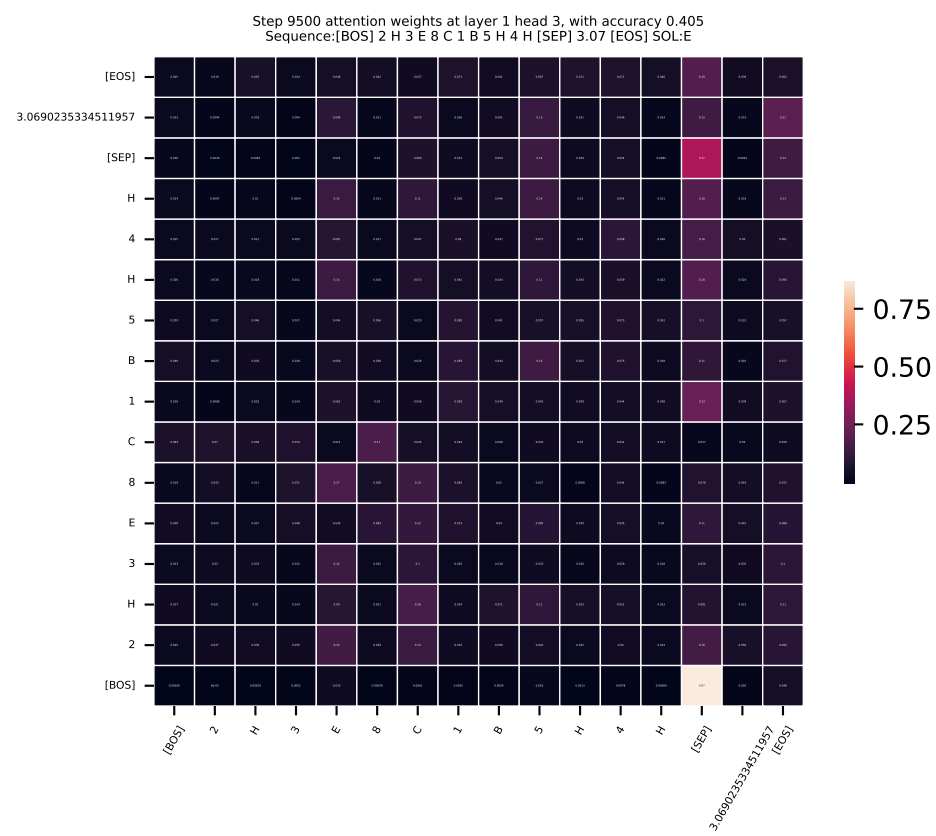

Step 9500 attention weights at layer 1 head 2, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

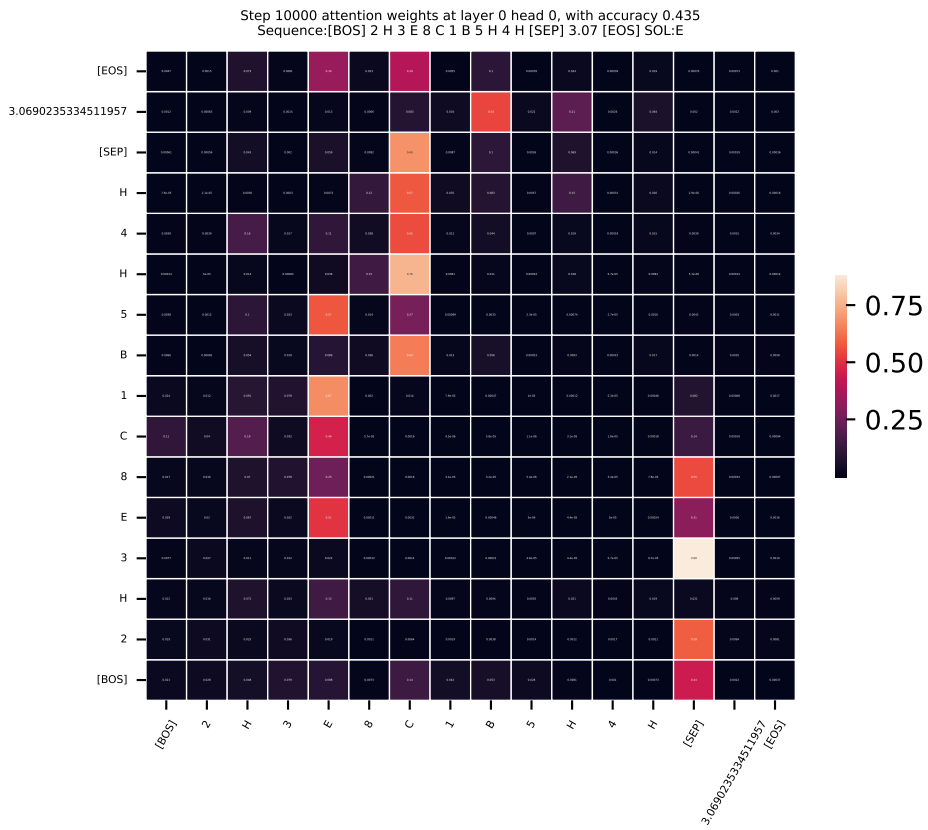

Step 9500 attention weights at layer 1 head 3, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

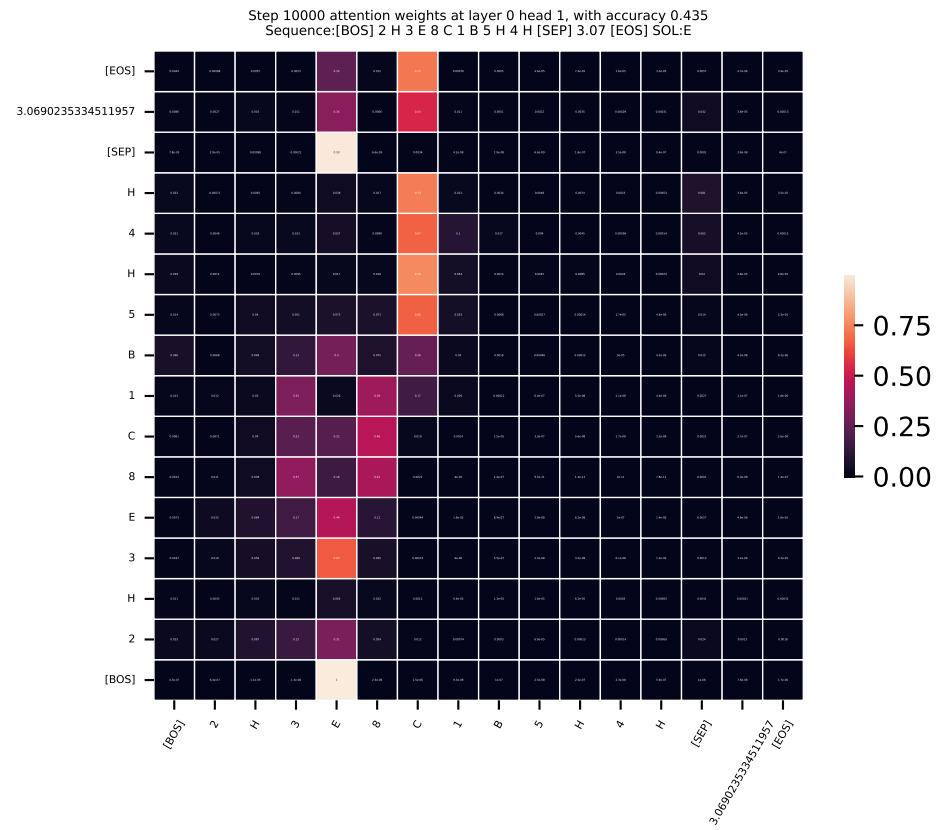

Step 9500 attention weights at layer 1 head 0, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.4
0.2

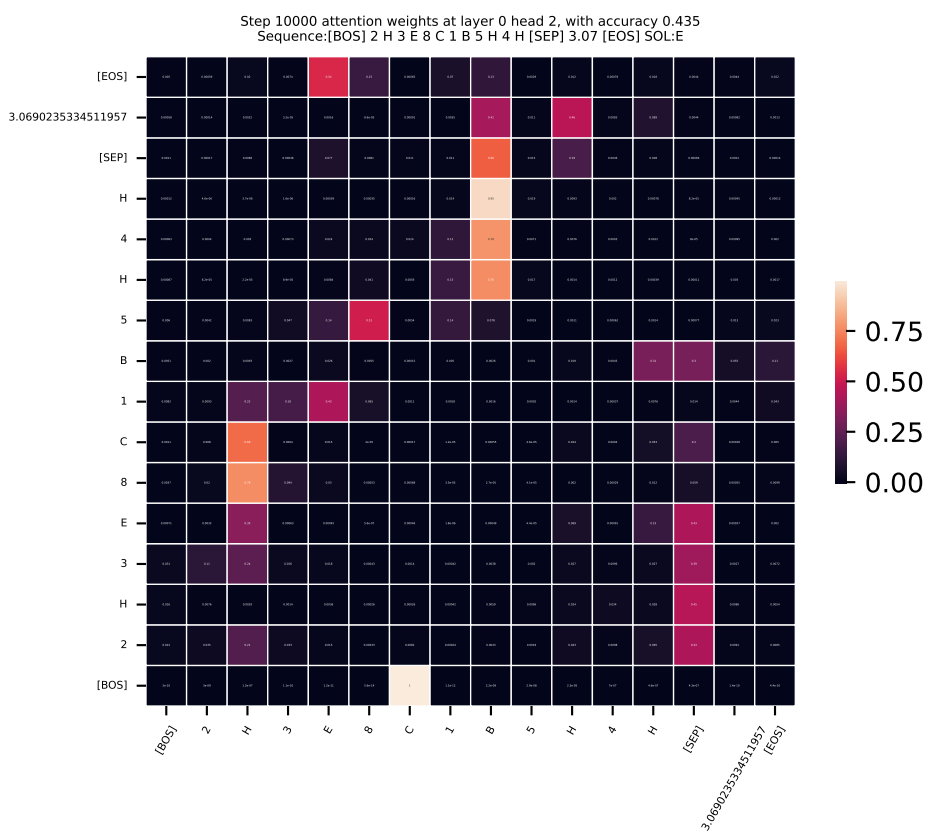

Step 9500 attention weights at layer 1 head 1, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

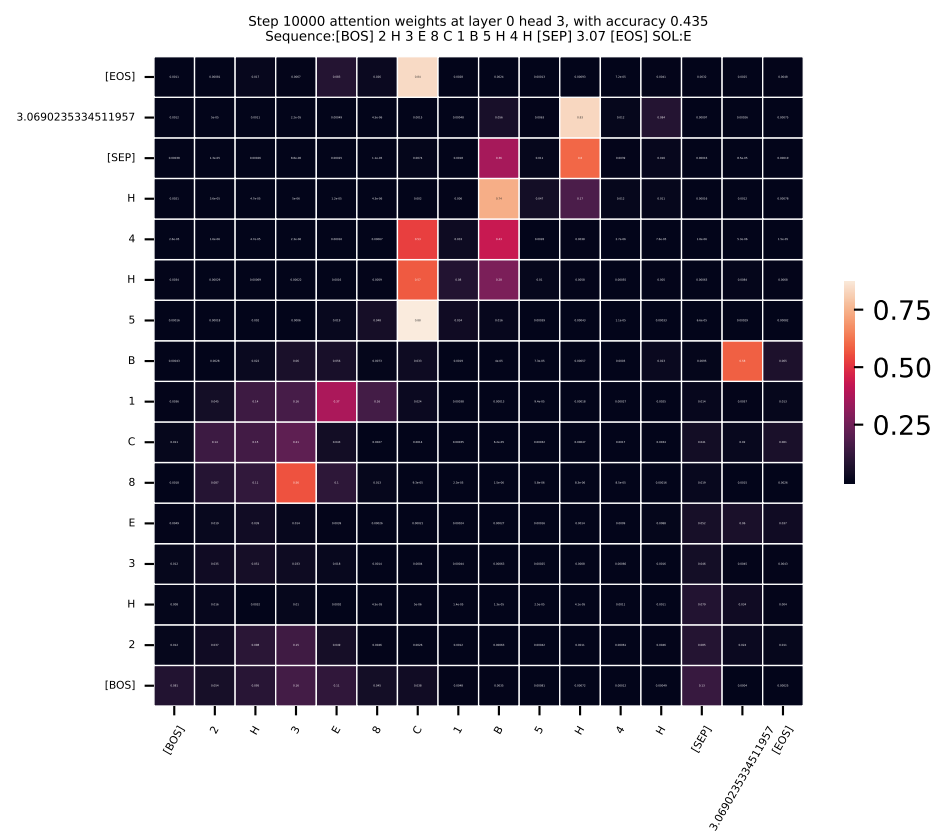

Step 9500 attention weights at layer 1 head 2, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

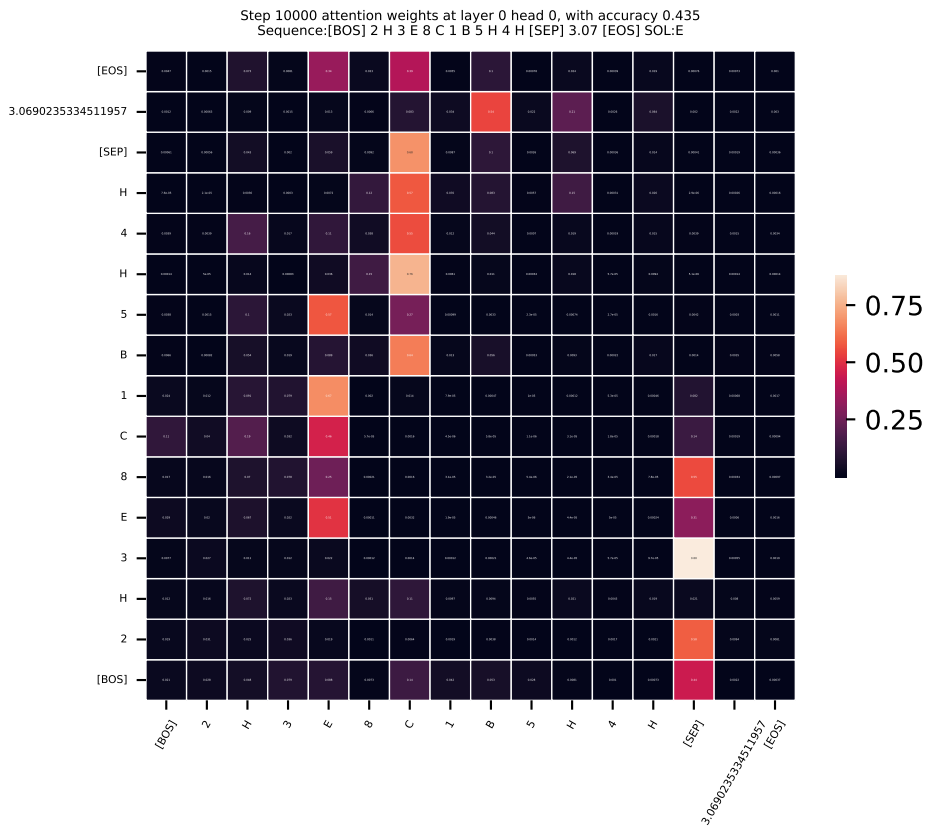

Step 9500 attention weights at layer 1 head 3, with accuracy 0.405
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

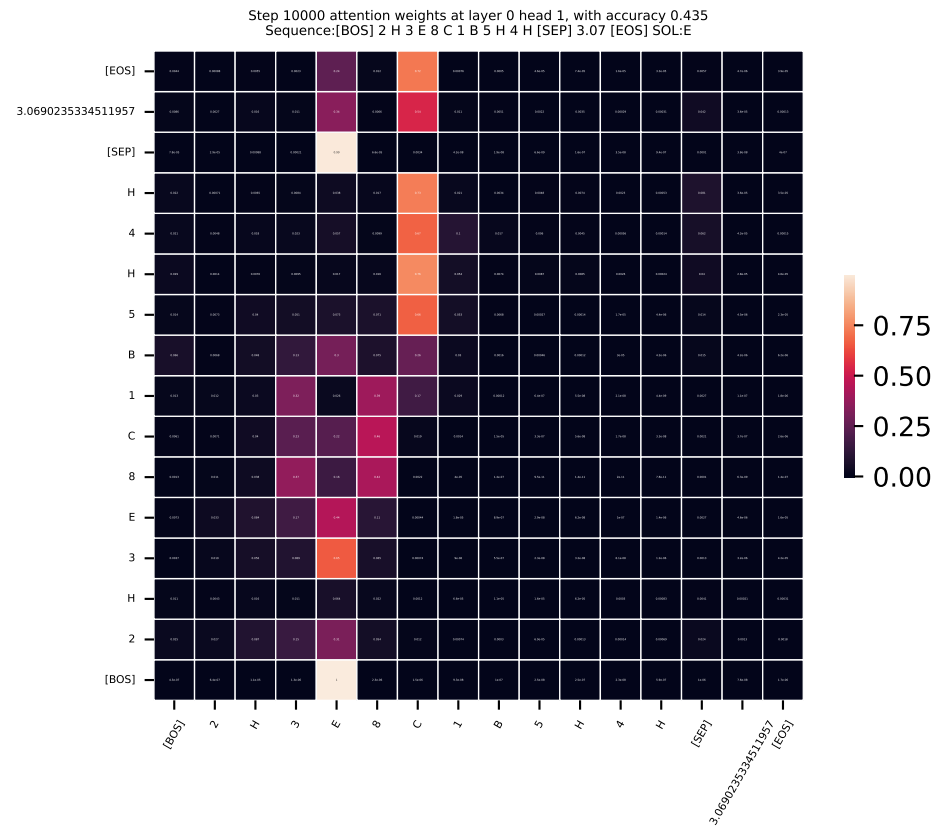

Step 10000 attention weights at layer 0 head 0, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

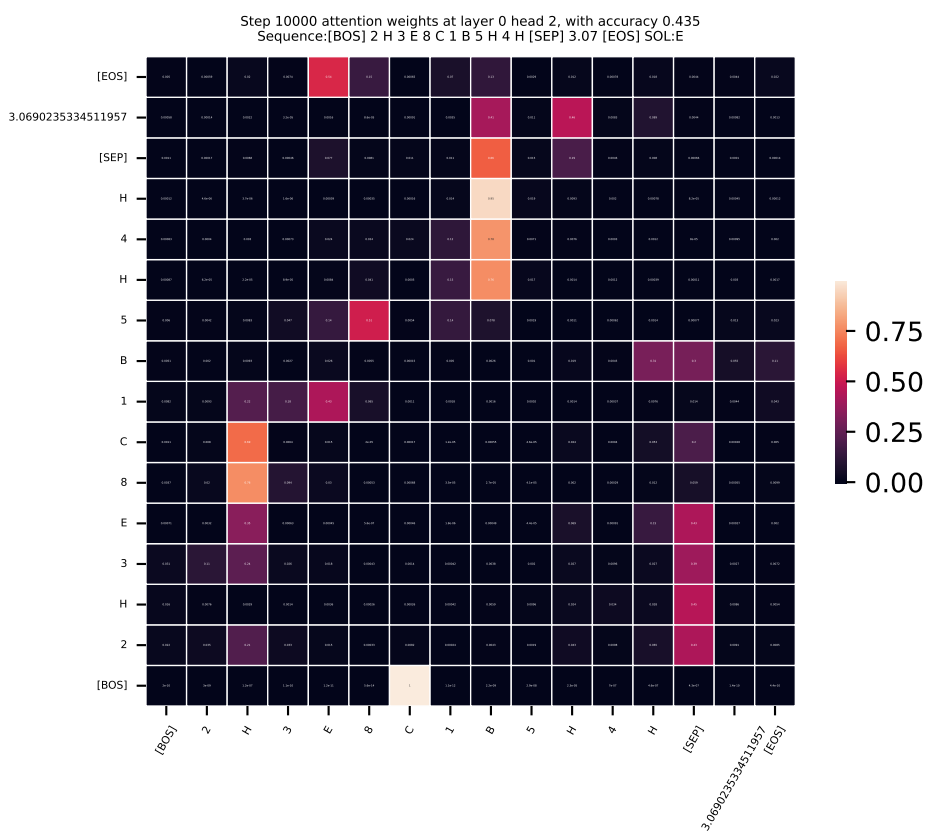

Step 10000 attention weights at layer 0 head 1, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

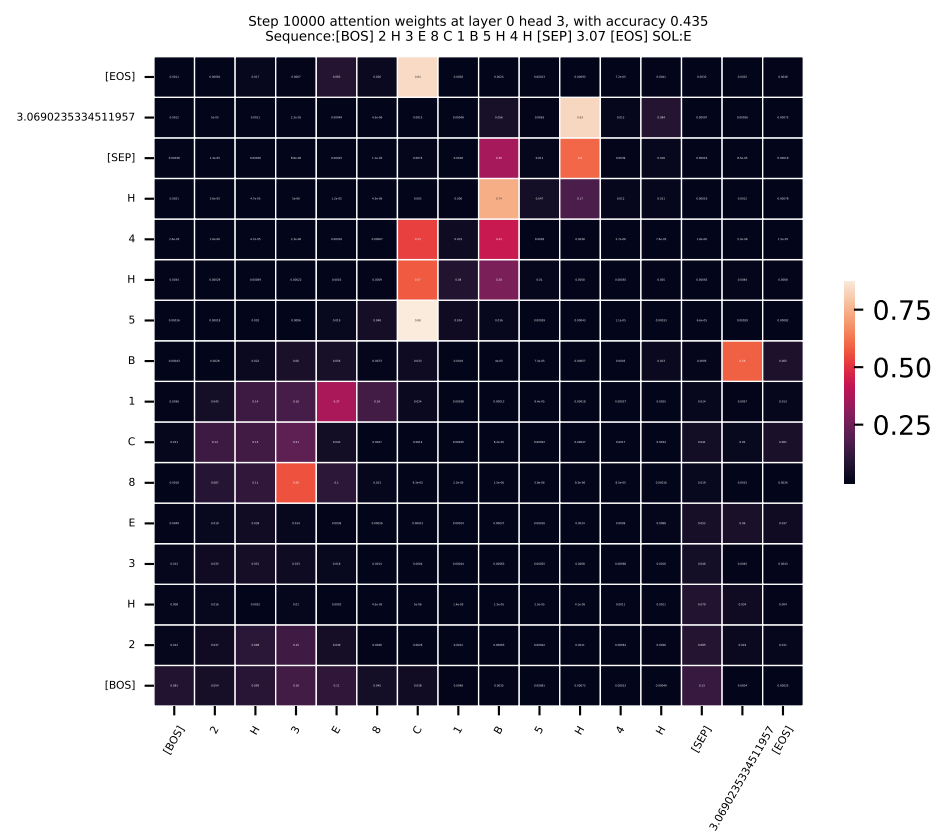

Step 10000 attention weights at layer 0 head 2, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

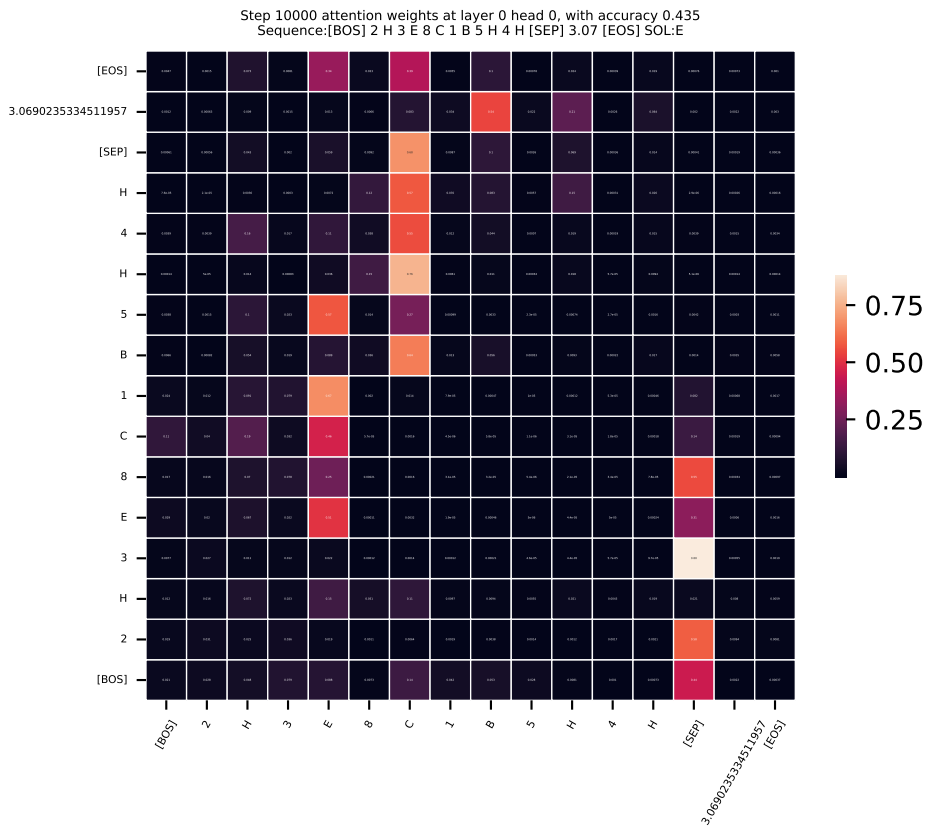

Step 10000 attention weights at layer 0 head 3, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

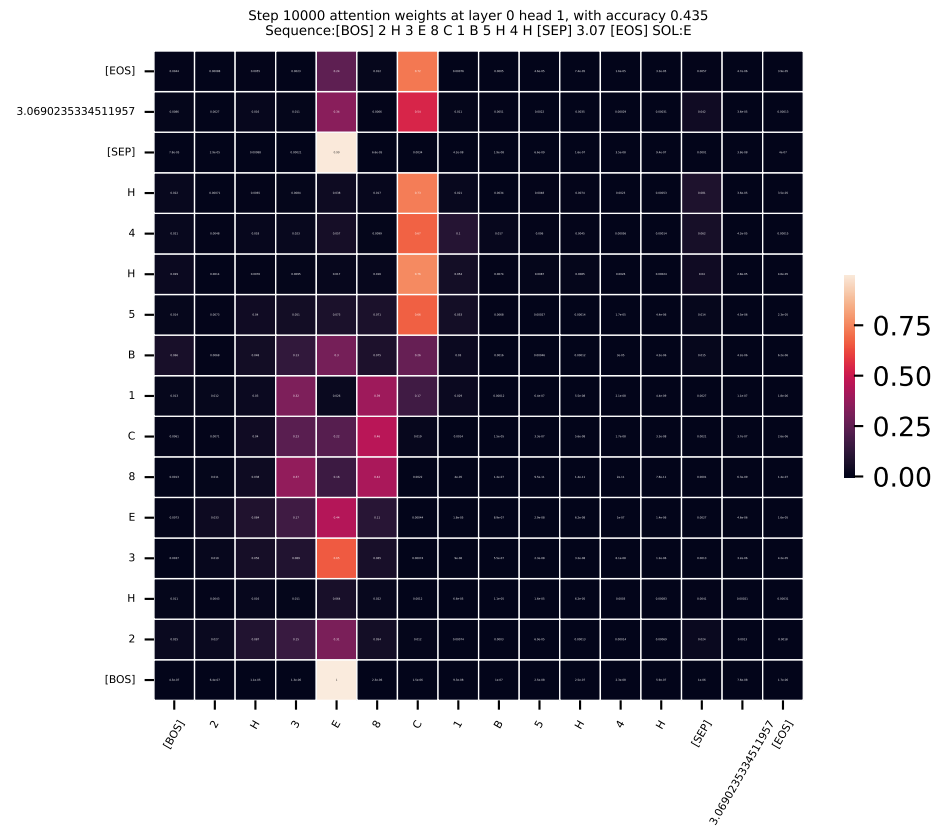

Step 10000 attention weights at layer 0 head 0, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

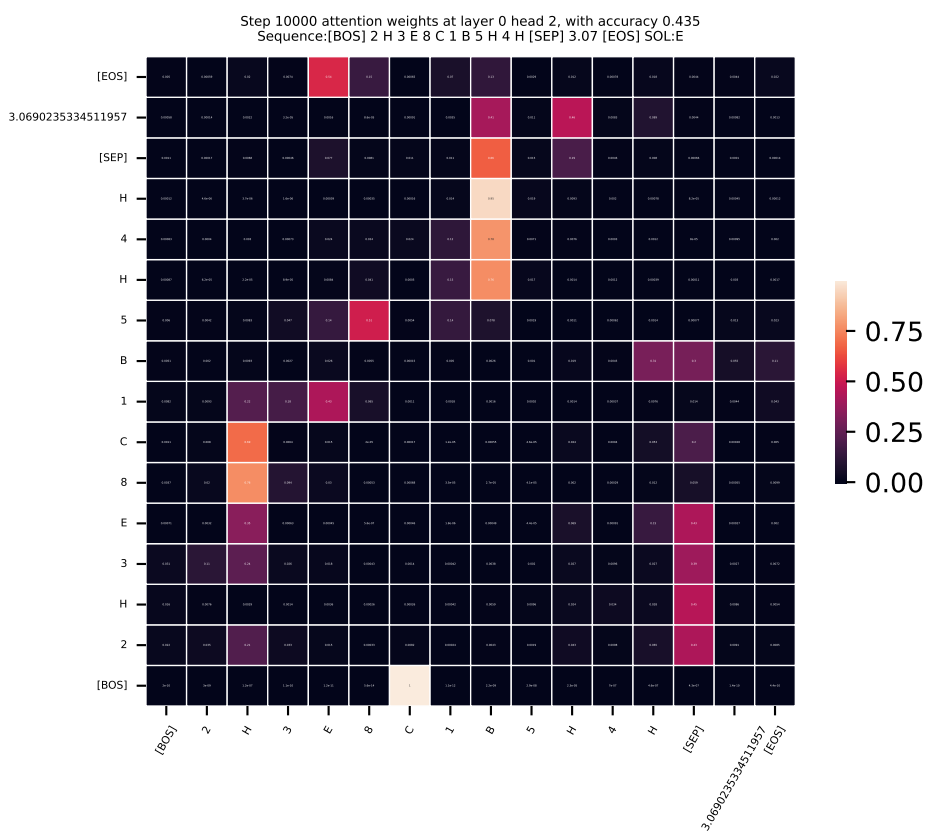

Step 10000 attention weights at layer 0 head 1, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

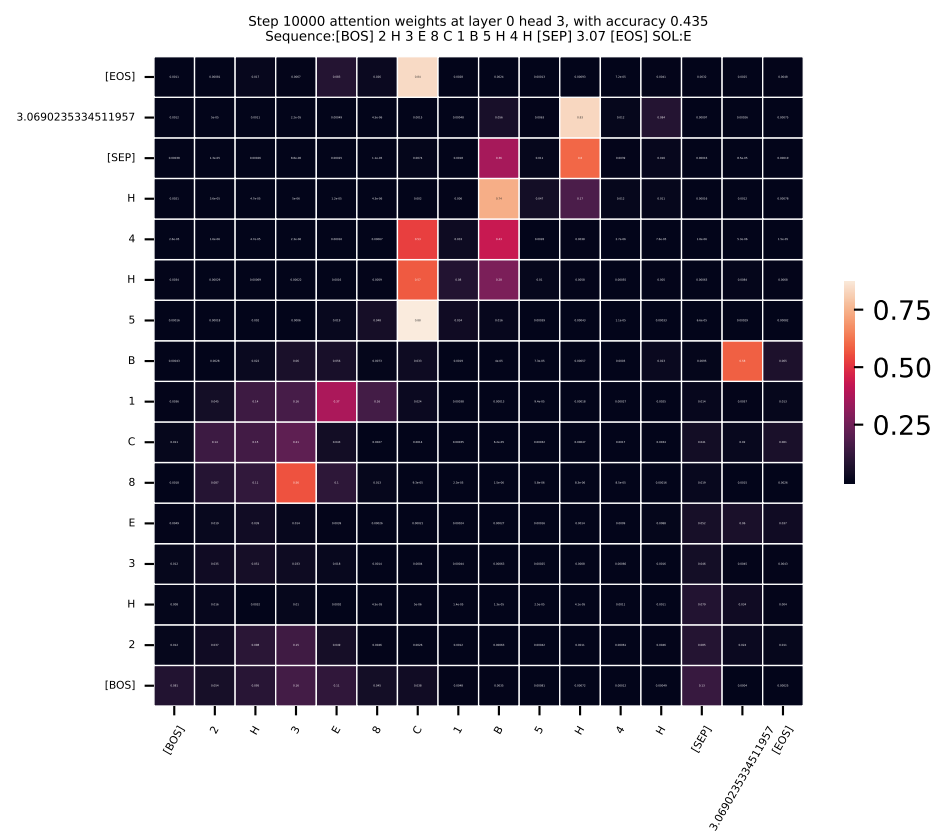

Step 10000 attention weights at layer 0 head 2, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

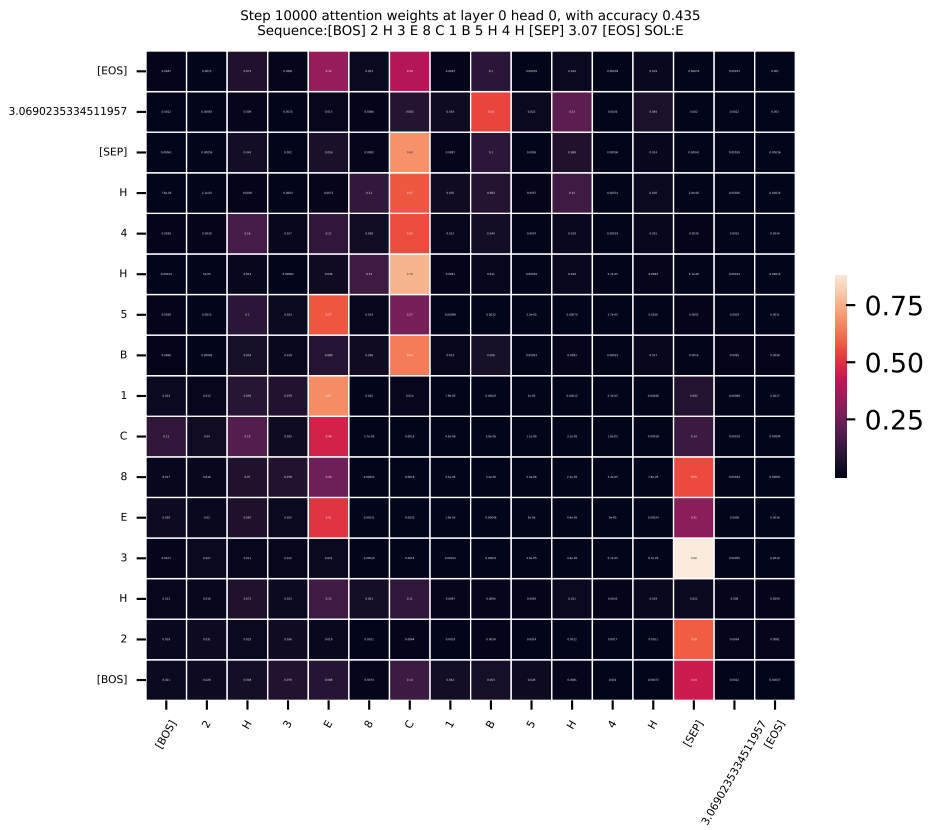

Step 10000 attention weights at layer 0 head 3, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

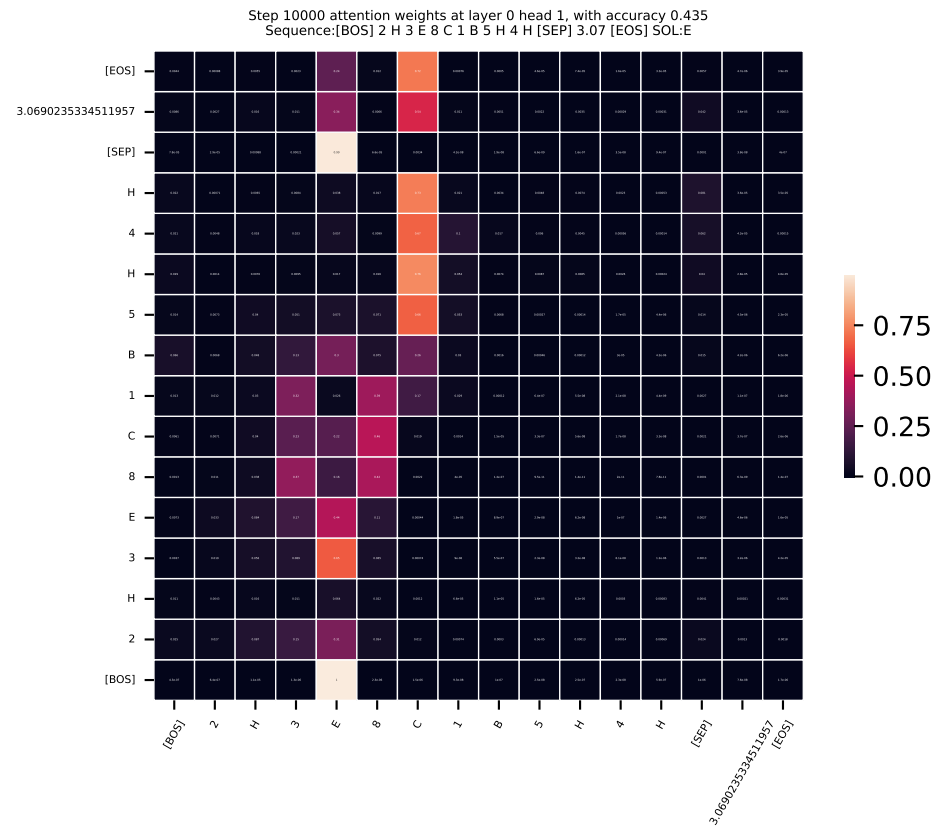

Step 10000 attention weights at layer 0 head 0, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

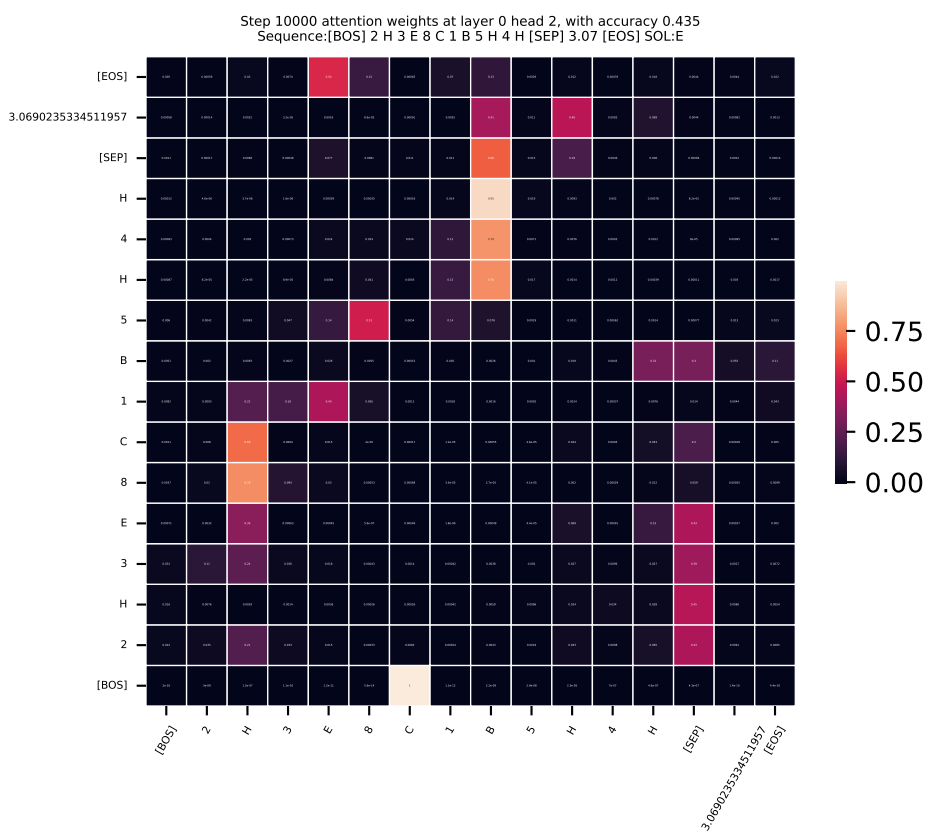

Step 10000 attention weights at layer 0 head 1, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

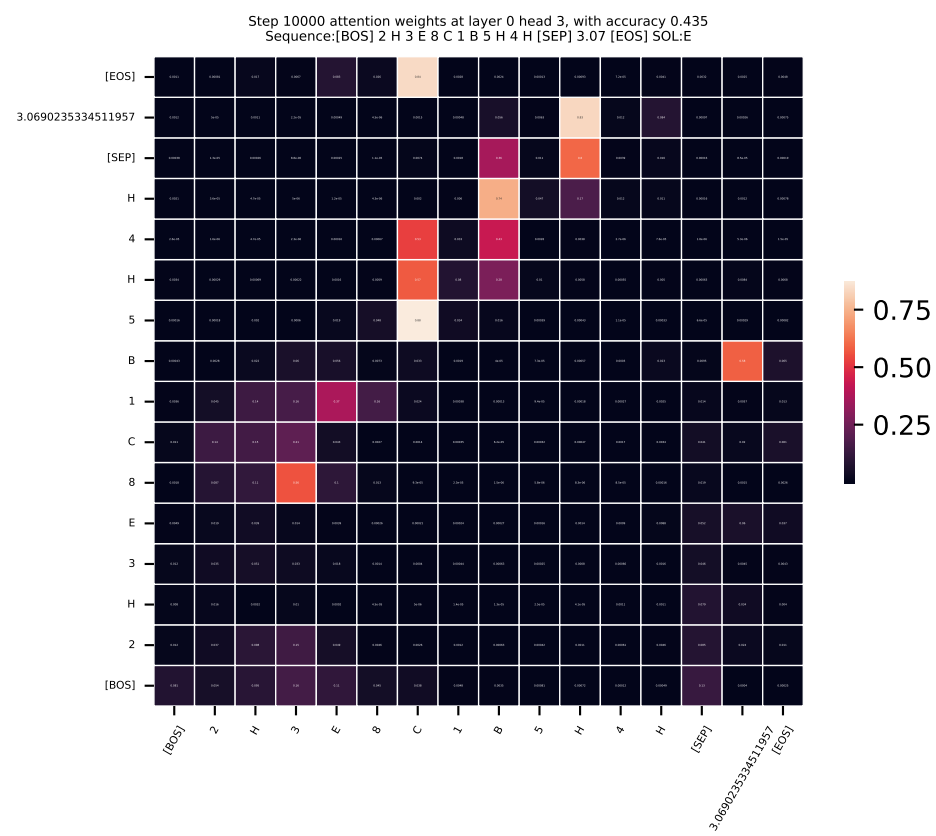

Step 10000 attention weights at layer 0 head 2, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

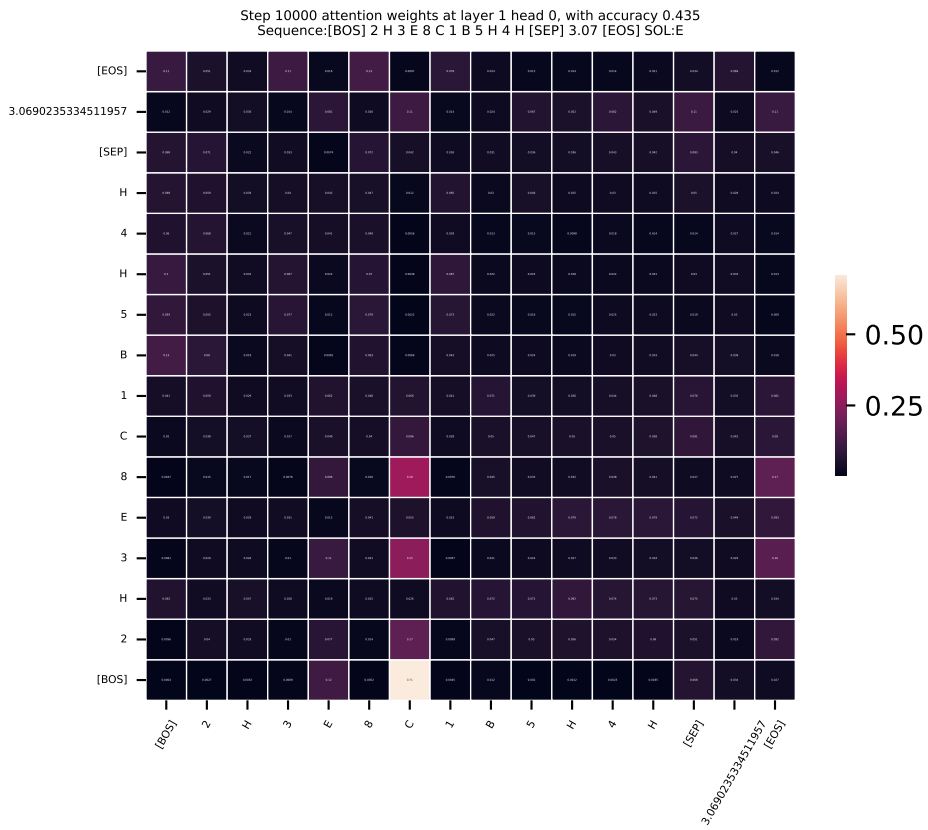

Step 10000 attention weights at layer 0 head 3, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

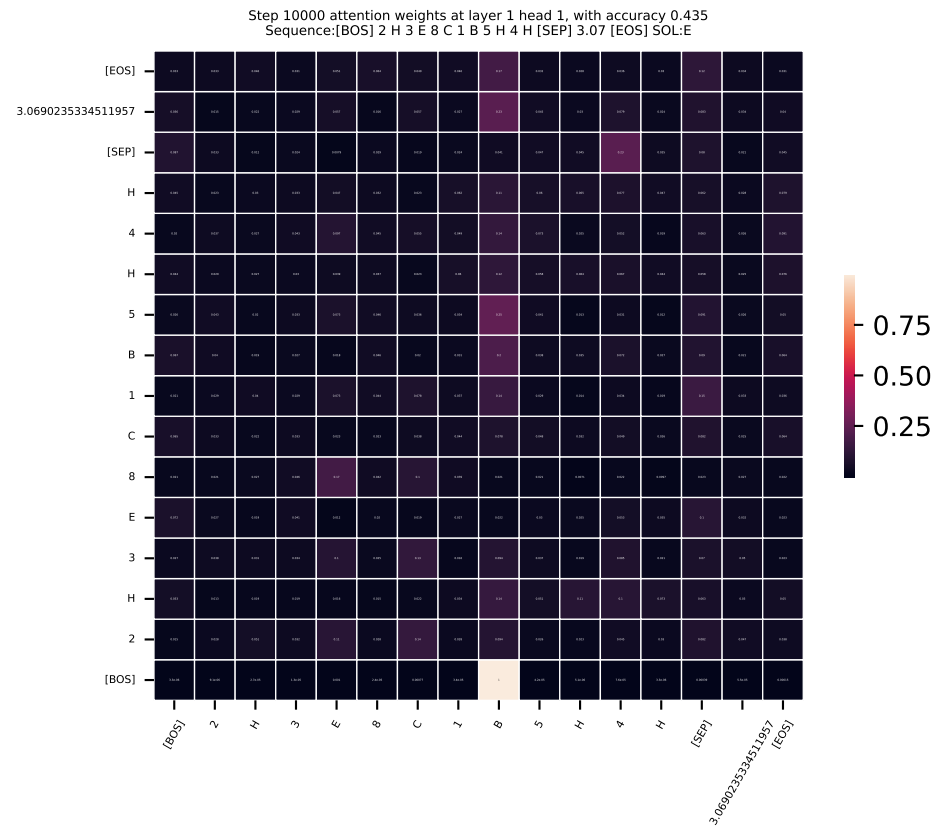

Step 10000 attention weights at layer 0 head 0, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

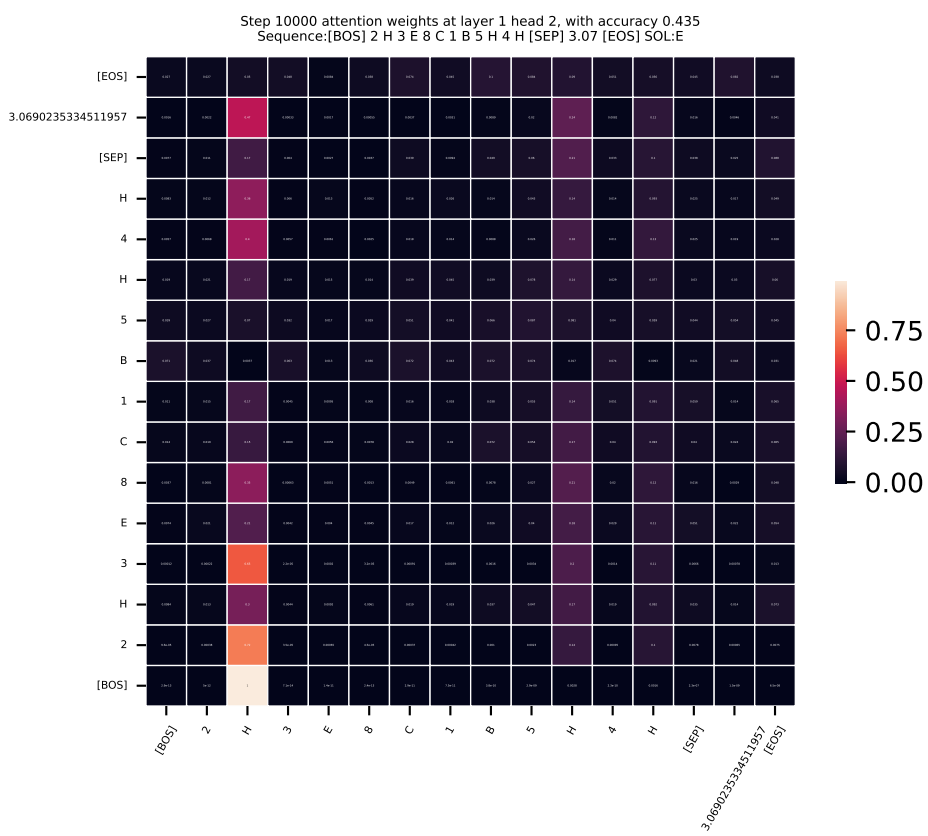

Step 10000 attention weights at layer 0 head 1, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

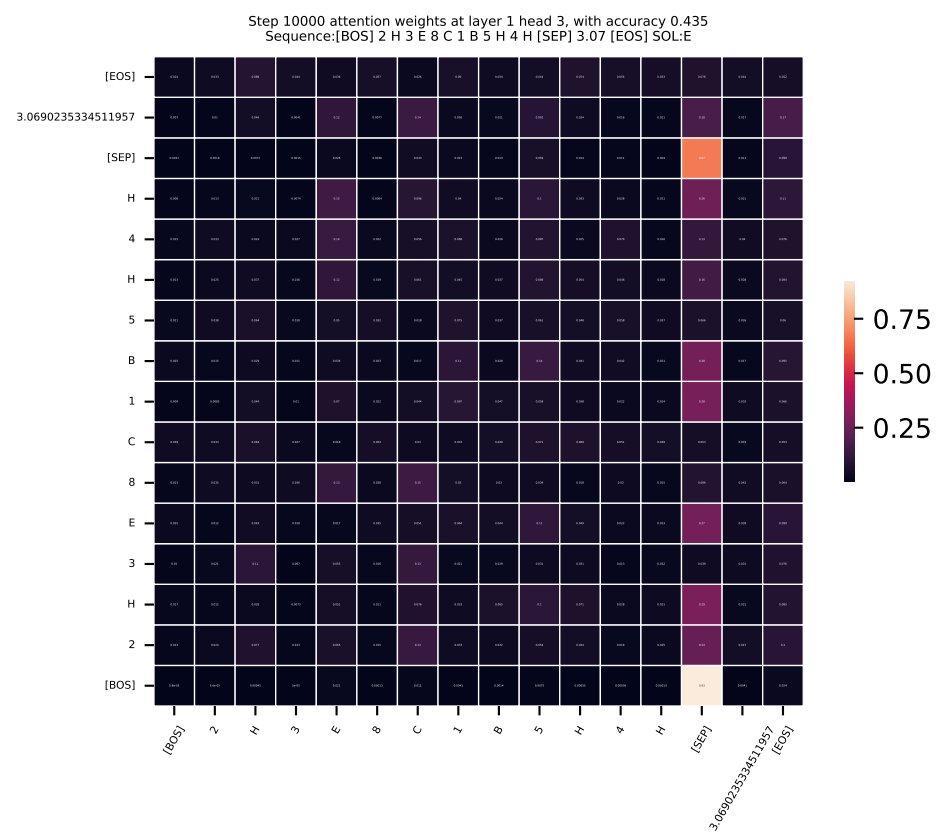

Step 10000 attention weights at layer 0 head 2, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

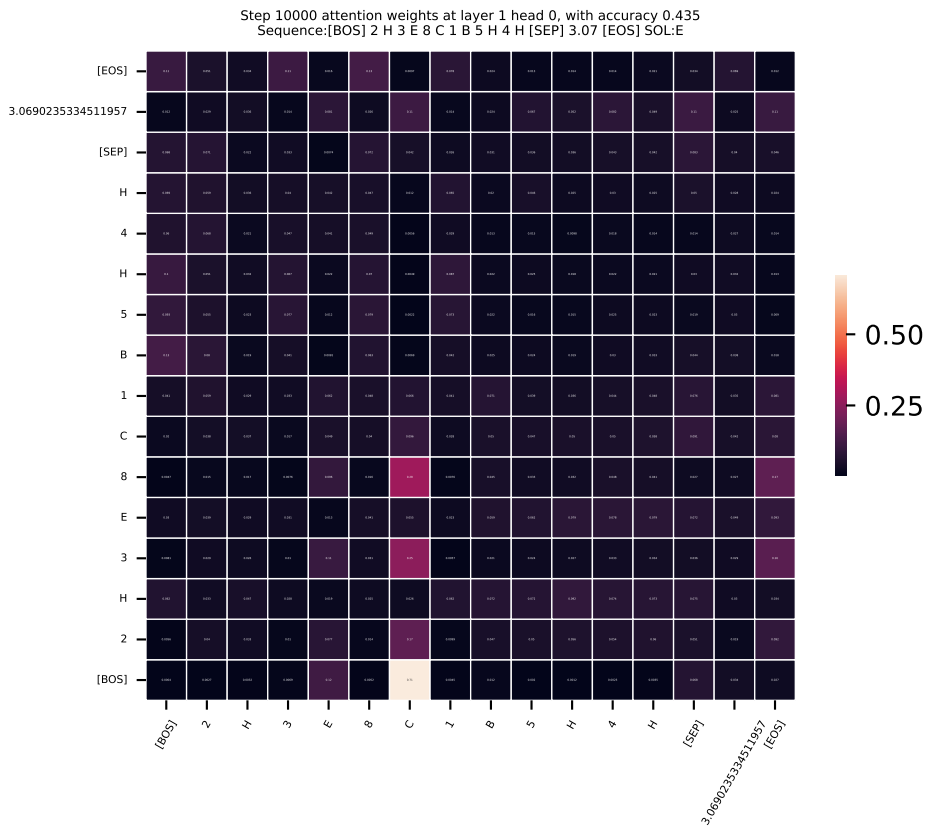

Step 10000 attention weights at layer 0 head 3, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

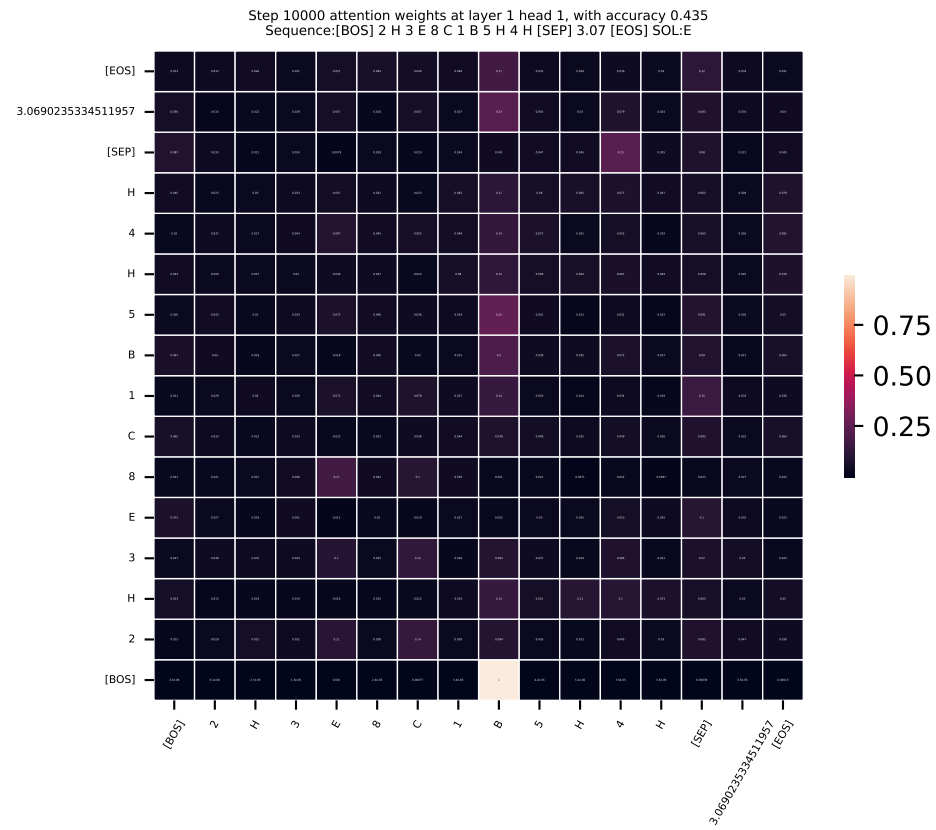

Step 10000 attention weights at layer 1 head 0, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.50
0.25

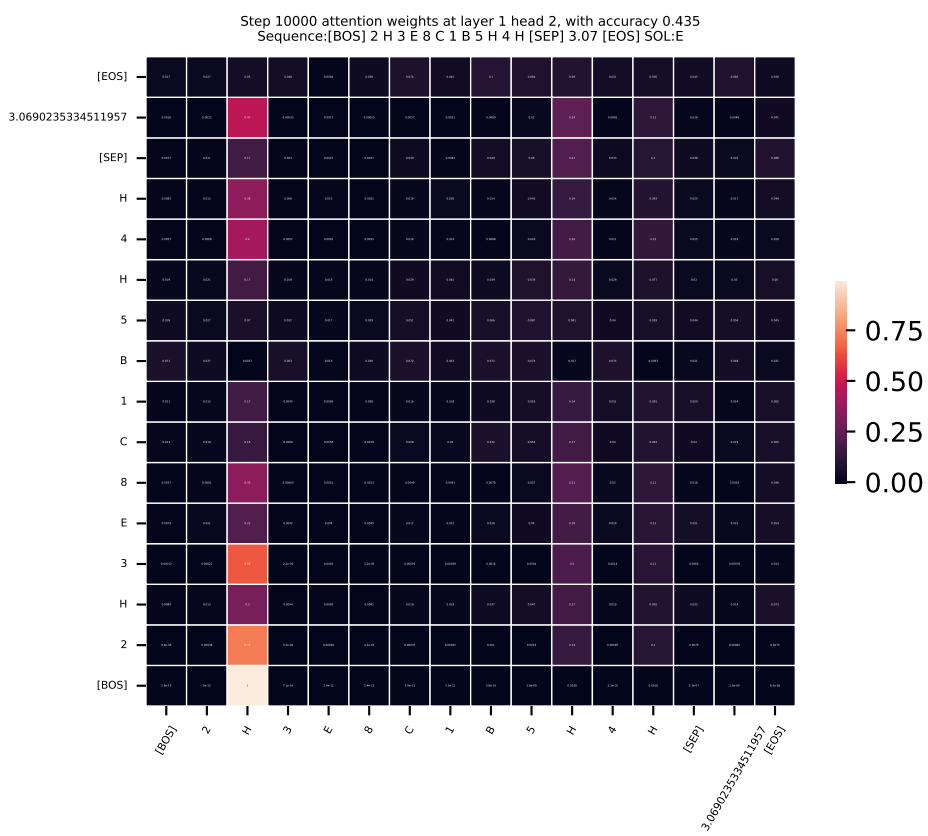

Step 10000 attention weights at layer 1 head 1, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

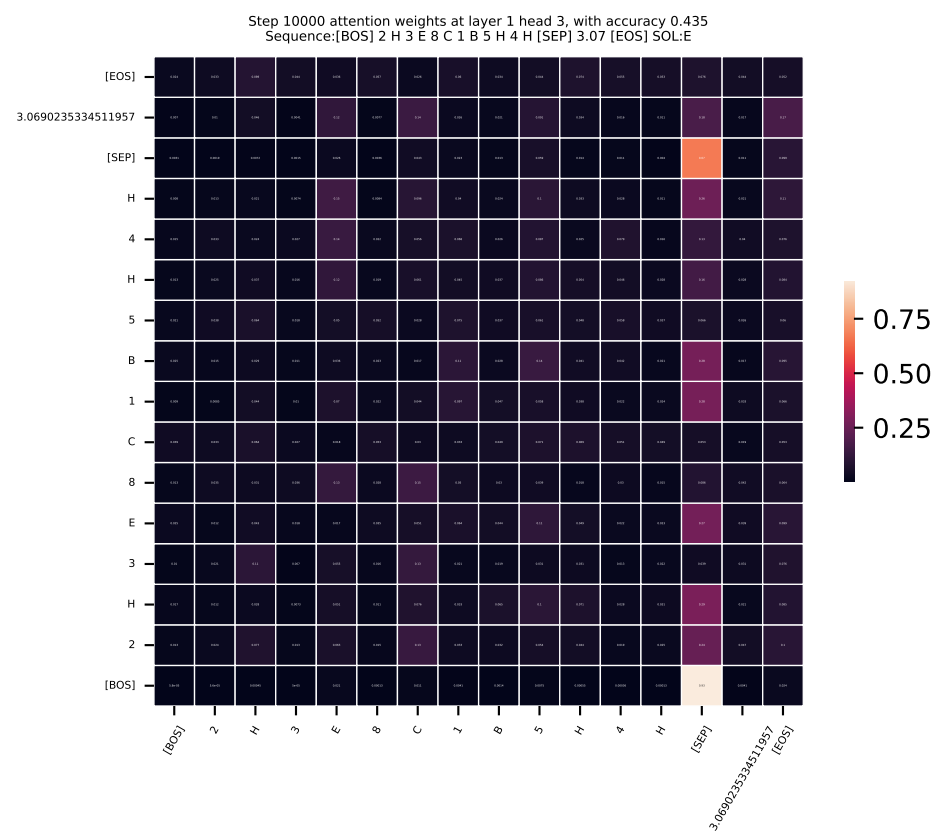

Step 10000 attention weights at layer 1 head 2, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

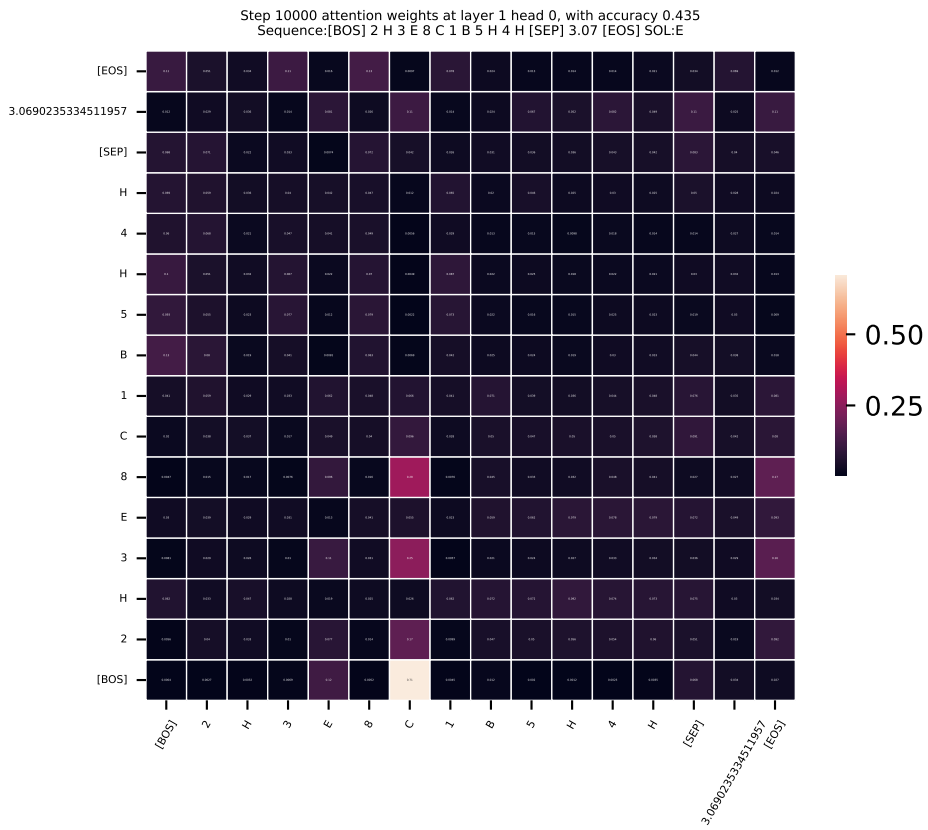

Step 10000 attention weights at layer 1 head 3, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

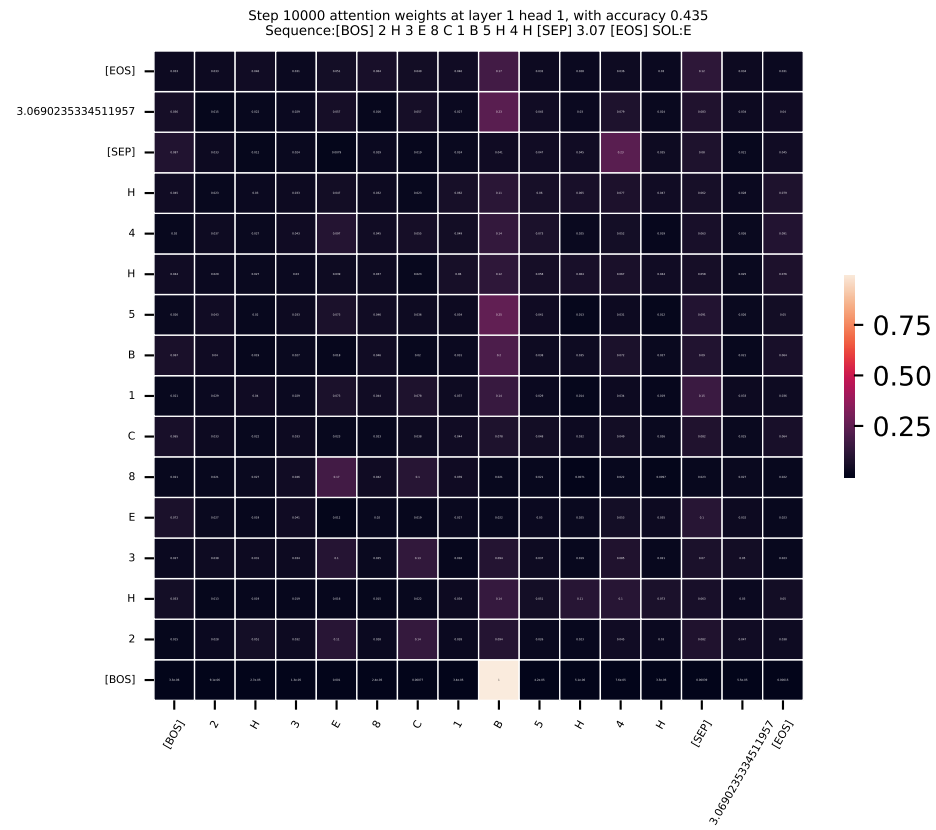

Step 10000 attention weights at layer 1 head 0, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.50
0.25

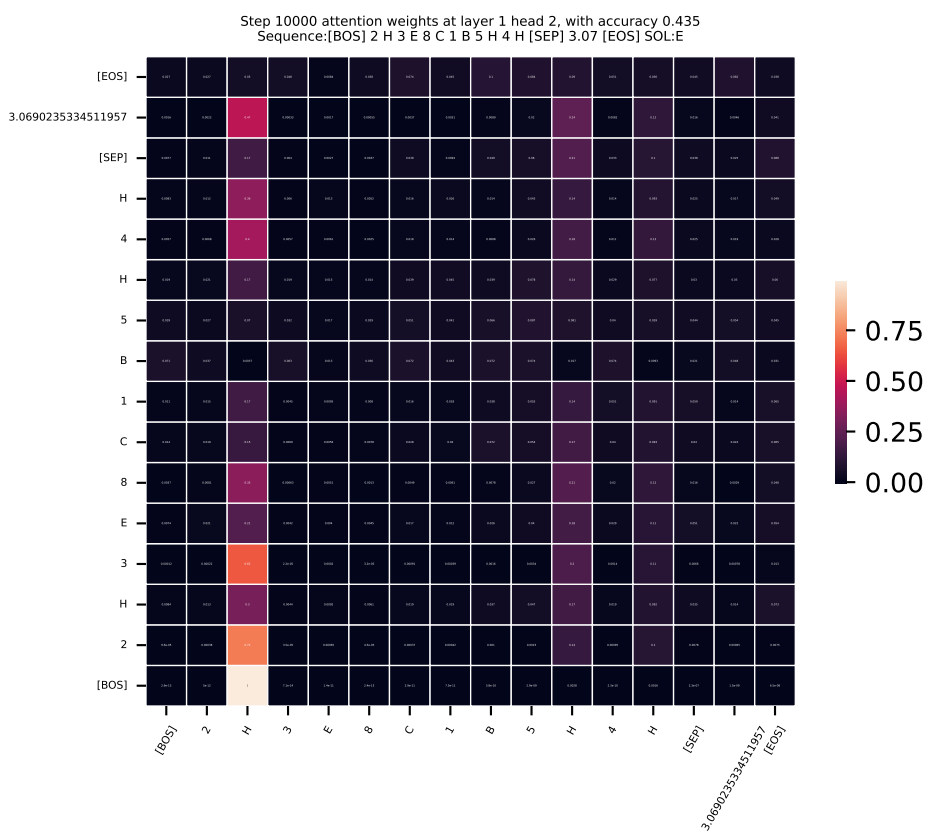

Step 10000 attention weights at layer 1 head 1, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

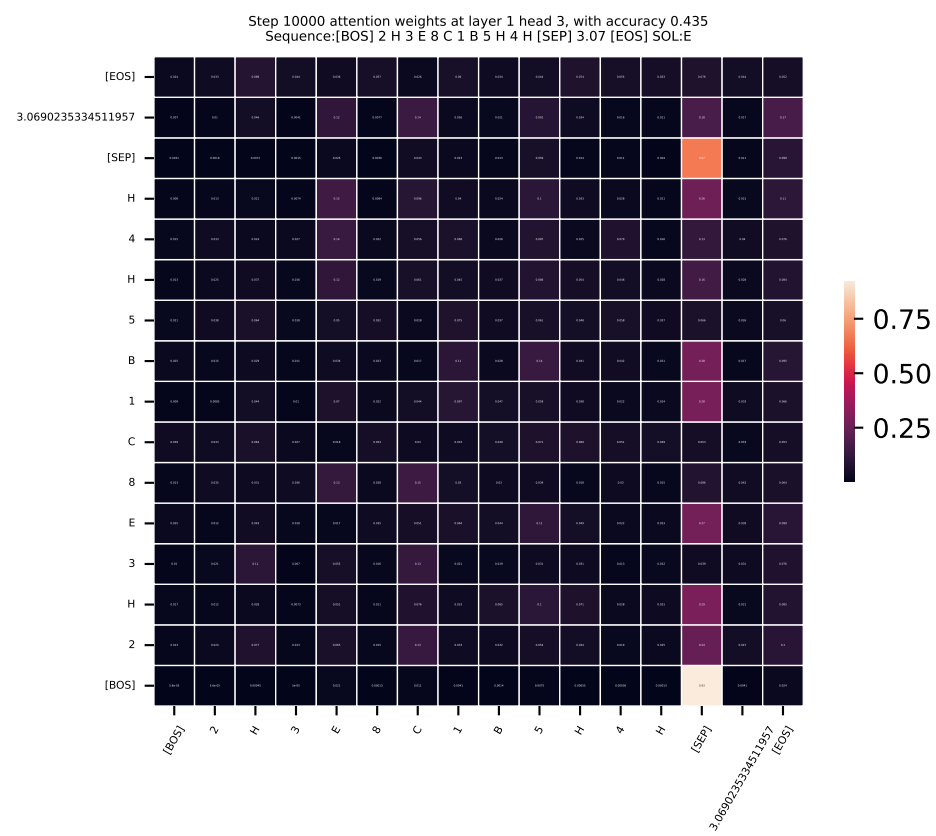

Step 10000 attention weights at layer 1 head 2, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

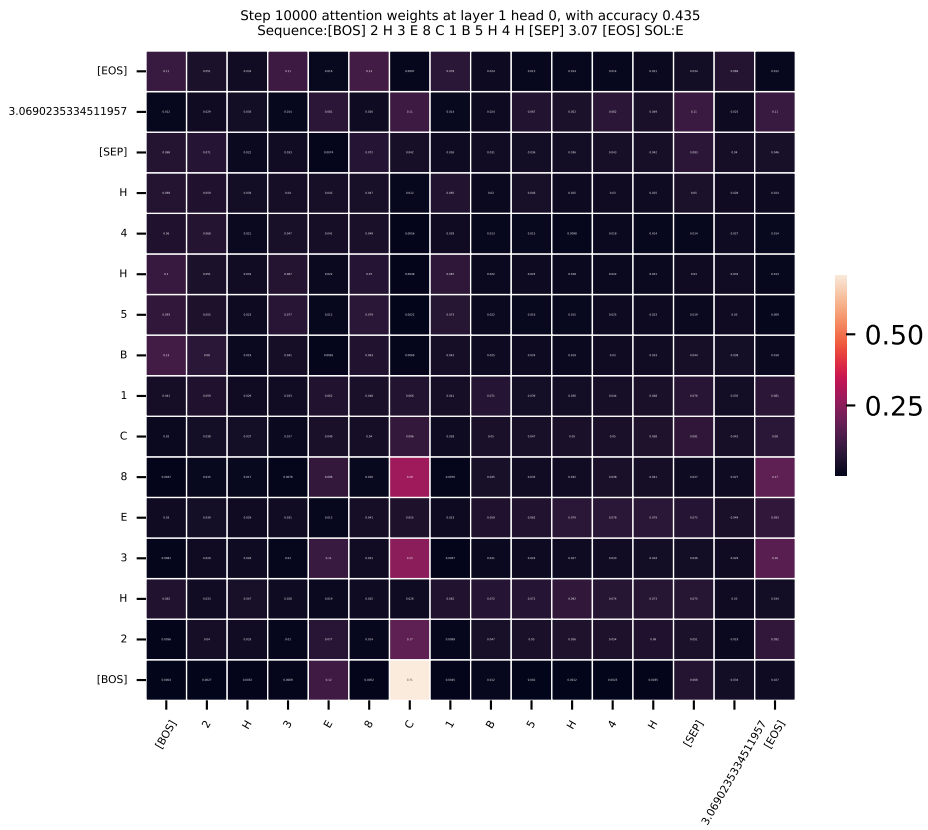

Step 10000 attention weights at layer 1 head 3, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

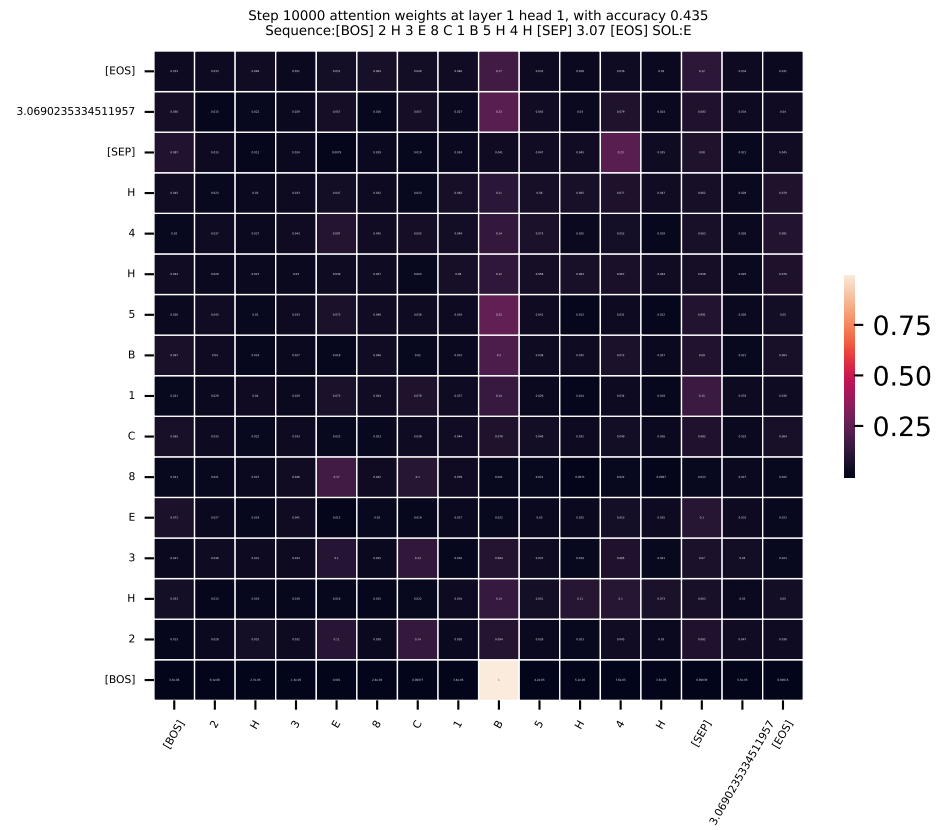

Step 10000 attention weights at layer 1 head 0, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.50
0.25

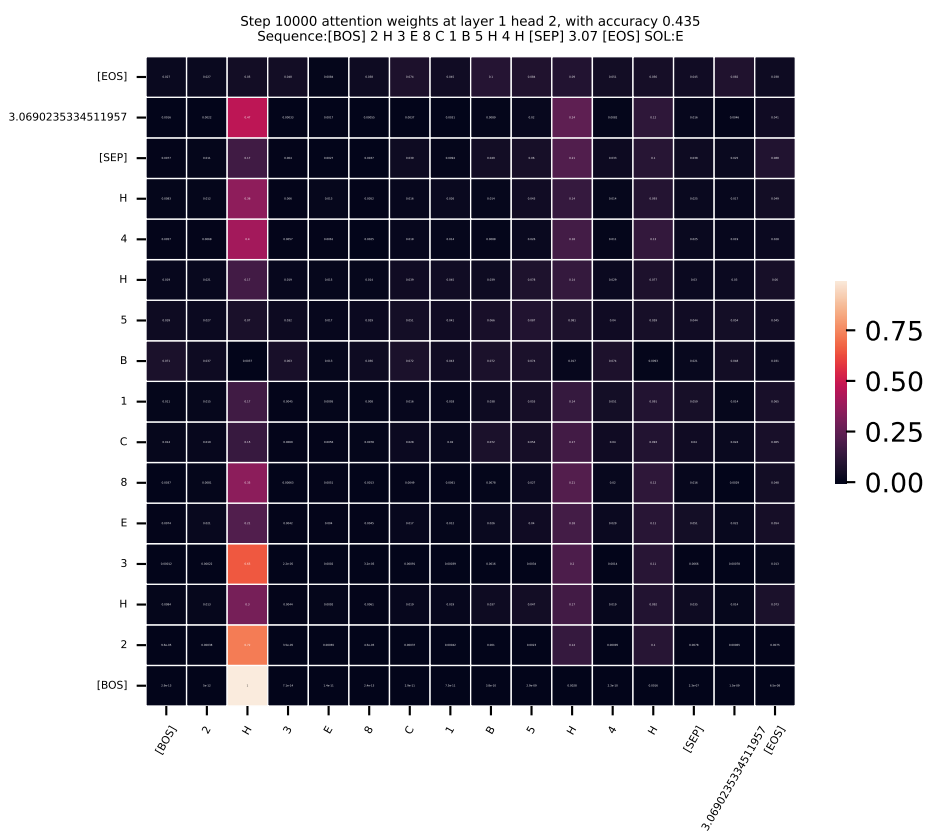

Step 10000 attention weights at layer 1 head 1, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

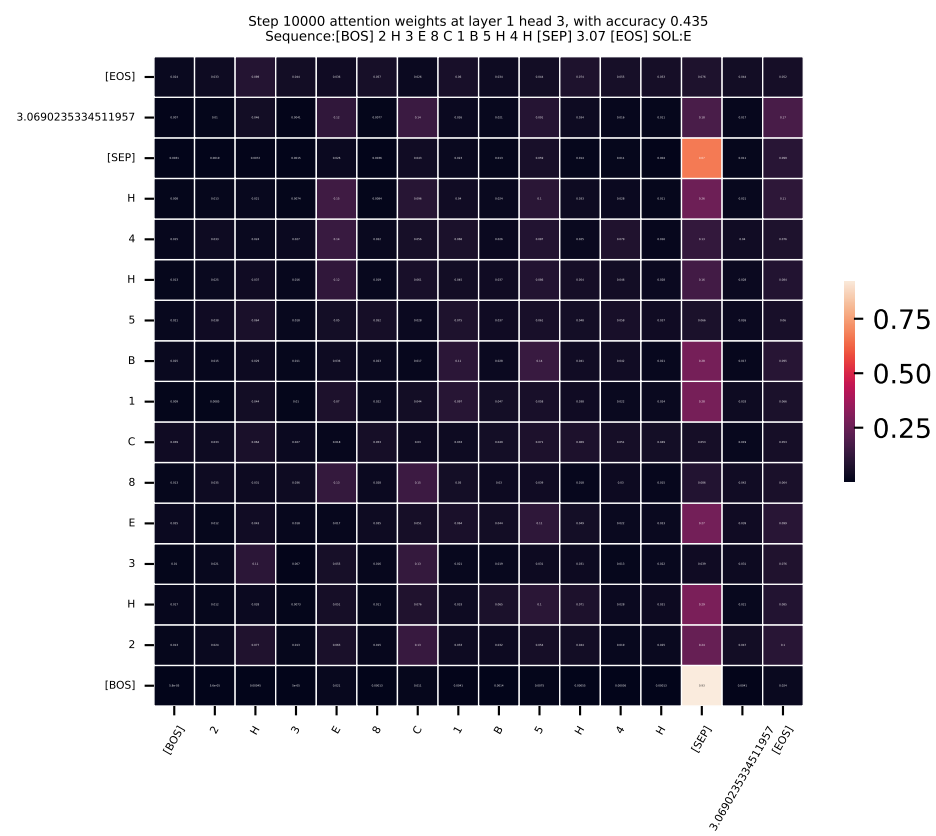

Step 10000 attention weights at layer 1 head 2, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

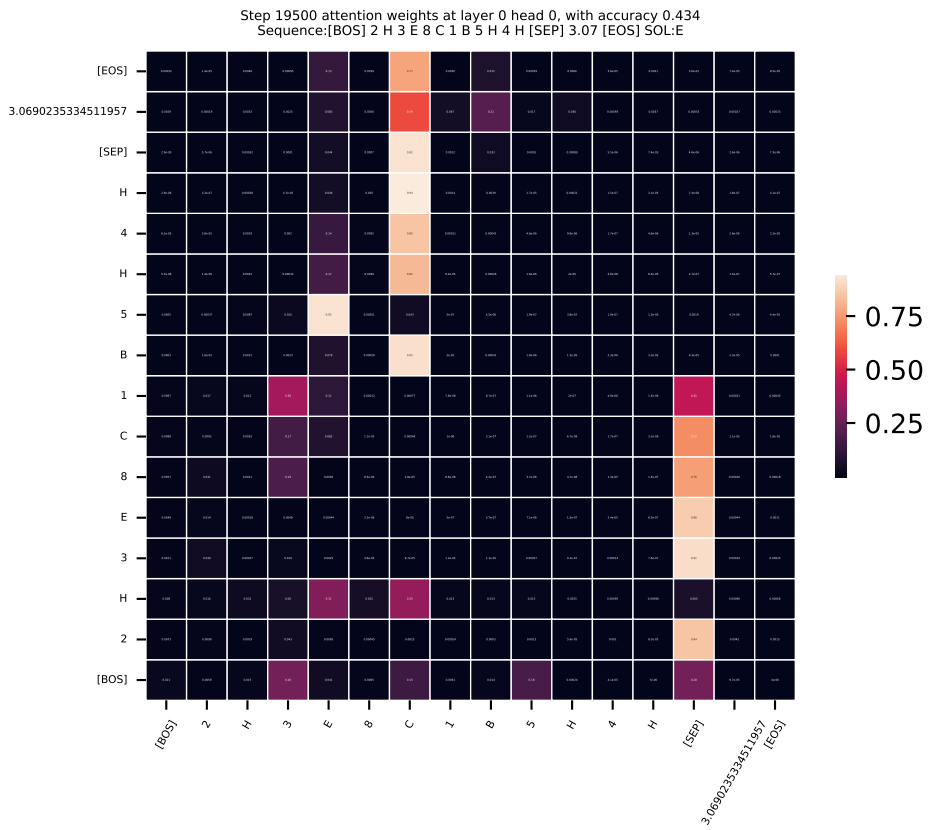

Step 10000 attention weights at layer 1 head 3, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

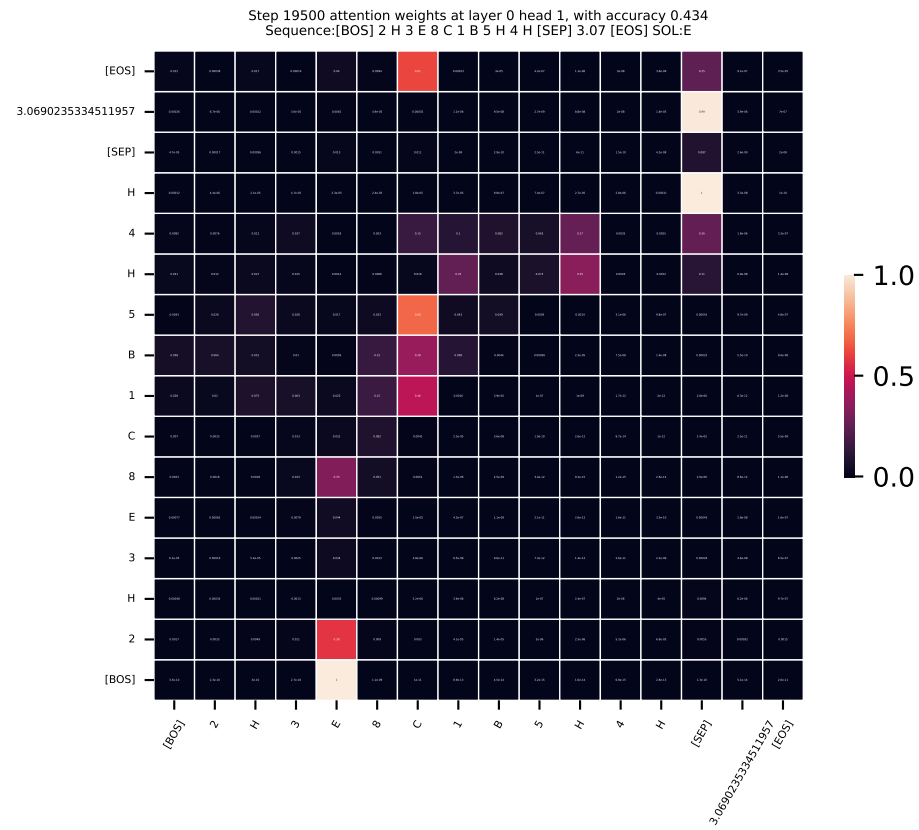

Step 10000 attention weights at layer 1 head 0, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.50
0.25

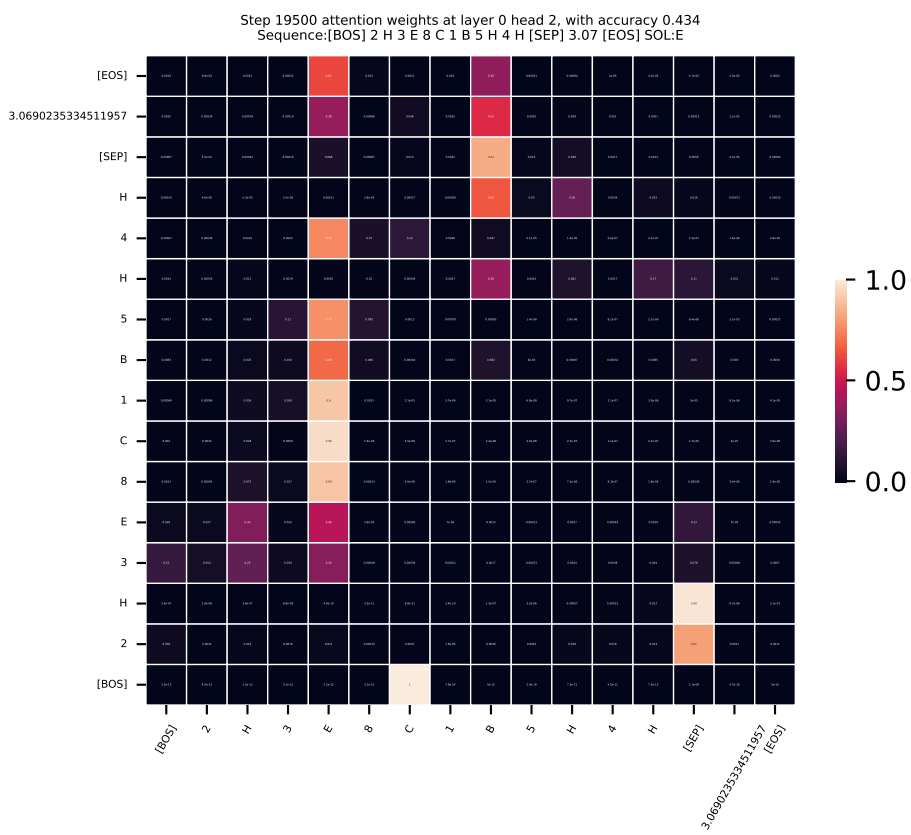

Step 10000 attention weights at layer 1 head 1, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

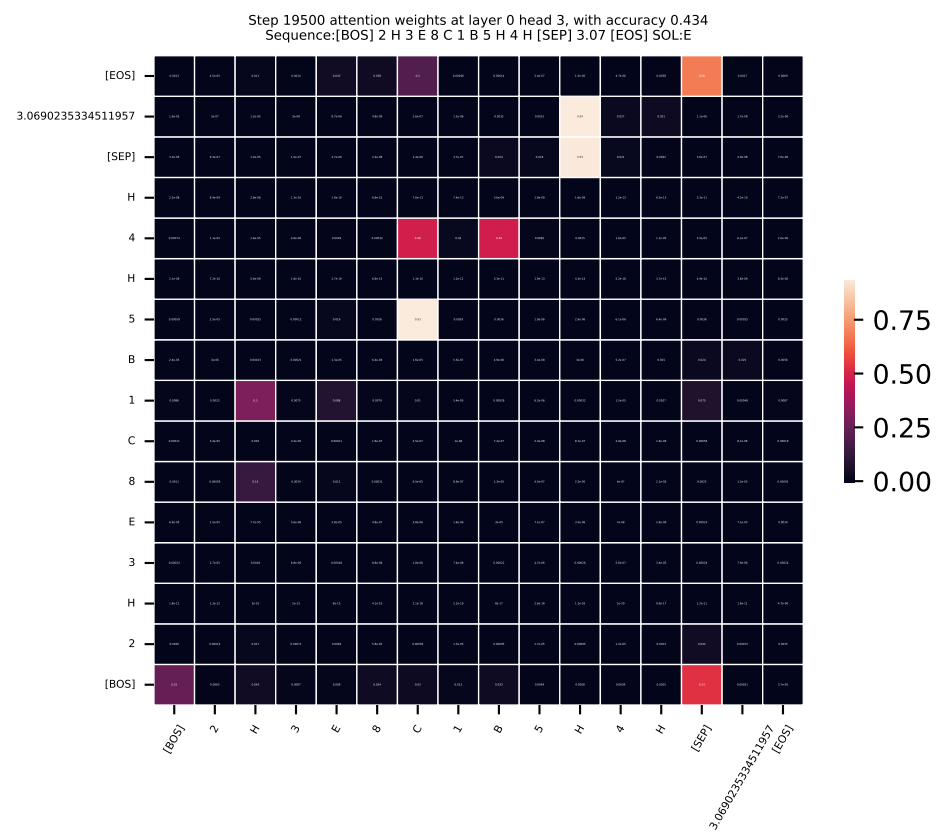

Step 10000 attention weights at layer 1 head 2, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

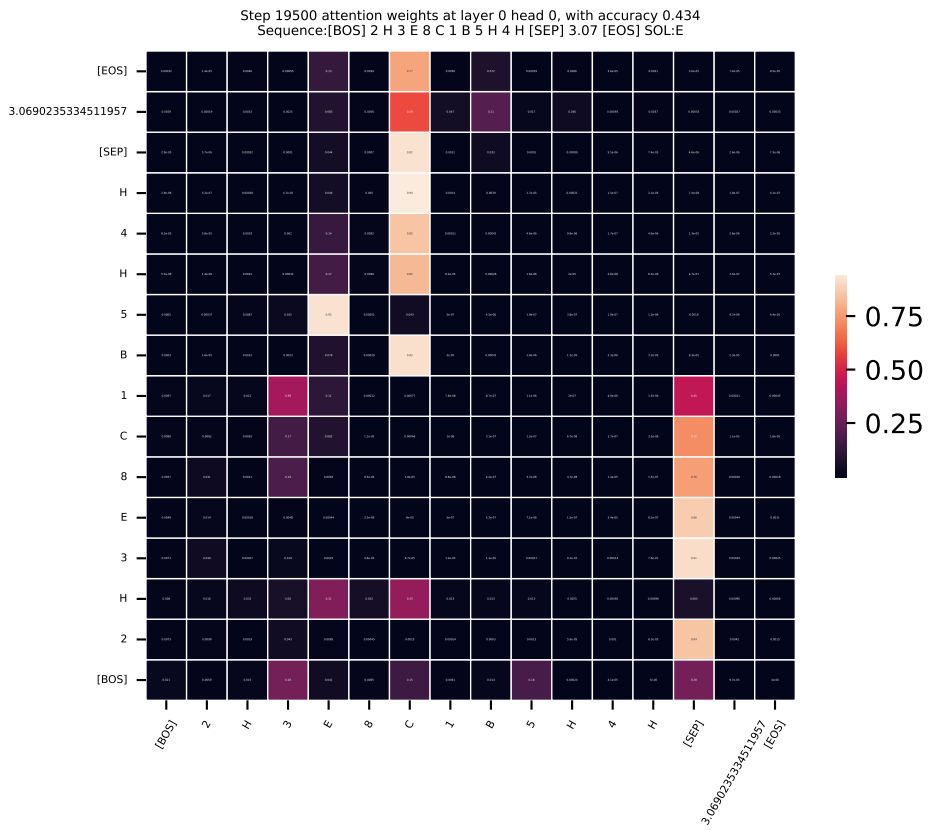

Step 10000 attention weights at layer 1 head 3, with accuracy 0.435
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

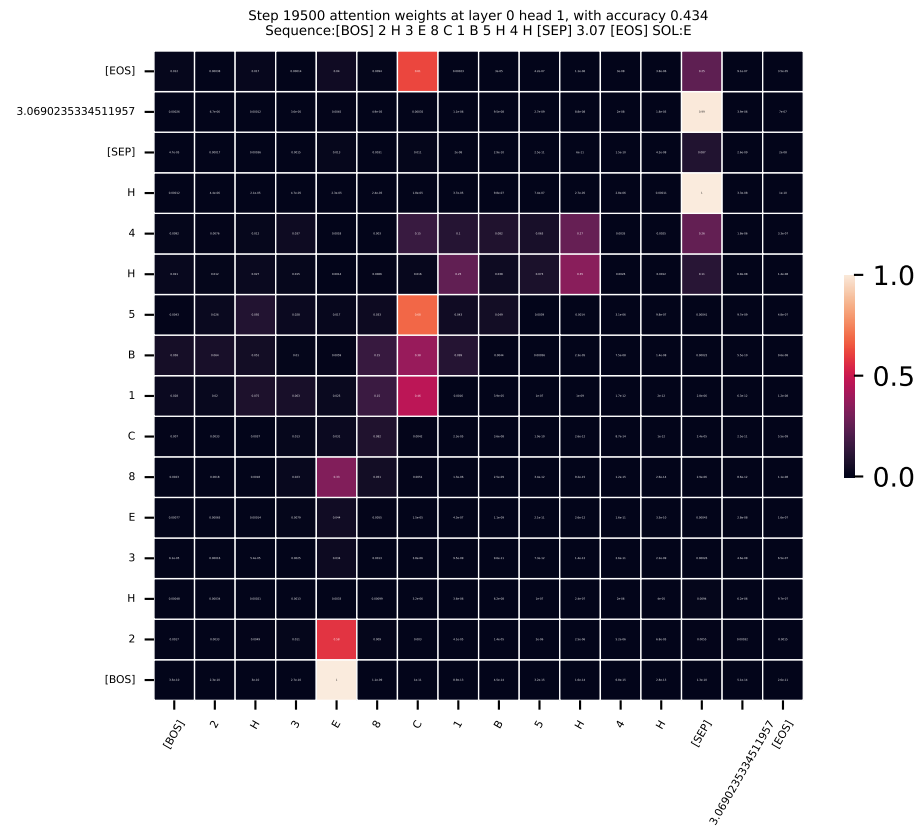

Step 19500 attention weights at layer 0 head 0, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

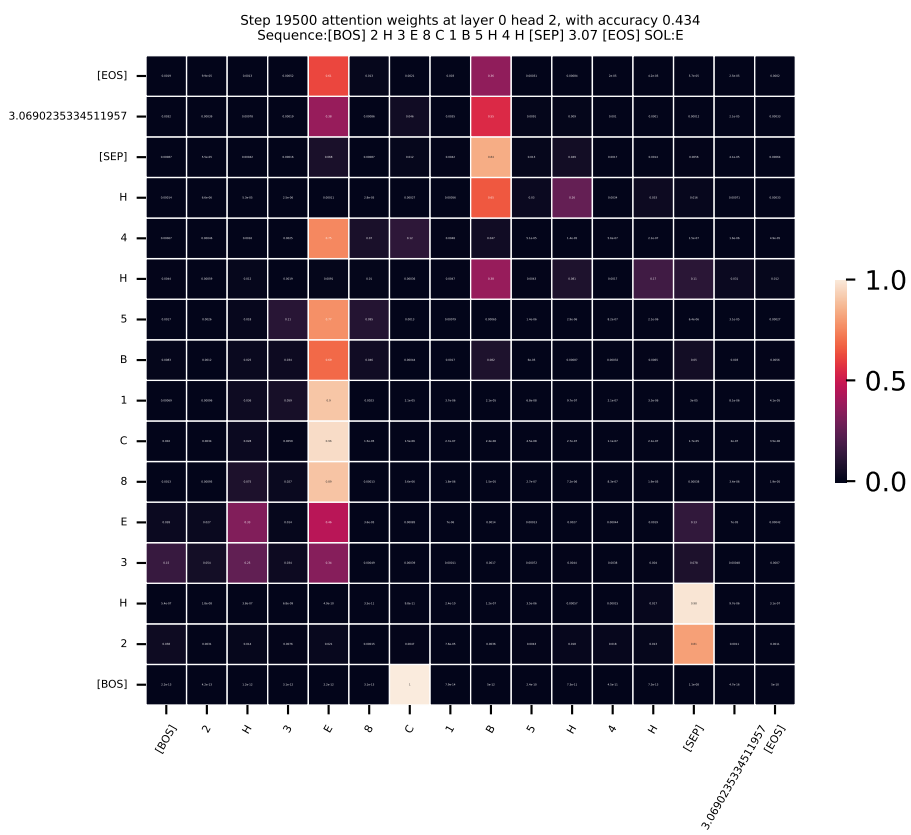

Step 19500 attention weights at layer 0 head 1, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

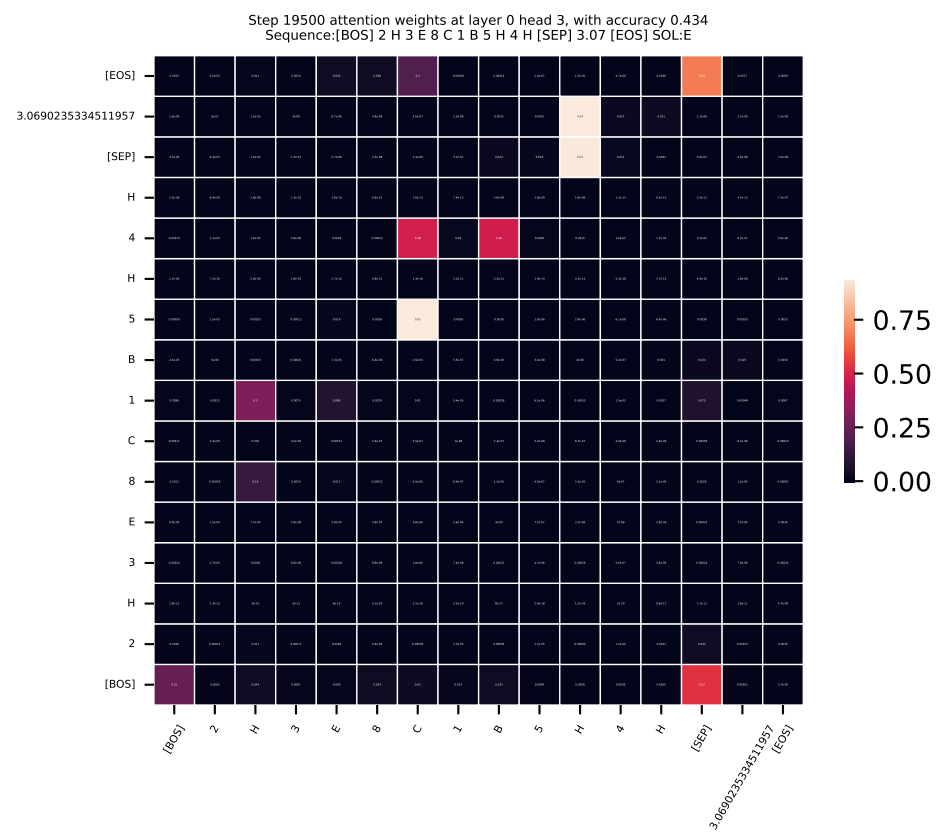

Step 19500 attention weights at layer 0 head 2, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

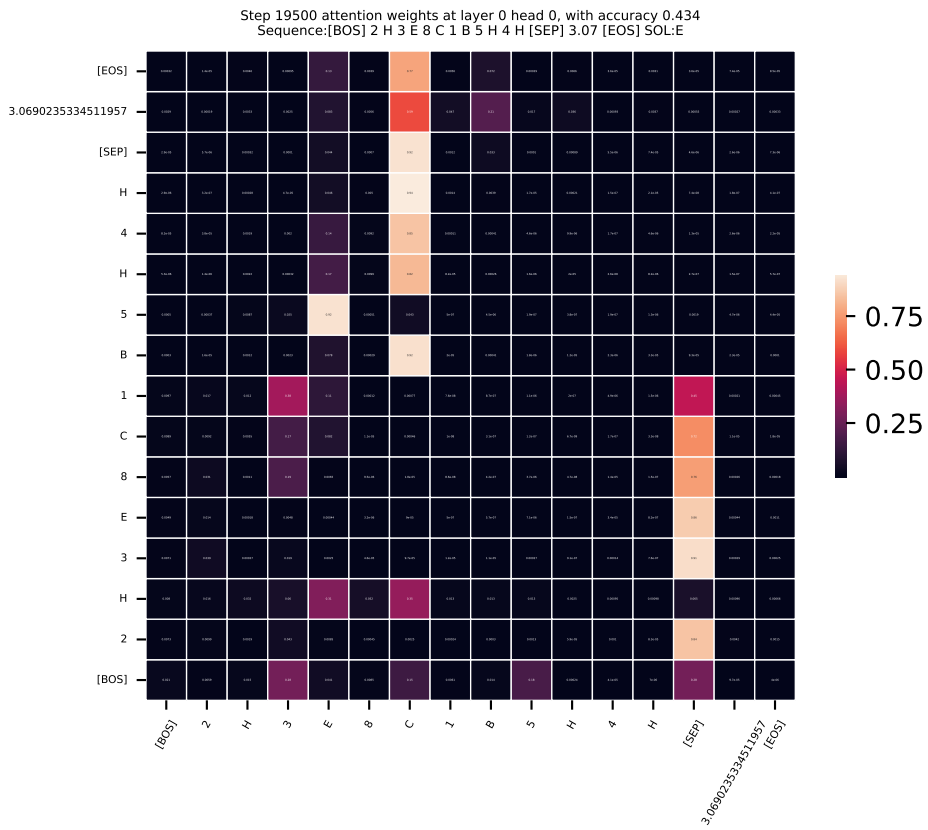

Step 19500 attention weights at layer 0 head 3, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

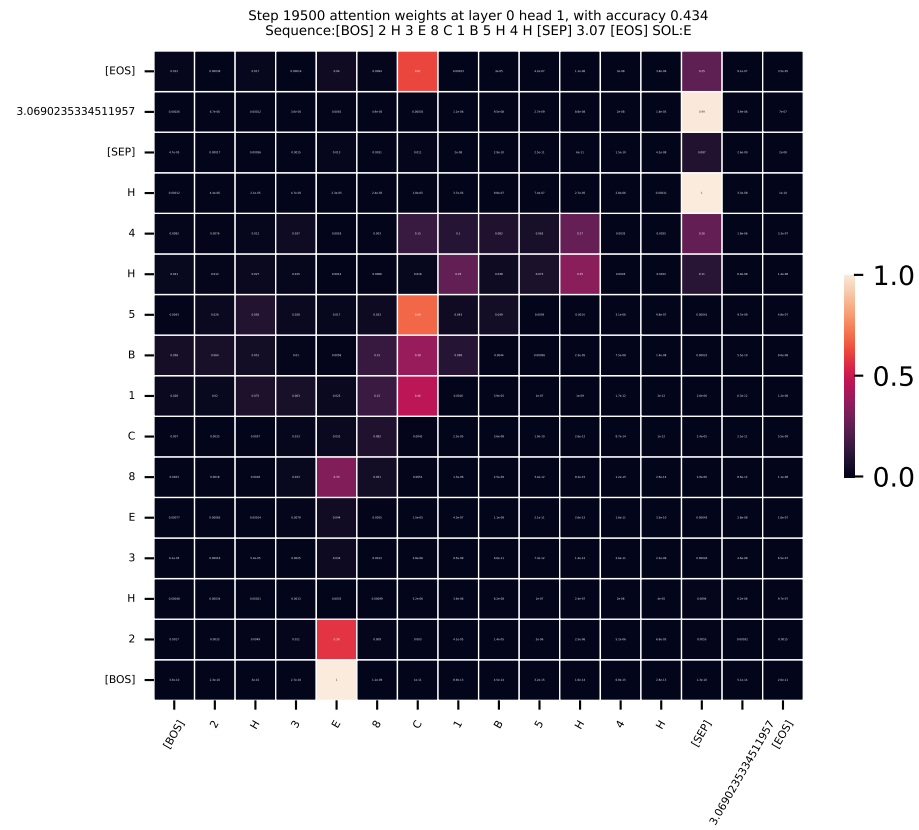

Step 19500 attention weights at layer 0 head 0, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

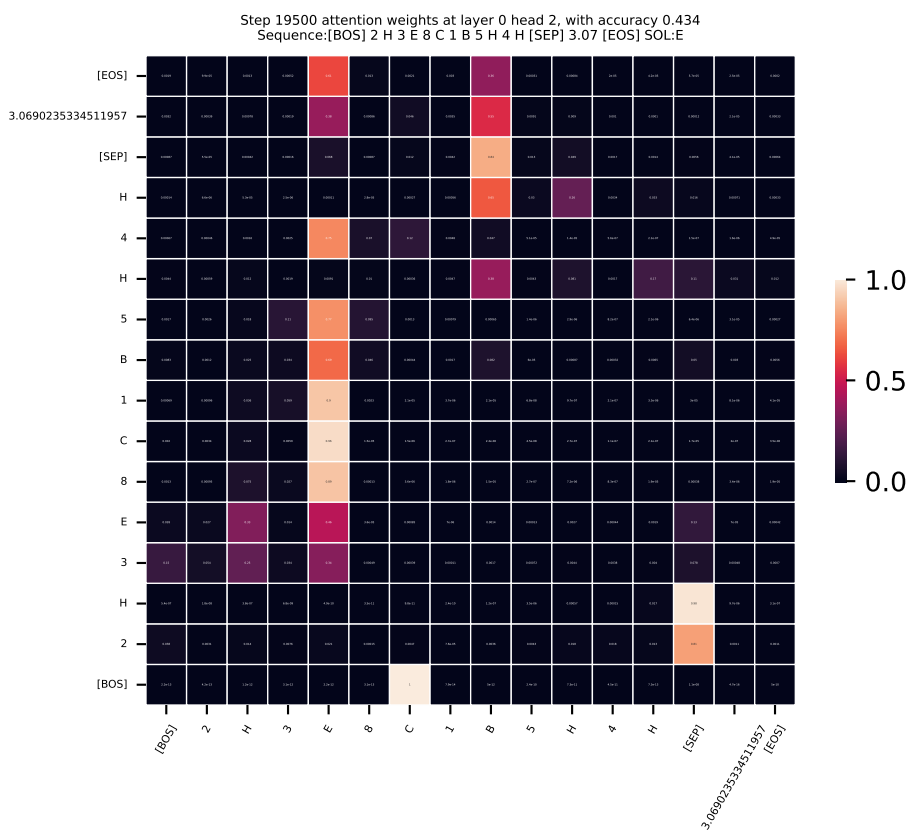

Step 19500 attention weights at layer 0 head 1, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

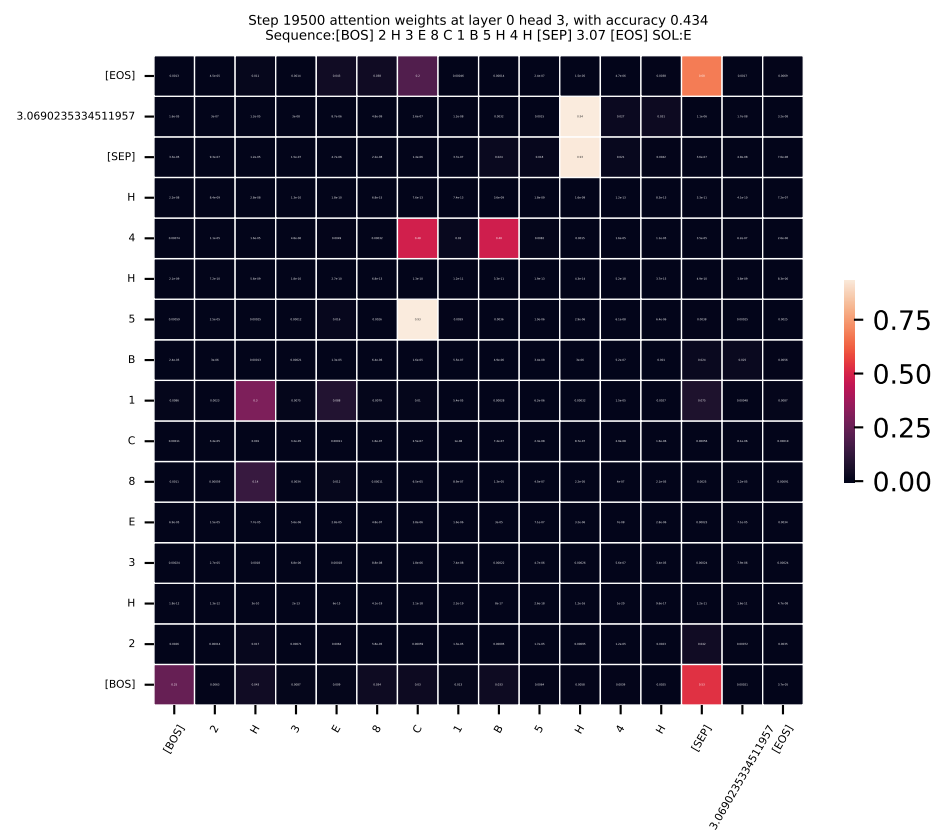

Step 19500 attention weights at layer 0 head 2, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

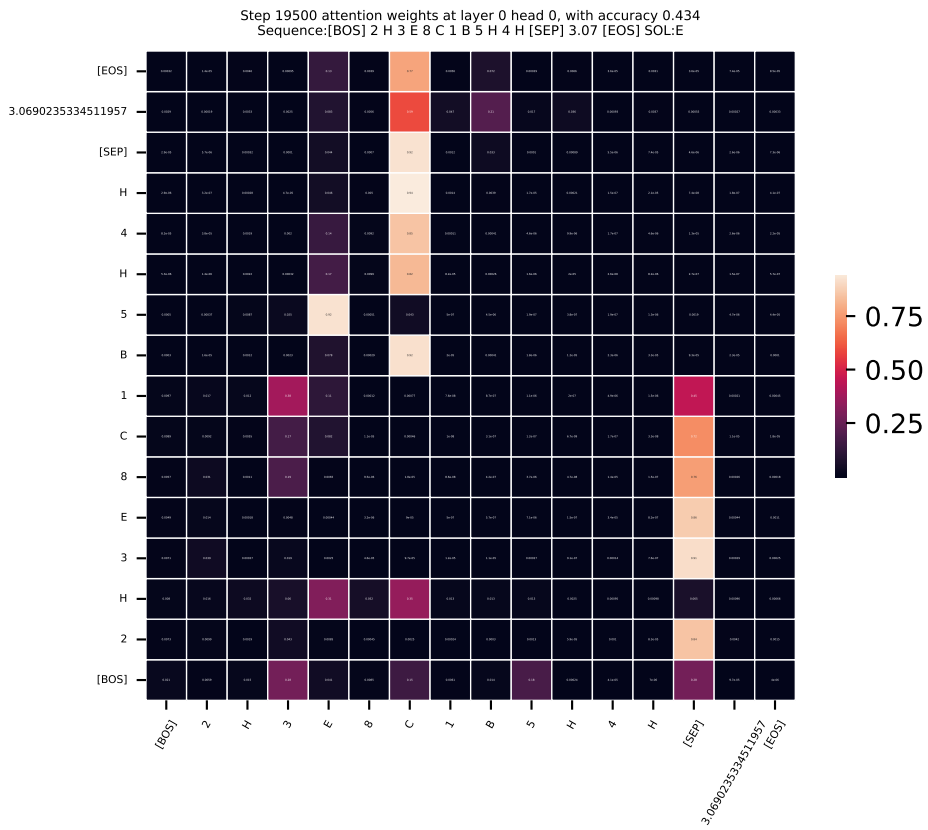

Step 19500 attention weights at layer 0 head 3, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

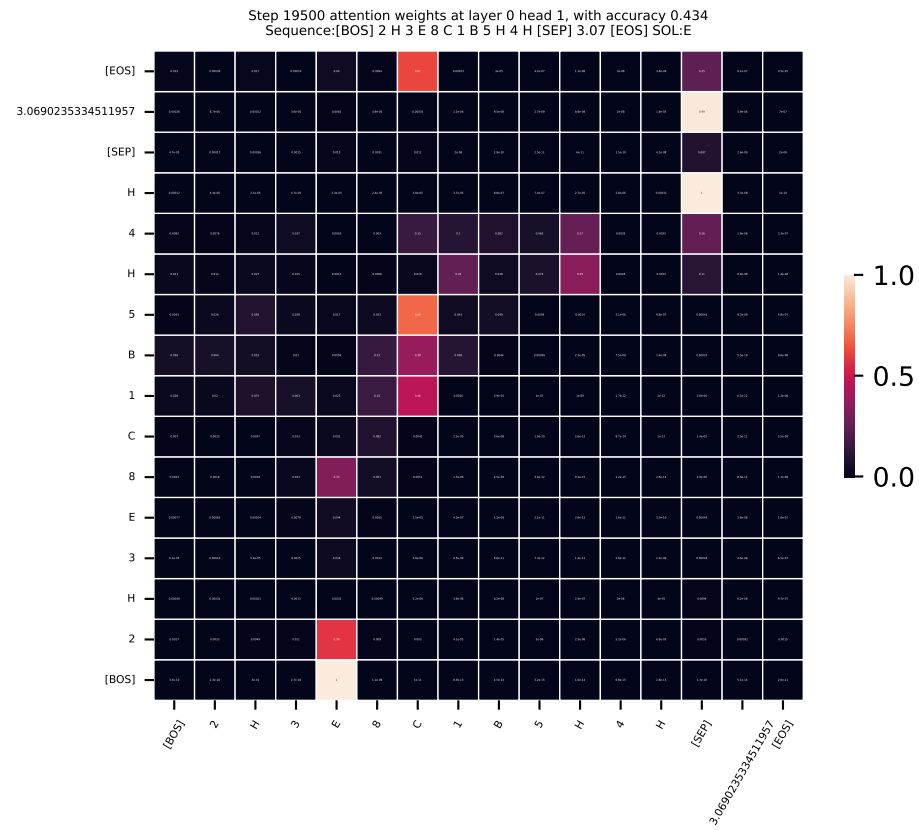

Step 19500 attention weights at layer 0 head 0, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

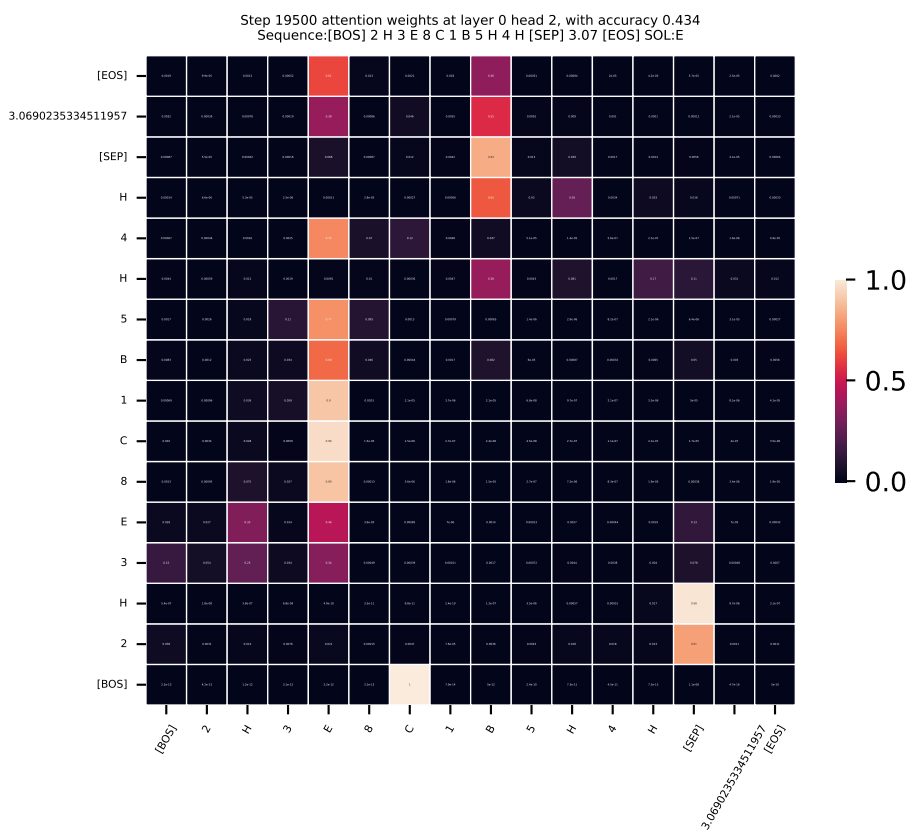

Step 19500 attention weights at layer 0 head 1, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

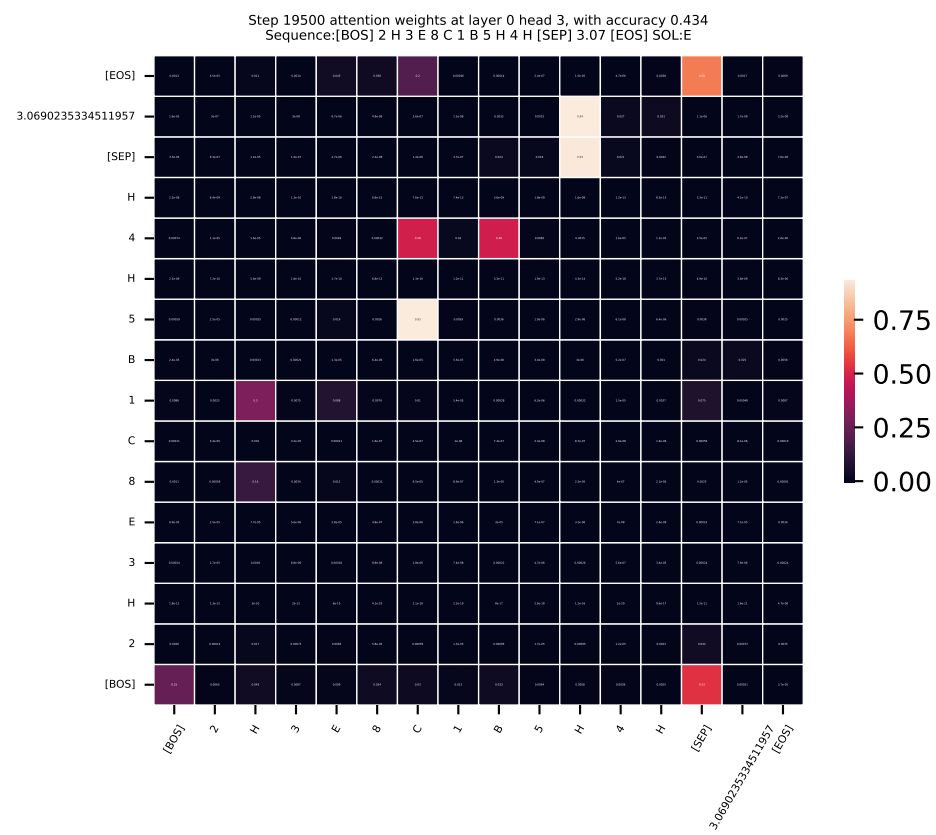

Step 19500 attention weights at layer 0 head 2, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

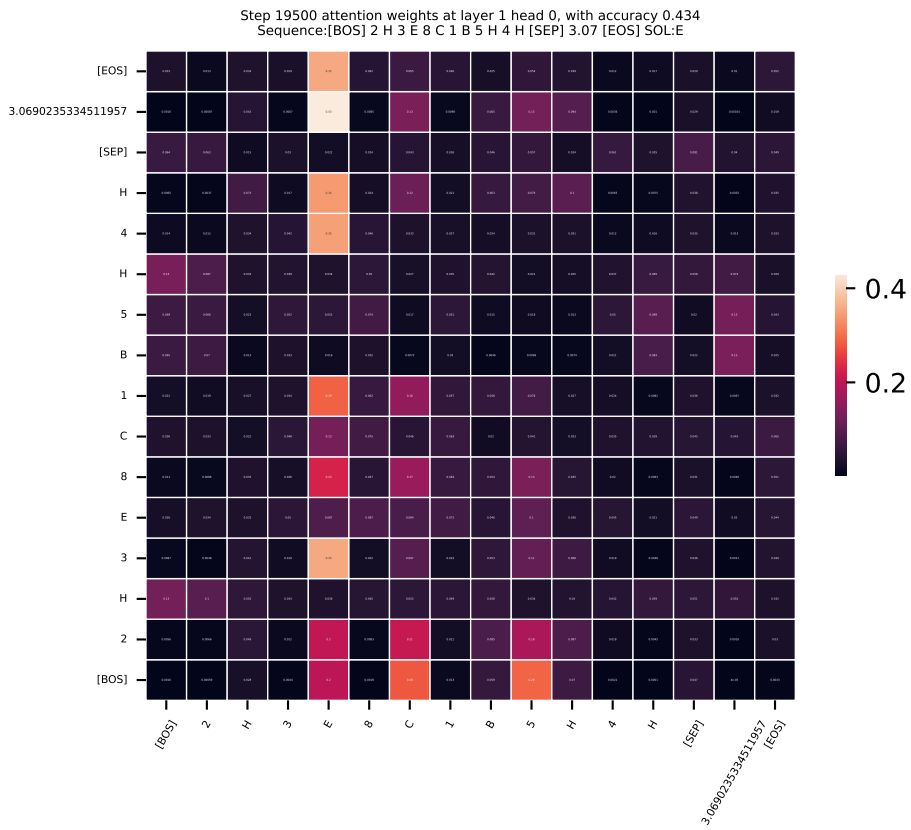

Step 19500 attention weights at layer 0 head 3, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

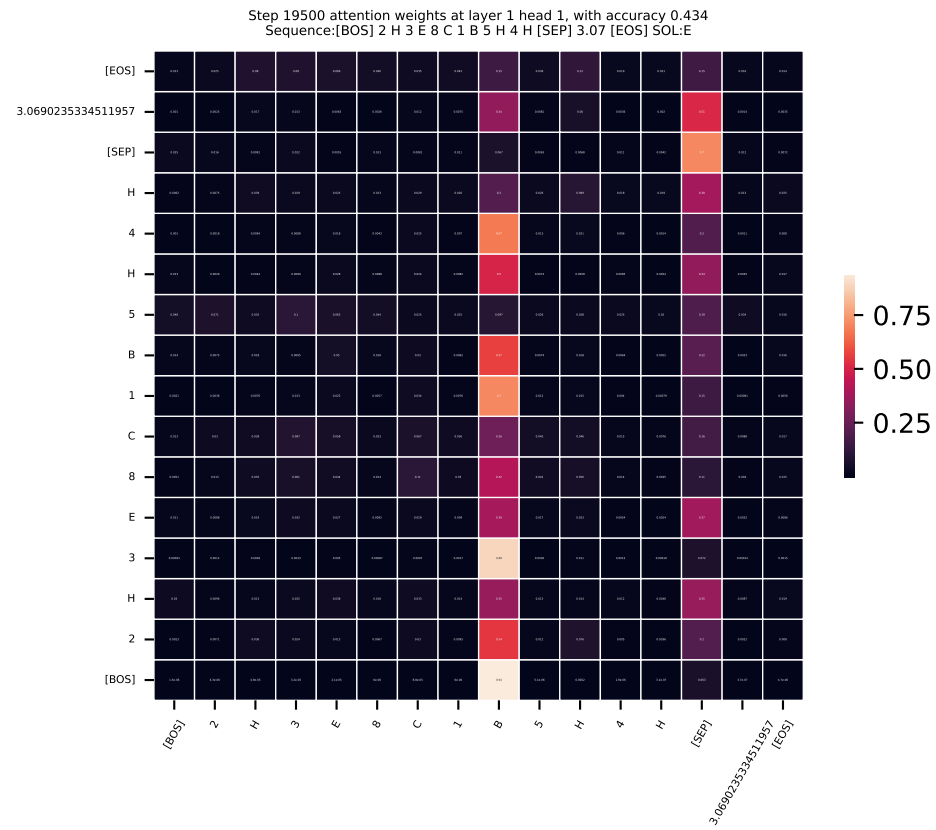

Step 19500 attention weights at layer 0 head 0, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

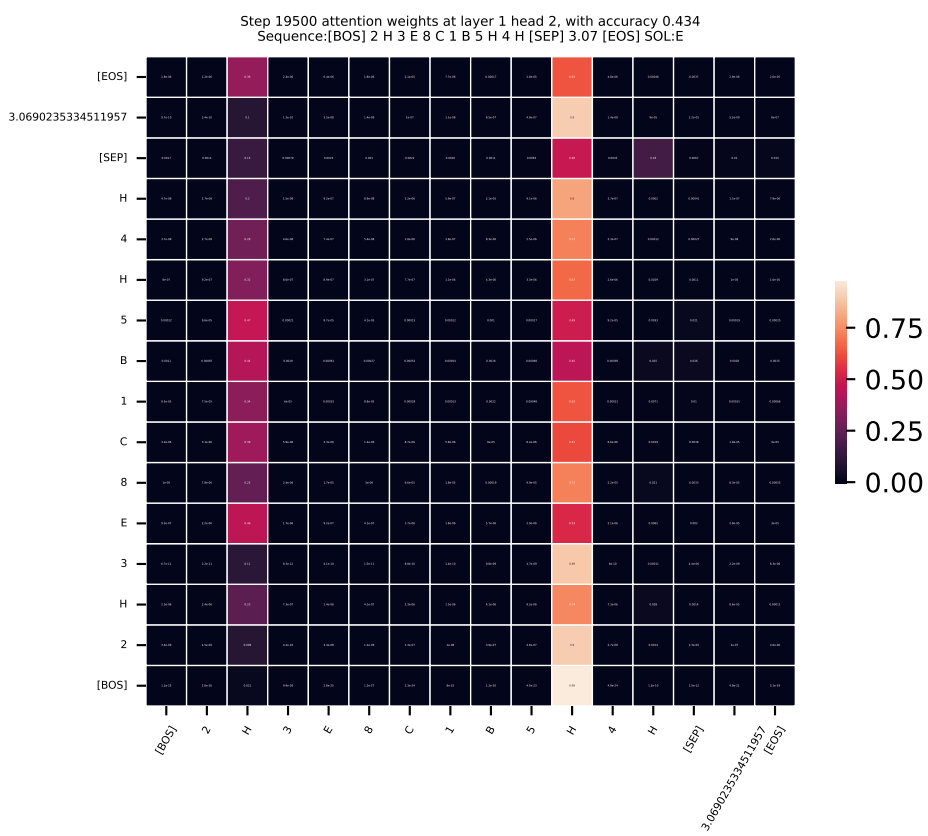

Step 19500 attention weights at layer 0 head 1, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

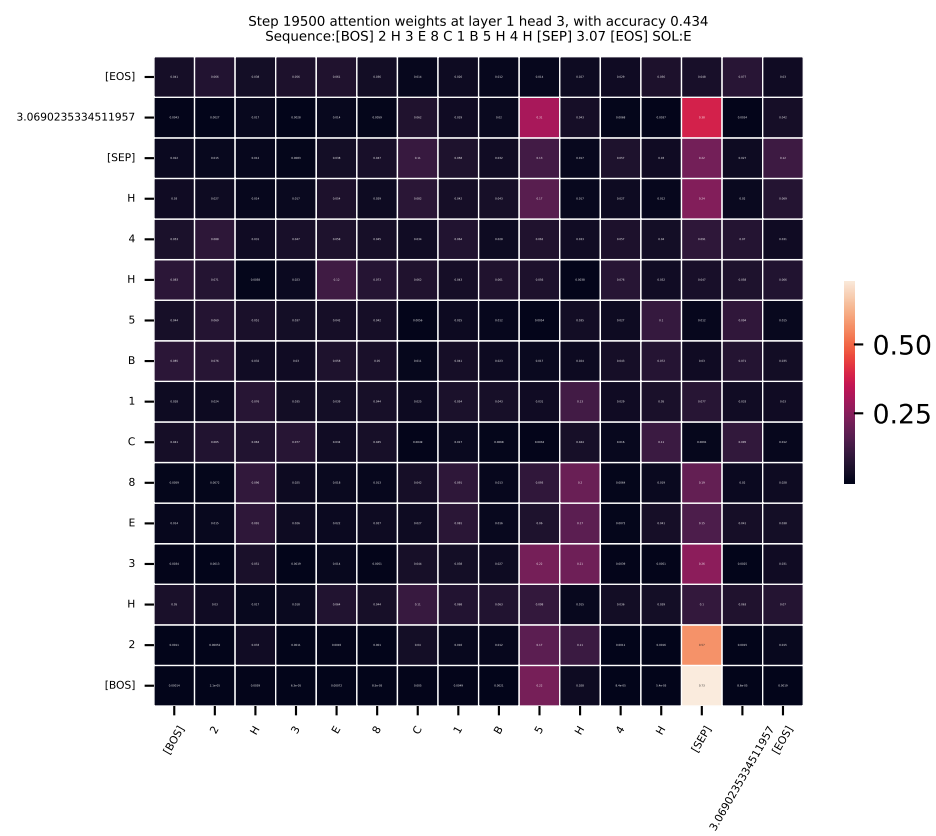

Step 19500 attention weights at layer 0 head 2, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

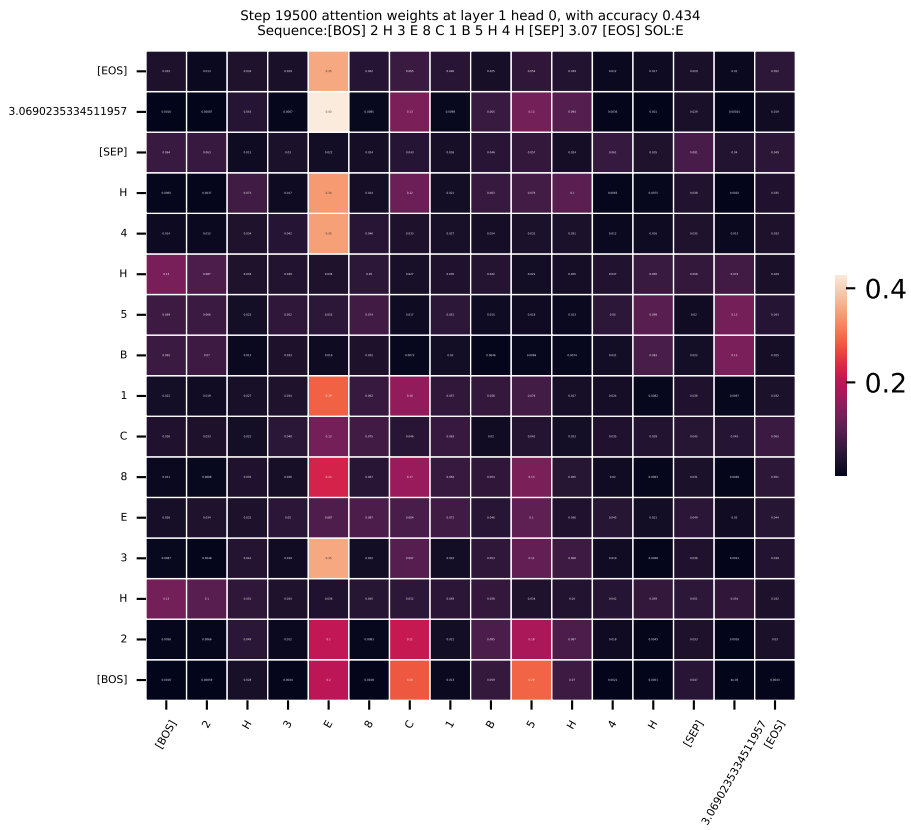

Step 19500 attention weights at layer 0 head 3, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

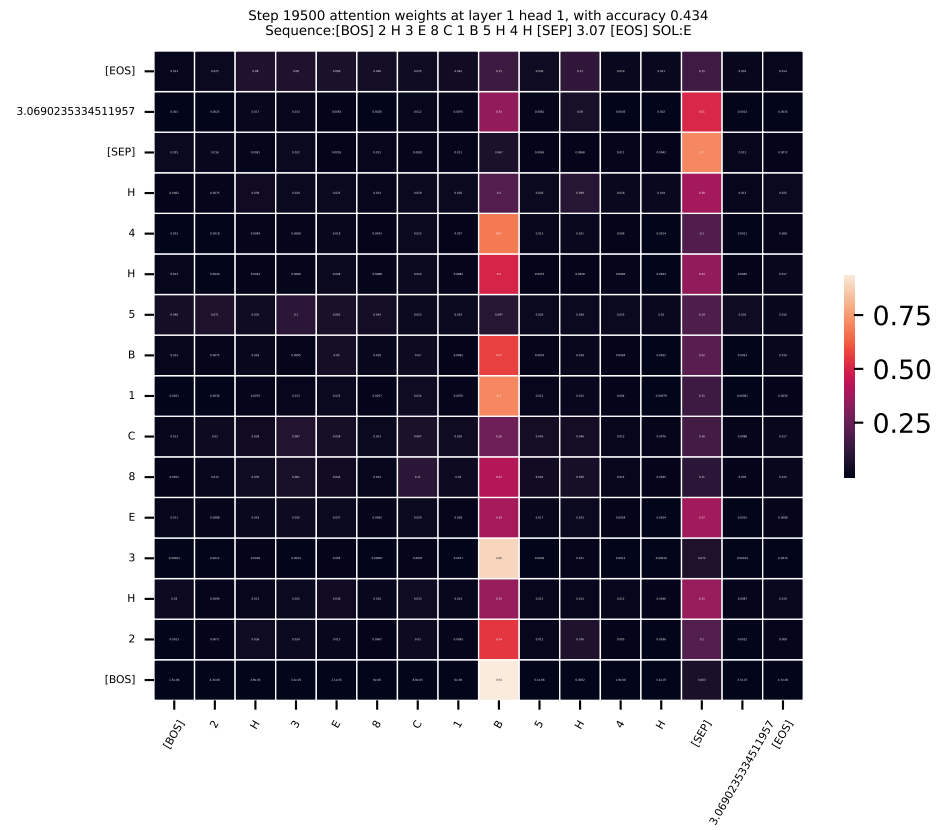

Step 19500 attention weights at layer 1 head 0, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.4
0.2

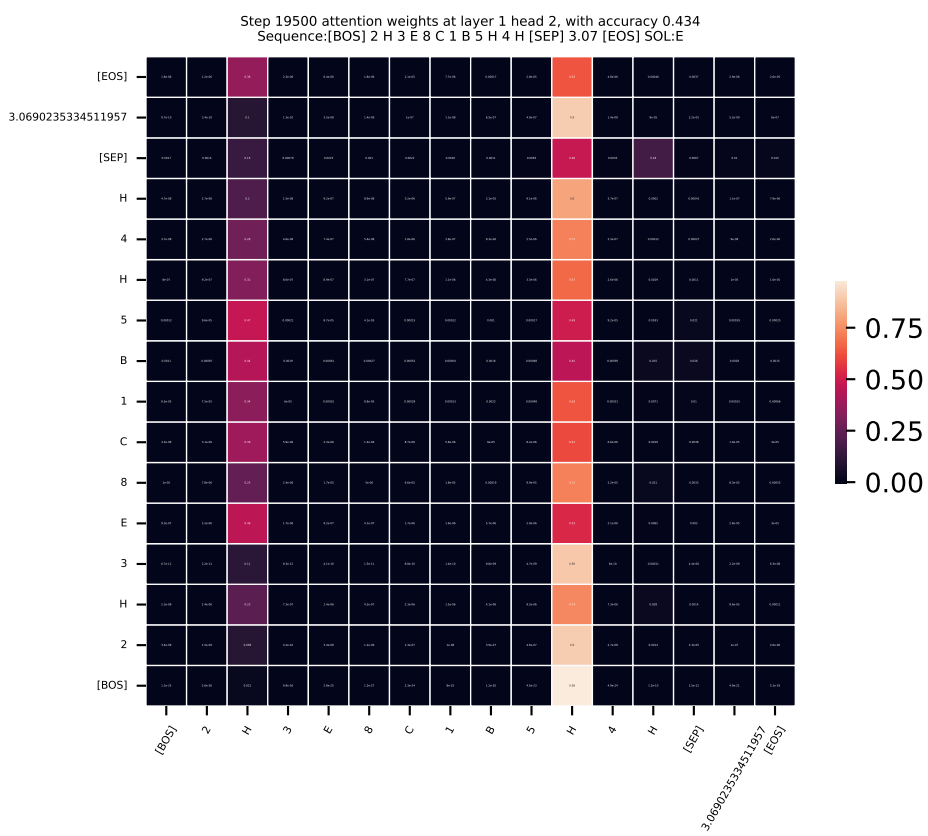

Step 19500 attention weights at layer 1 head 1, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

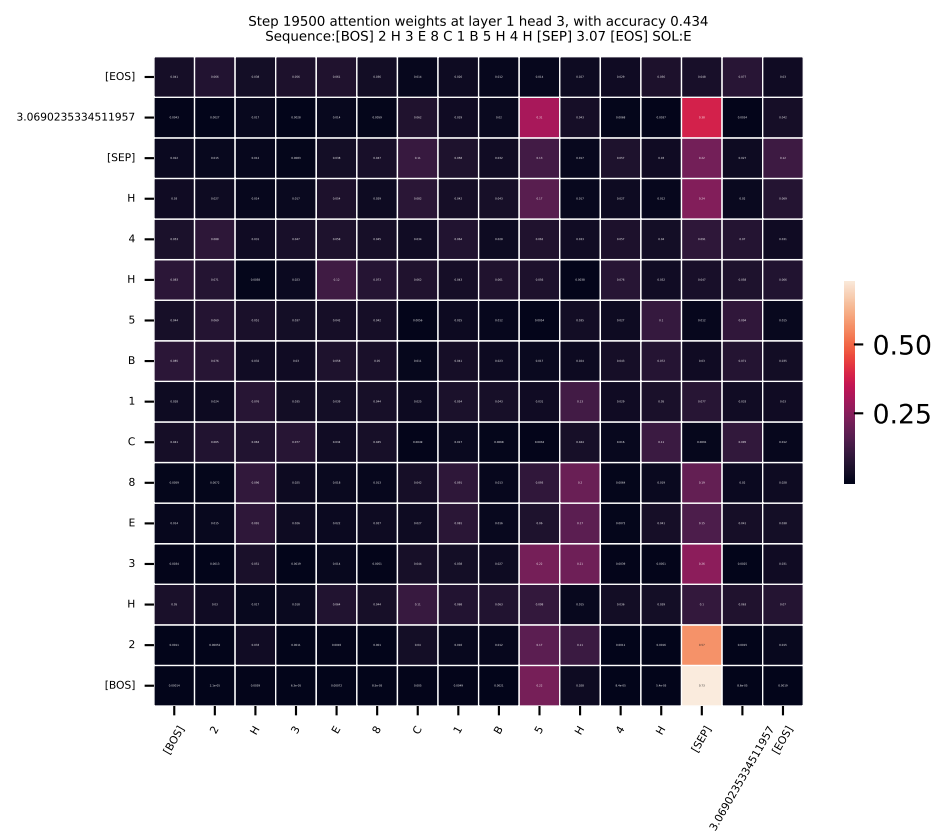

Step 19500 attention weights at layer 1 head 2, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

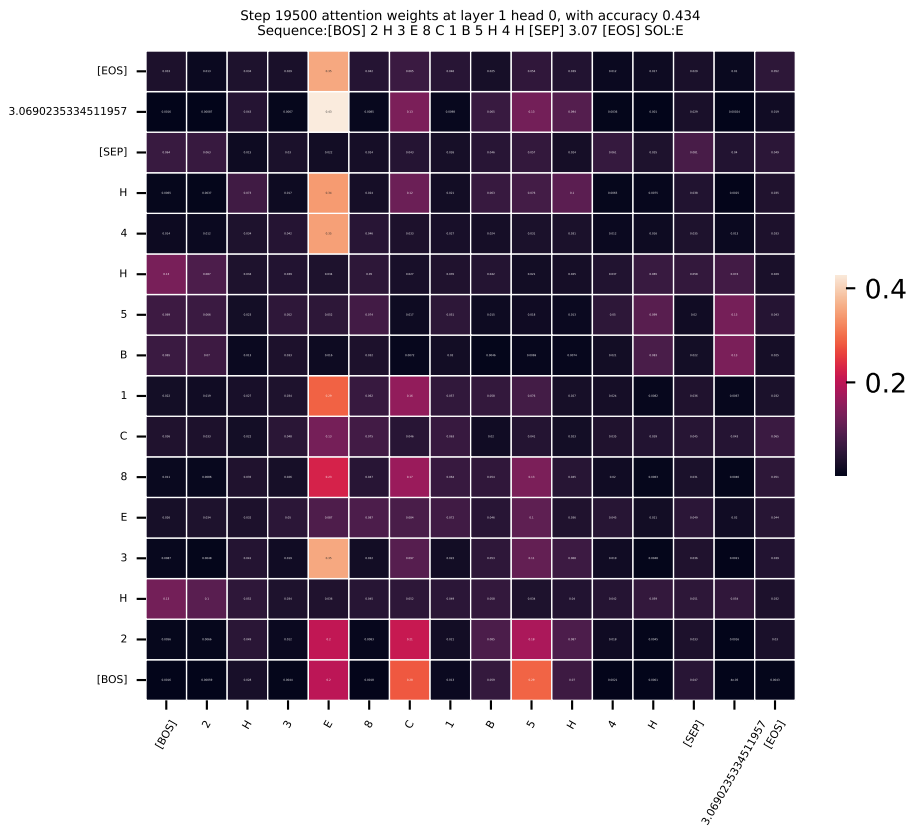

Step 19500 attention weights at layer 1 head 3, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.50
0.25

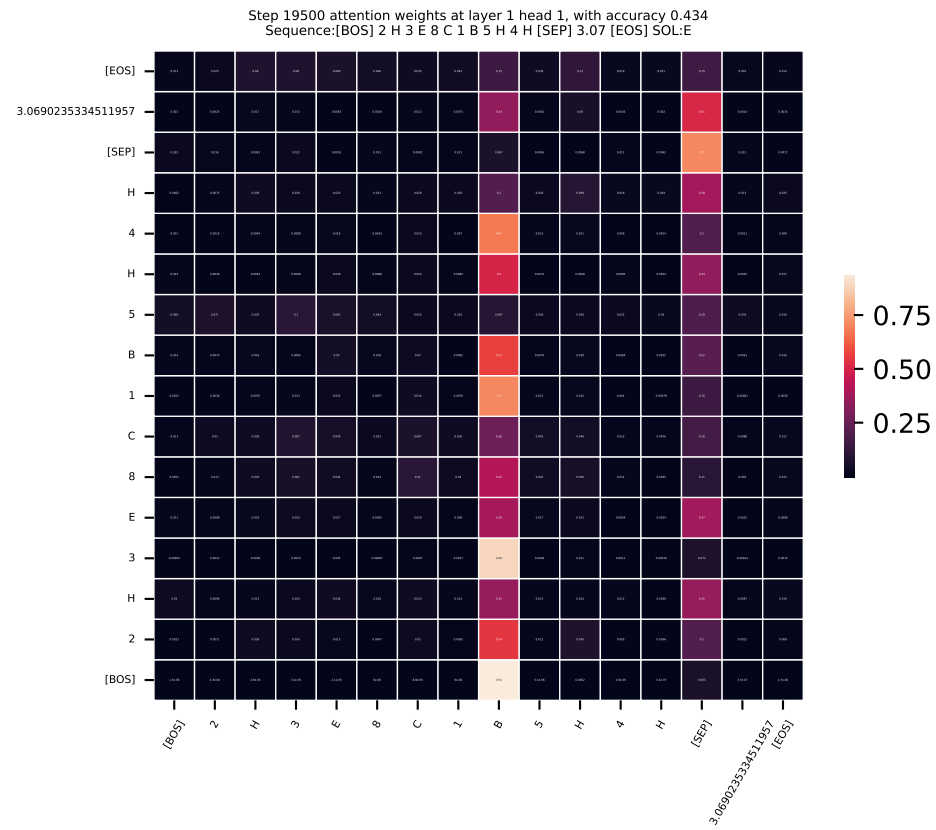

Step 19500 attention weights at layer 1 head 0, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.4
0.2

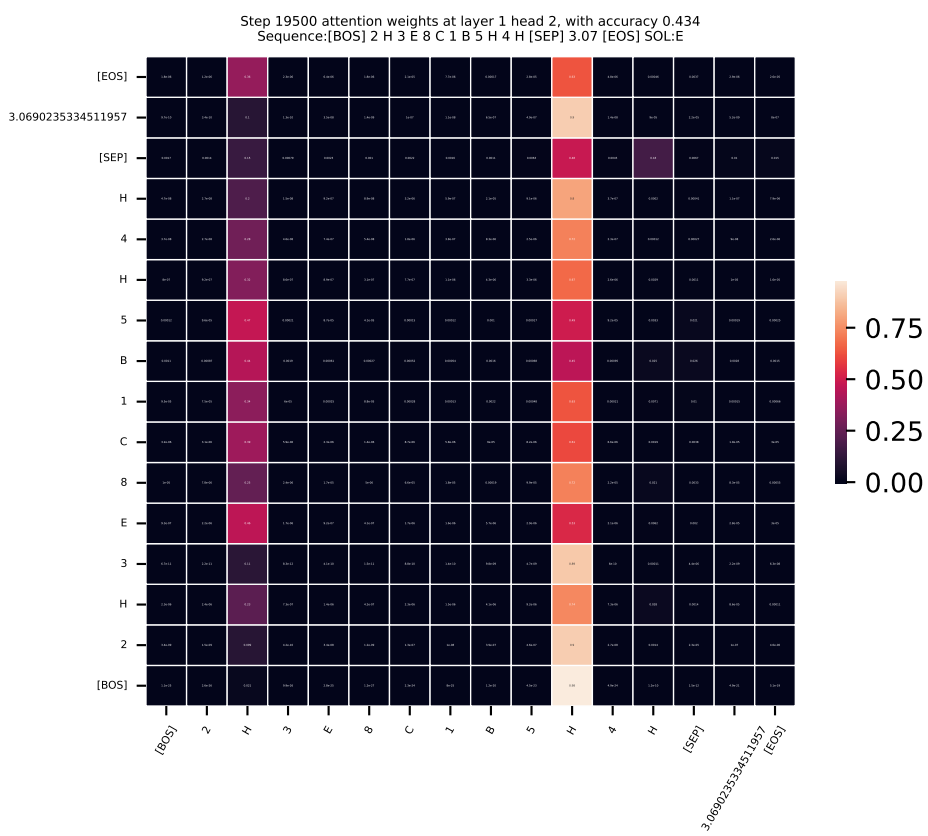

Step 19500 attention weights at layer 1 head 1, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

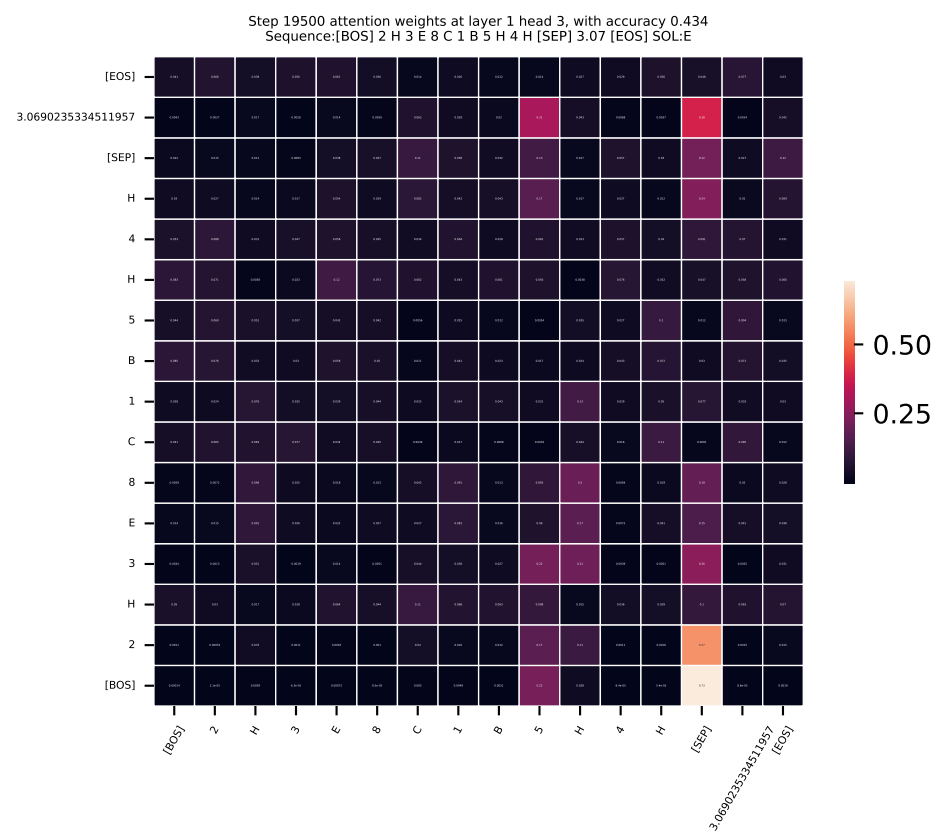

Step 19500 attention weights at layer 1 head 2, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

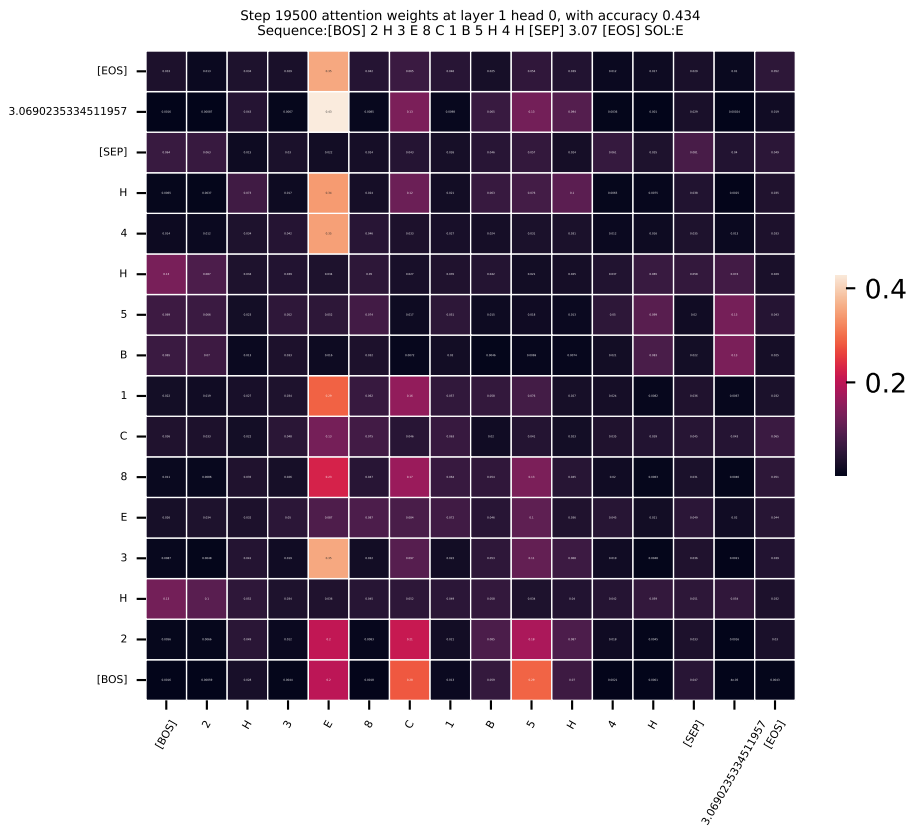

Step 19500 attention weights at layer 1 head 3, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.50
0.25

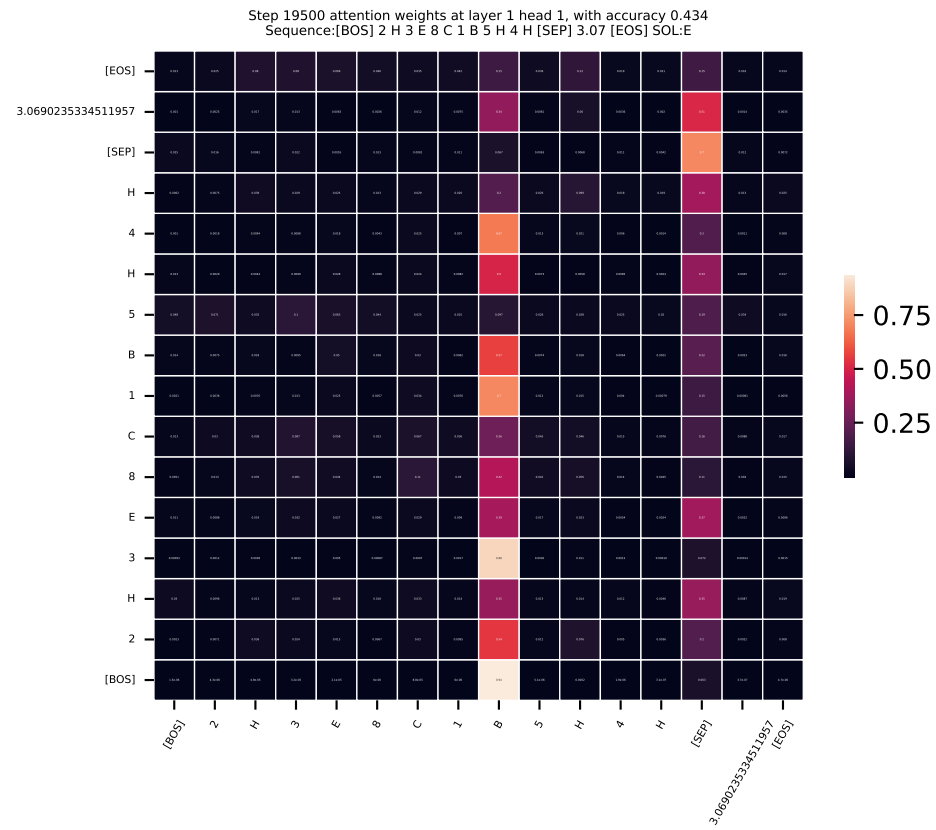

Step 19500 attention weights at layer 1 head 0, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.4
0.2

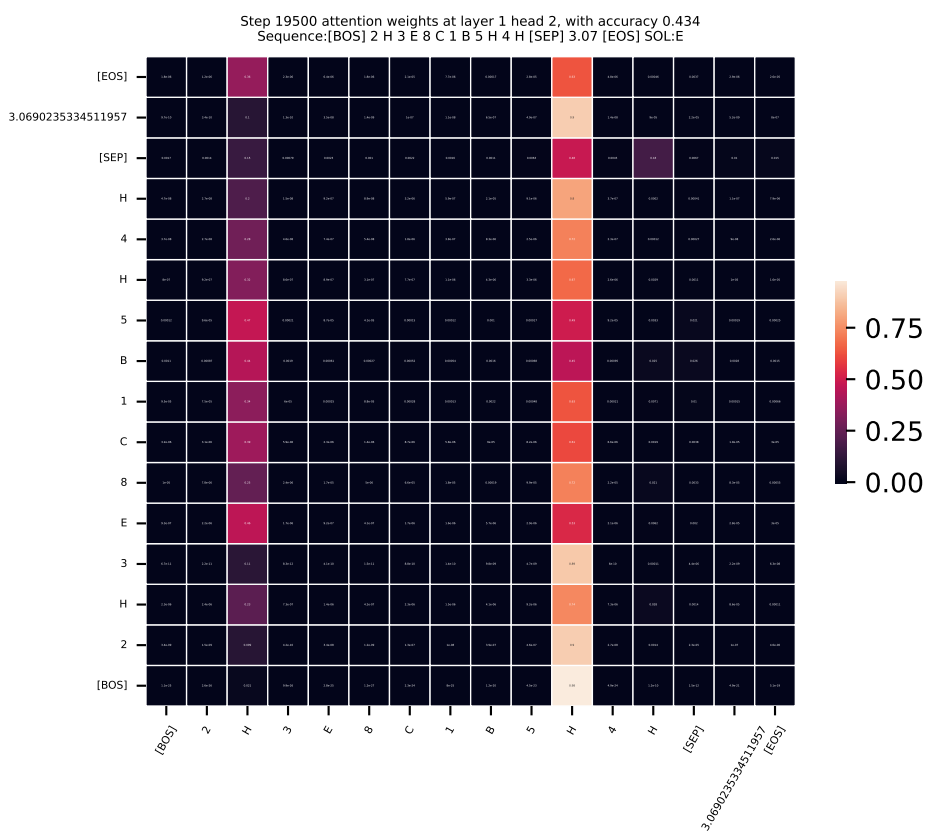

Step 19500 attention weights at layer 1 head 1, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

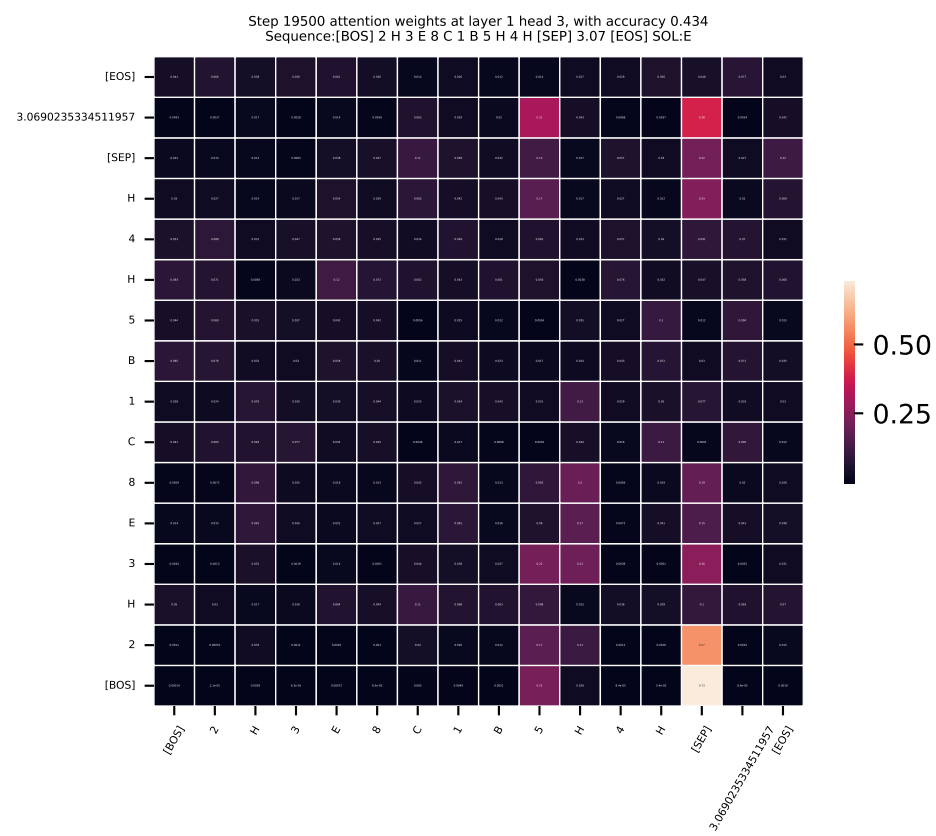

Step 19500 attention weights at layer 1 head 2, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

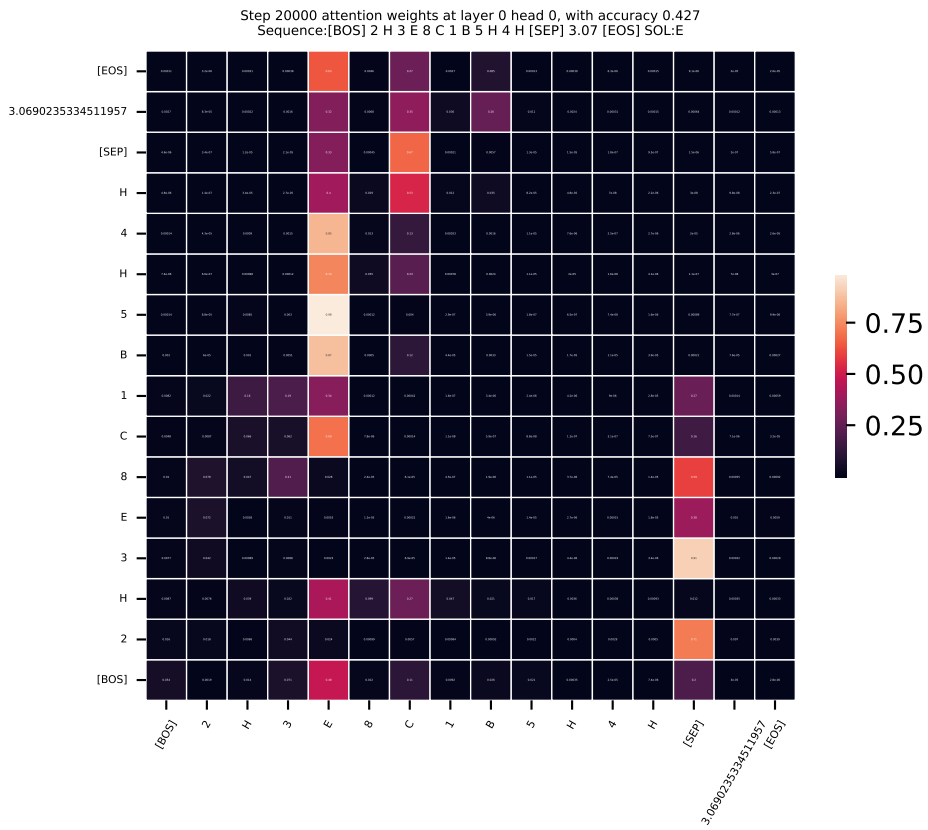

Step 19500 attention weights at layer 1 head 3, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.50
0.25

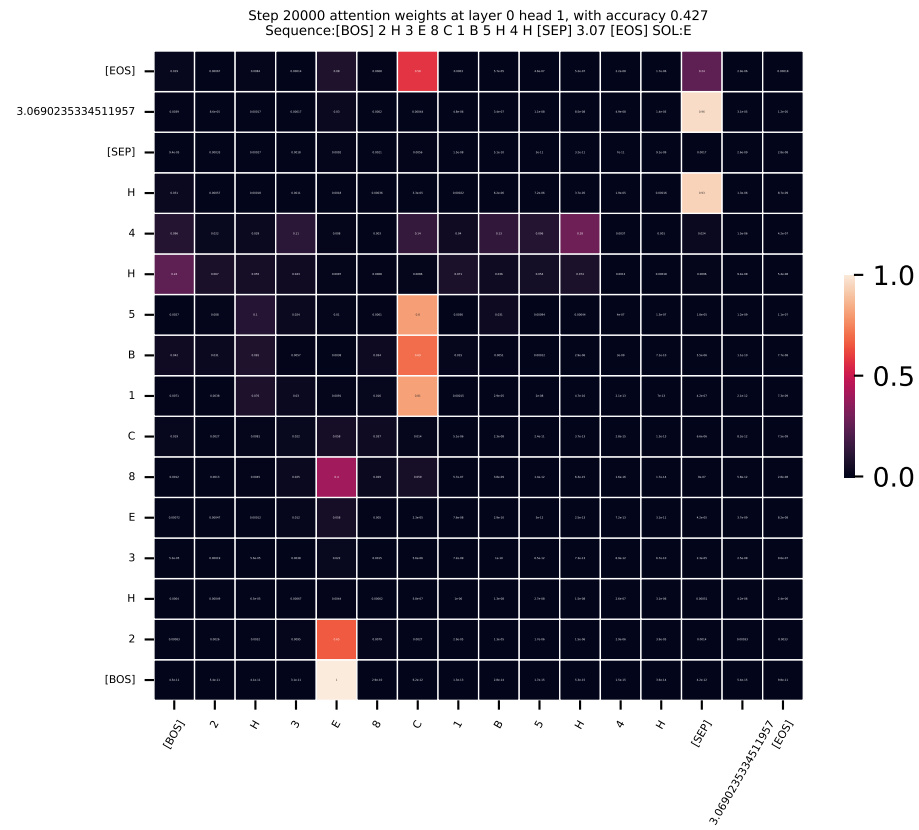

Step 19500 attention weights at layer 1 head 0, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.4
0.2

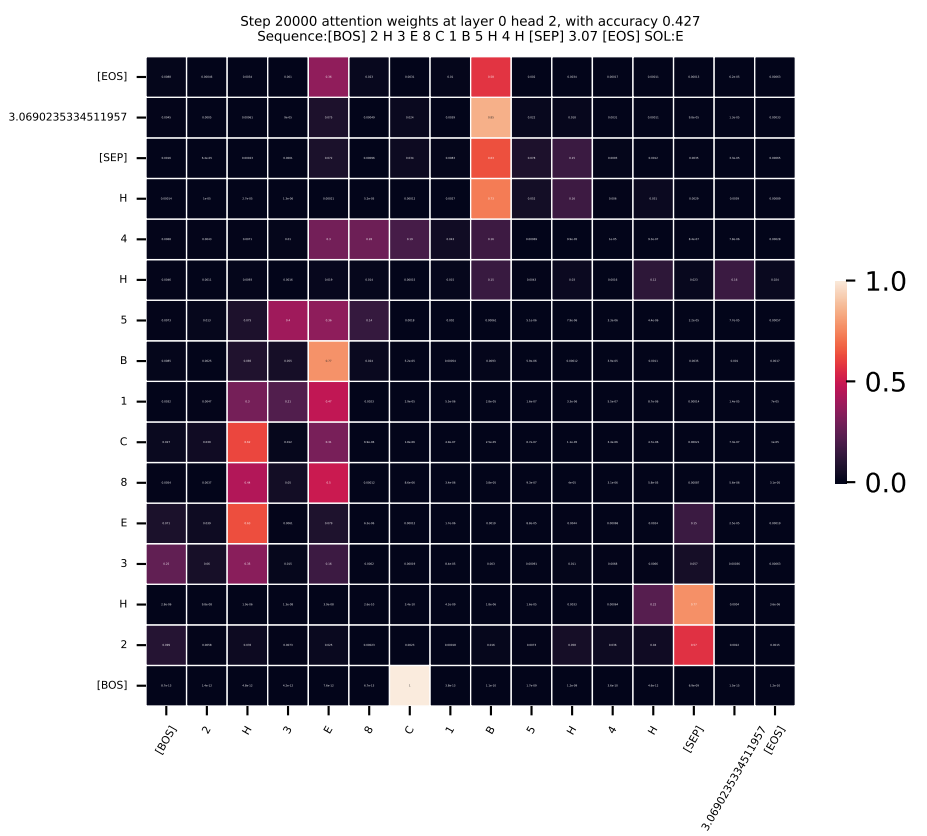

Step 19500 attention weights at layer 1 head 1, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

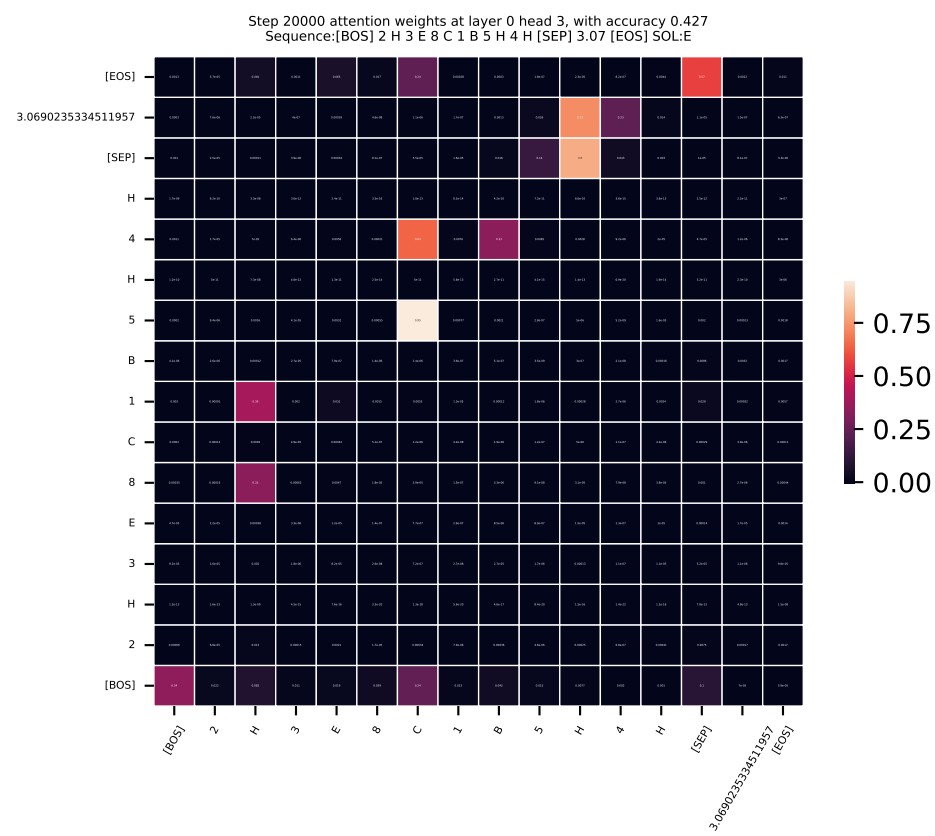

Step 19500 attention weights at layer 1 head 2, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

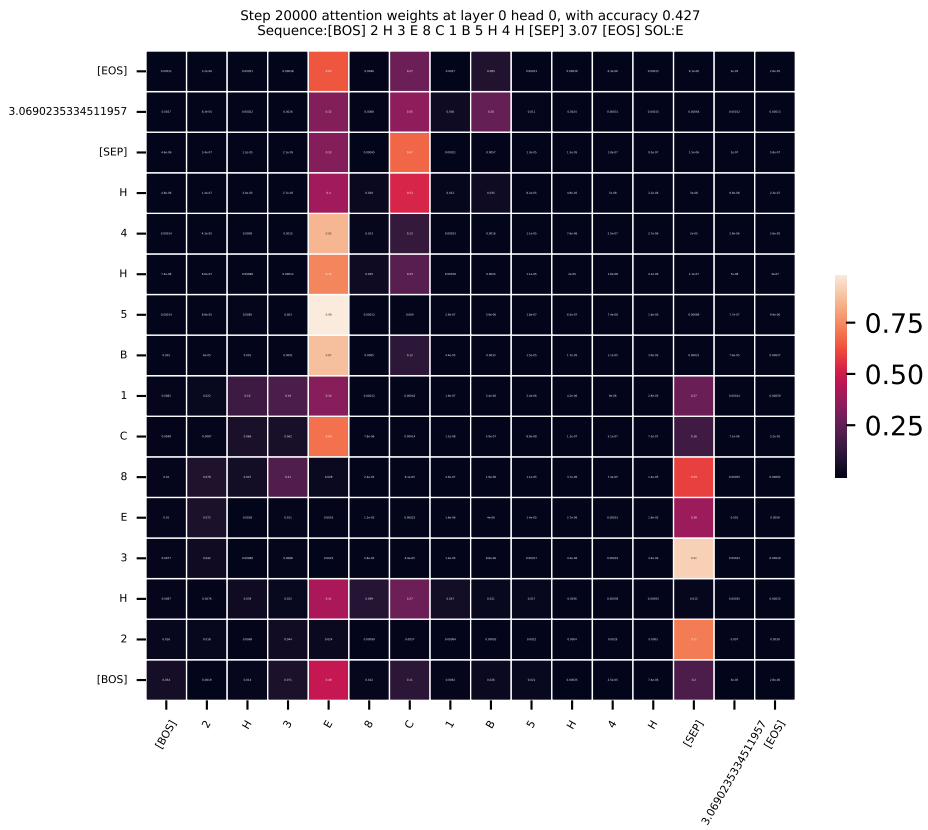

Step 19500 attention weights at layer 1 head 3, with accuracy 0.434
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.50
0.25

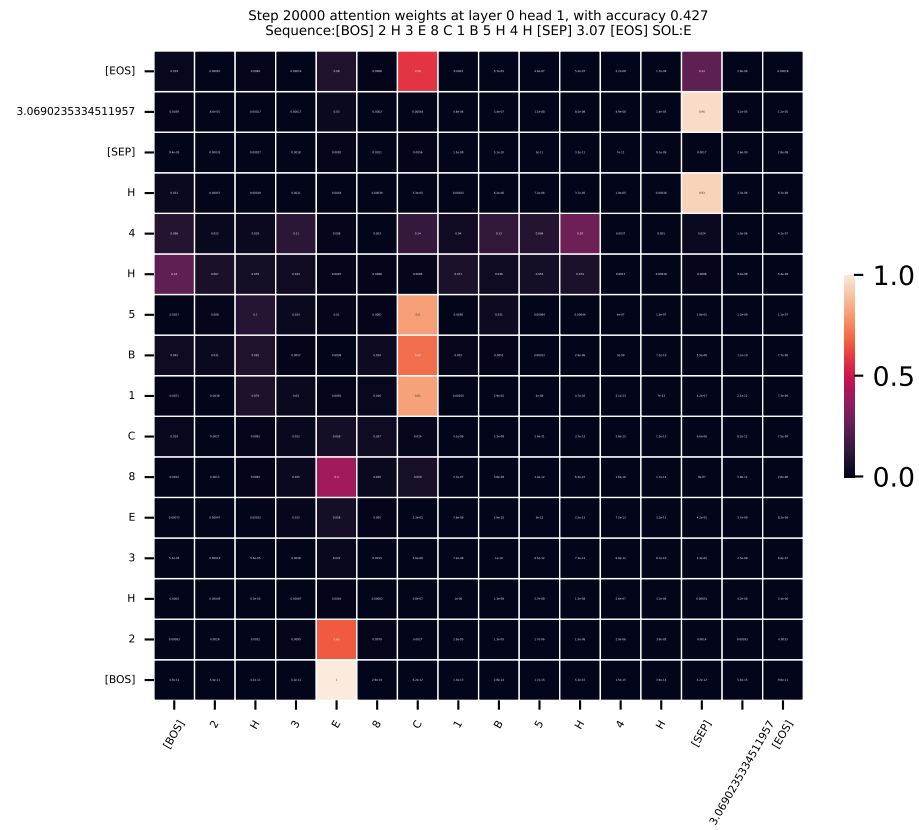

Step 20000 attention weights at layer 0 head 0, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

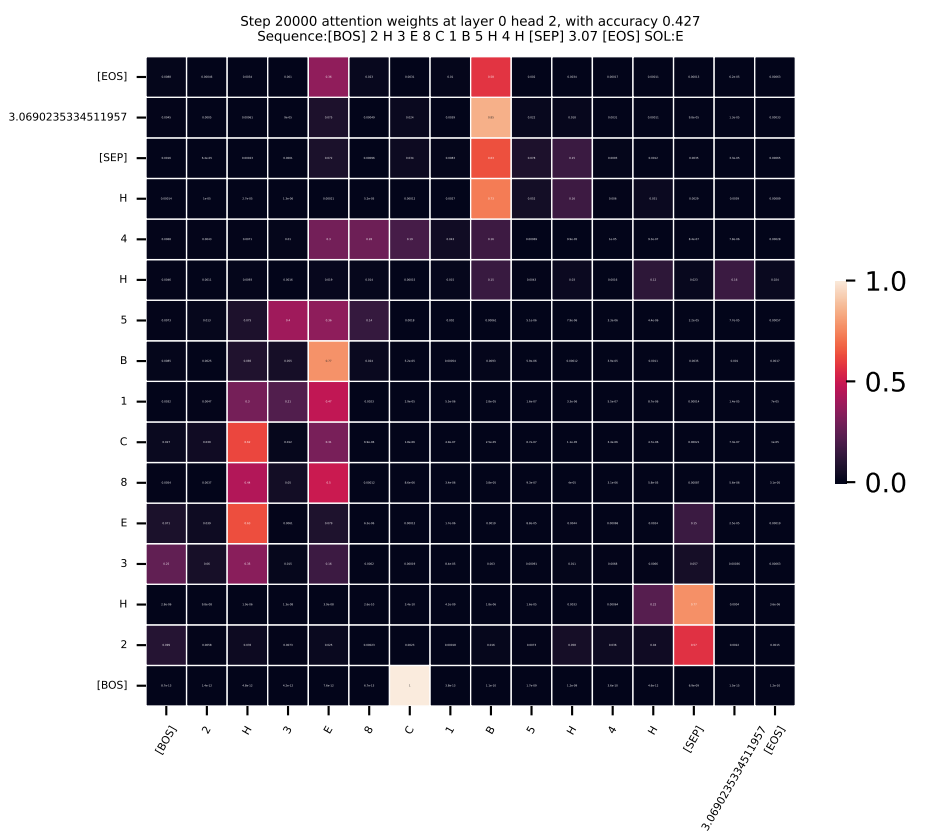

Step 20000 attention weights at layer 0 head 1, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

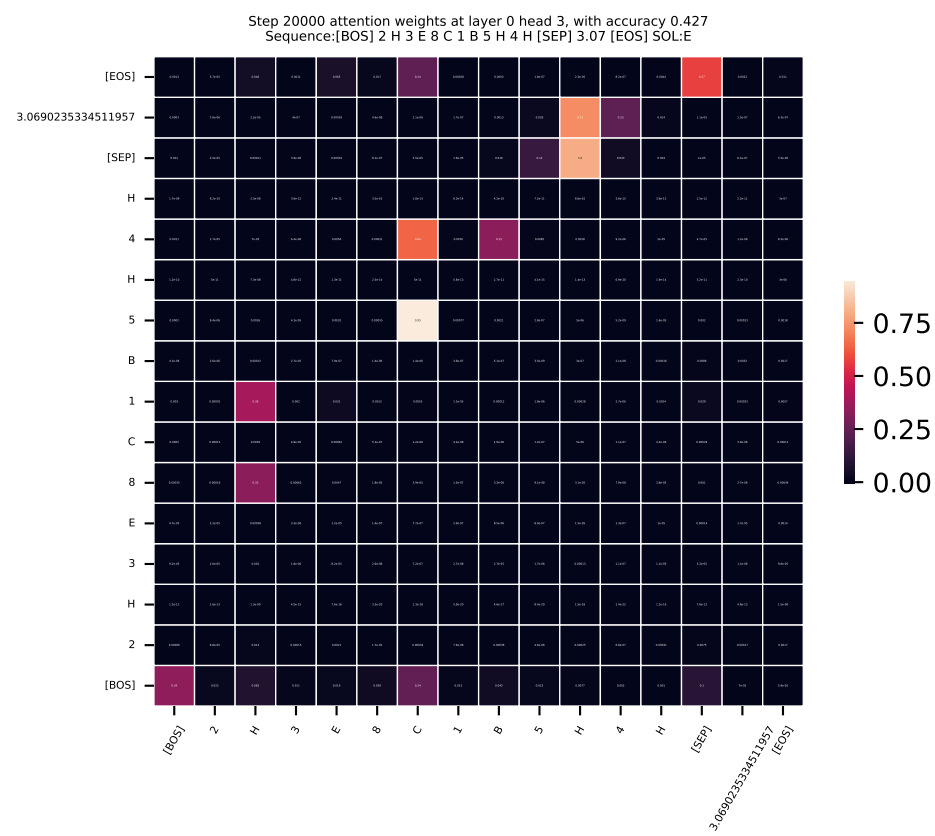

Step 20000 attention weights at layer 0 head 2, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

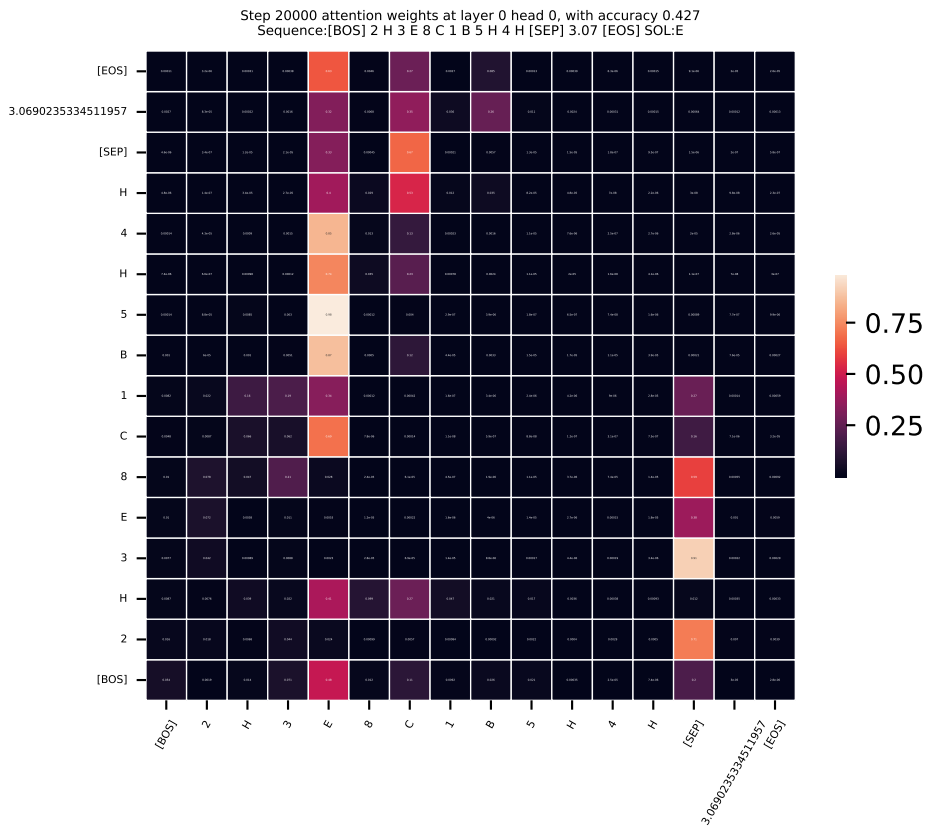

Step 20000 attention weights at layer 0 head 3, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

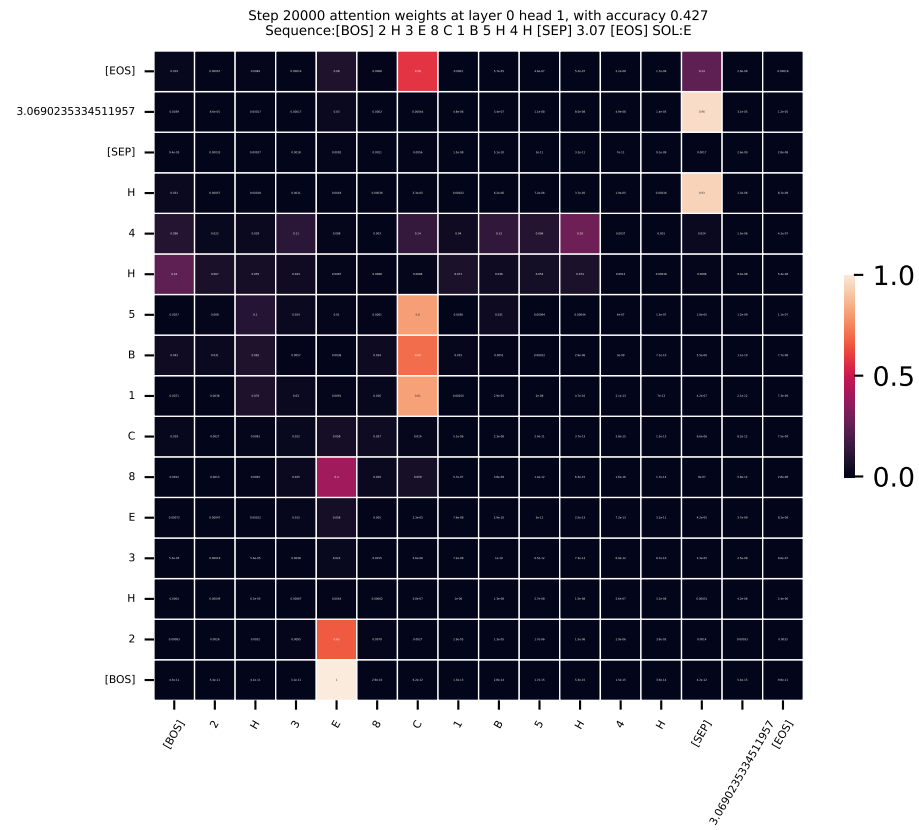

Step 20000 attention weights at layer 0 head 0, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

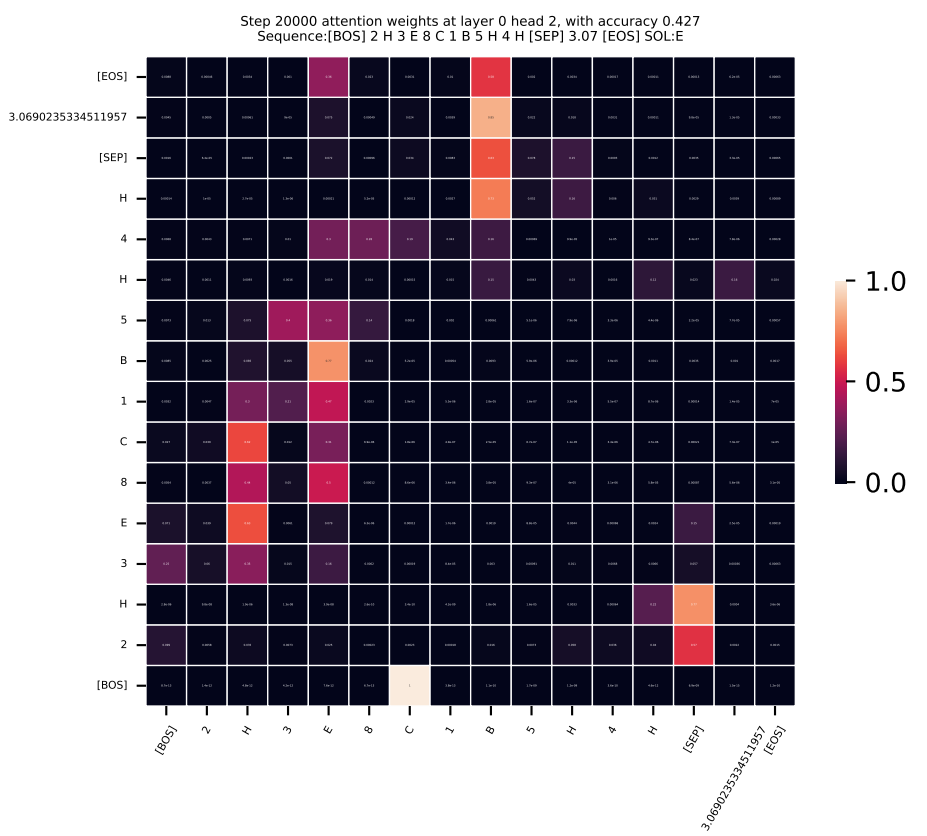

Step 20000 attention weights at layer 0 head 1, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

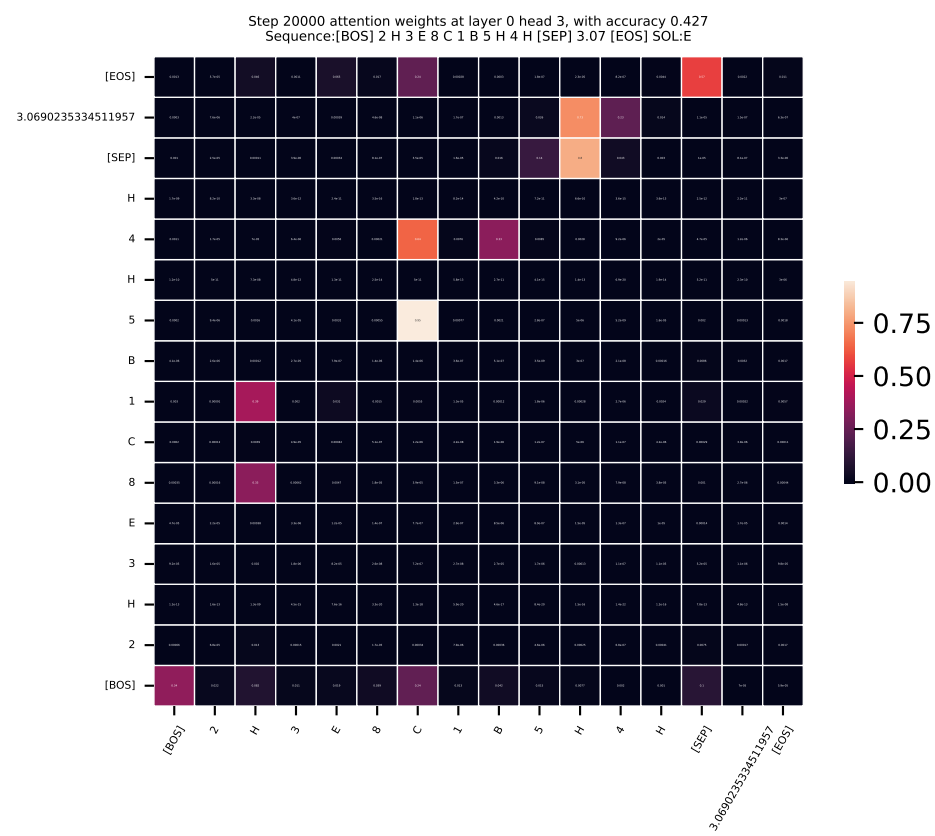

Step 20000 attention weights at layer 0 head 2, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

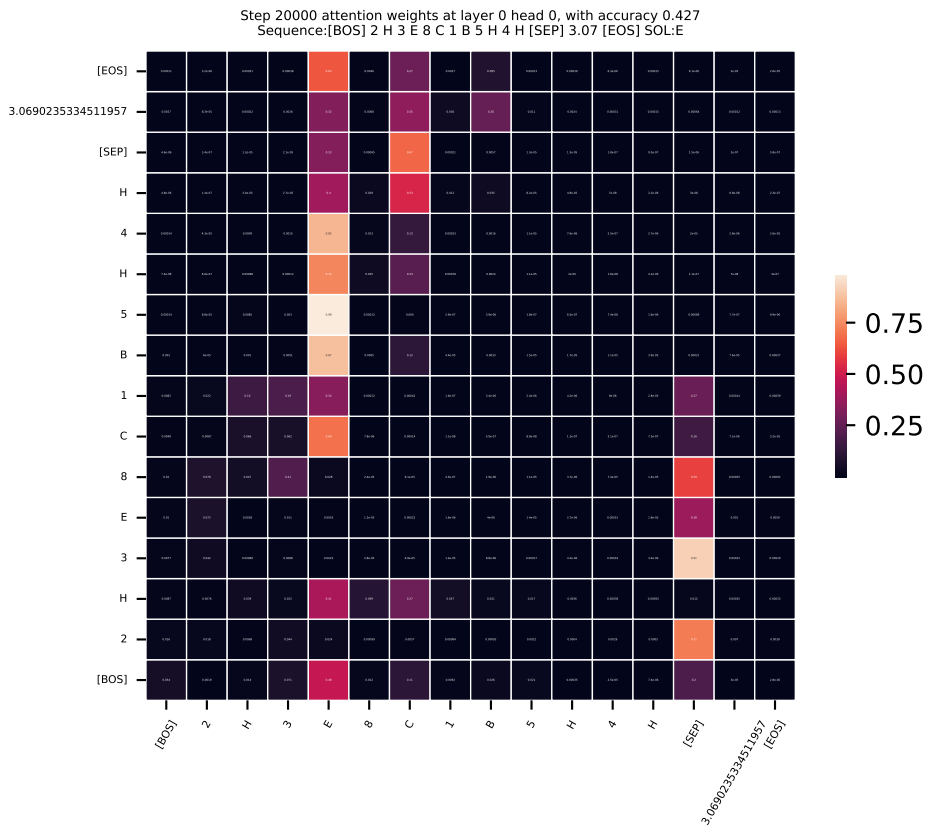

Step 20000 attention weights at layer 0 head 3, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

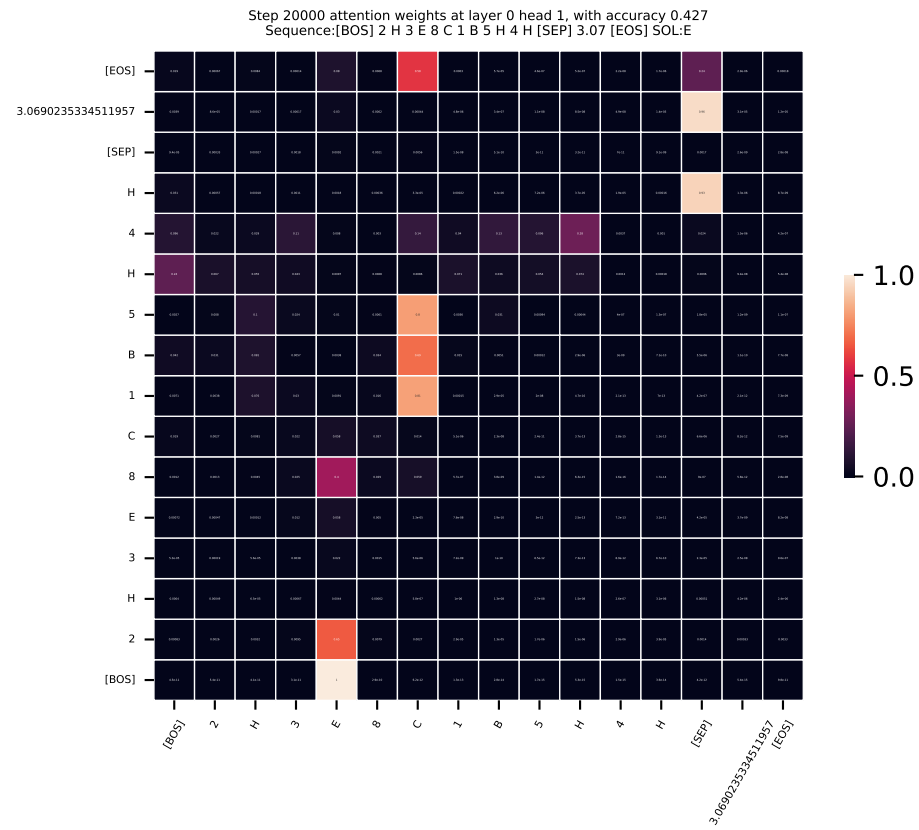

Step 20000 attention weights at layer 0 head 0, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

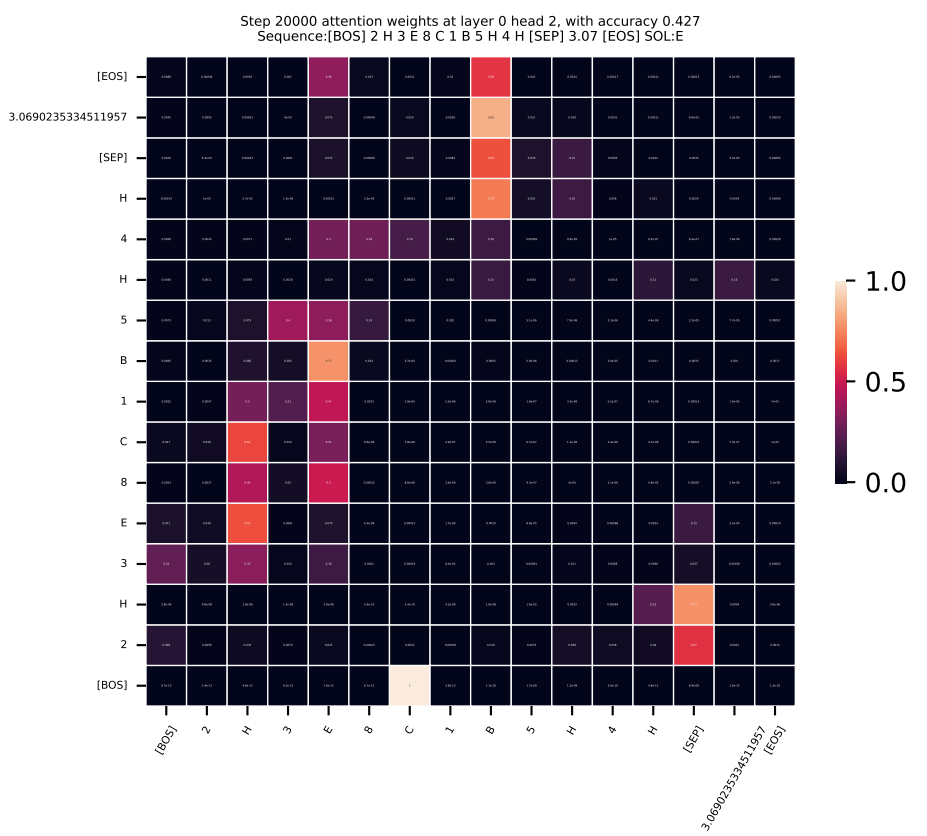

Step 20000 attention weights at layer 0 head 1, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

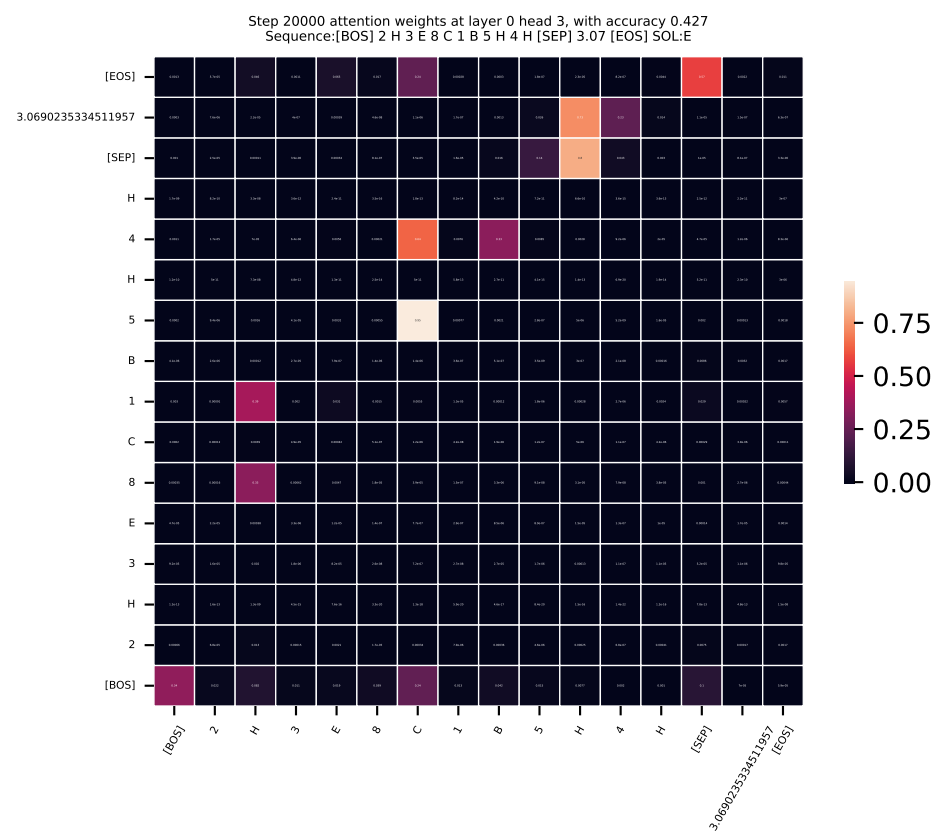

Step 20000 attention weights at layer 0 head 2, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

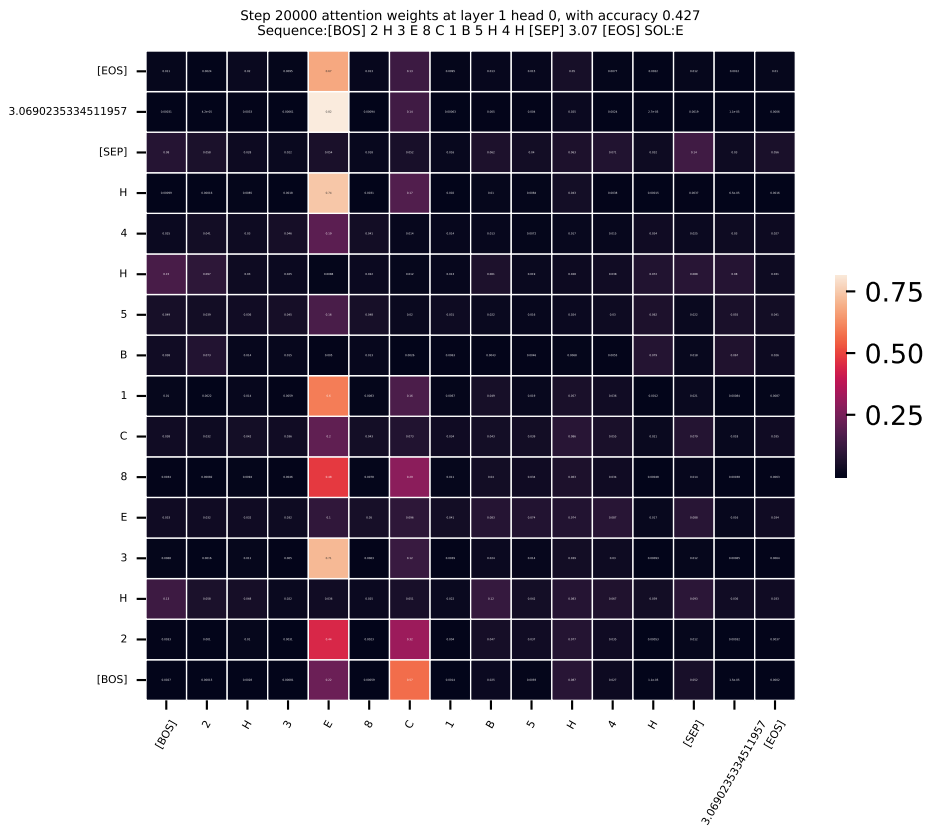

Step 20000 attention weights at layer 0 head 3, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

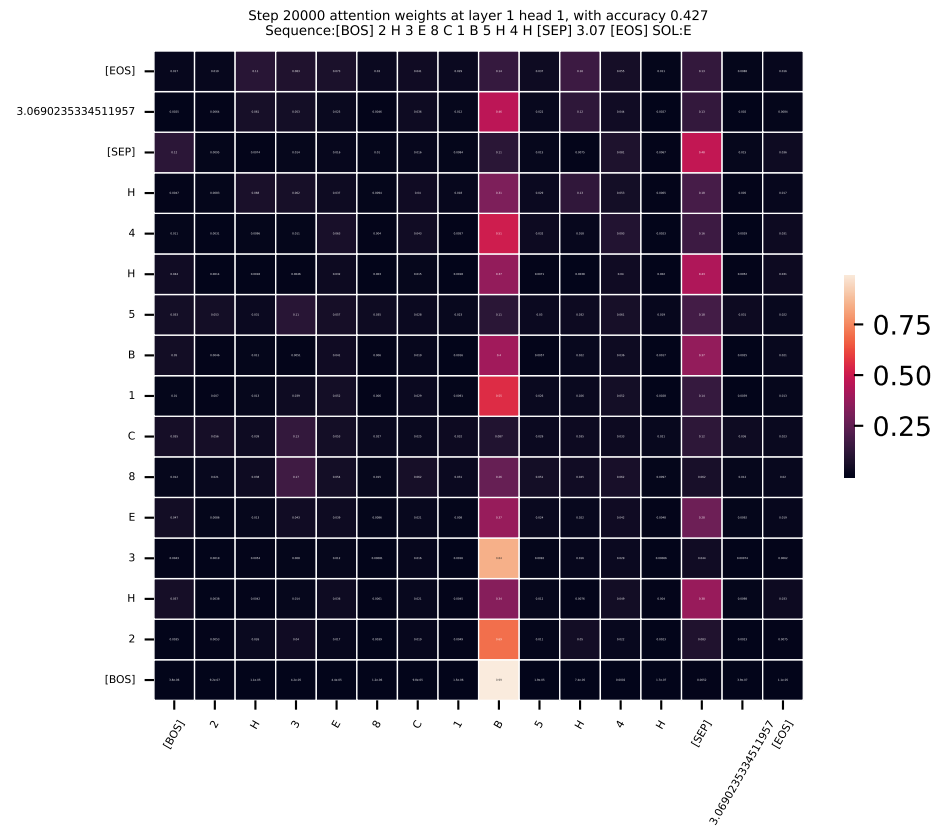

Step 20000 attention weights at layer 0 head 0, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

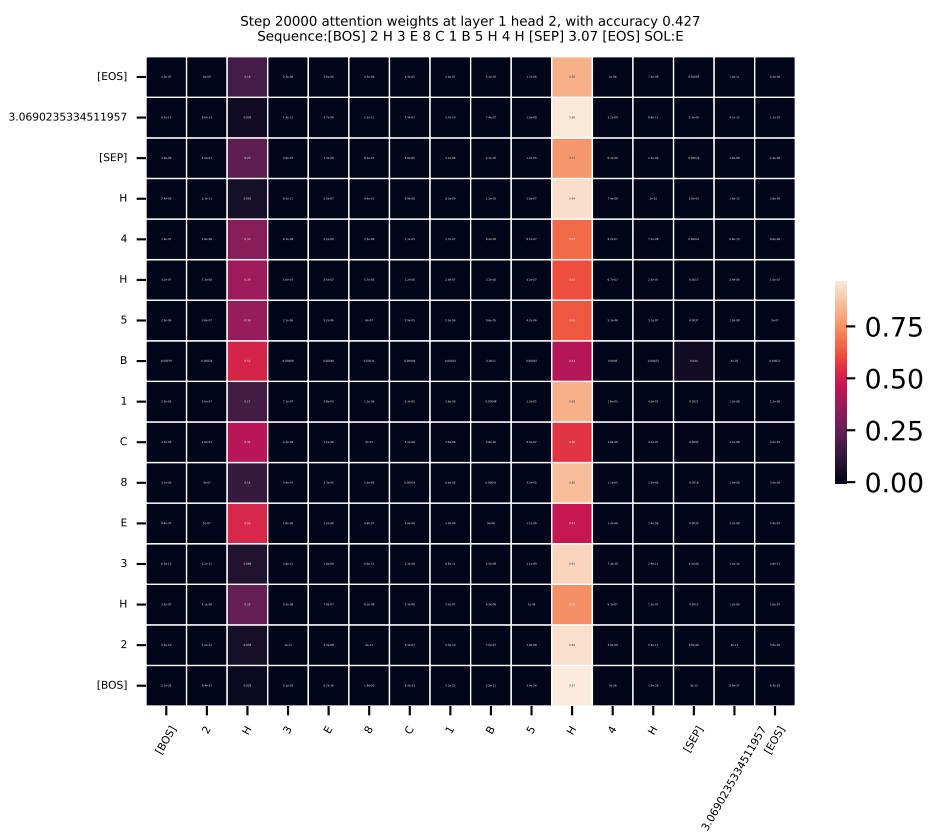

Step 20000 attention weights at layer 0 head 1, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

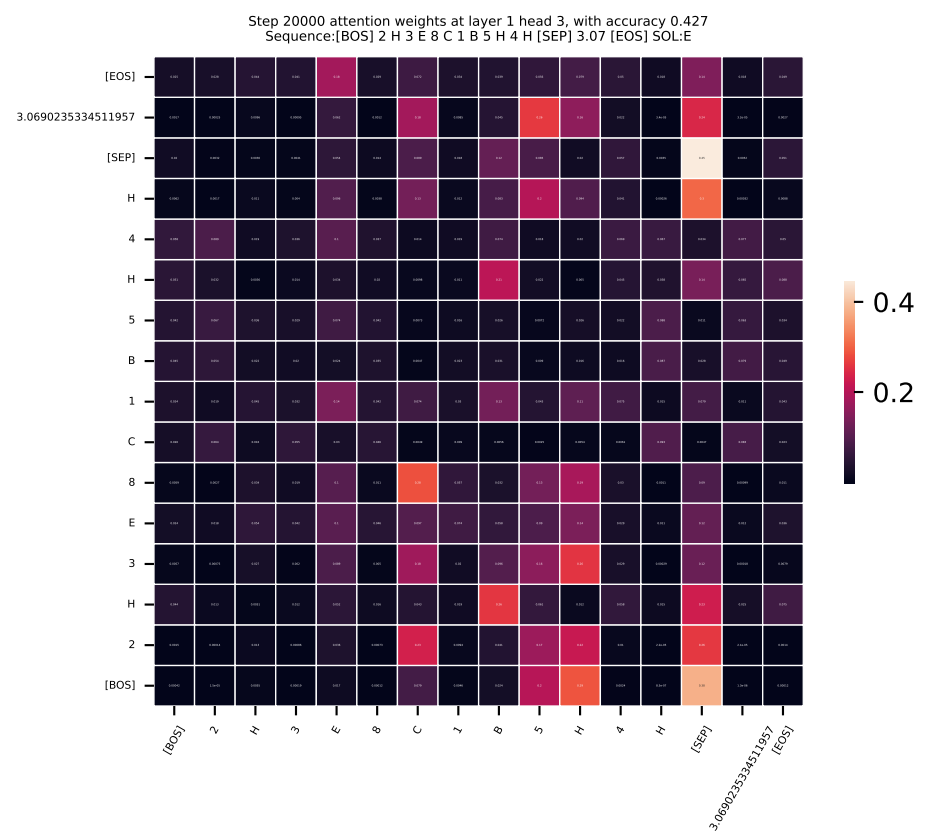

Step 20000 attention weights at layer 0 head 2, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

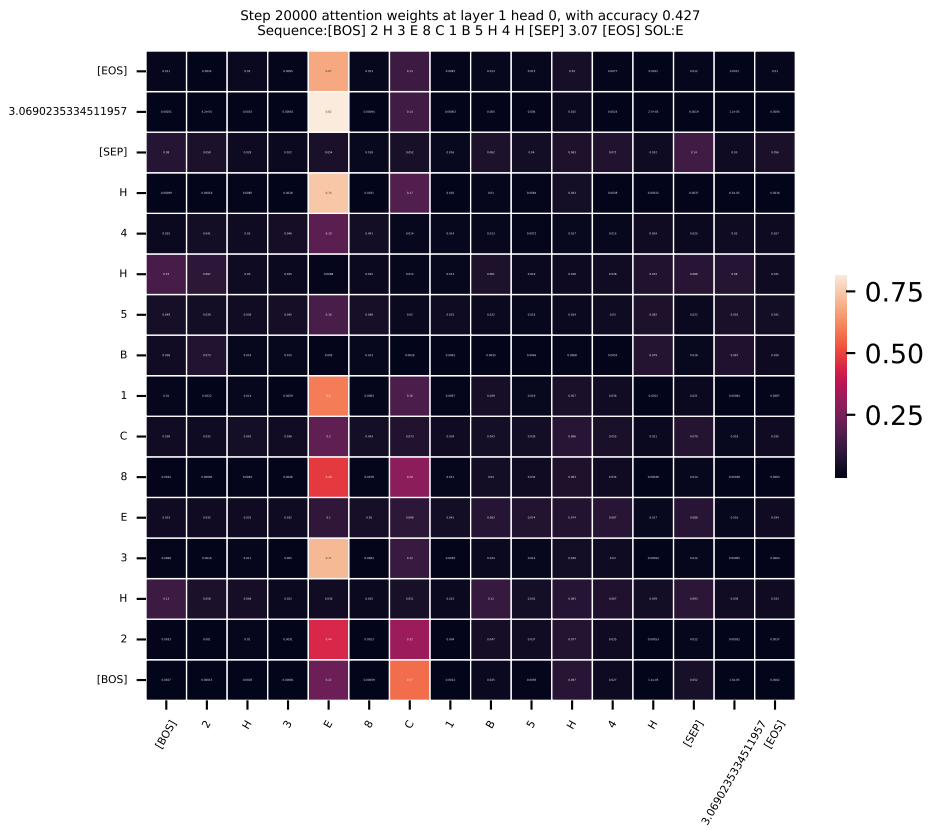

Step 20000 attention weights at layer 0 head 3, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

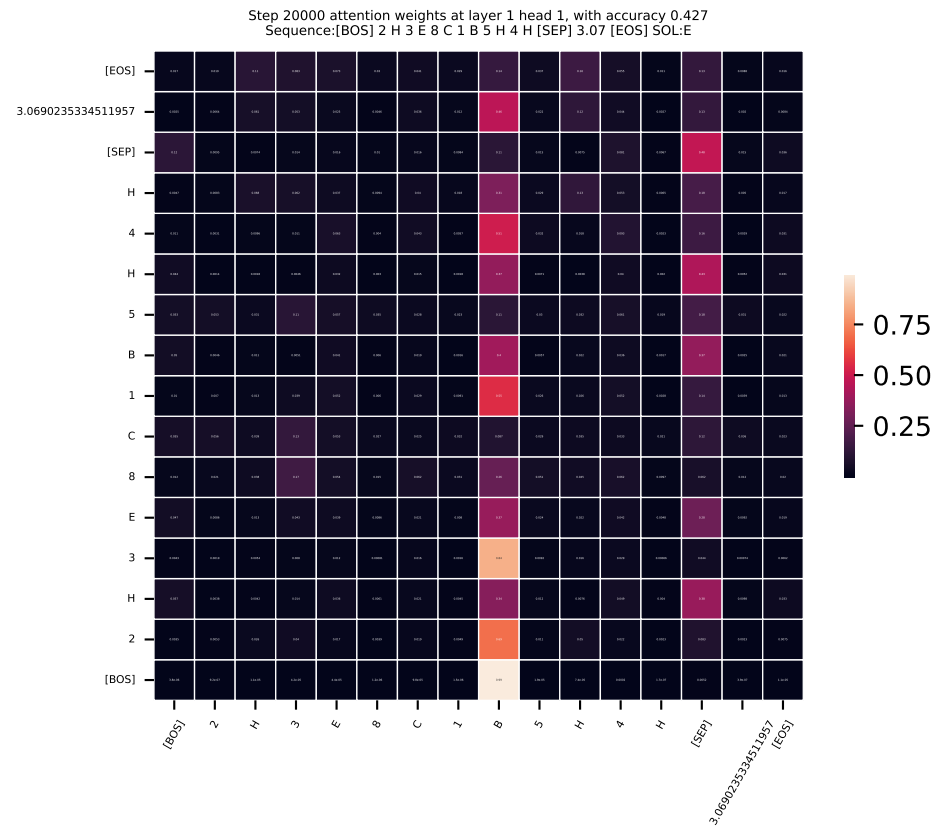

Step 20000 attention weights at layer 1 head 0, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

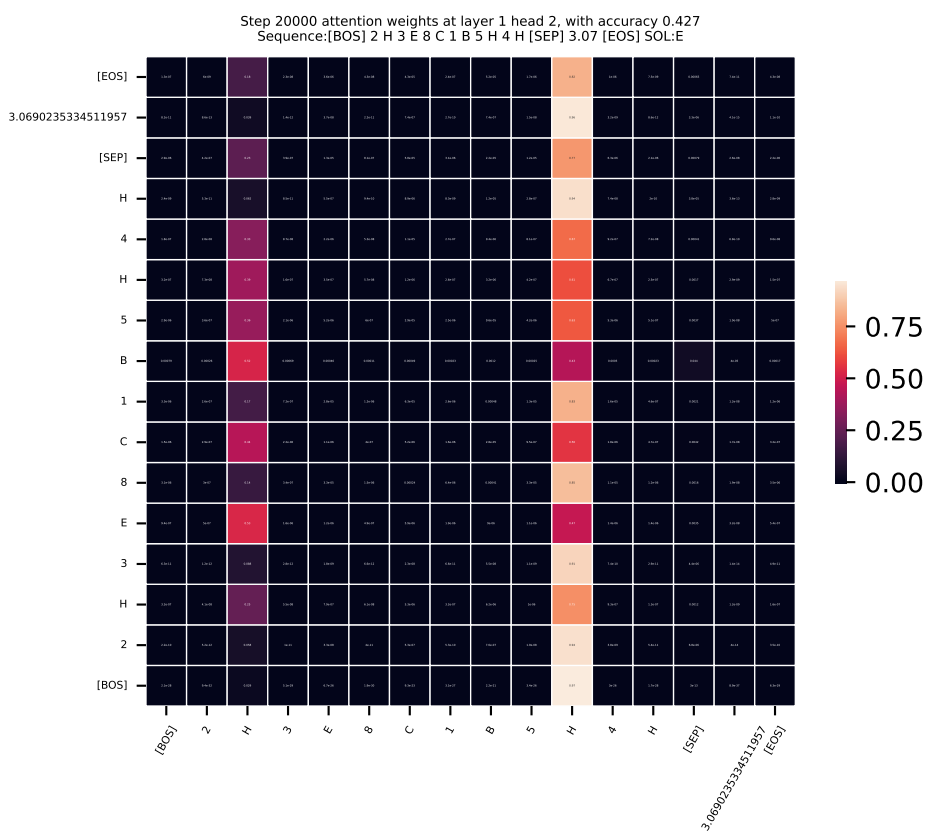

Step 20000 attention weights at layer 1 head 1, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

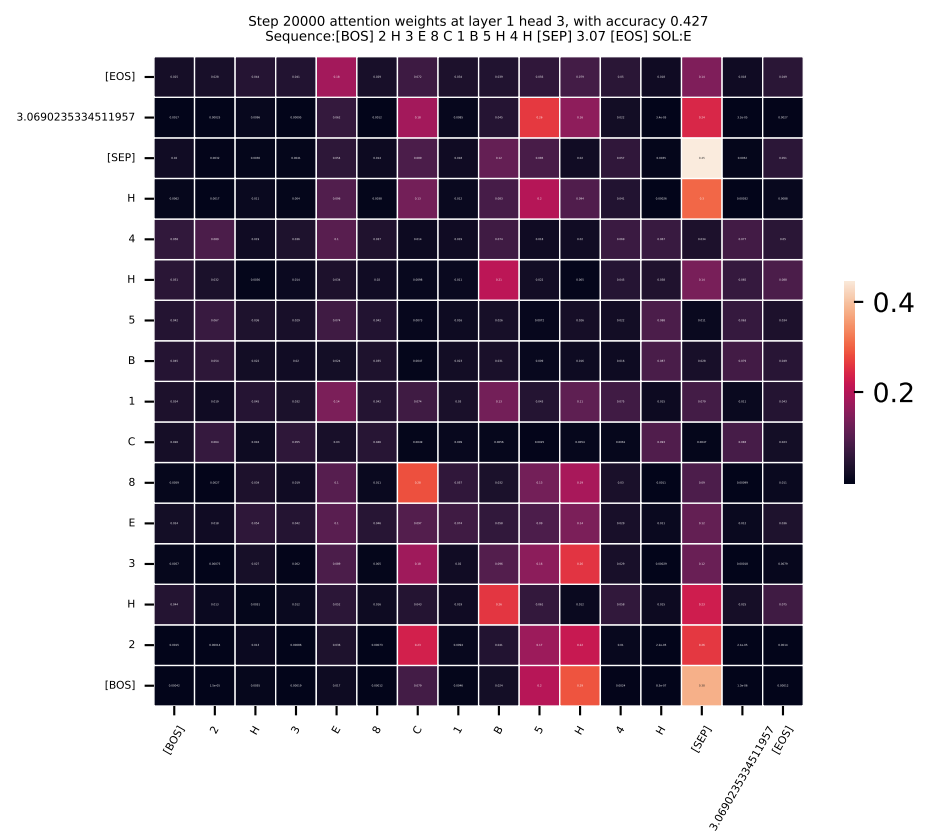

Step 20000 attention weights at layer 1 head 2, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

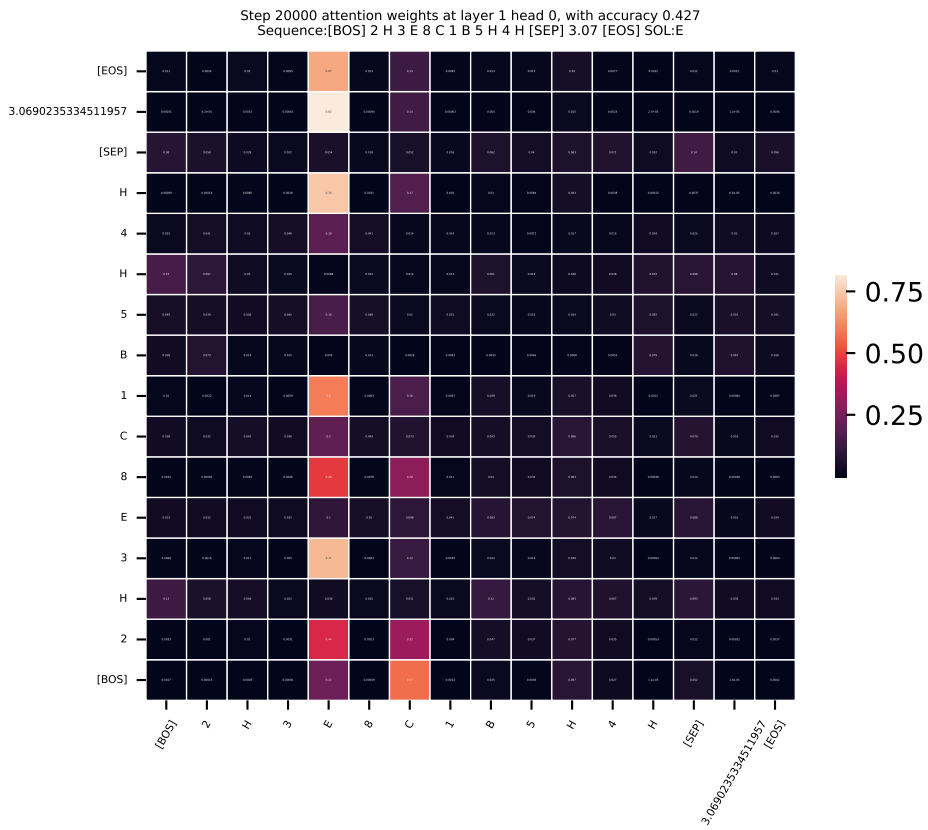

Step 20000 attention weights at layer 1 head 3, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.4
0.2

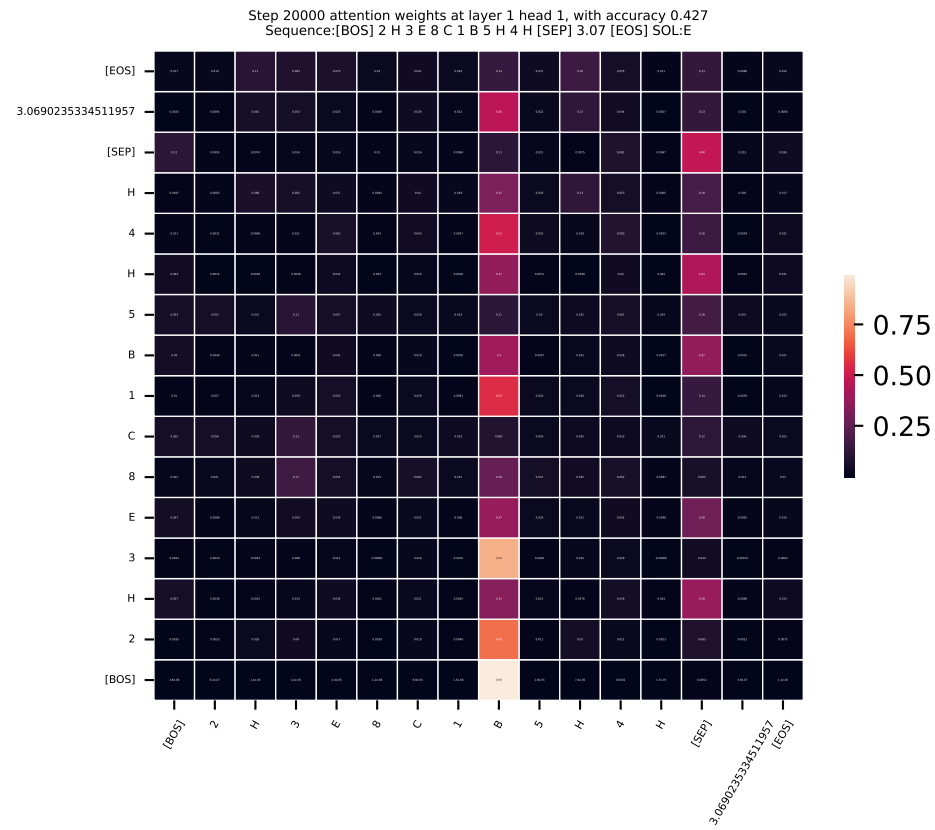

Step 20000 attention weights at layer 1 head 0, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

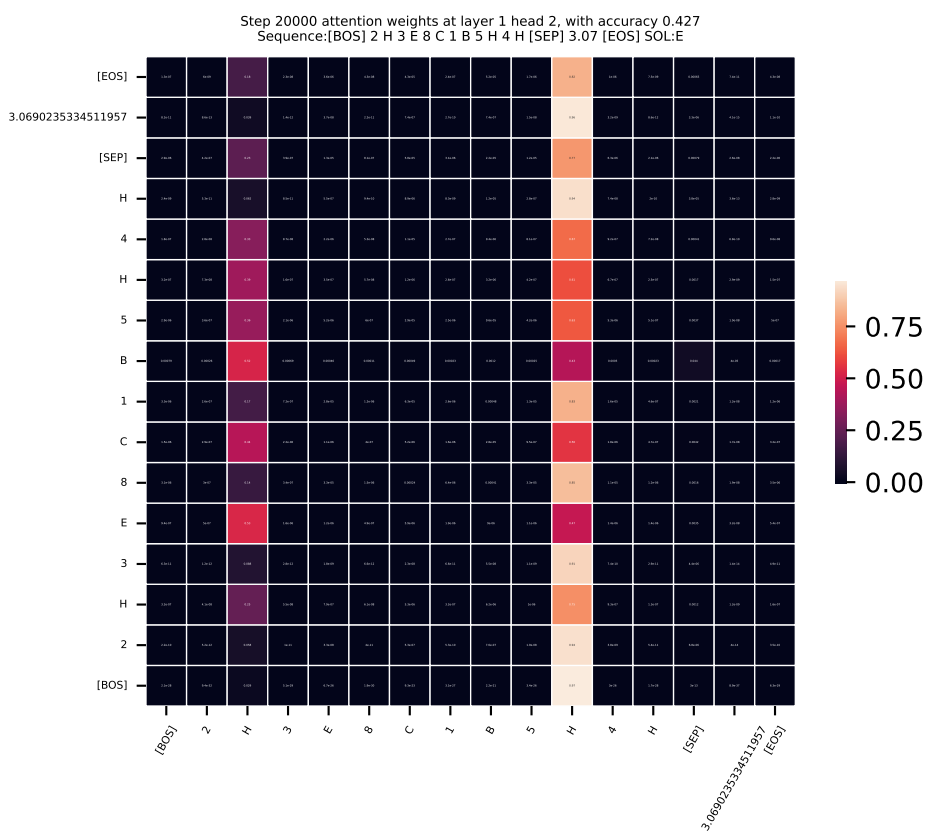

Step 20000 attention weights at layer 1 head 1, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

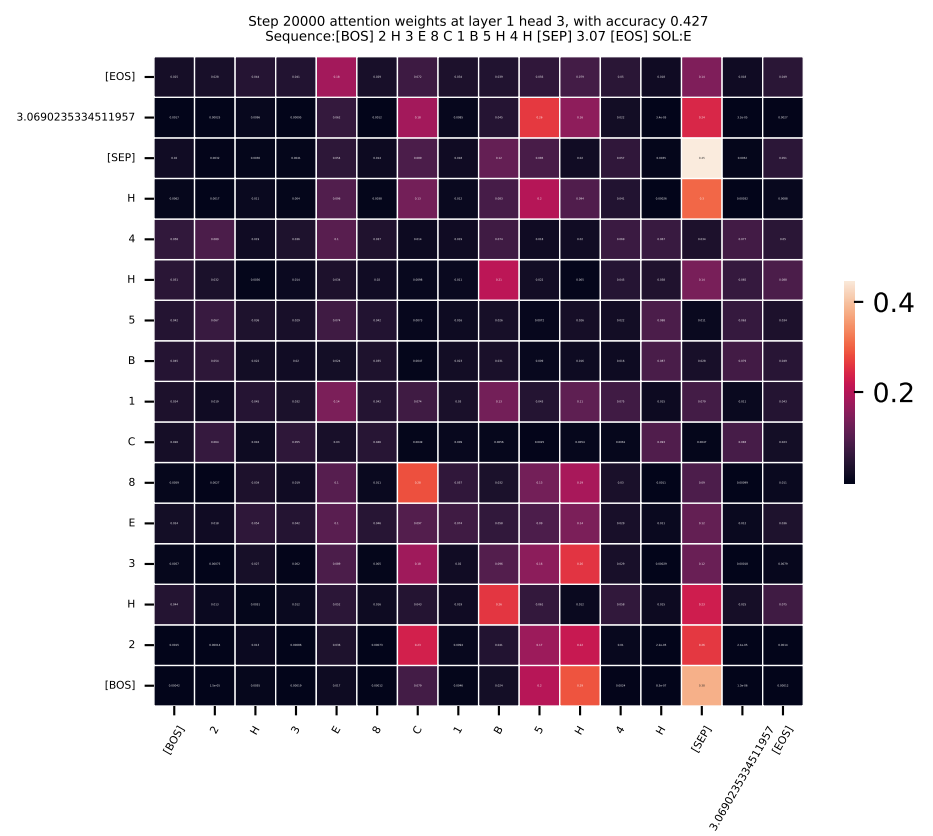

Step 20000 attention weights at layer 1 head 2, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

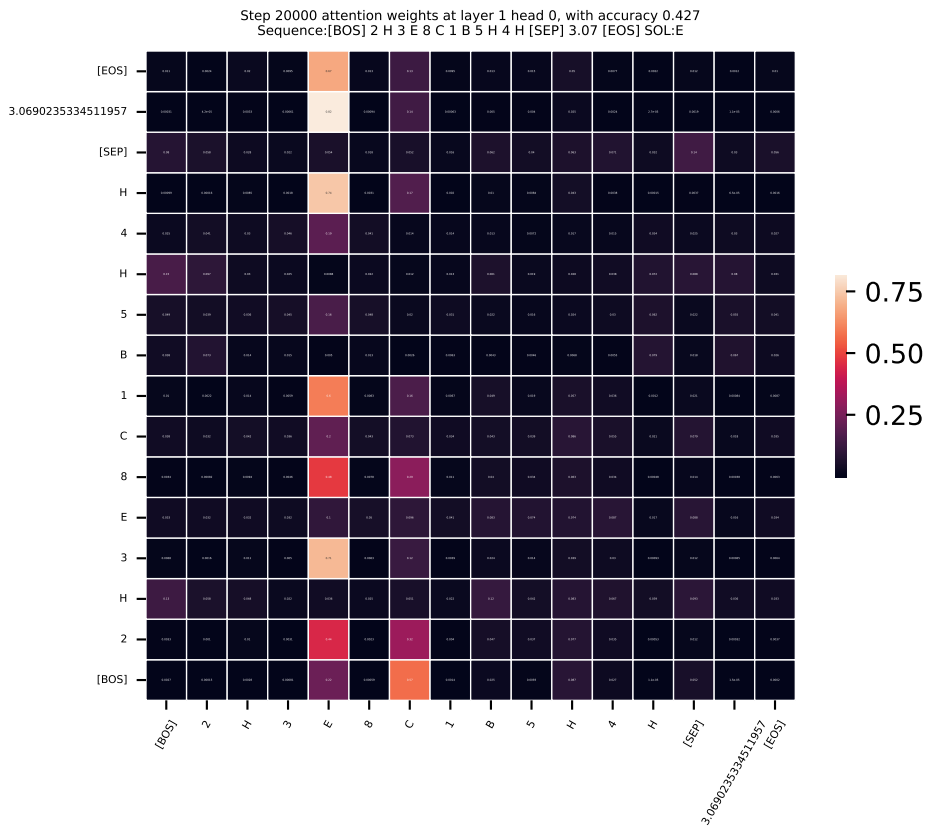

Step 20000 attention weights at layer 1 head 3, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.4
0.2

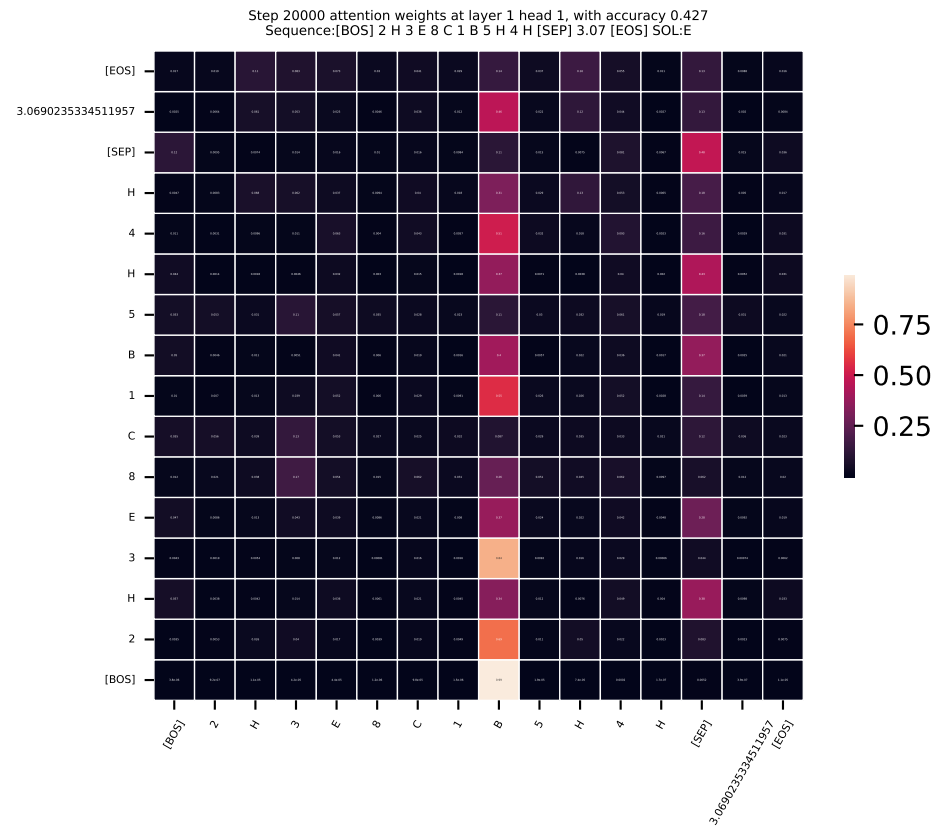

Step 20000 attention weights at layer 1 head 0, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

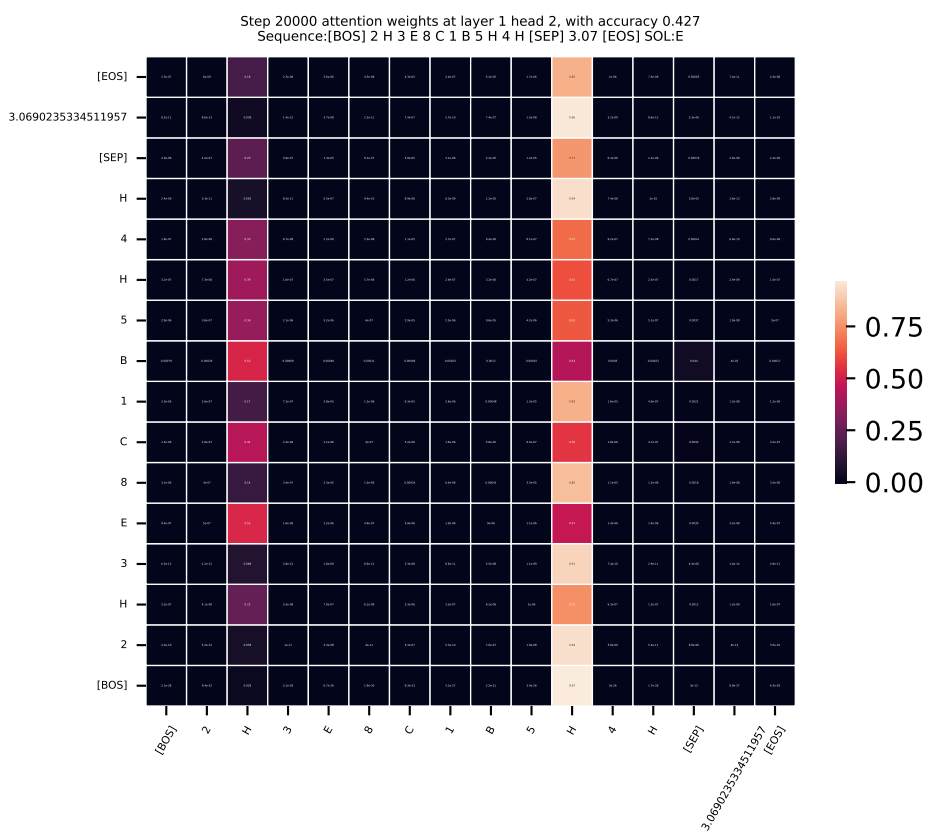

Step 20000 attention weights at layer 1 head 1, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

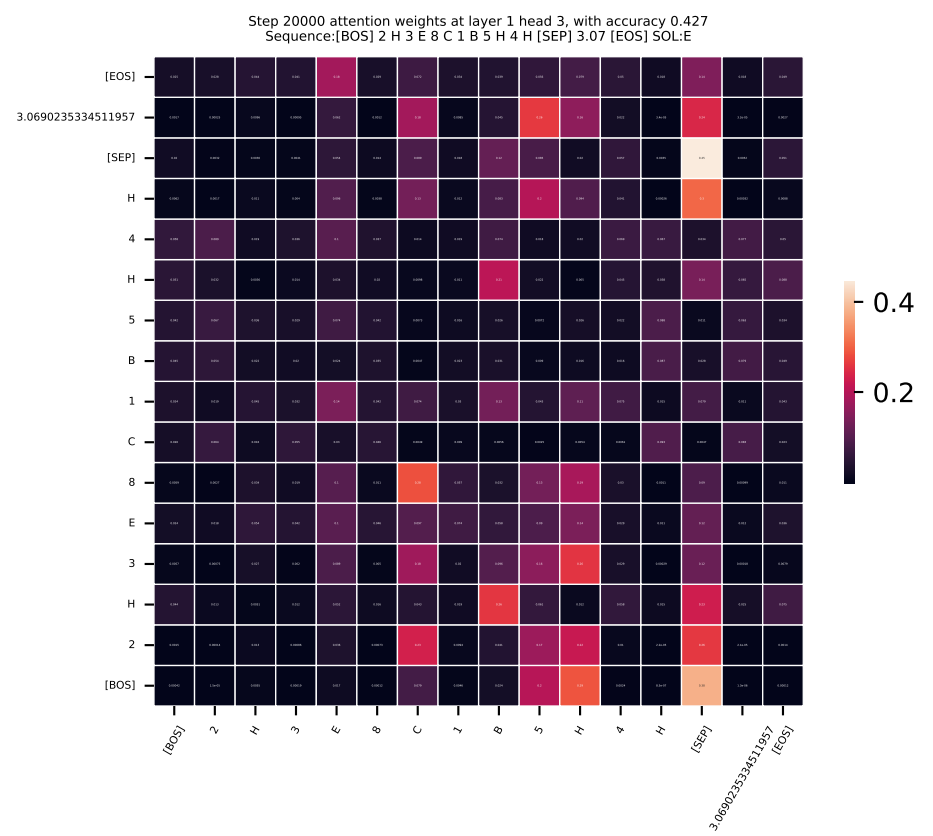

Step 20000 attention weights at layer 1 head 2, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

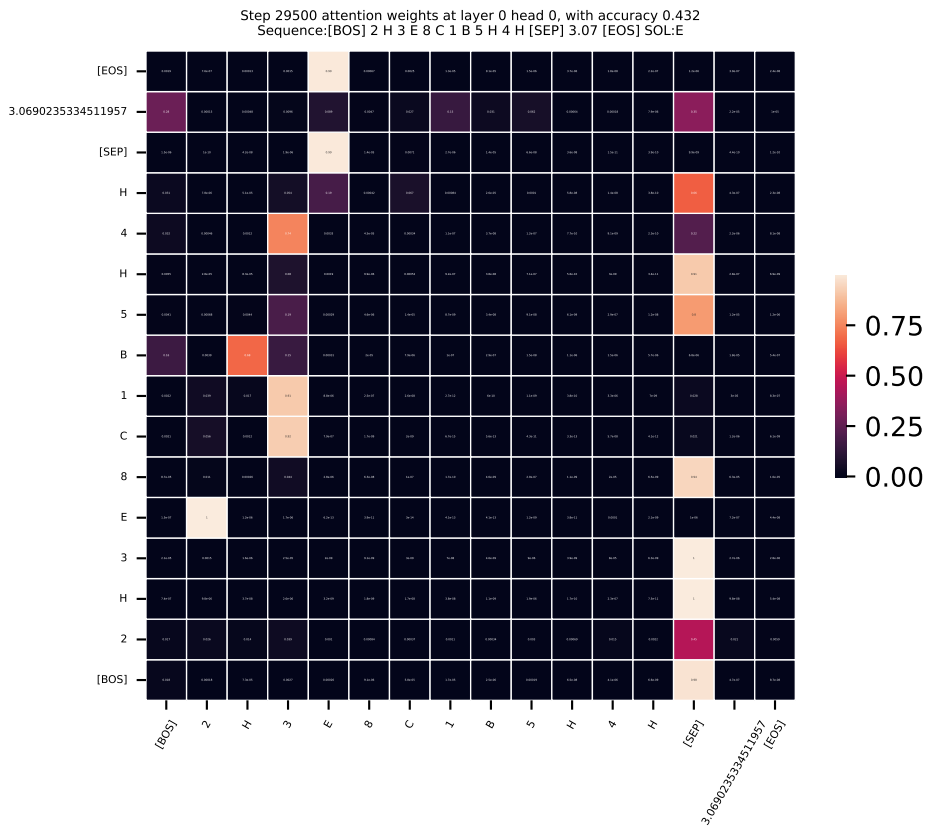

Step 20000 attention weights at layer 1 head 3, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.4
0.2

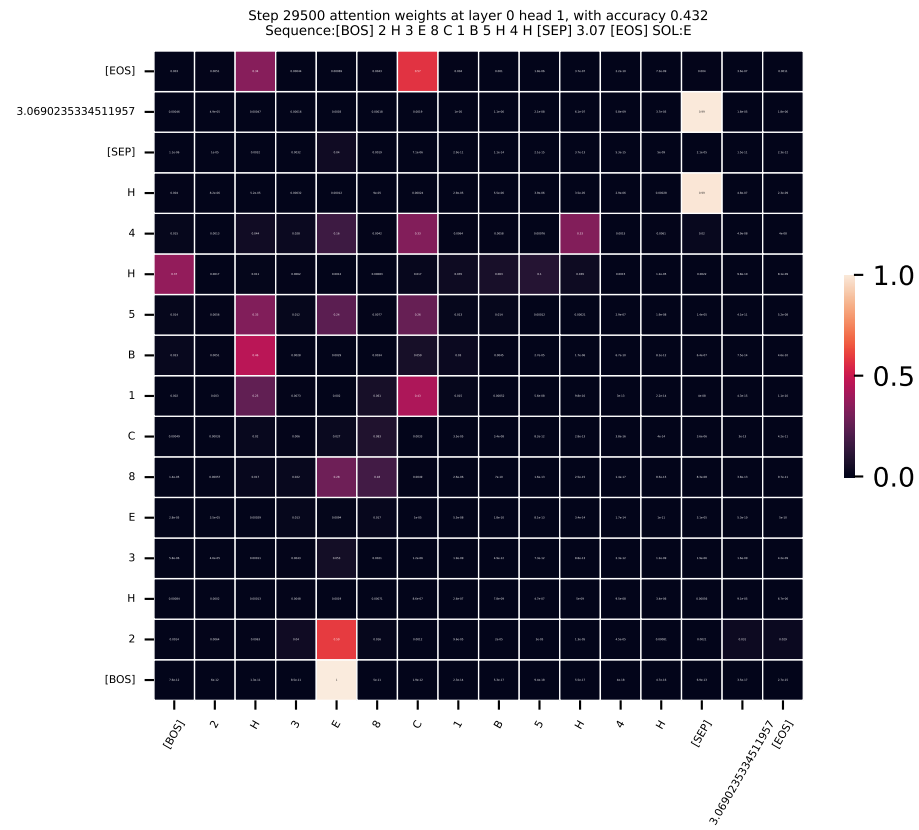

Step 20000 attention weights at layer 1 head 0, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

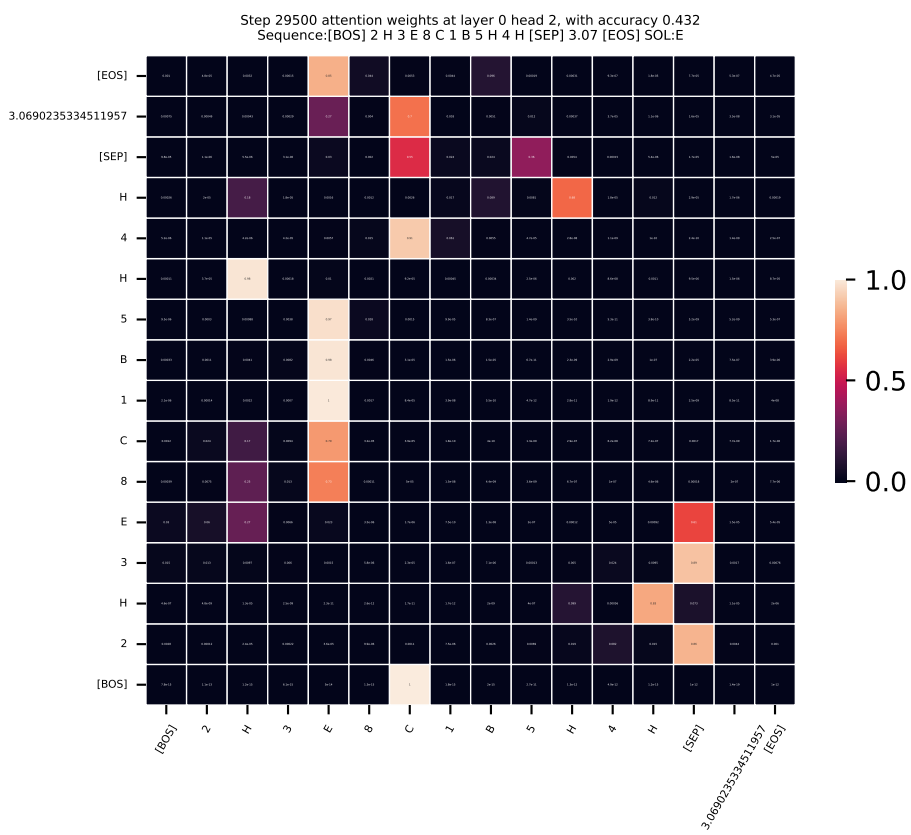

Step 20000 attention weights at layer 1 head 1, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

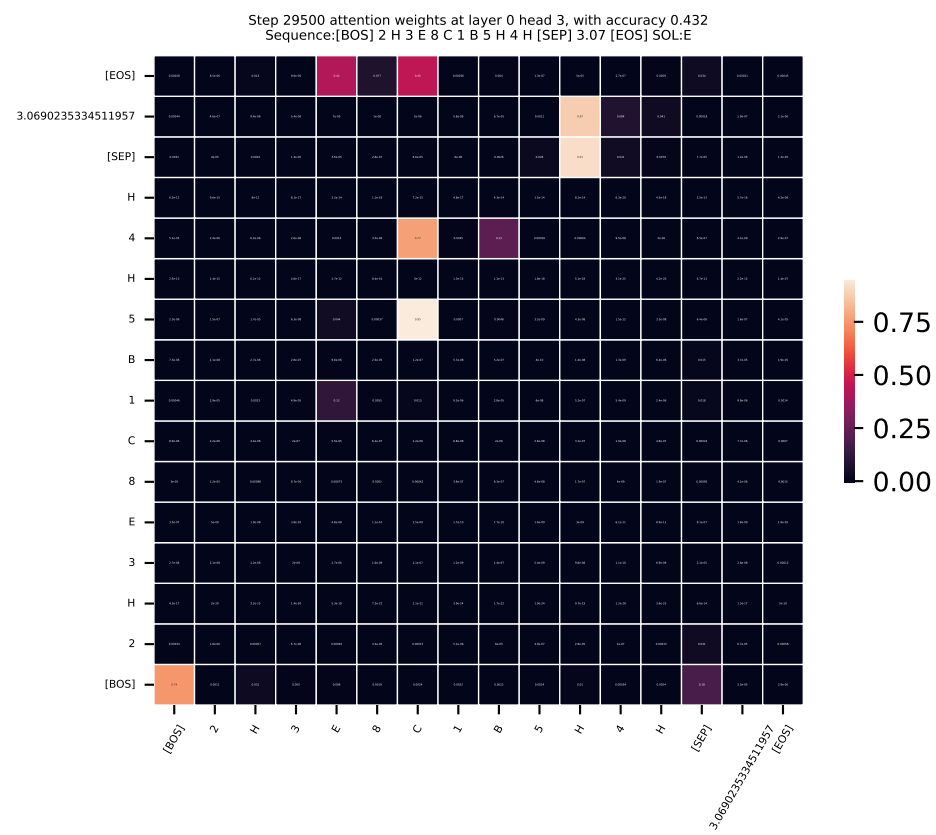

Step 20000 attention weights at layer 1 head 2, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

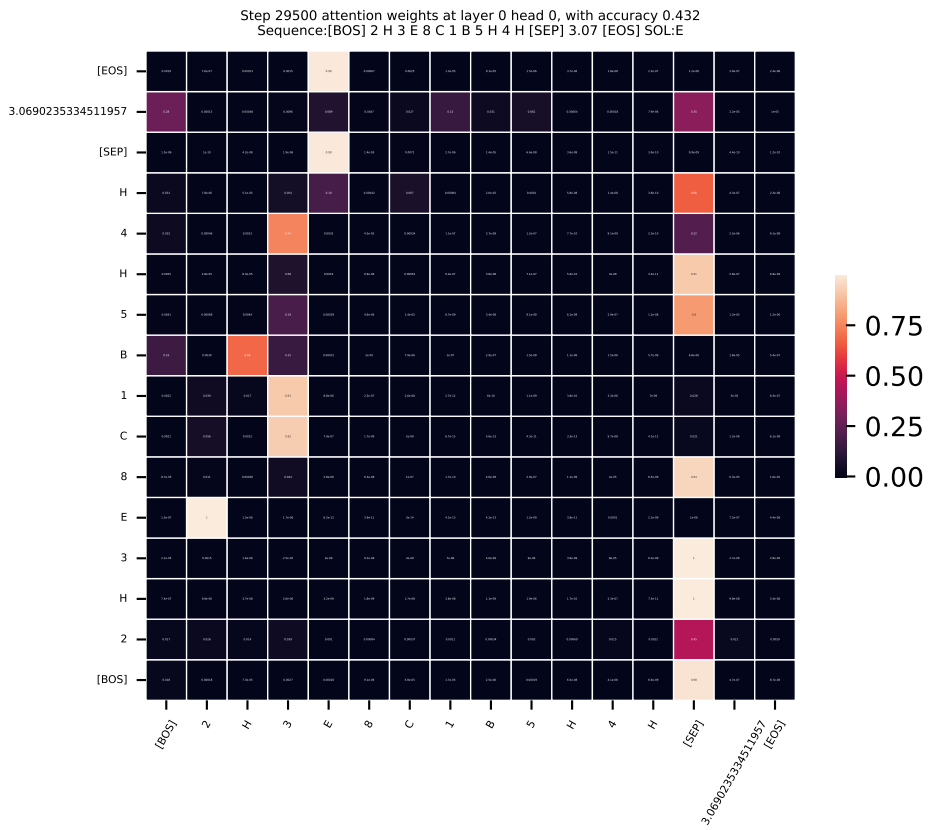

Step 20000 attention weights at layer 1 head 3, with accuracy 0.427
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.4
0.2

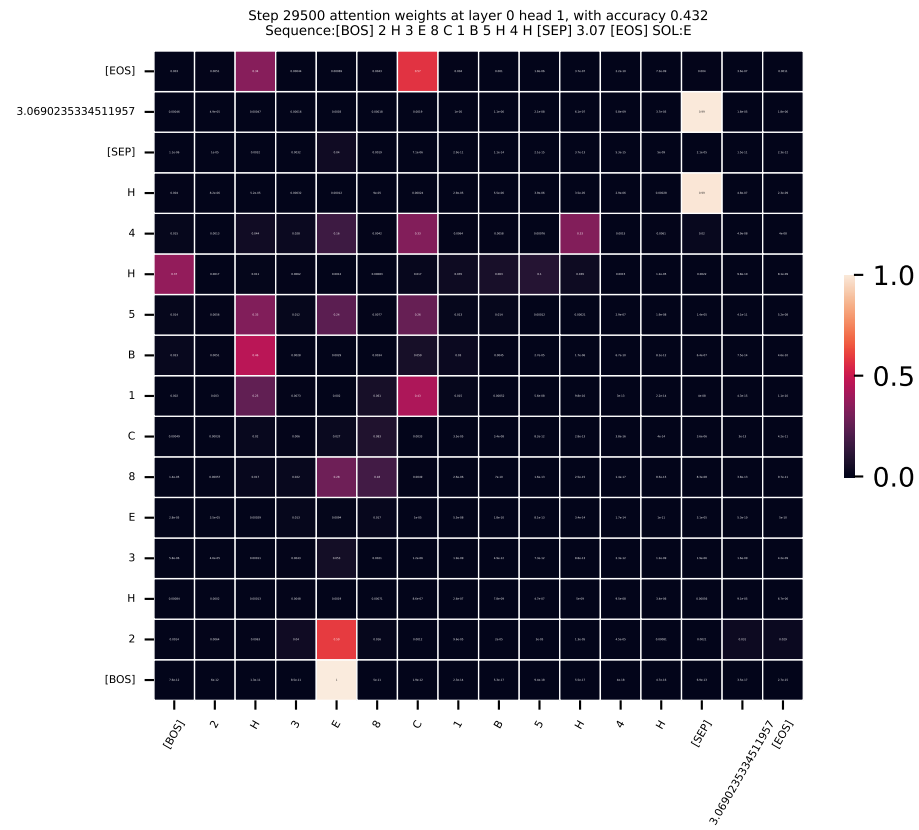

Step 29500 attention weights at layer 0 head 0, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

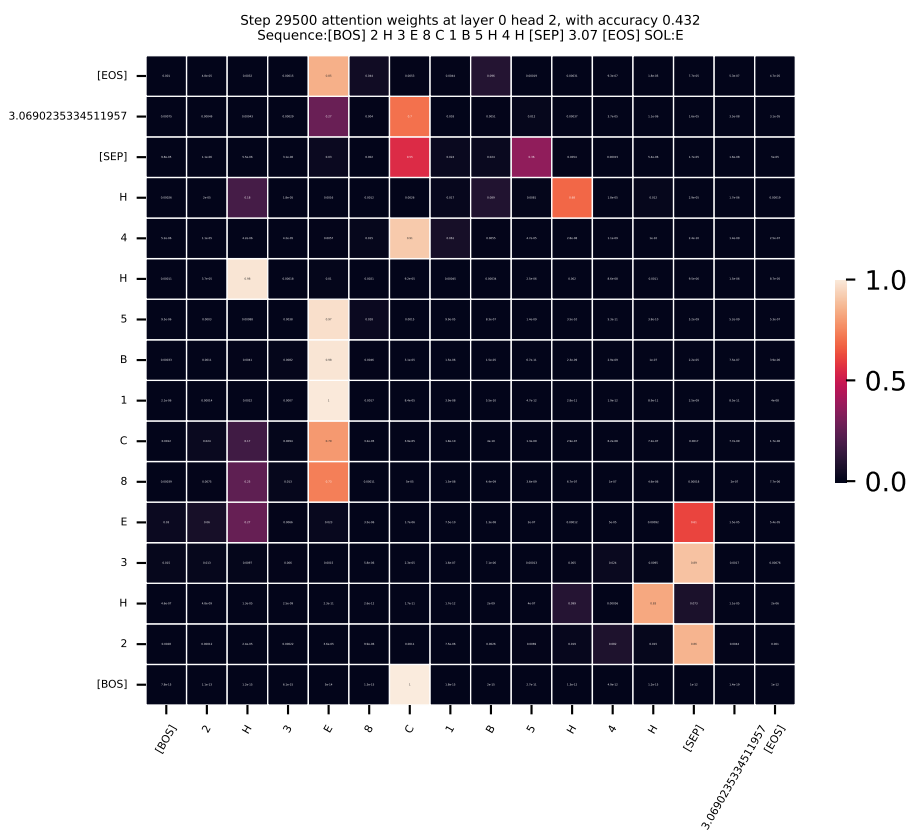

Step 29500 attention weights at layer 0 head 1, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

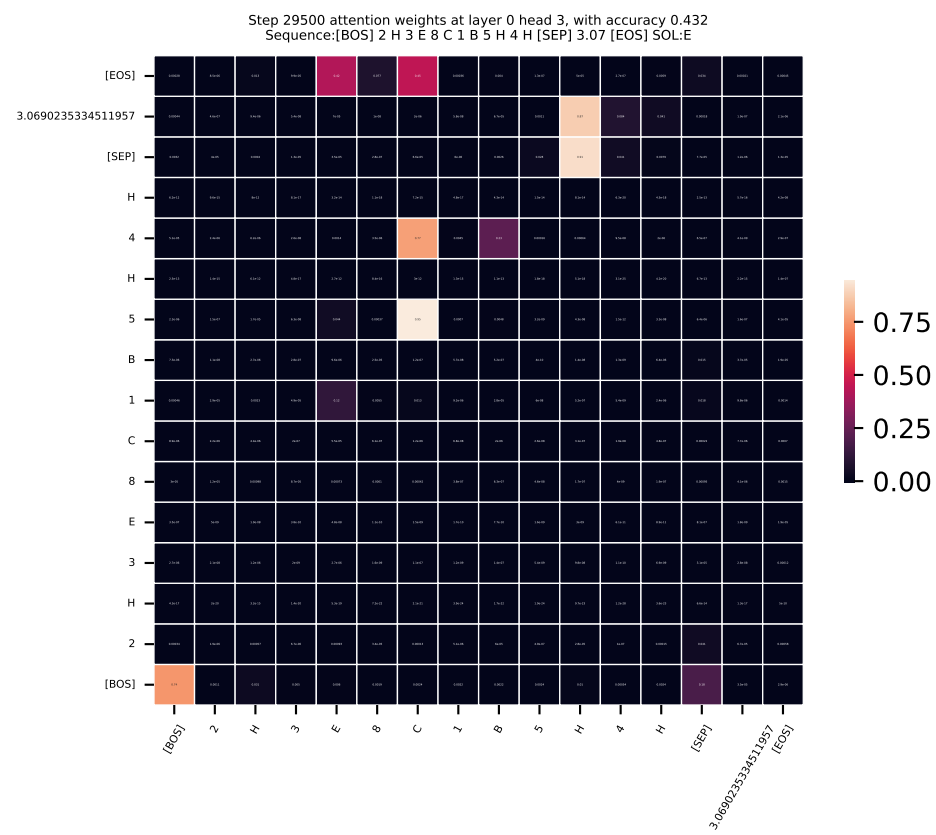

Step 29500 attention weights at layer 0 head 2, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

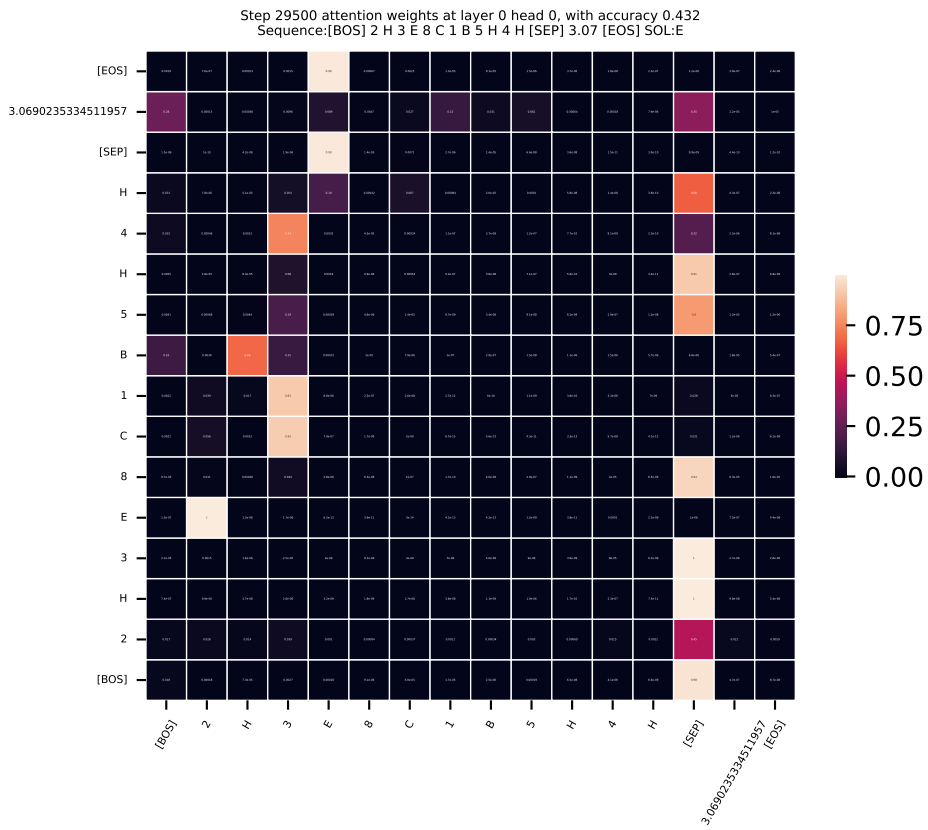

Step 29500 attention weights at layer 0 head 3, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

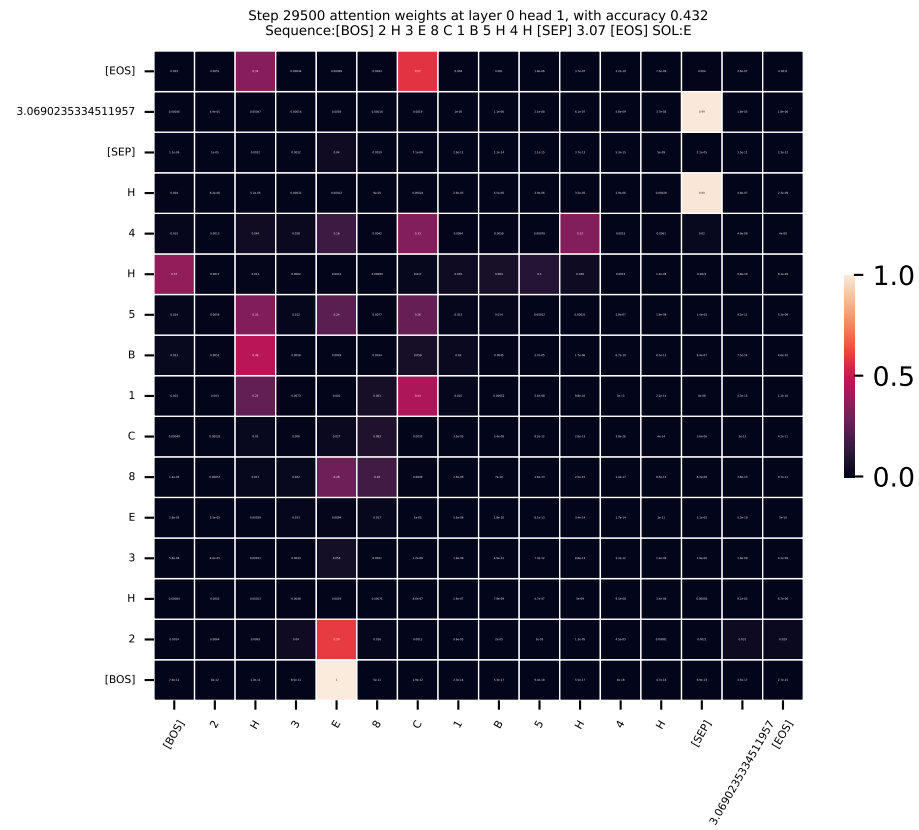

Step 29500 attention weights at layer 0 head 0, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

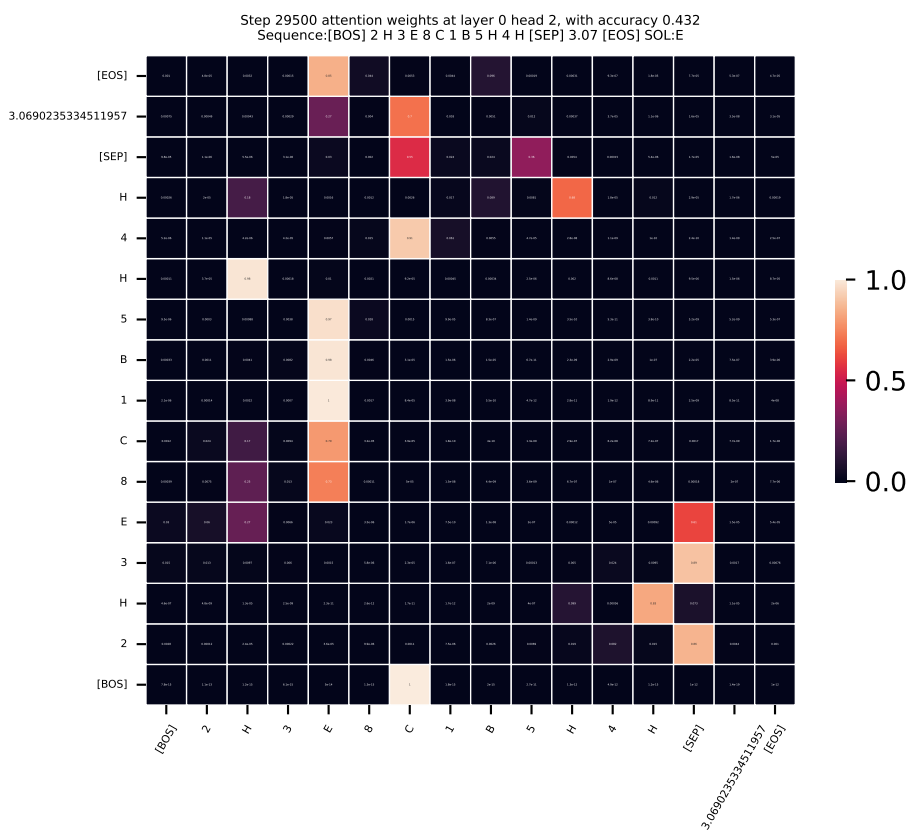

Step 29500 attention weights at layer 0 head 1, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

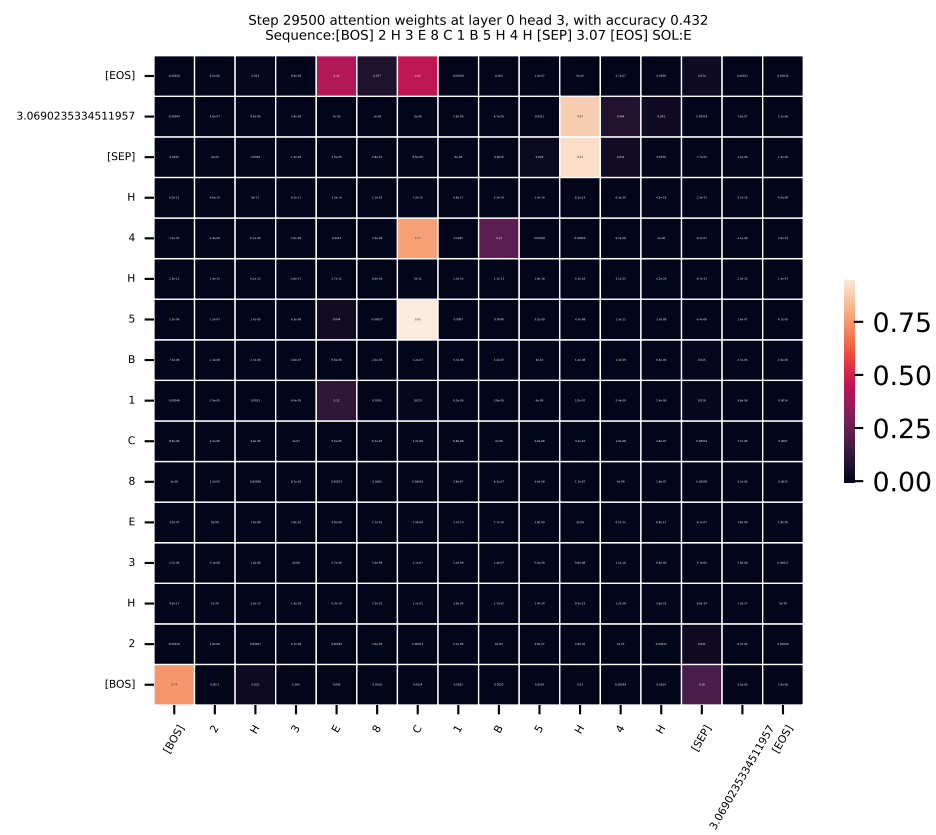

Step 29500 attention weights at layer 0 head 2, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

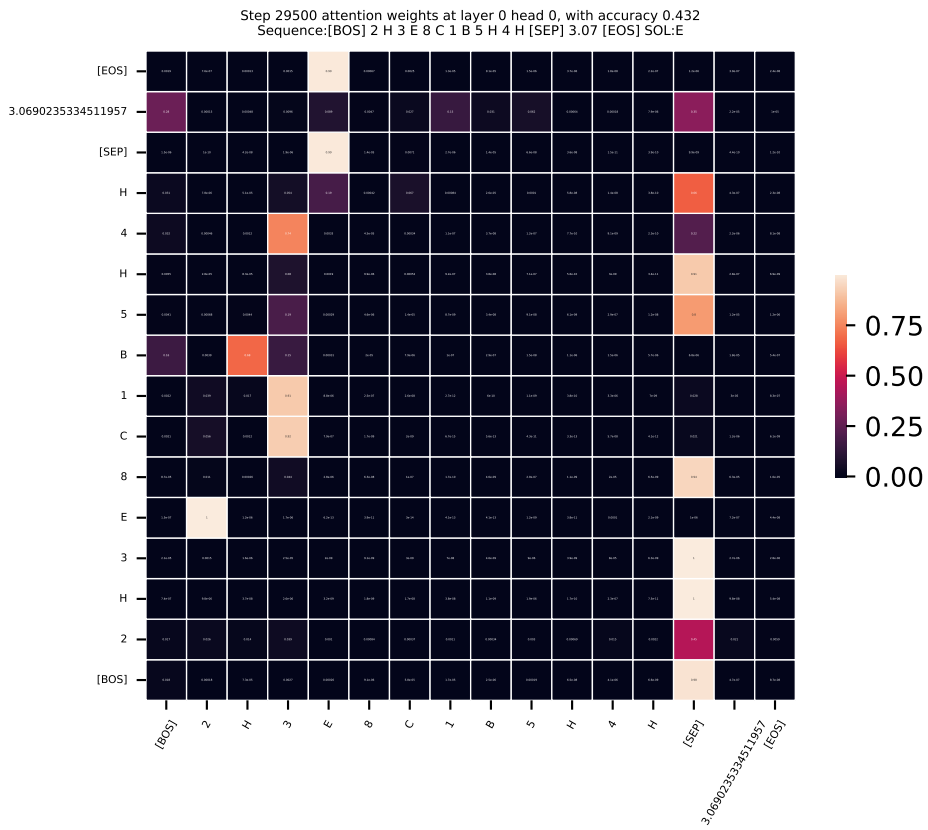

Step 29500 attention weights at layer 0 head 3, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

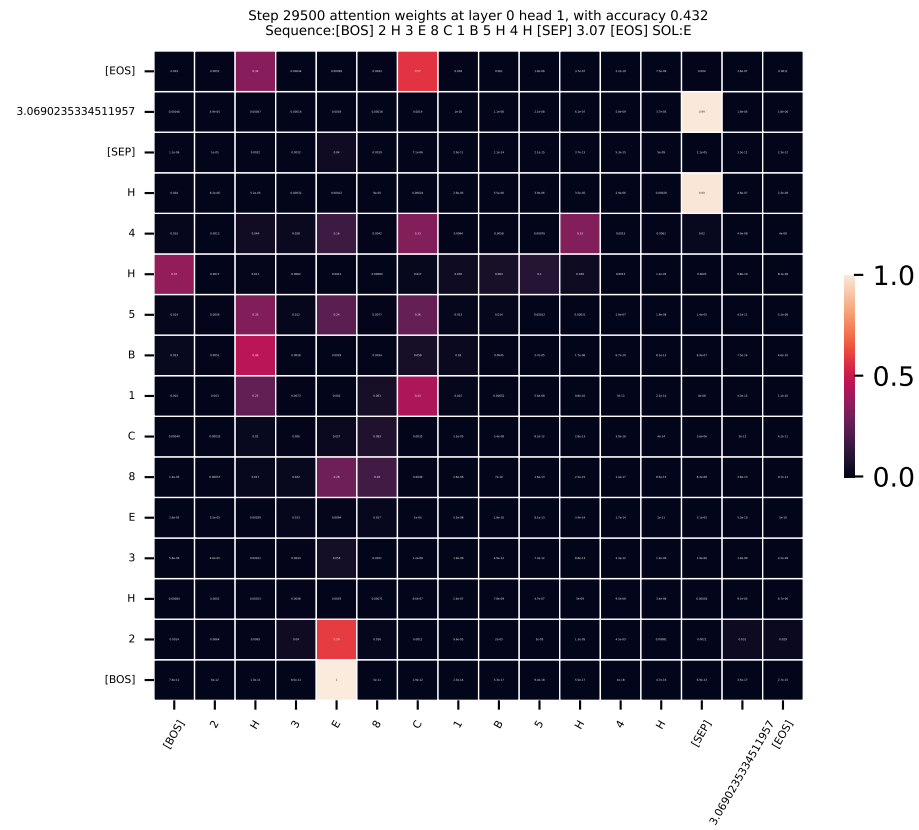

Step 29500 attention weights at layer 0 head 0, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

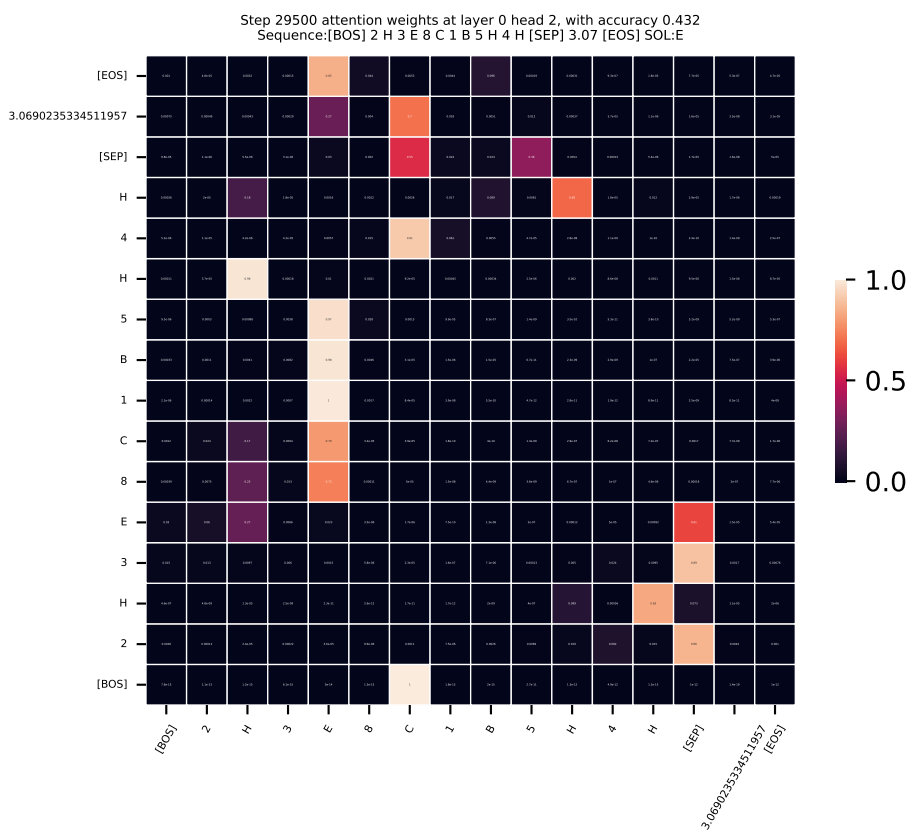

Step 29500 attention weights at layer 0 head 1, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

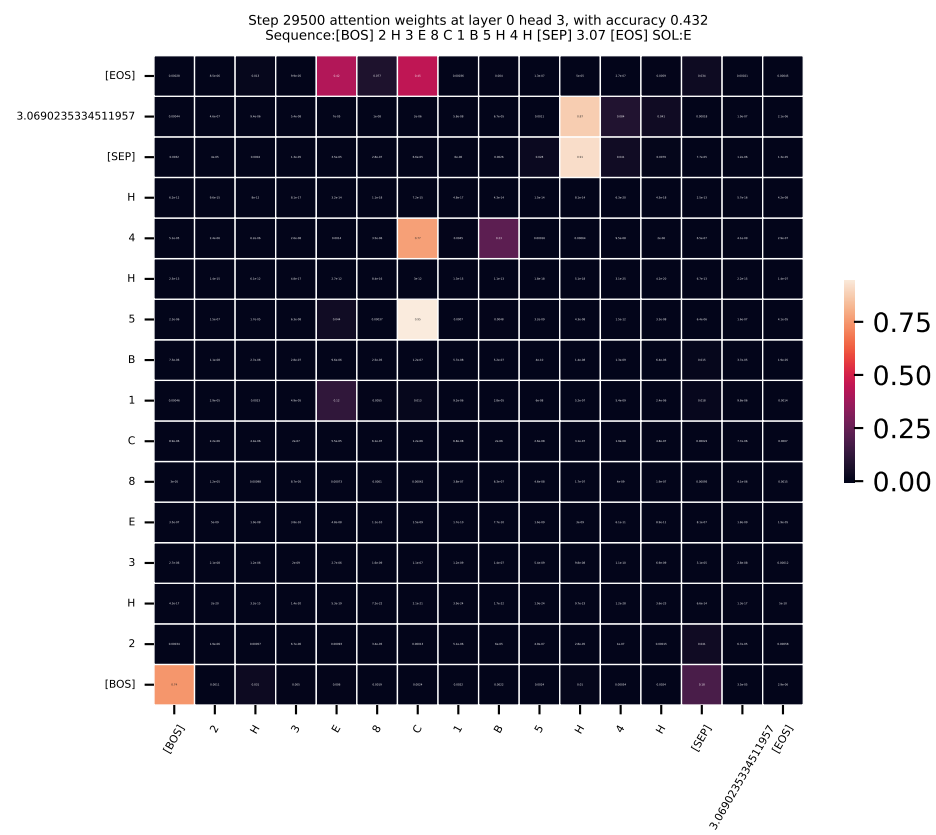

Step 29500 attention weights at layer 0 head 2, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

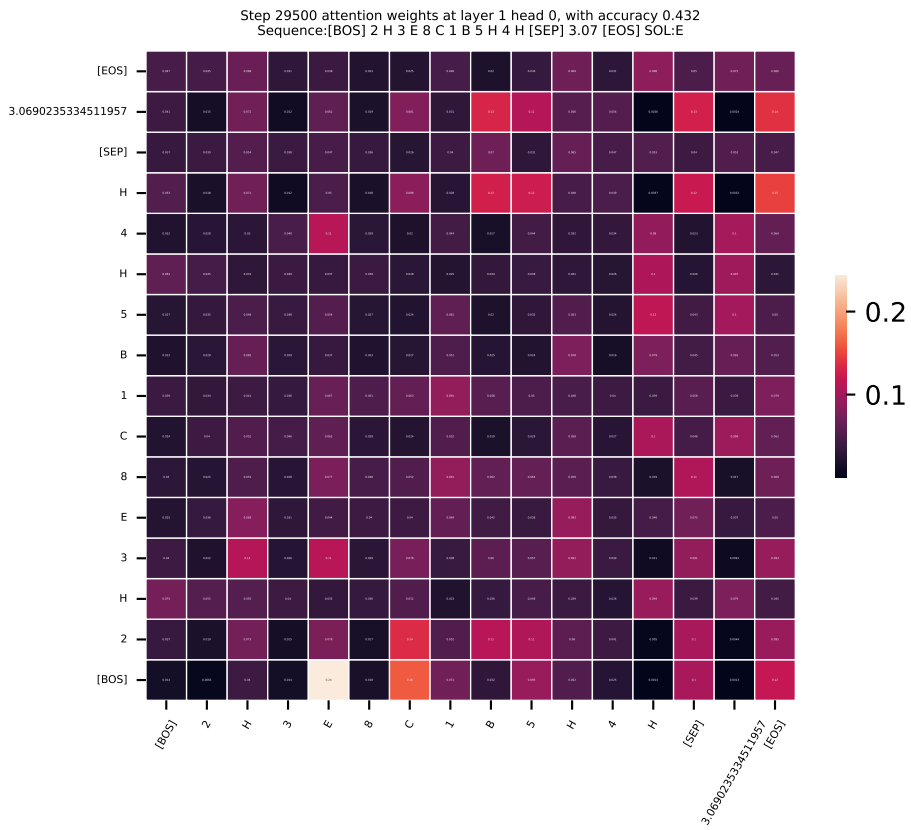

Step 29500 attention weights at layer 0 head 3, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

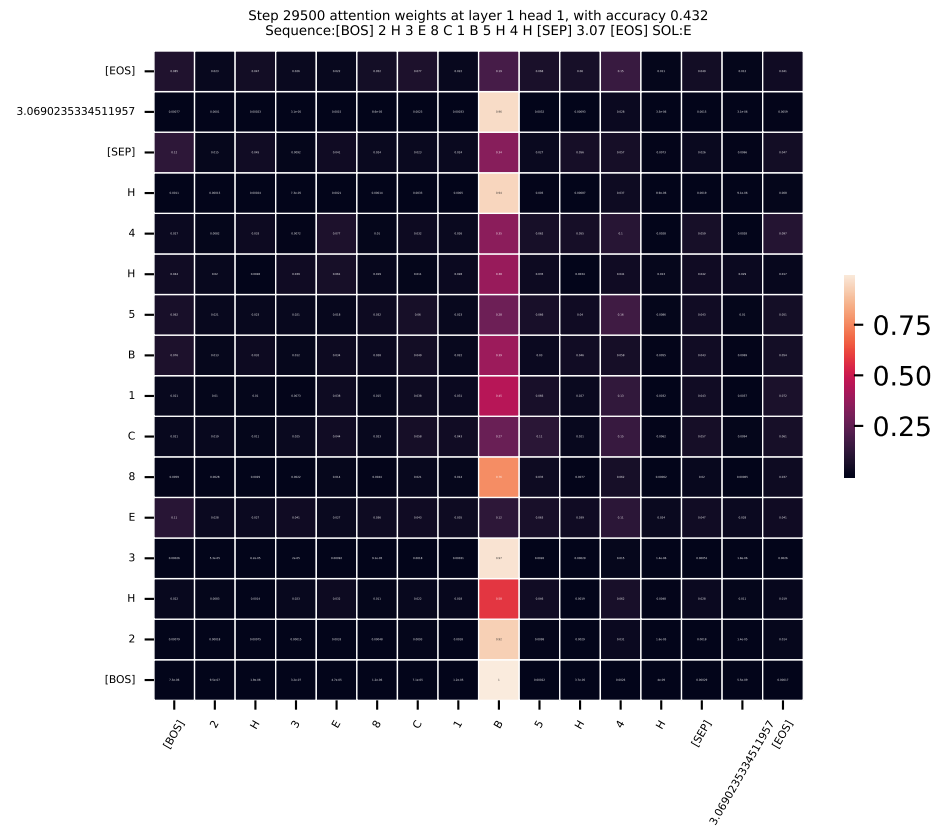

Step 29500 attention weights at layer 0 head 0, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

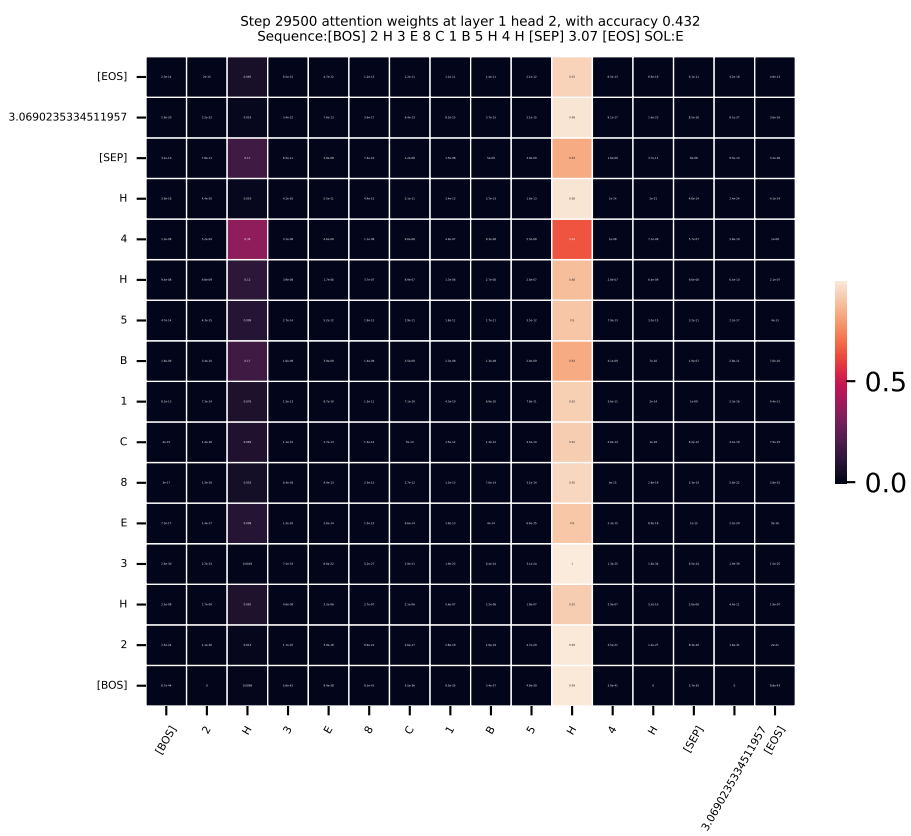

Step 29500 attention weights at layer 0 head 1, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

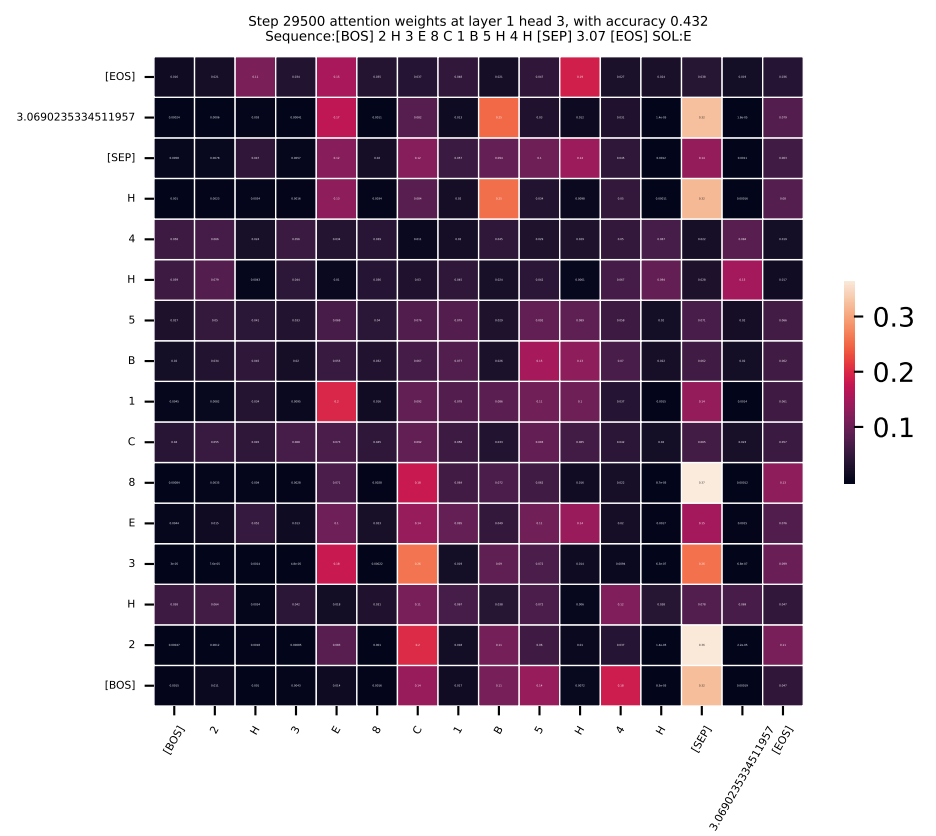

Step 29500 attention weights at layer 0 head 2, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

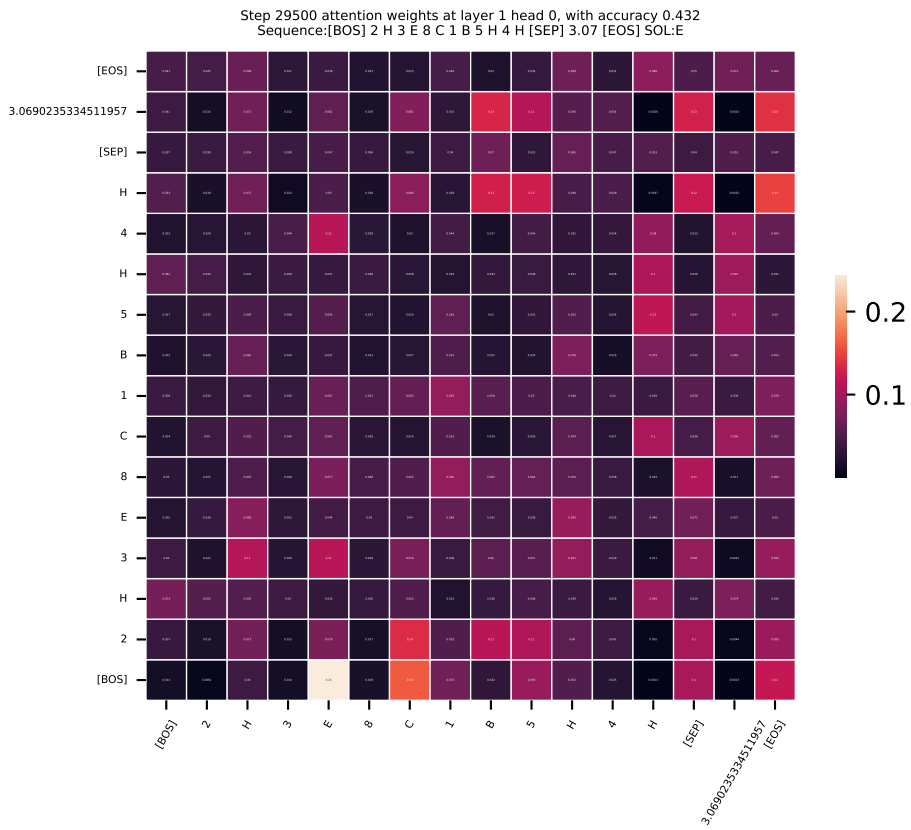

Step 29500 attention weights at layer 0 head 3, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

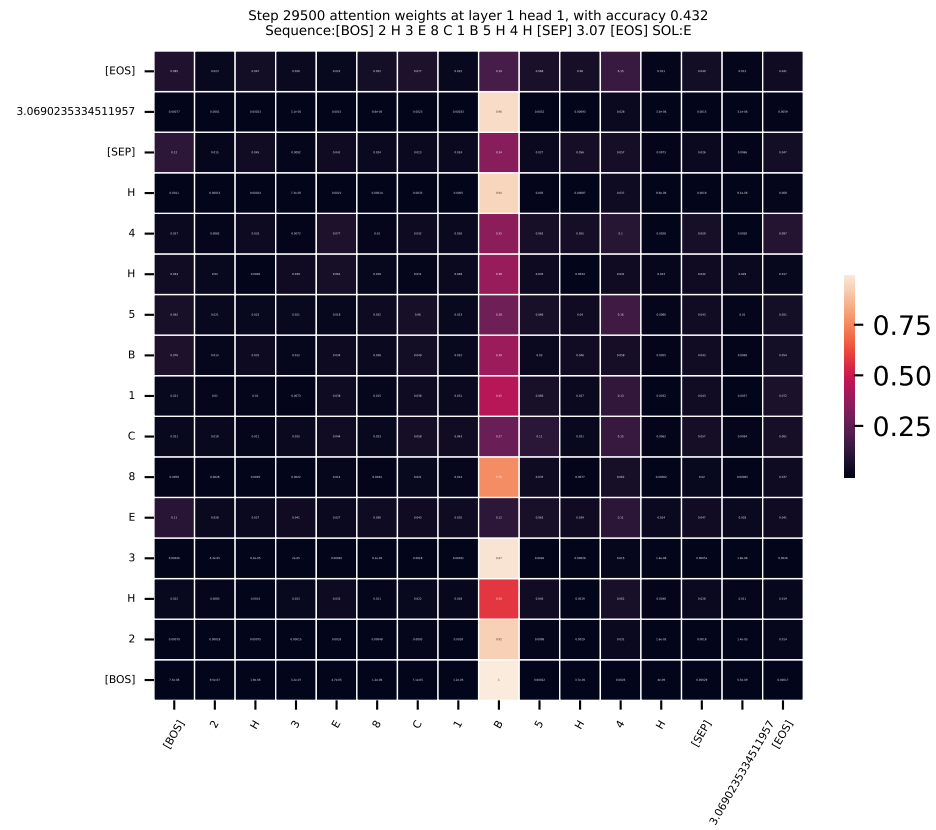

Step 29500 attention weights at layer 1 head 0, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.2
0.1

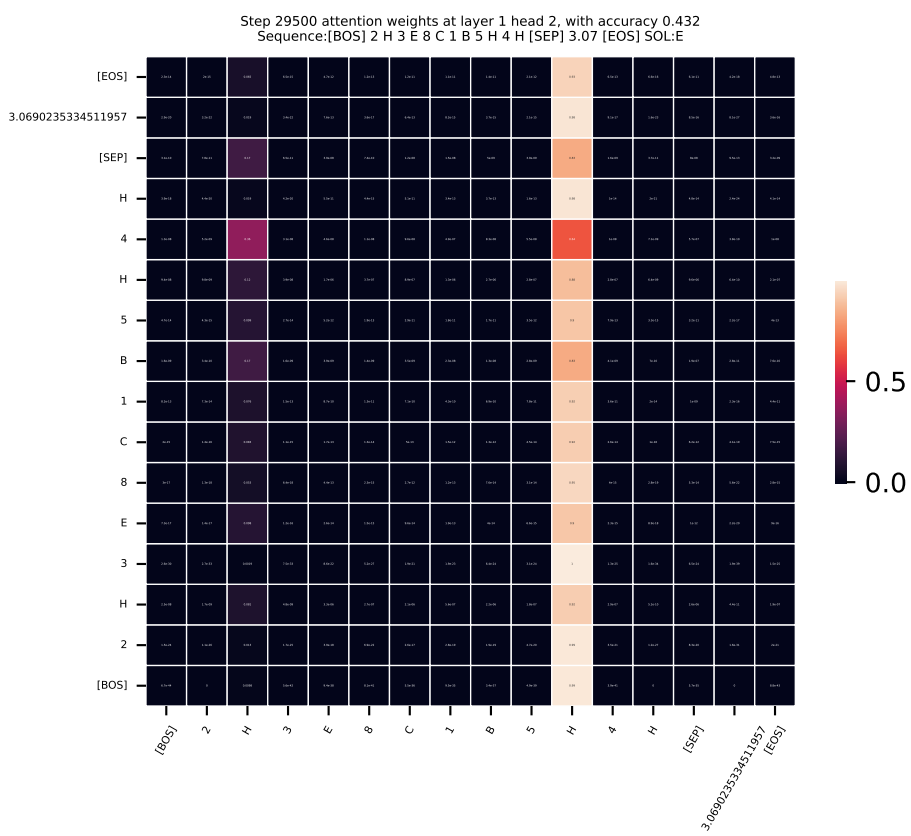

Step 29500 attention weights at layer 1 head 1, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

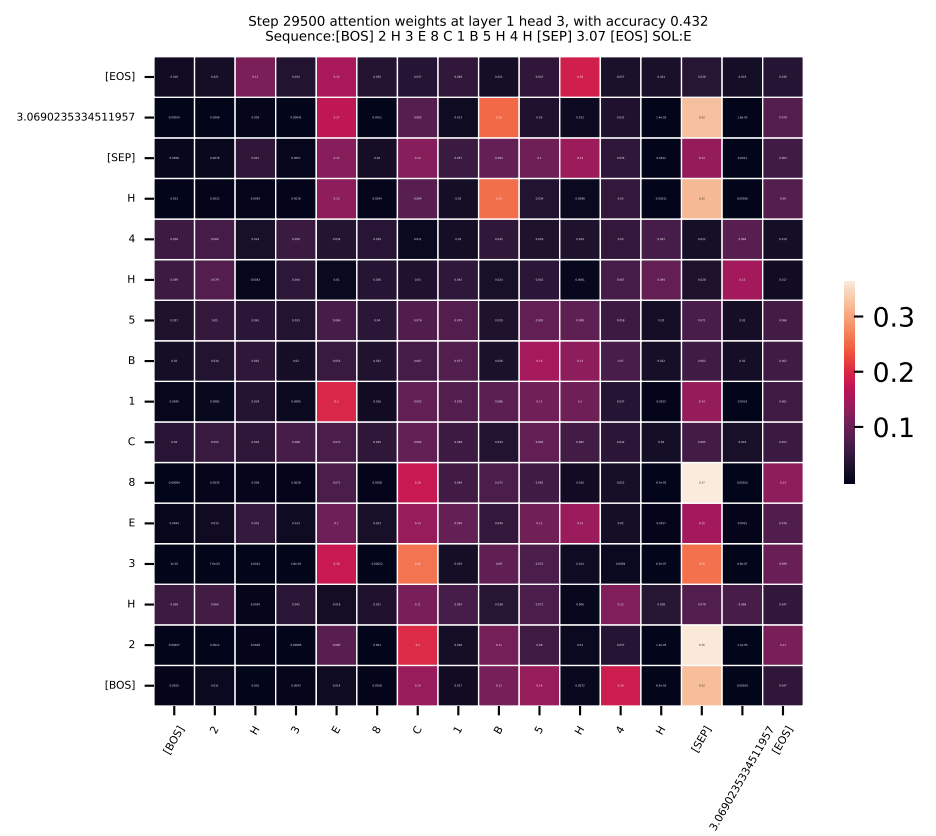

Step 29500 attention weights at layer 1 head 2, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.5
0.0

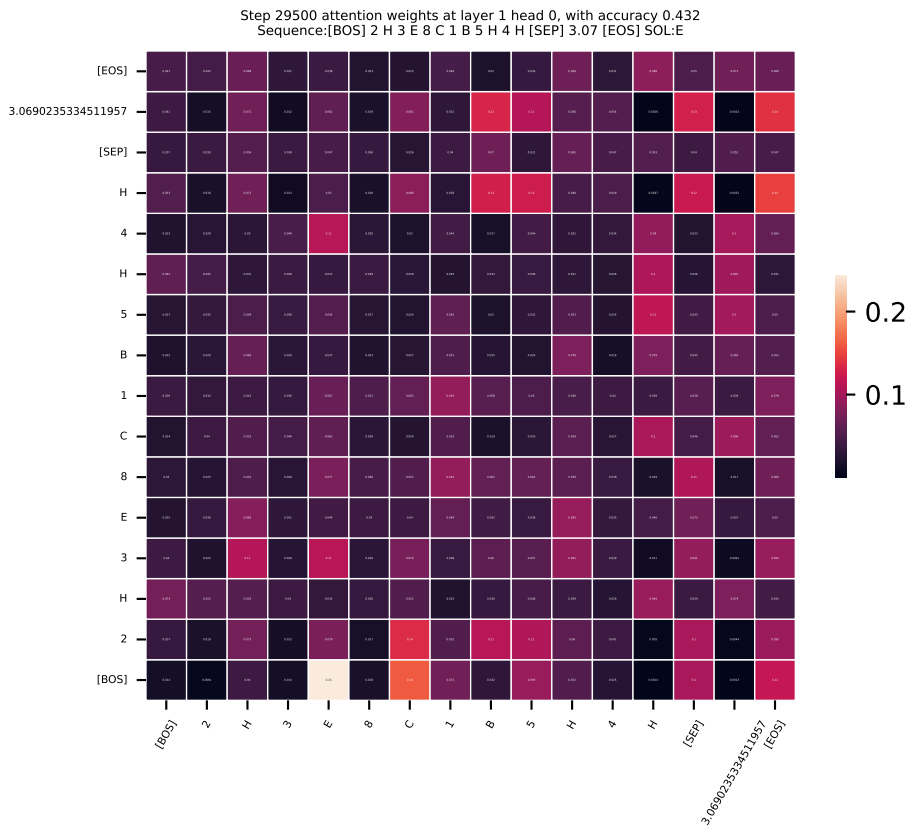

Step 29500 attention weights at layer 1 head 3, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.3
0.2
0.1

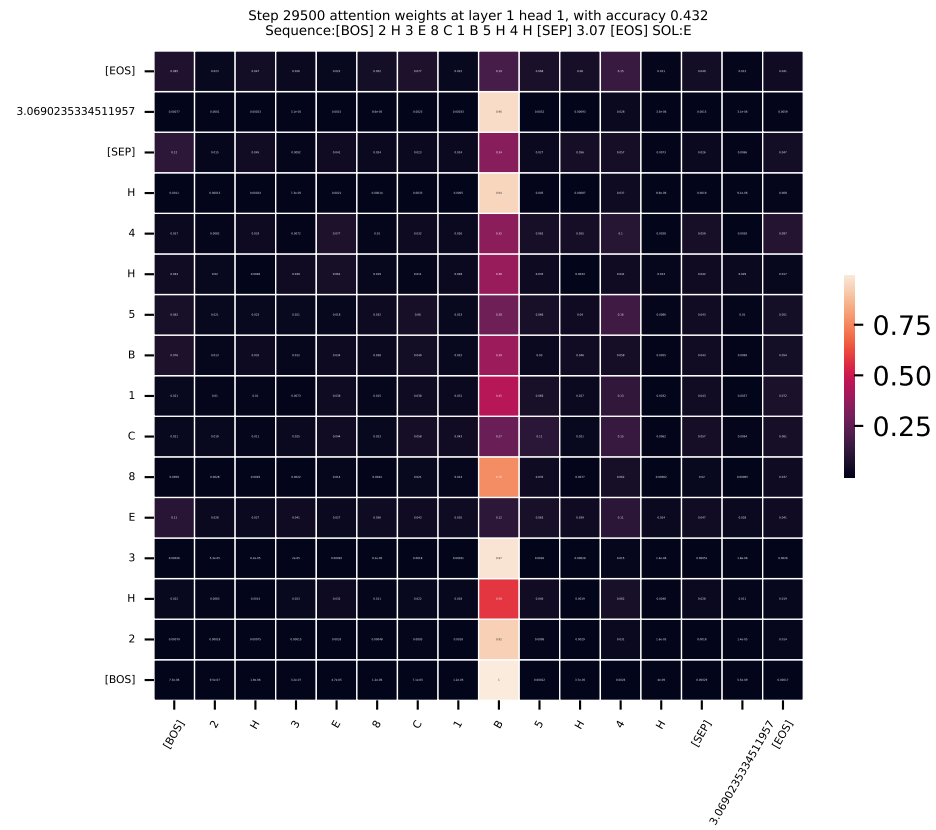

Step 29500 attention weights at layer 1 head 0, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.2
0.1

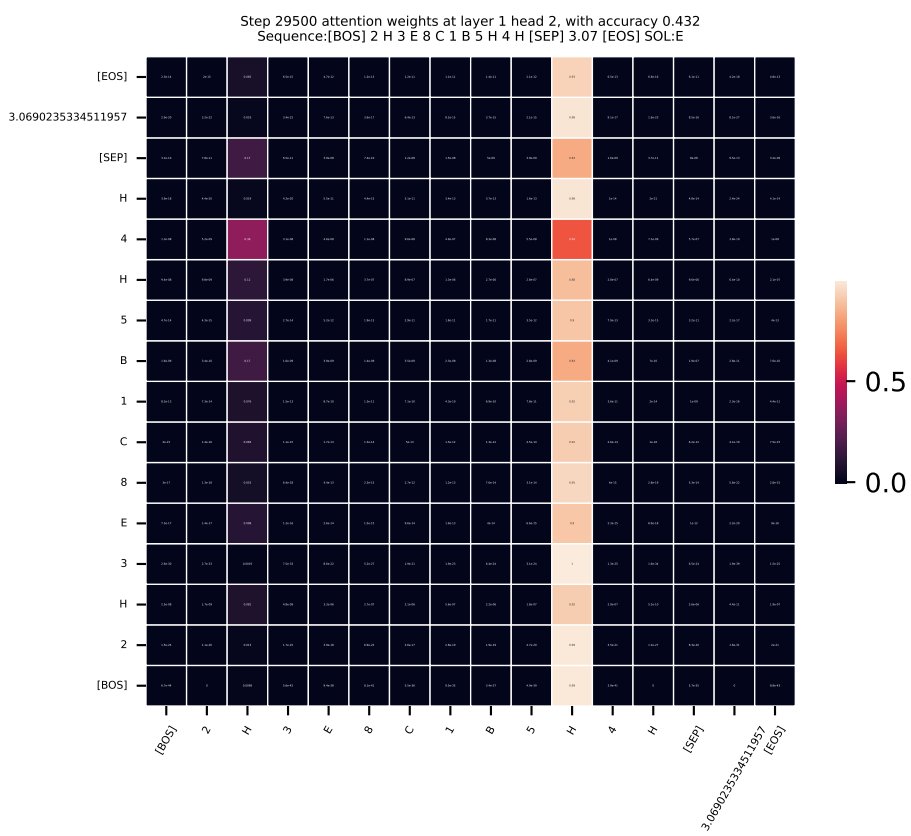

Step 29500 attention weights at layer 1 head 1, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

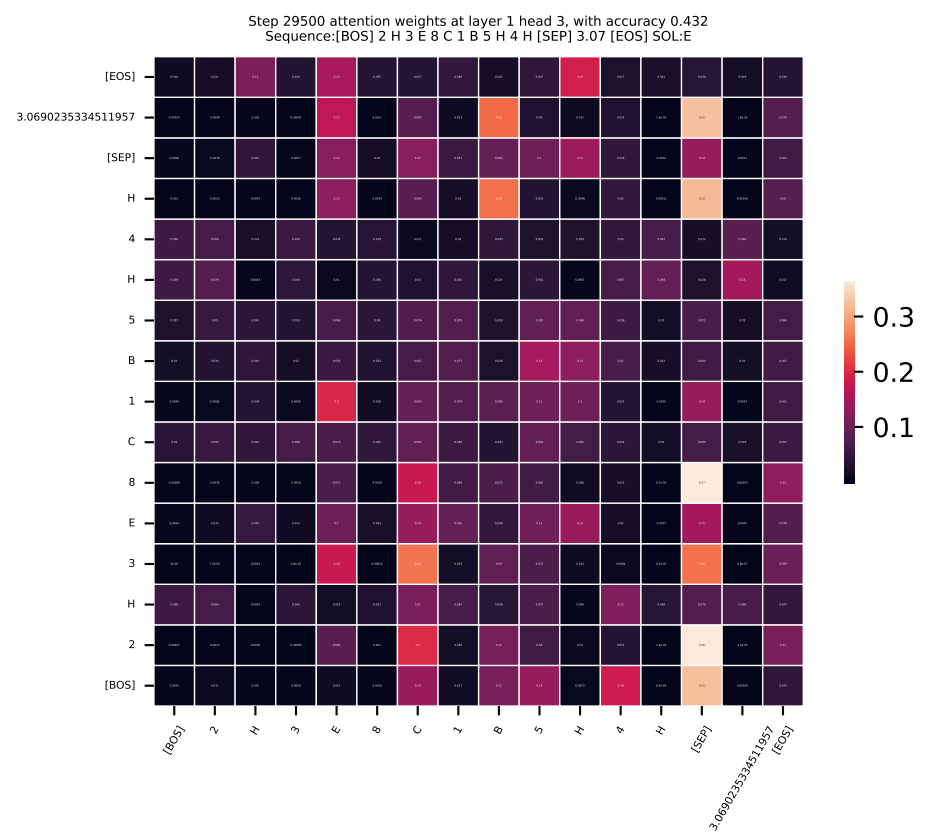

Step 29500 attention weights at layer 1 head 2, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.5
0.0

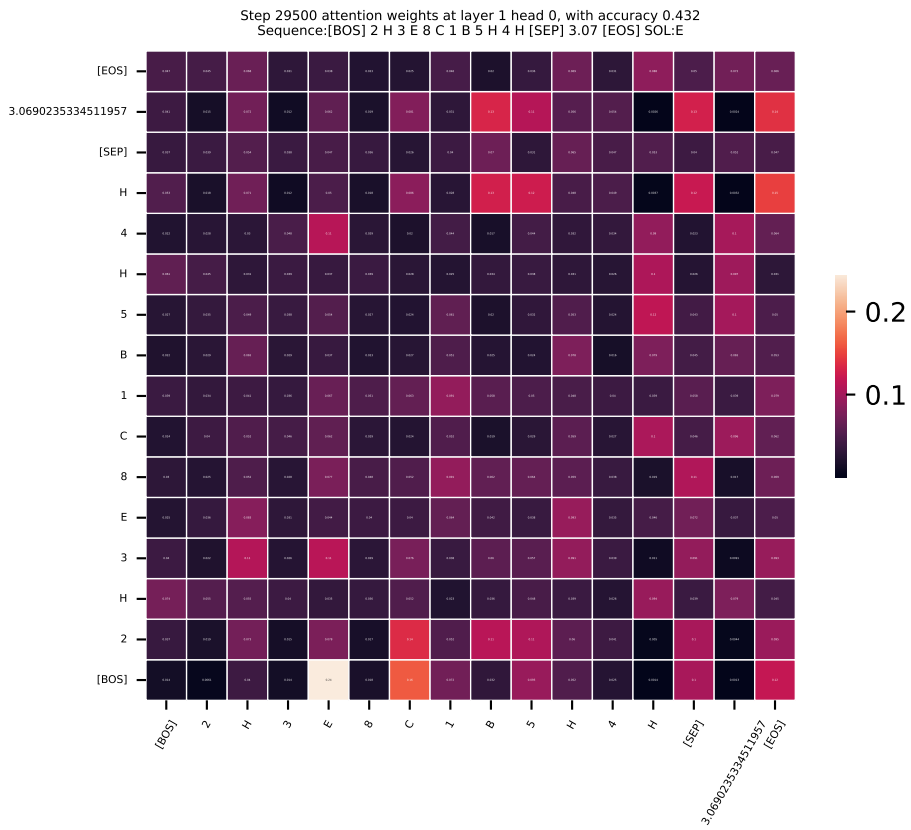

Step 29500 attention weights at layer 1 head 3, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.3
0.2
0.1

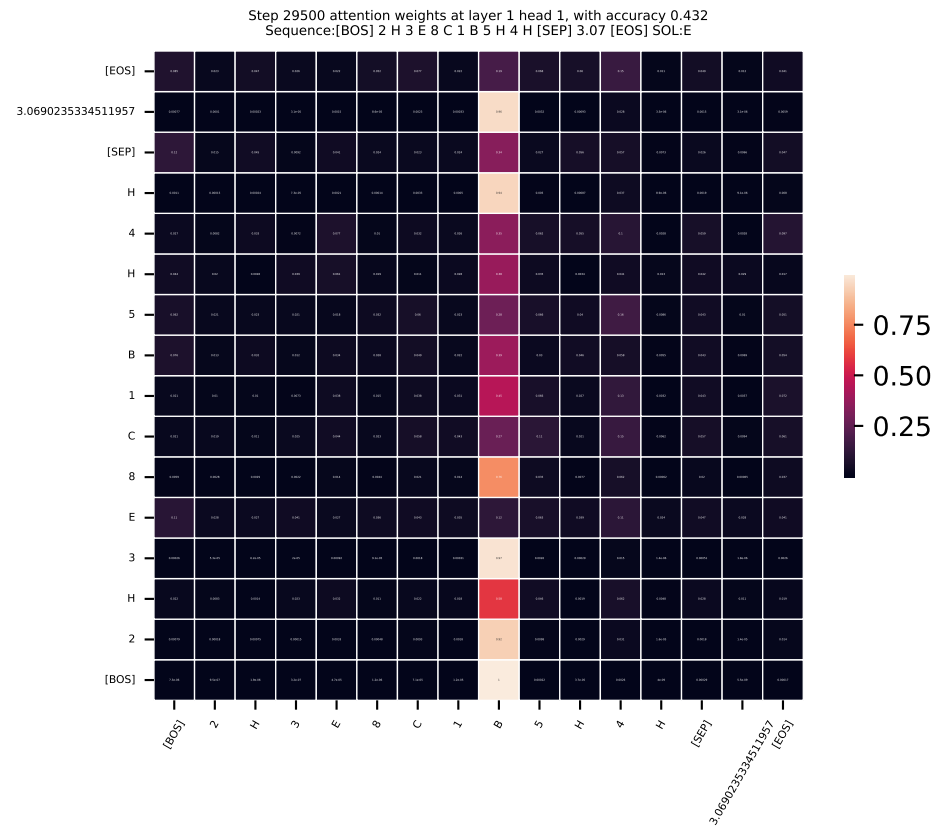

Step 29500 attention weights at layer 1 head 0, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.2
0.1

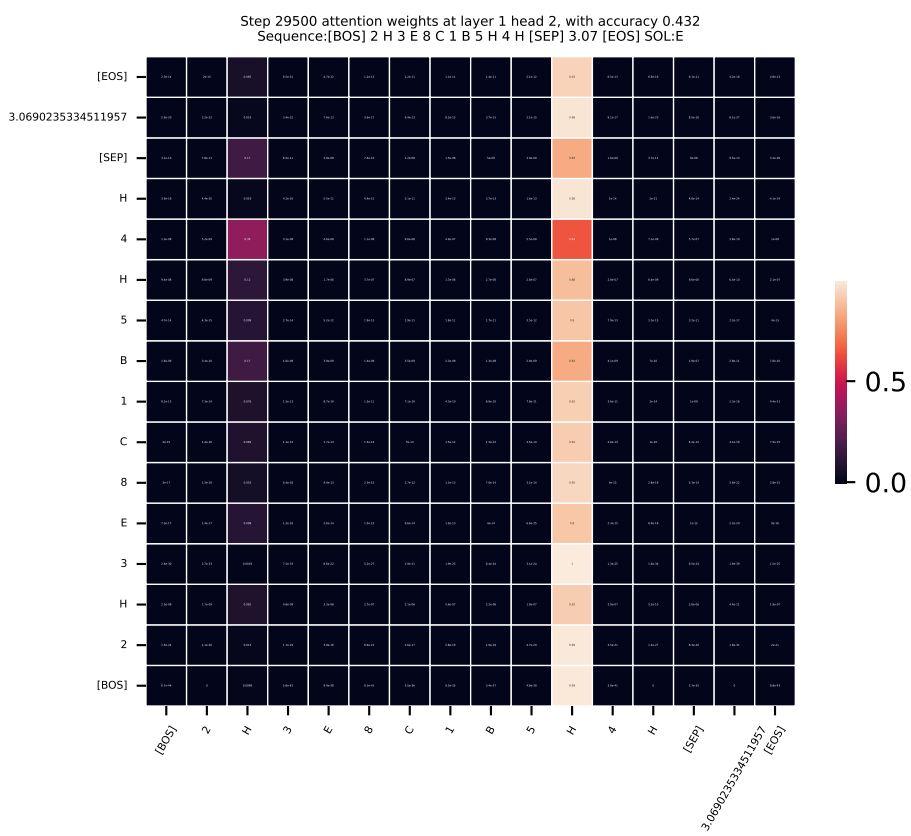

Step 29500 attention weights at layer 1 head 1, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

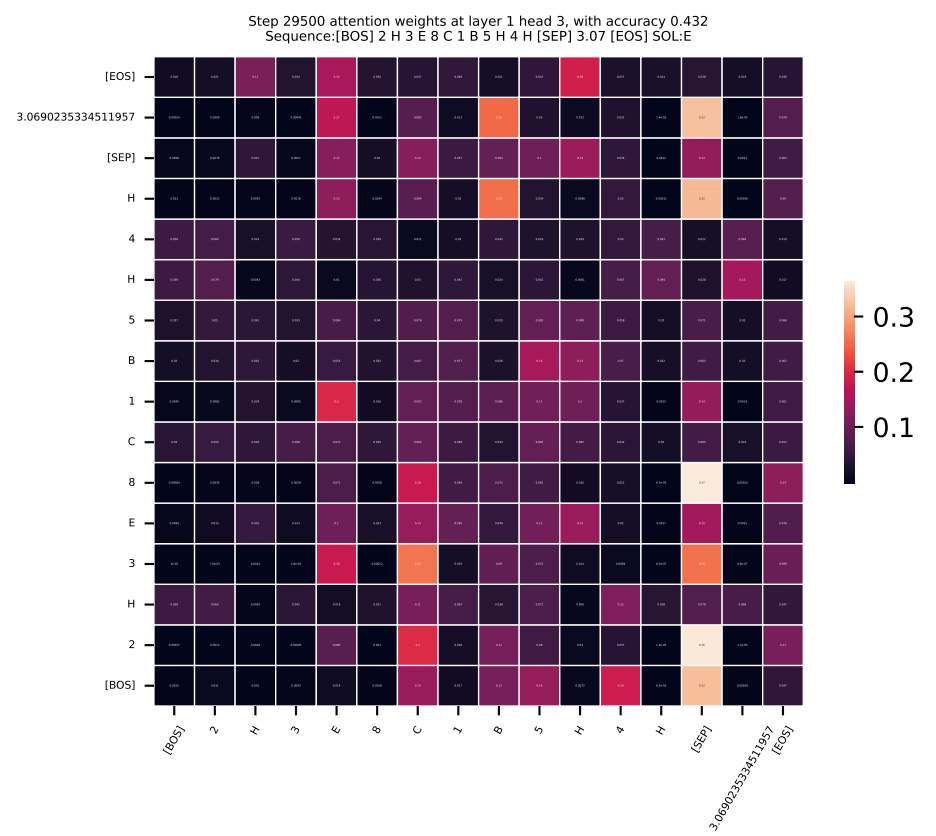

Step 29500 attention weights at layer 1 head 2, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.5
0.0

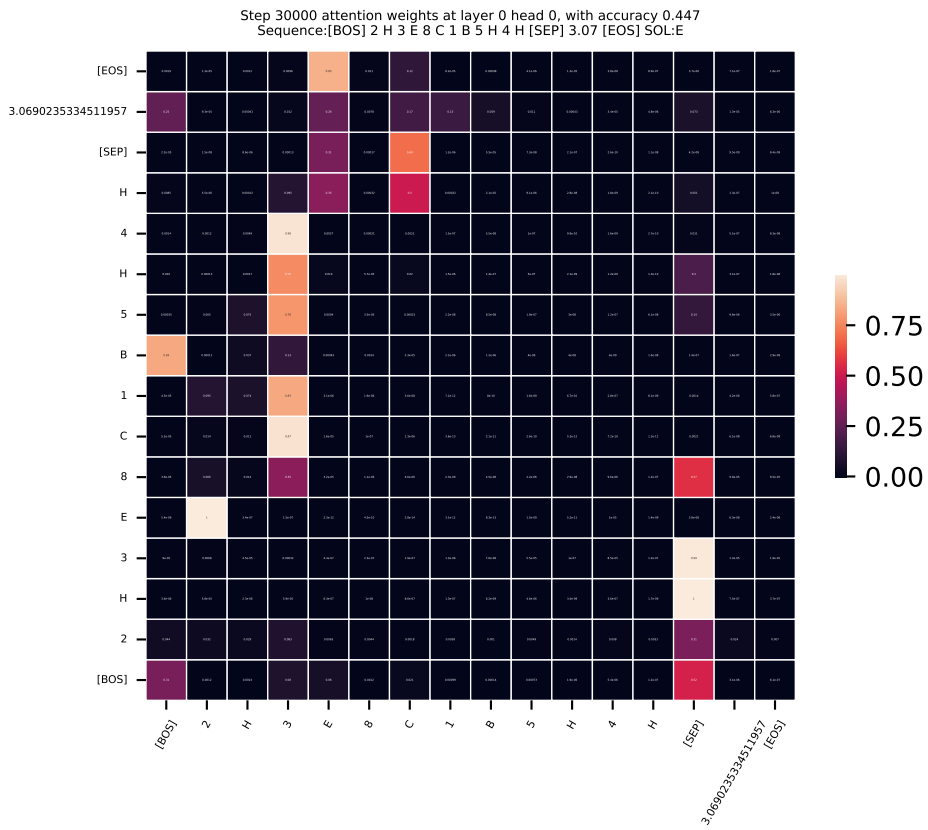

Step 29500 attention weights at layer 1 head 3, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.3
0.2
0.1

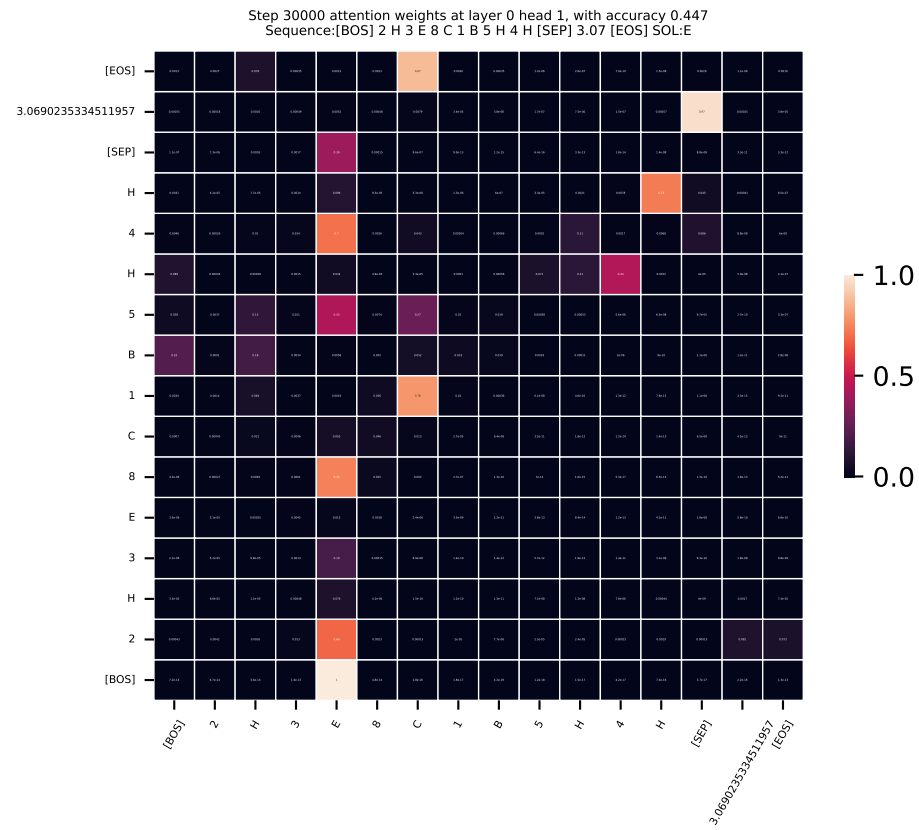

Step 29500 attention weights at layer 1 head 0, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.2
0.1

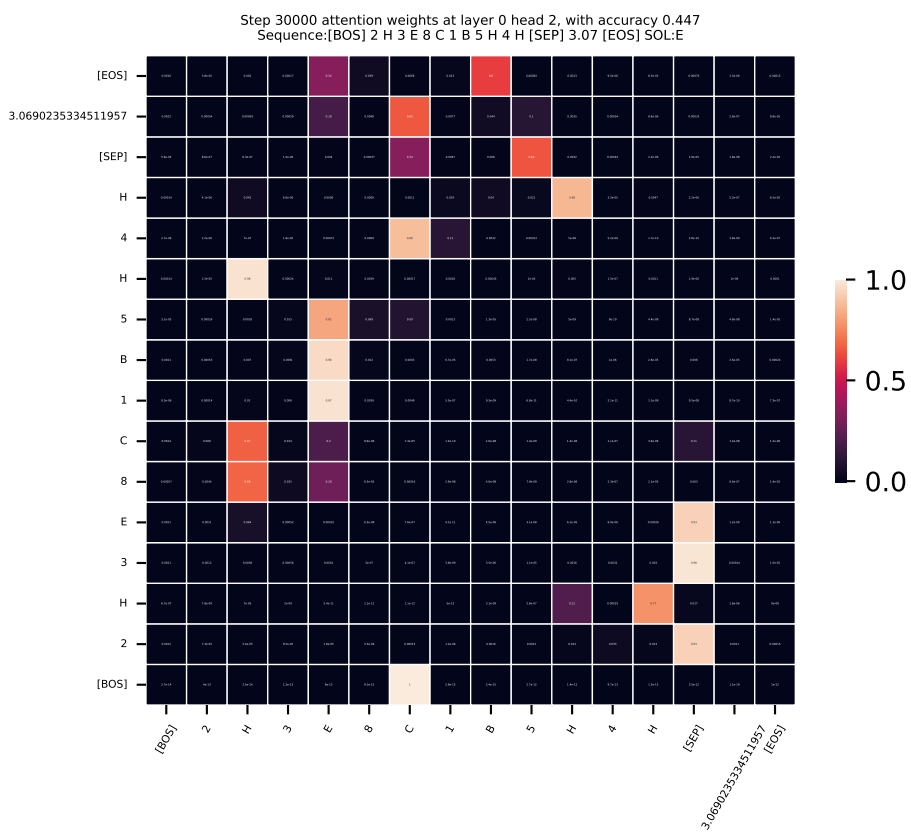

Step 29500 attention weights at layer 1 head 1, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

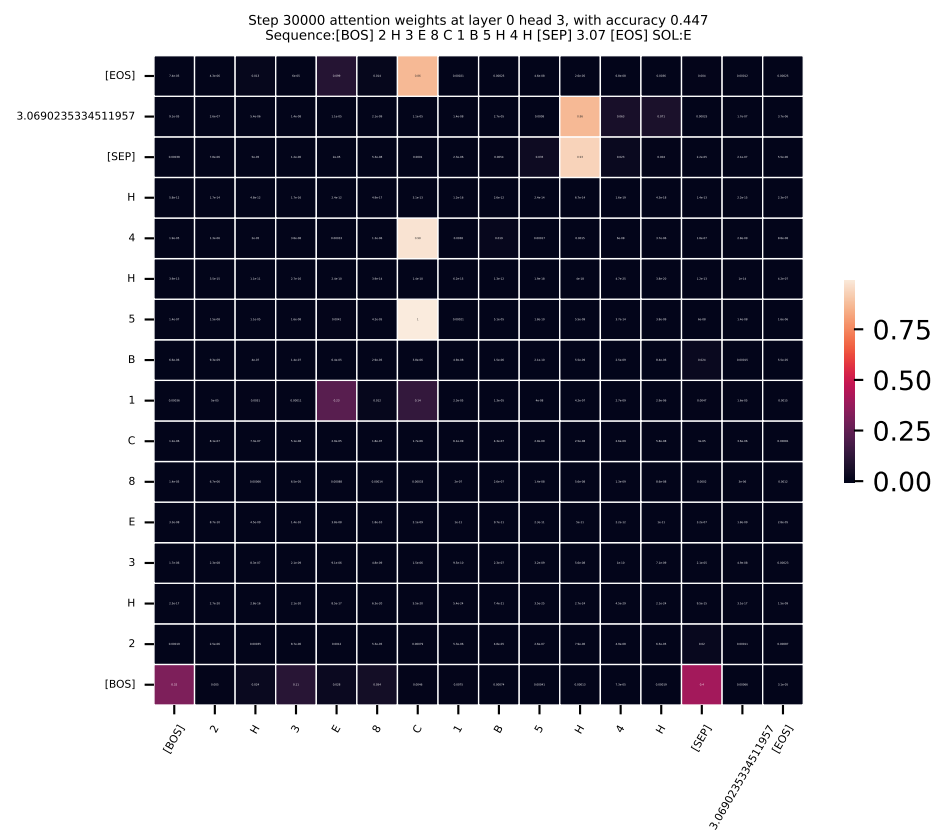

Step 29500 attention weights at layer 1 head 2, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.5
0.0

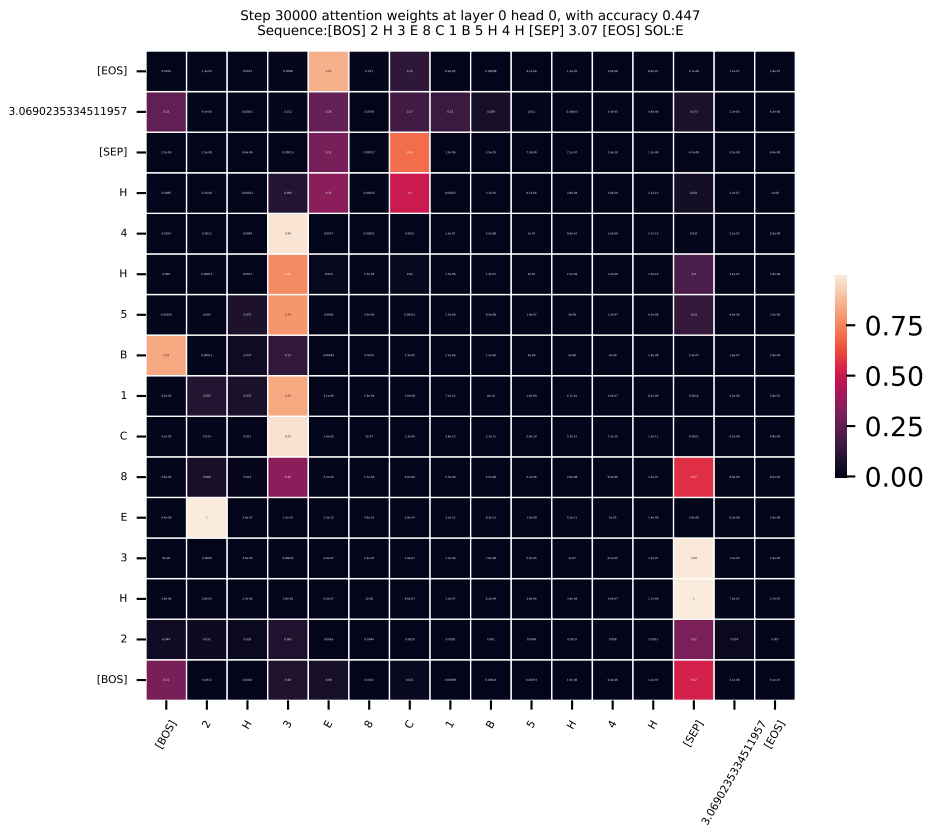

Step 29500 attention weights at layer 1 head 3, with accuracy 0.432
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.3
0.2
0.1

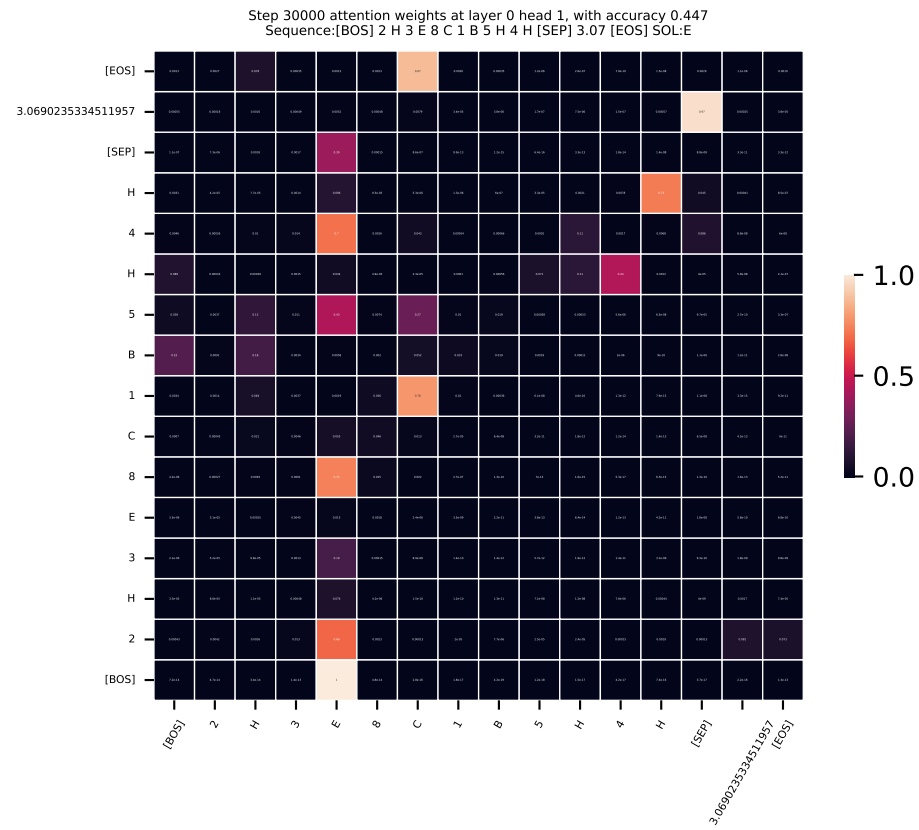

Step 30000 attention weights at layer 0 head 0, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

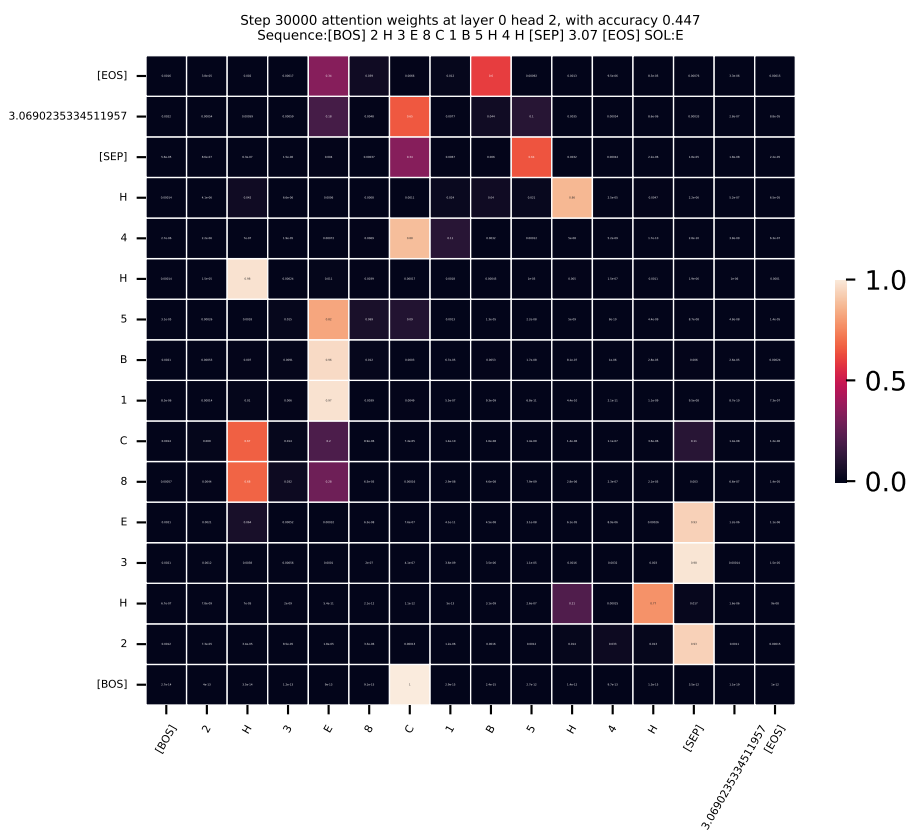

Step 30000 attention weights at layer 0 head 1, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

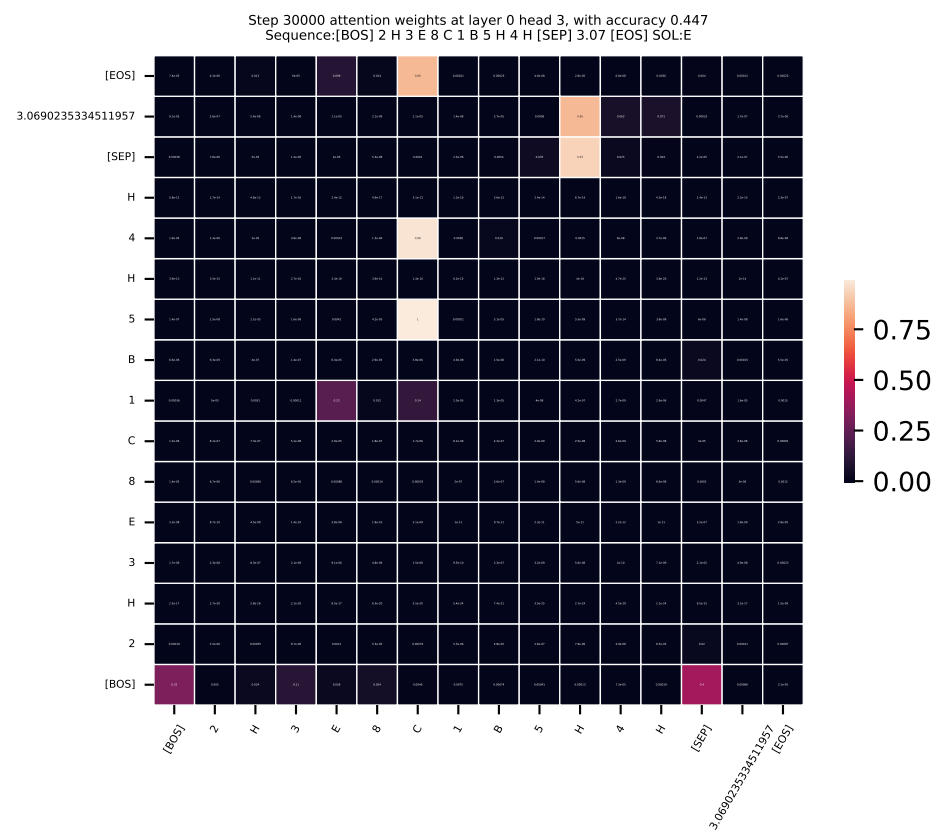

Step 30000 attention weights at layer 0 head 2, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

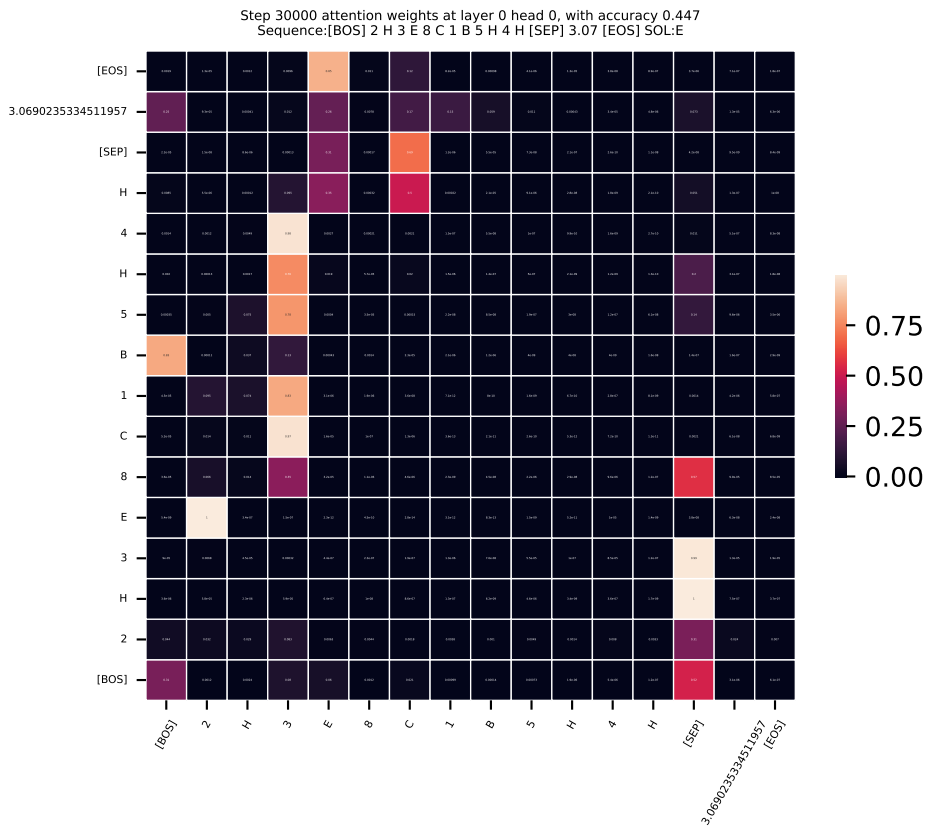

Step 30000 attention weights at layer 0 head 3, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

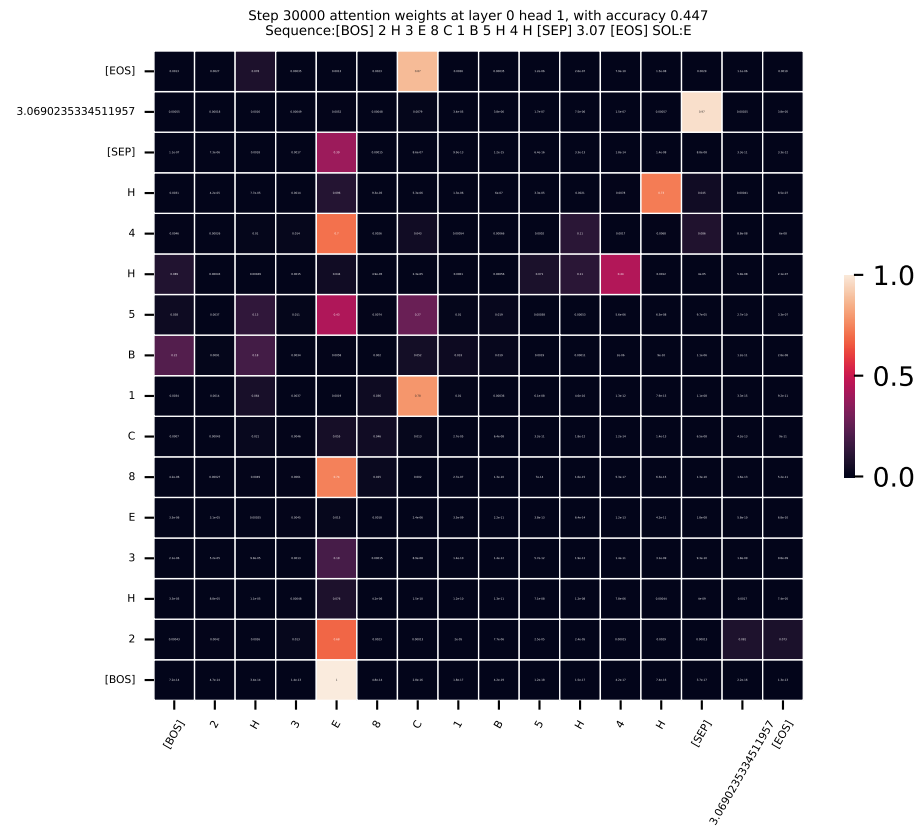

Step 30000 attention weights at layer 0 head 0, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

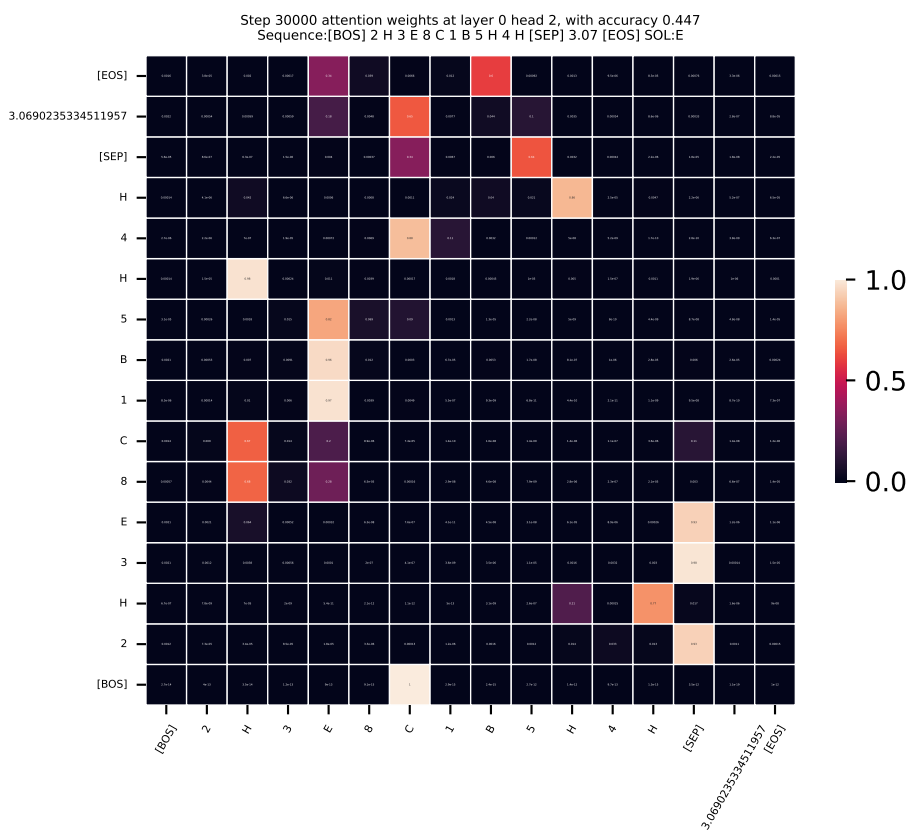

Step 30000 attention weights at layer 0 head 1, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

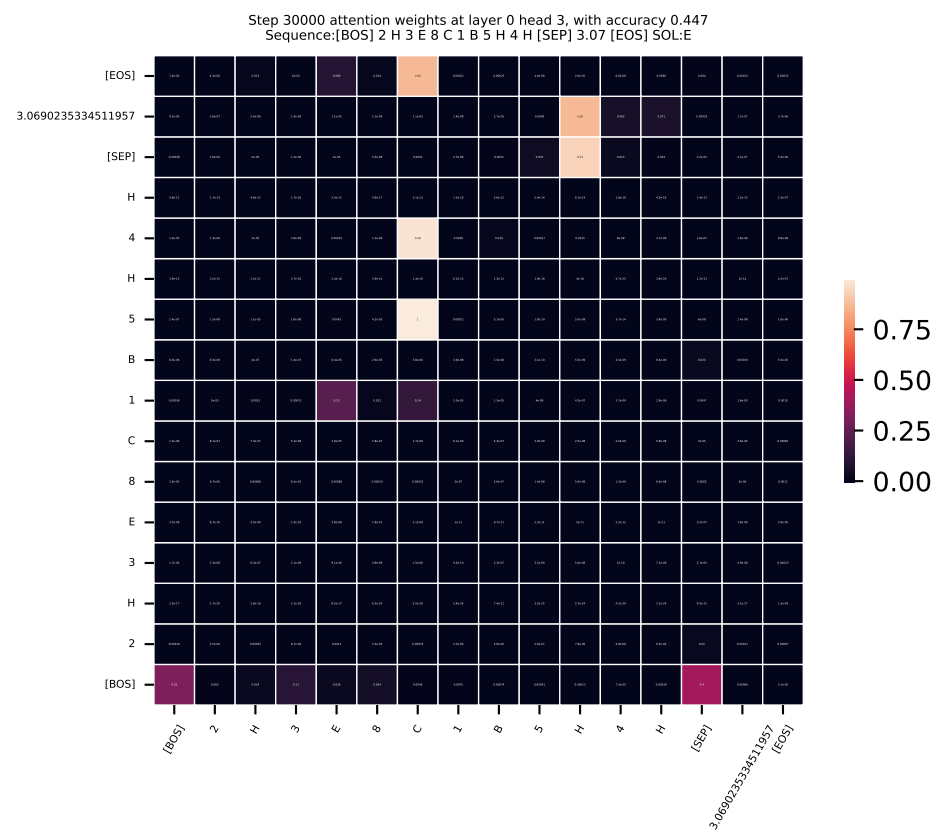

Step 30000 attention weights at layer 0 head 2, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

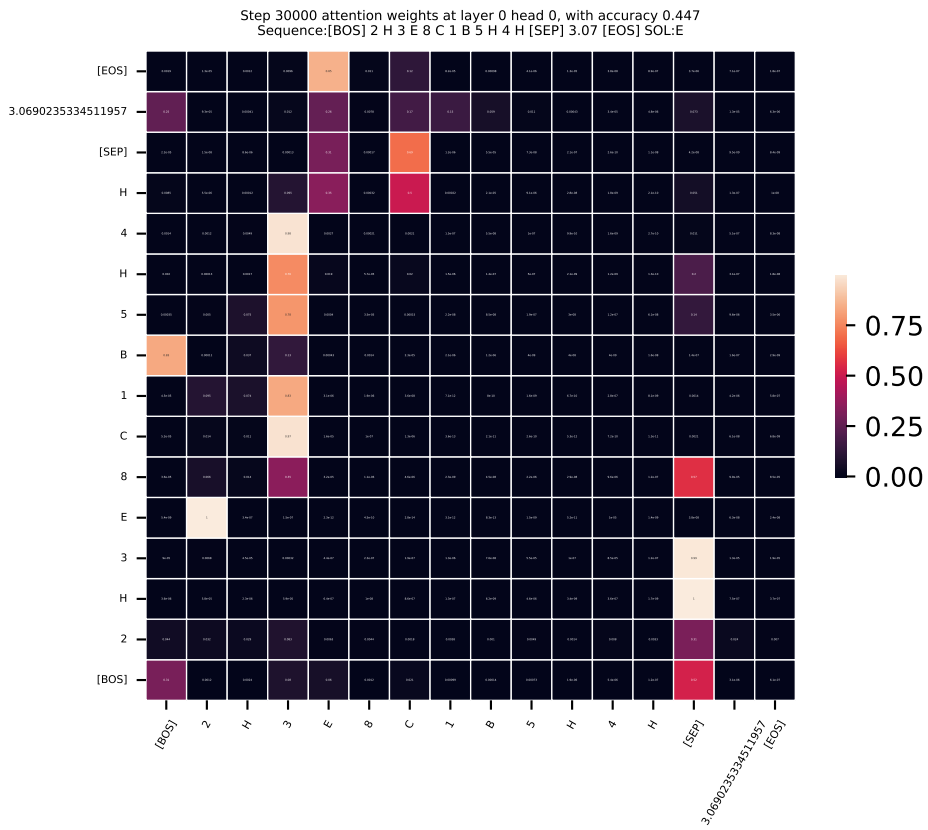

Step 30000 attention weights at layer 0 head 3, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

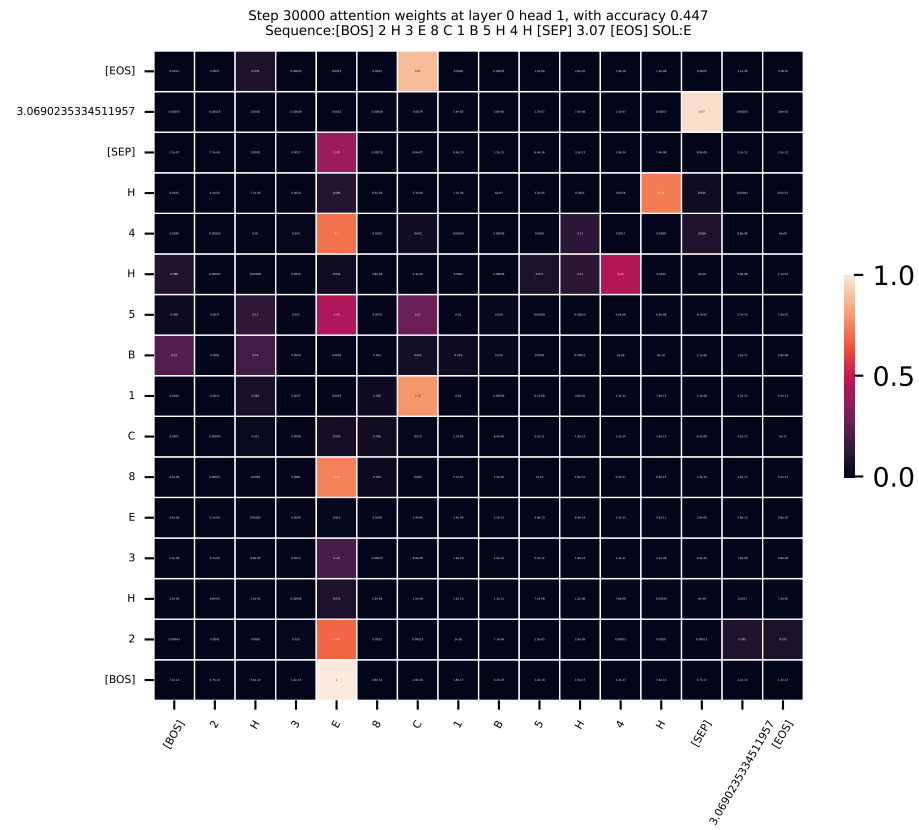

Step 30000 attention weights at layer 0 head 0, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

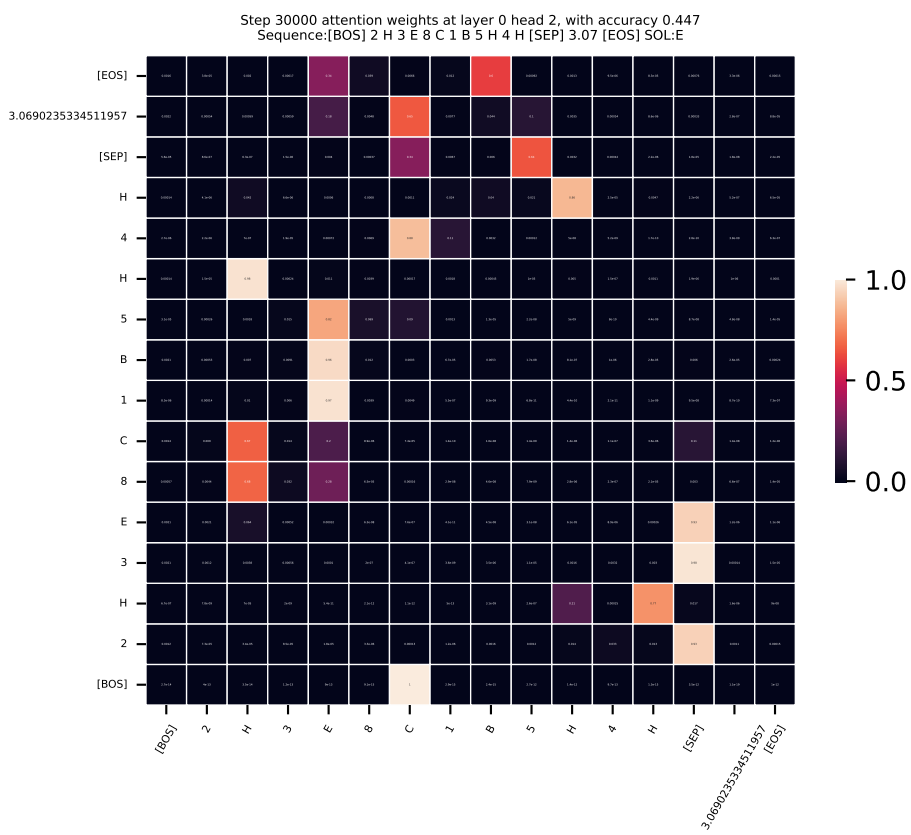

Step 30000 attention weights at layer 0 head 1, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

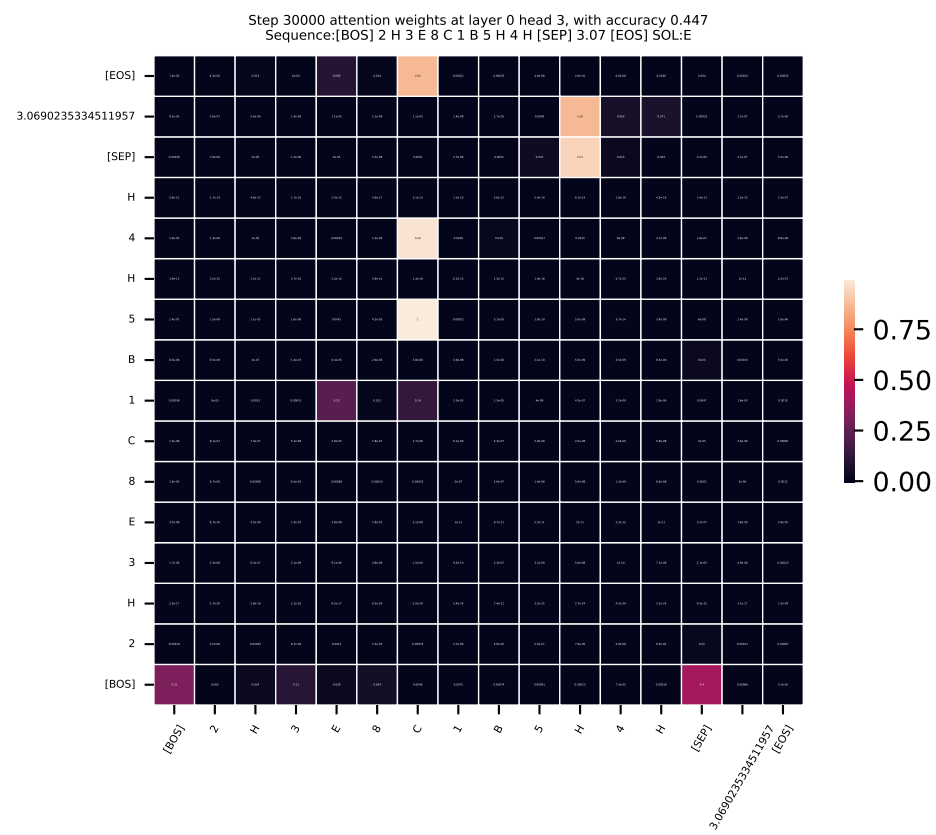

Step 30000 attention weights at layer 0 head 2, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

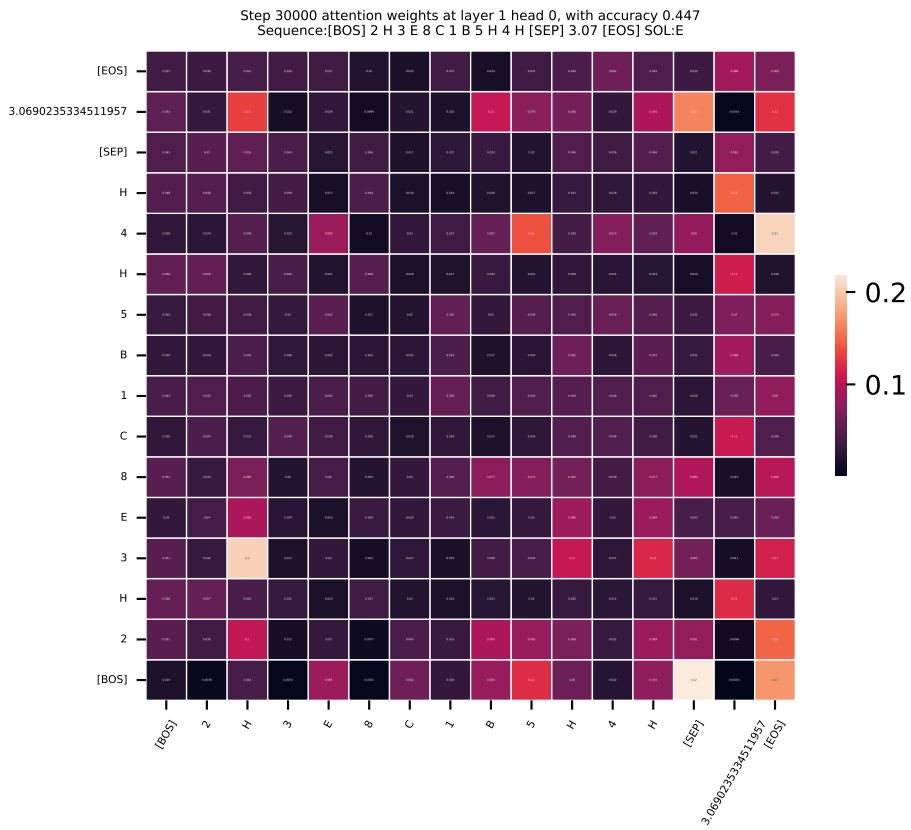

Step 30000 attention weights at layer 0 head 3, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

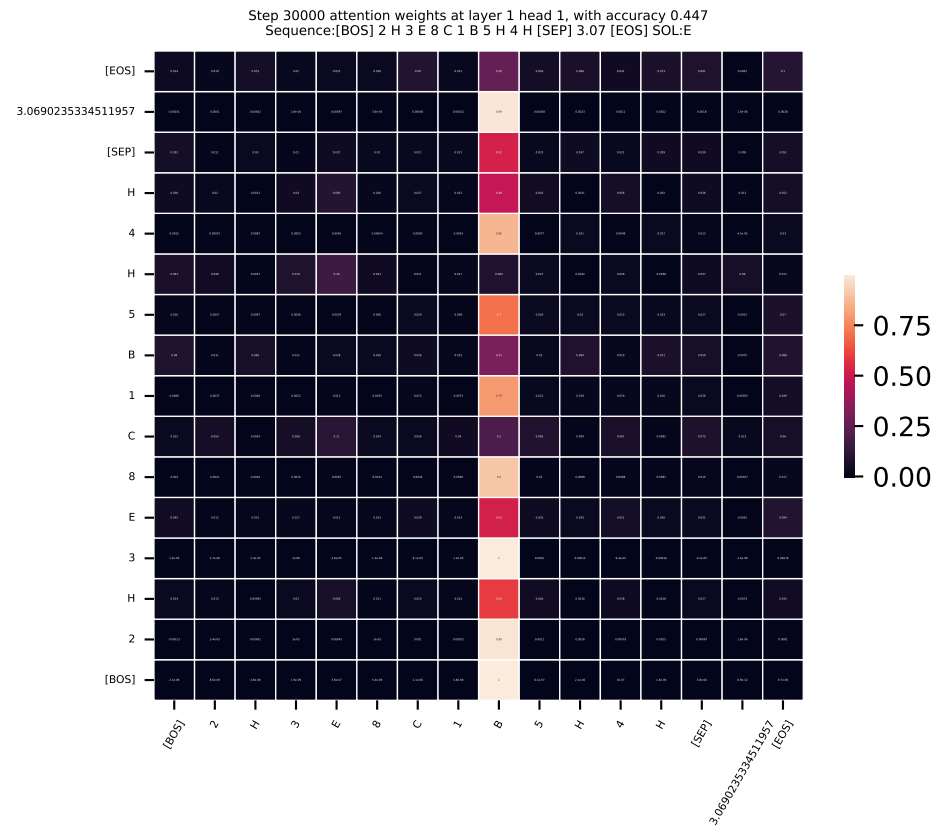

Step 30000 attention weights at layer 0 head 0, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

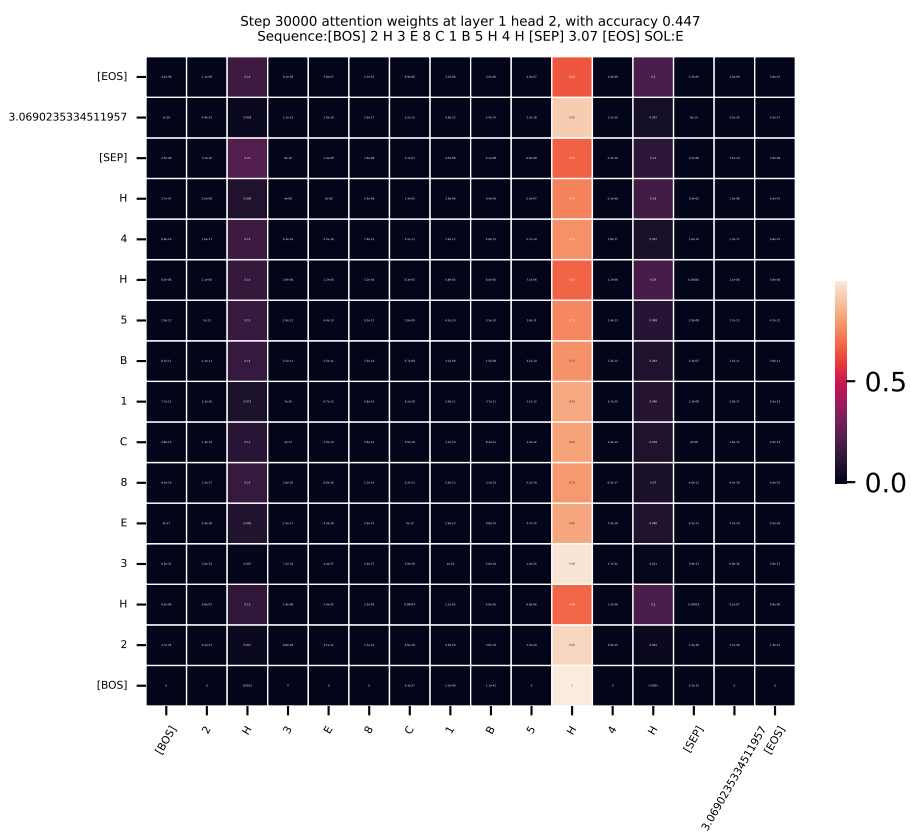

Step 30000 attention weights at layer 0 head 1, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

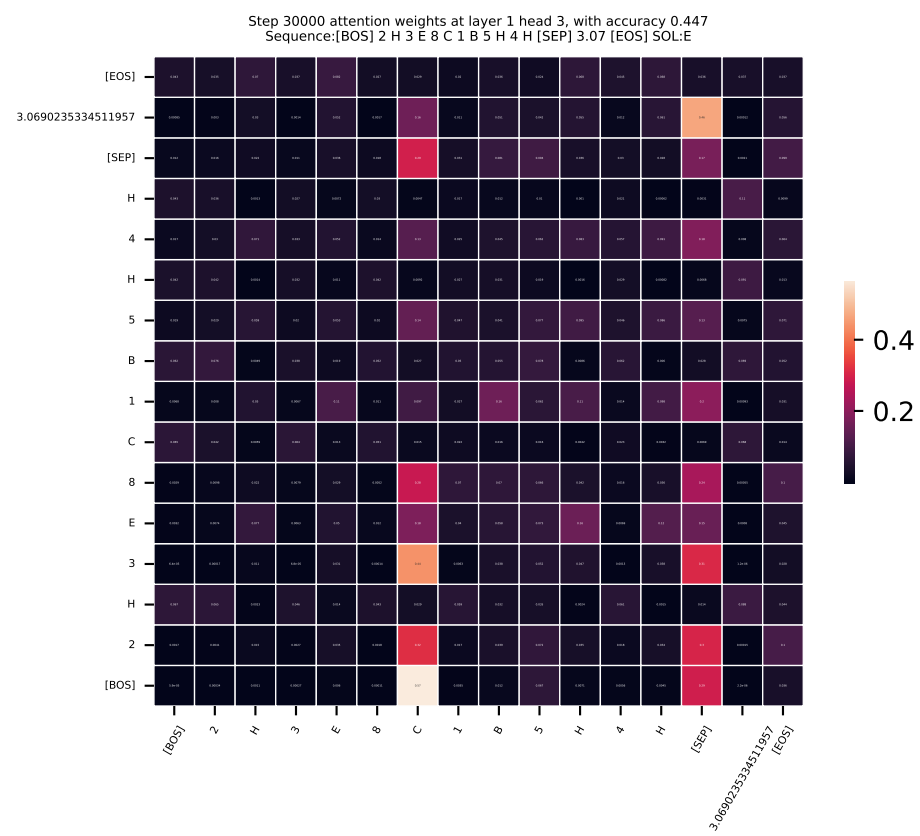

Step 30000 attention weights at layer 0 head 2, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

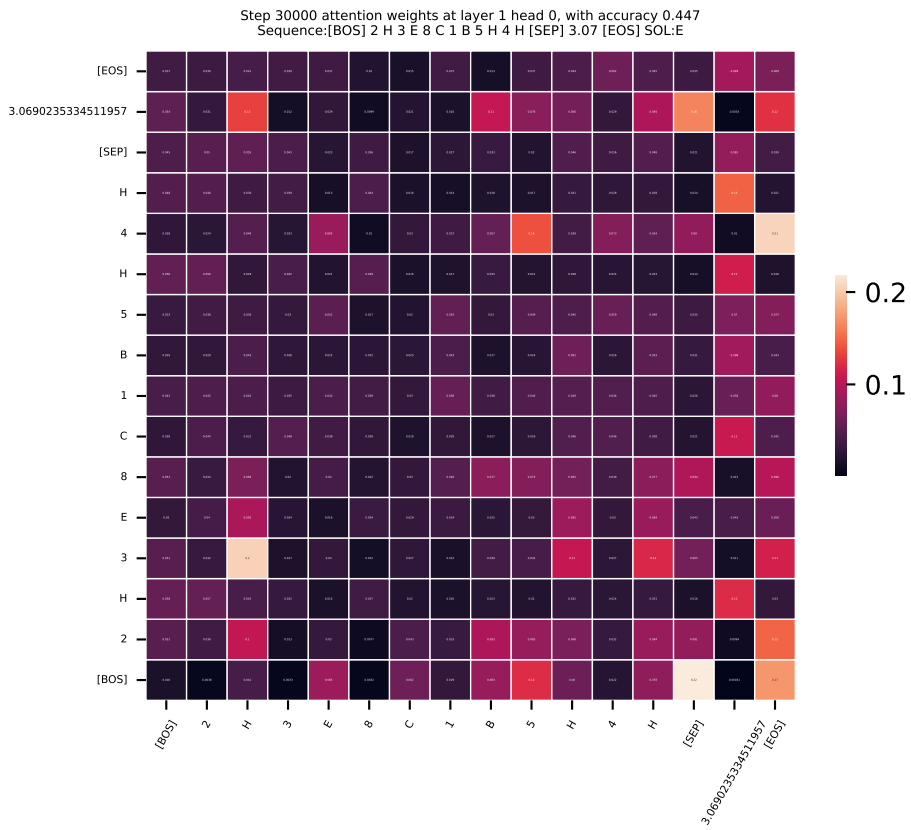

Step 30000 attention weights at layer 0 head 3, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

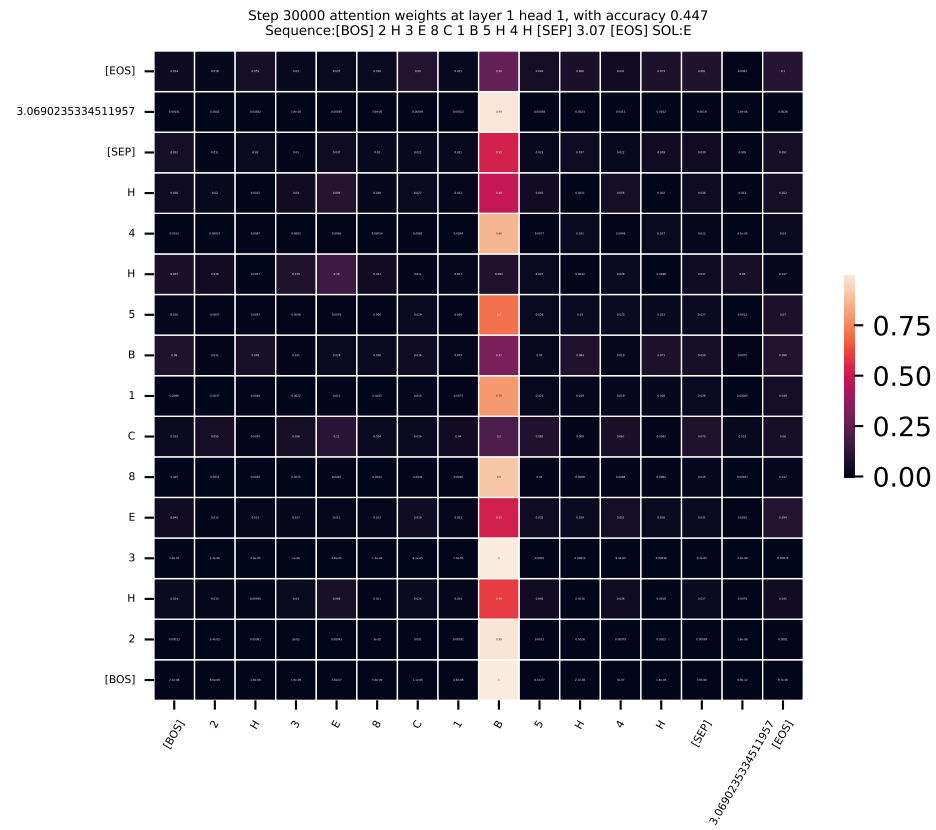

Step 30000 attention weights at layer 1 head 0, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.2
0.1

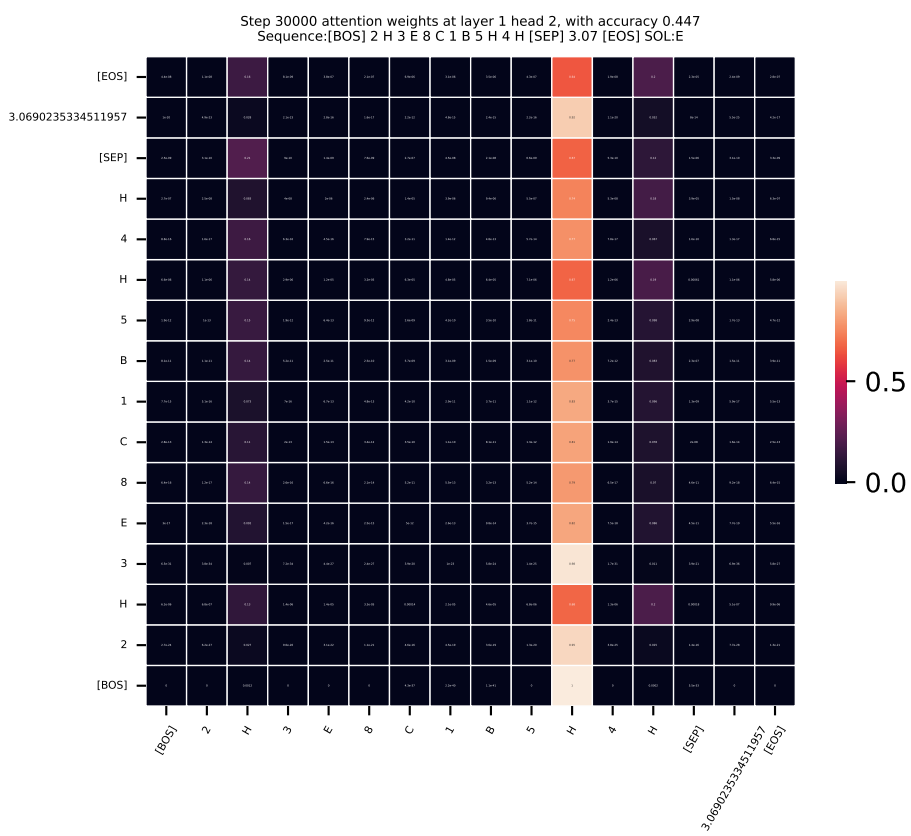

Step 30000 attention weights at layer 1 head 1, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

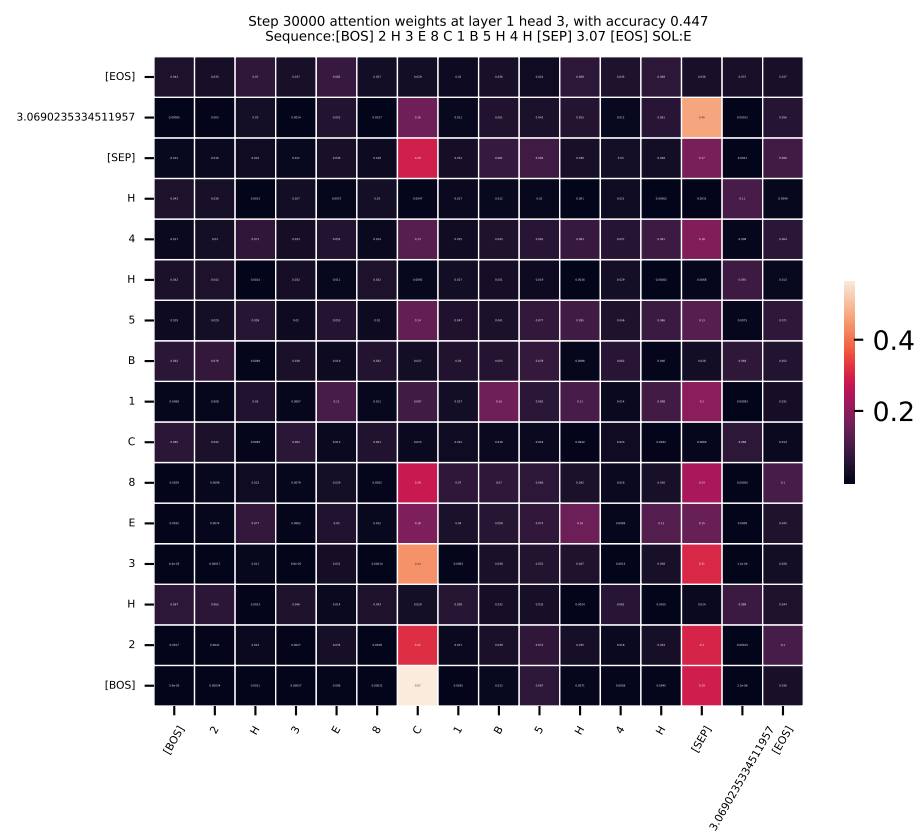

Step 30000 attention weights at layer 1 head 2, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.5
0.0

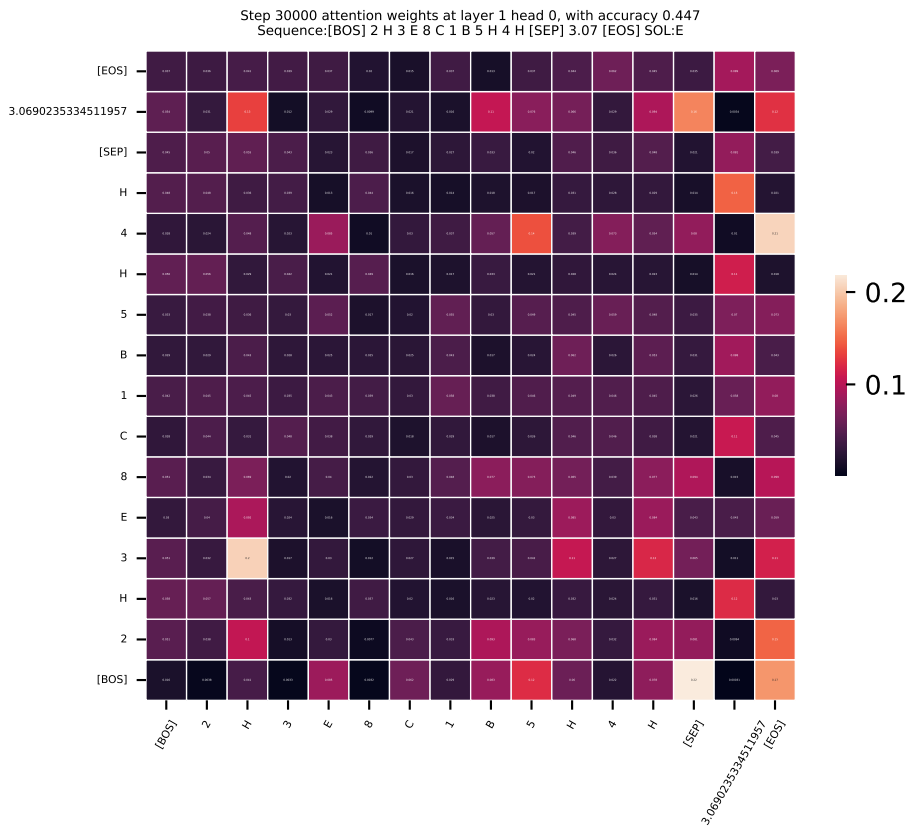

Step 30000 attention weights at layer 1 head 3, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.4
0.2

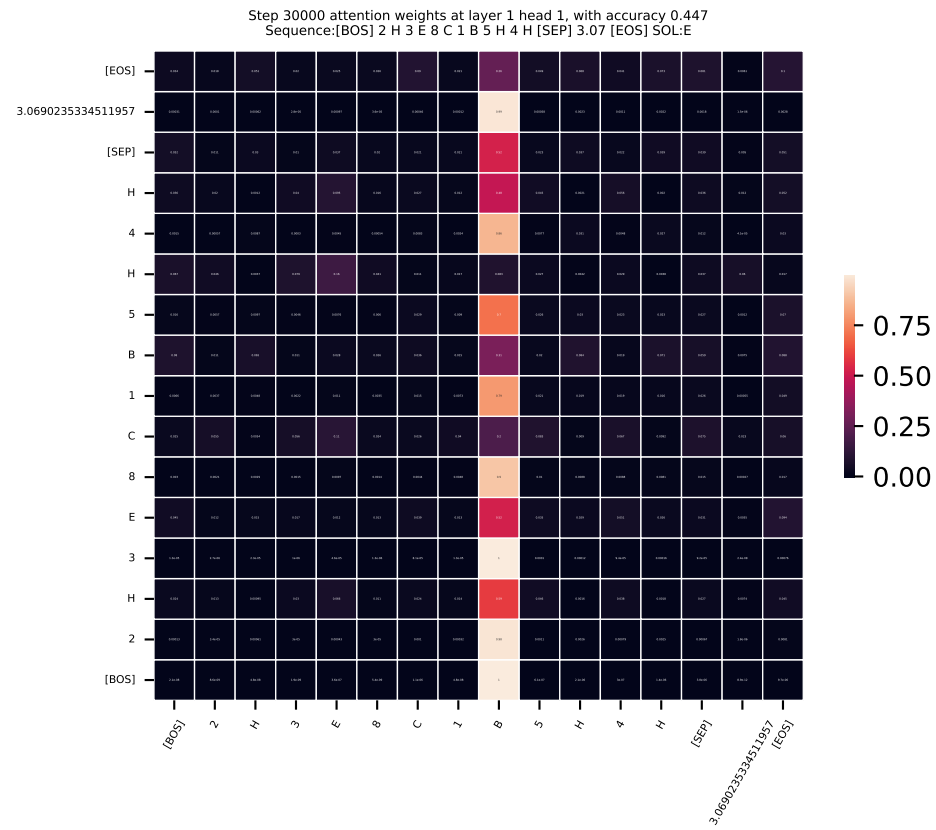

Step 30000 attention weights at layer 1 head 0, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.2
0.1

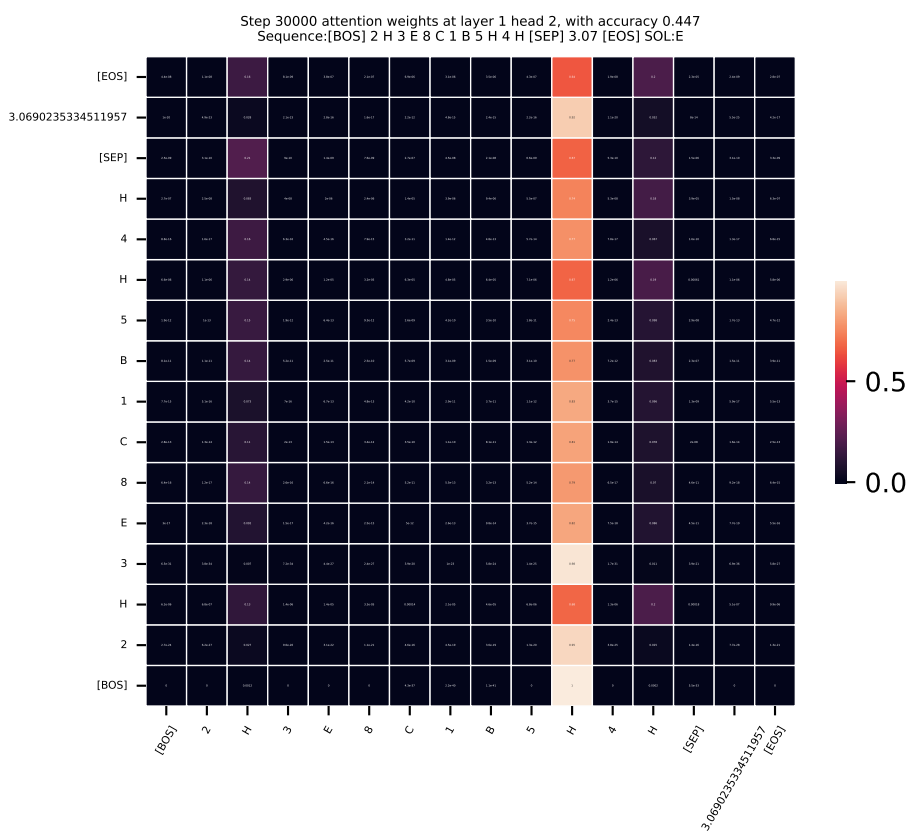

Step 30000 attention weights at layer 1 head 1, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

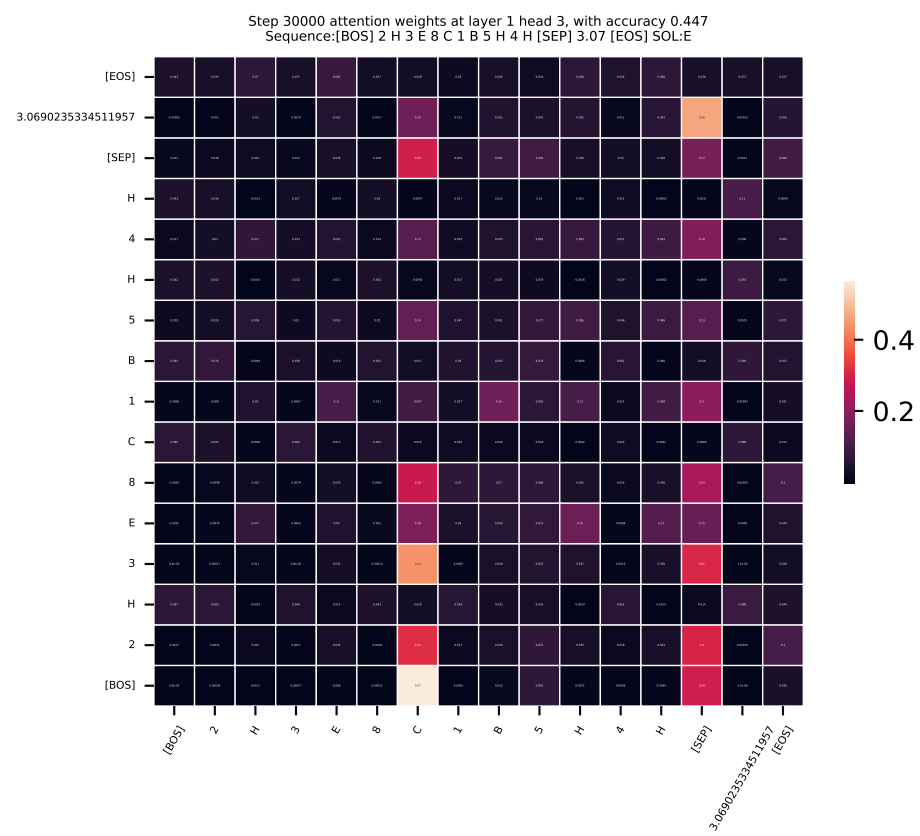

Step 30000 attention weights at layer 1 head 2, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.5
0.0

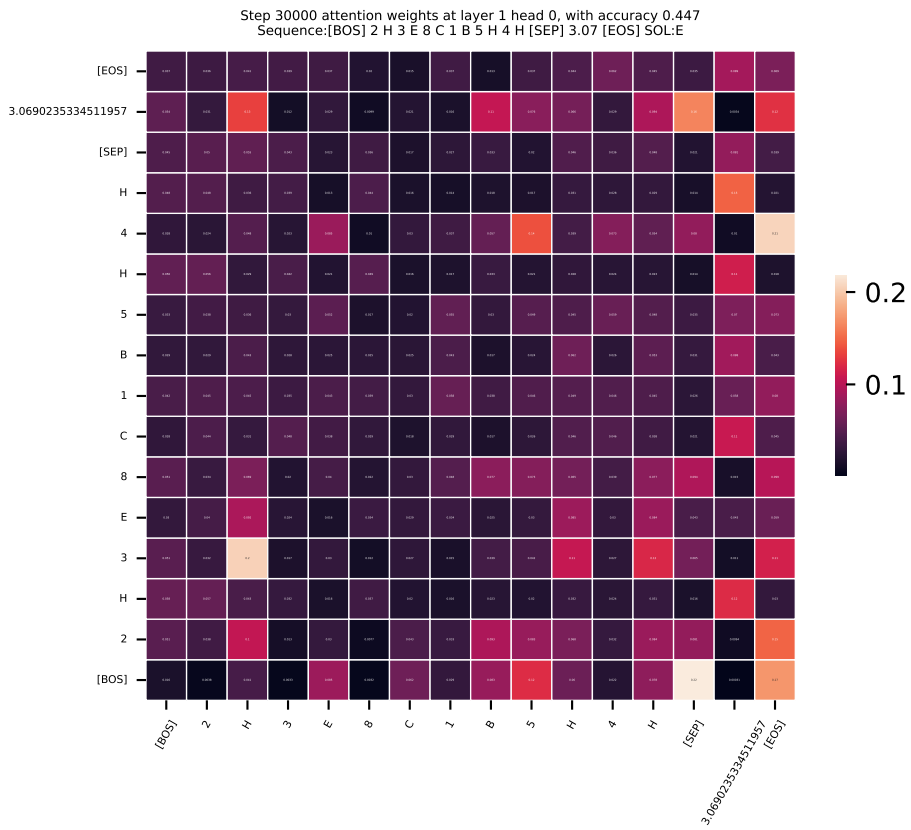

Step 30000 attention weights at layer 1 head 3, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.4
0.2

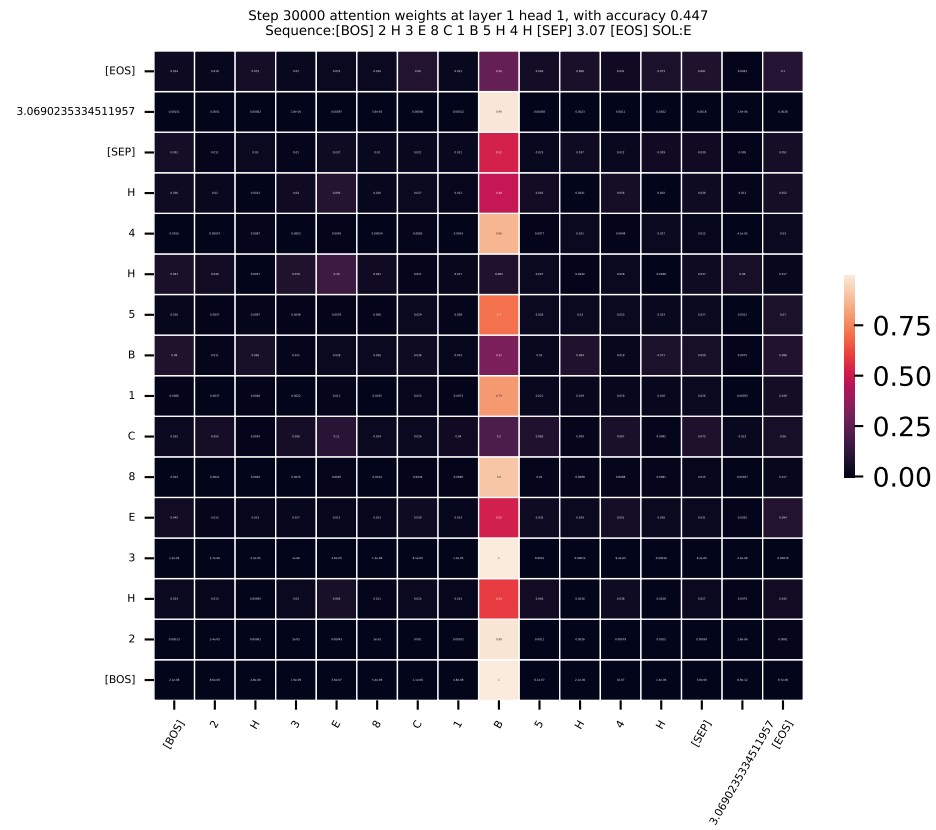

Step 30000 attention weights at layer 1 head 0, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.2
0.1

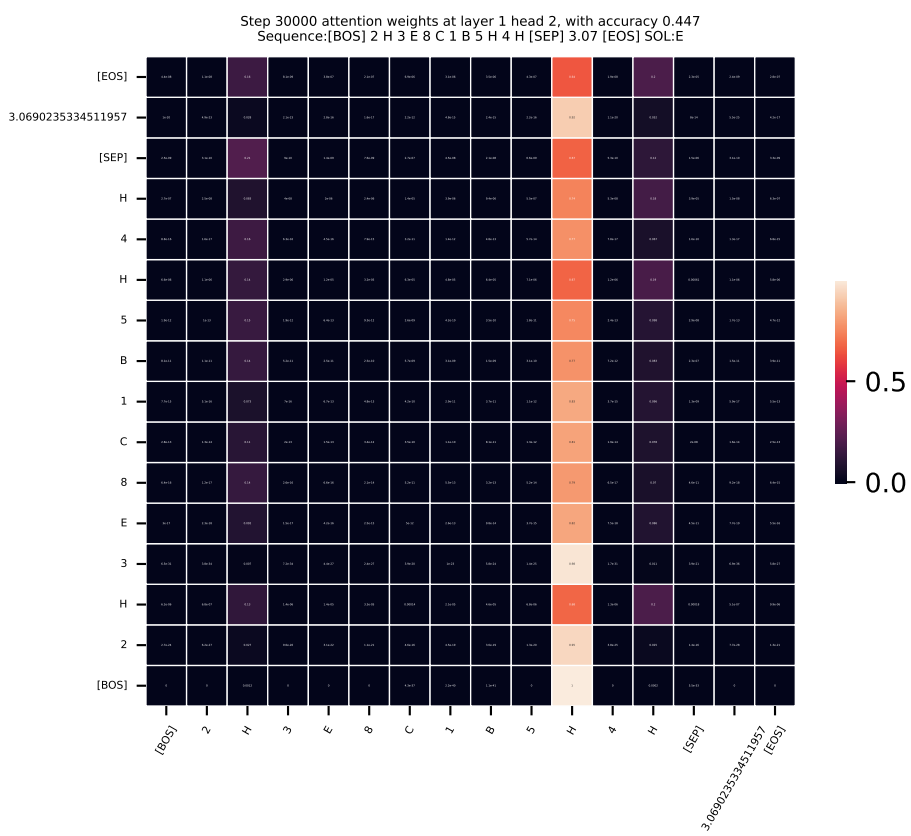

Step 30000 attention weights at layer 1 head 1, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

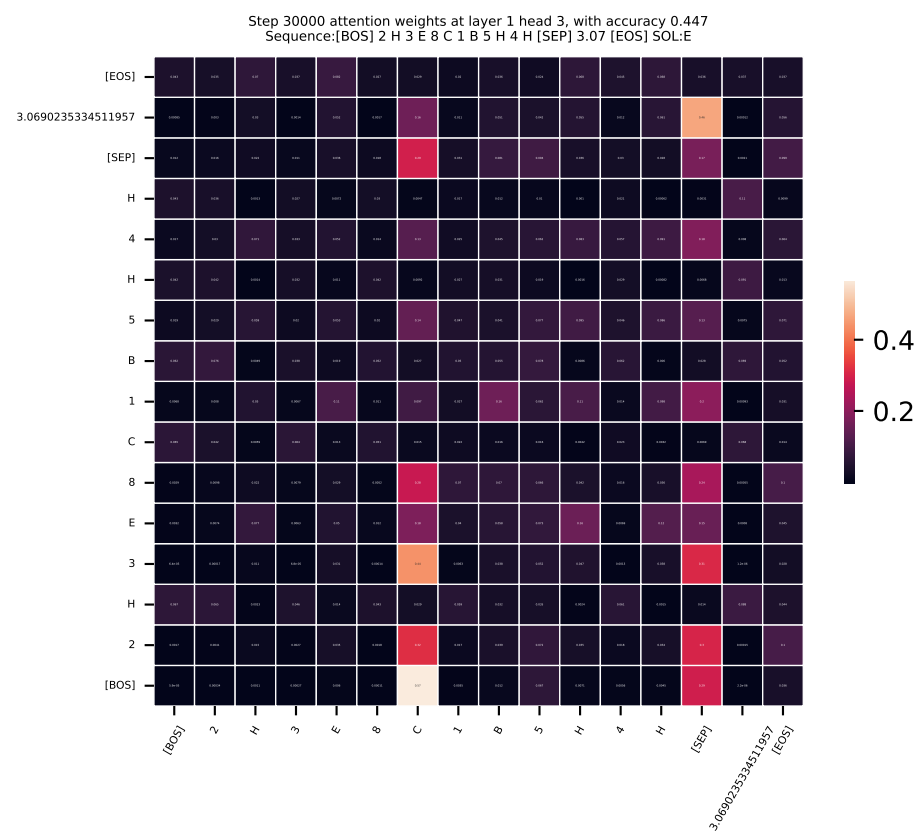

Step 30000 attention weights at layer 1 head 2, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.5
0.0

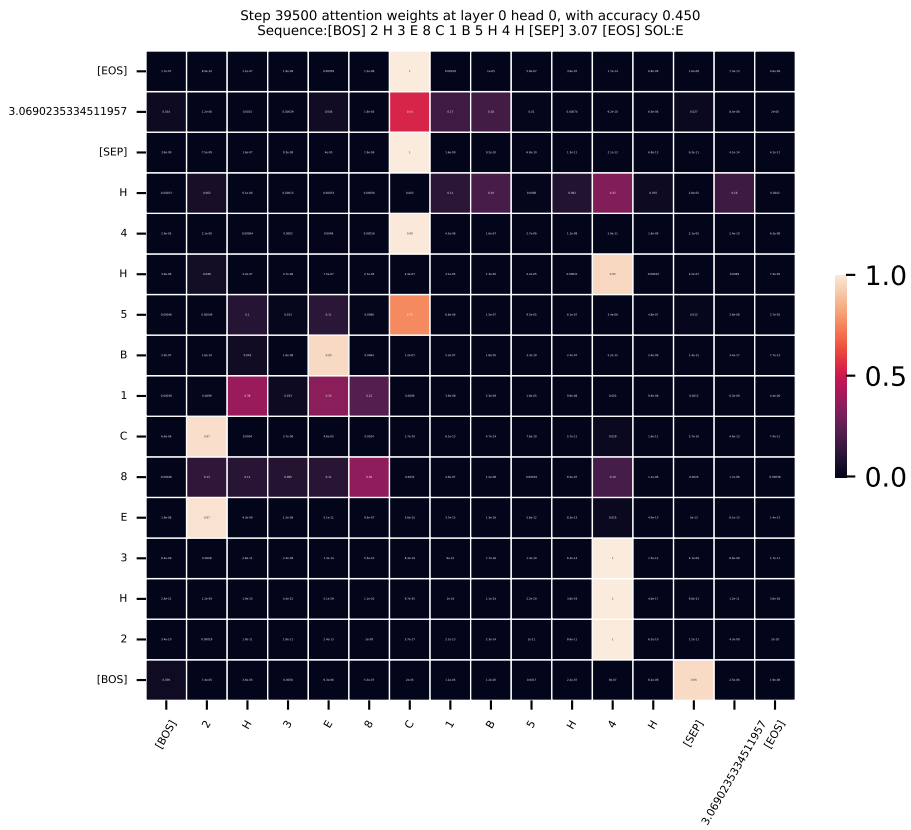

Step 30000 attention weights at layer 1 head 3, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.4
0.2

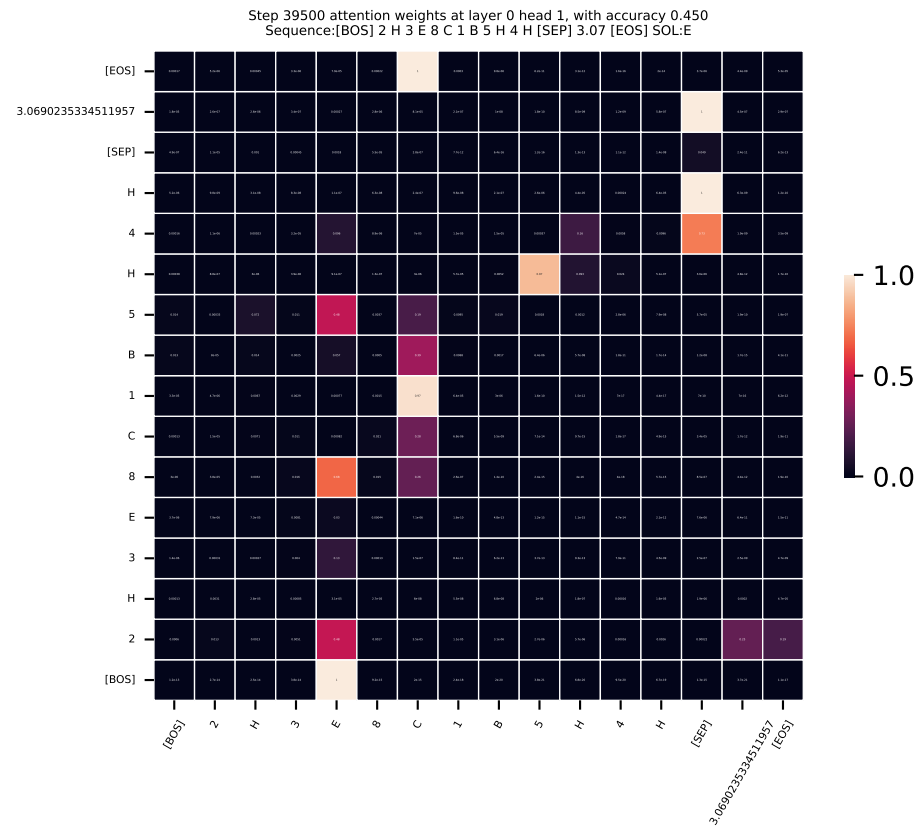

Step 30000 attention weights at layer 1 head 0, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.2
0.1

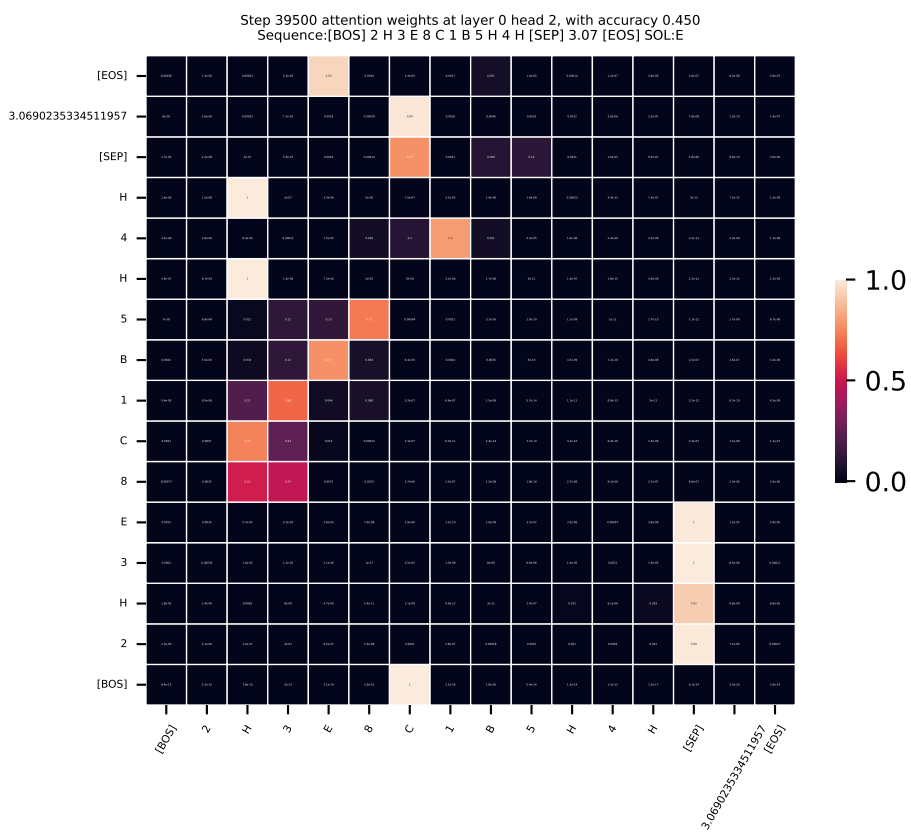

Step 30000 attention weights at layer 1 head 1, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

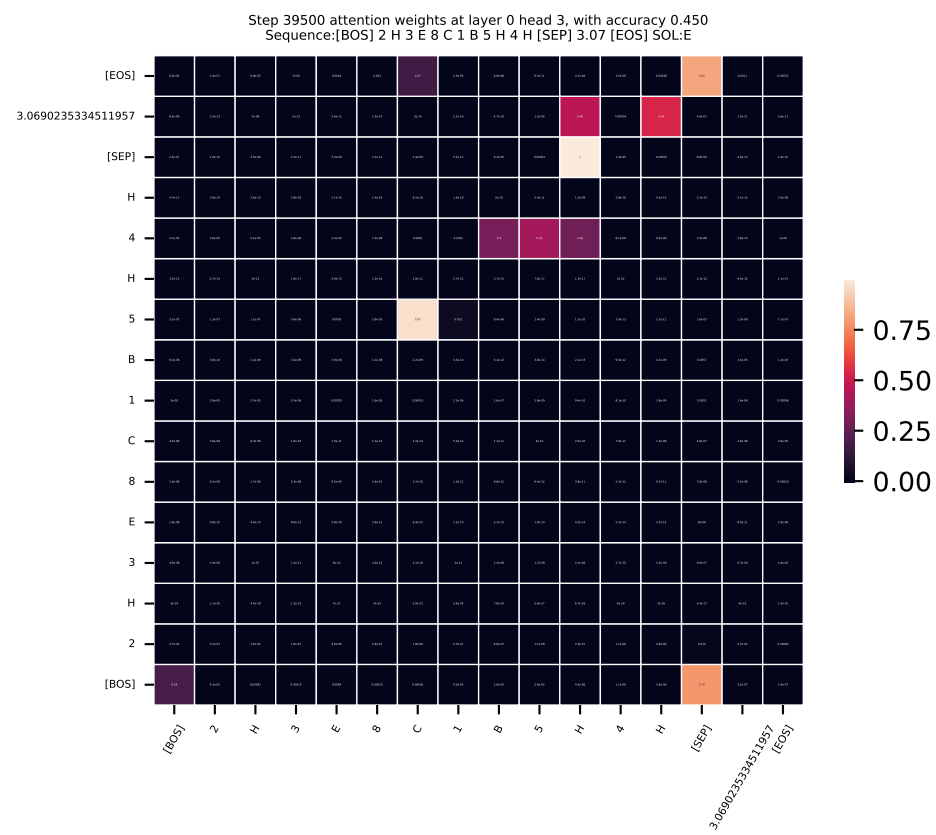

Step 30000 attention weights at layer 1 head 2, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.5
0.0

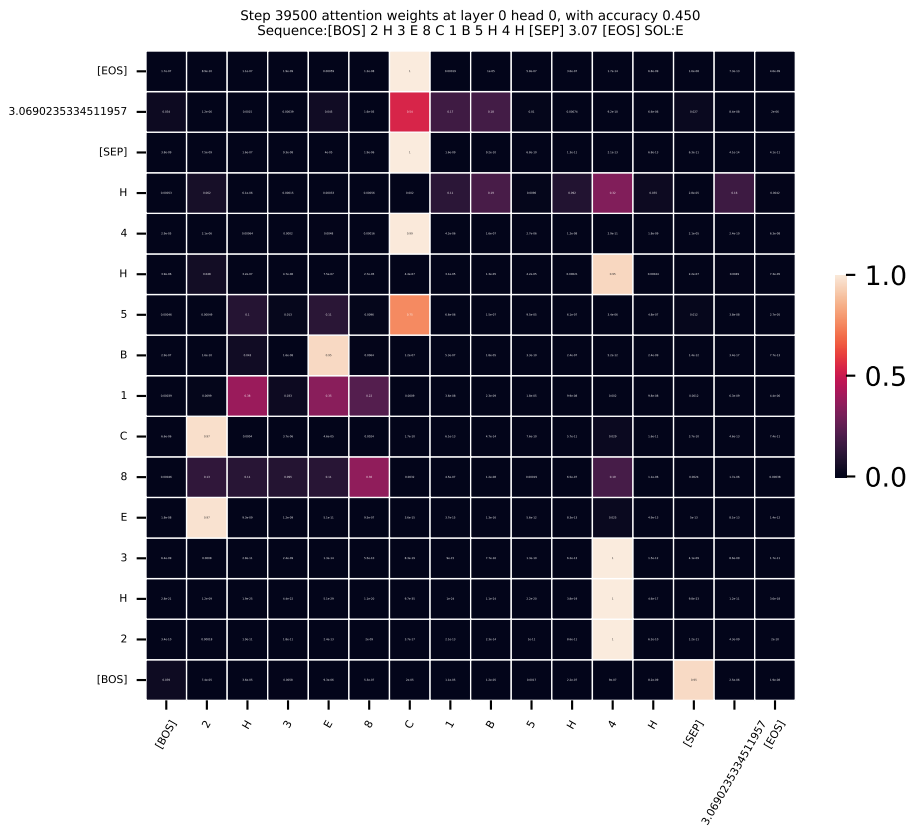

Step 30000 attention weights at layer 1 head 3, with accuracy 0.447
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.4
0.2

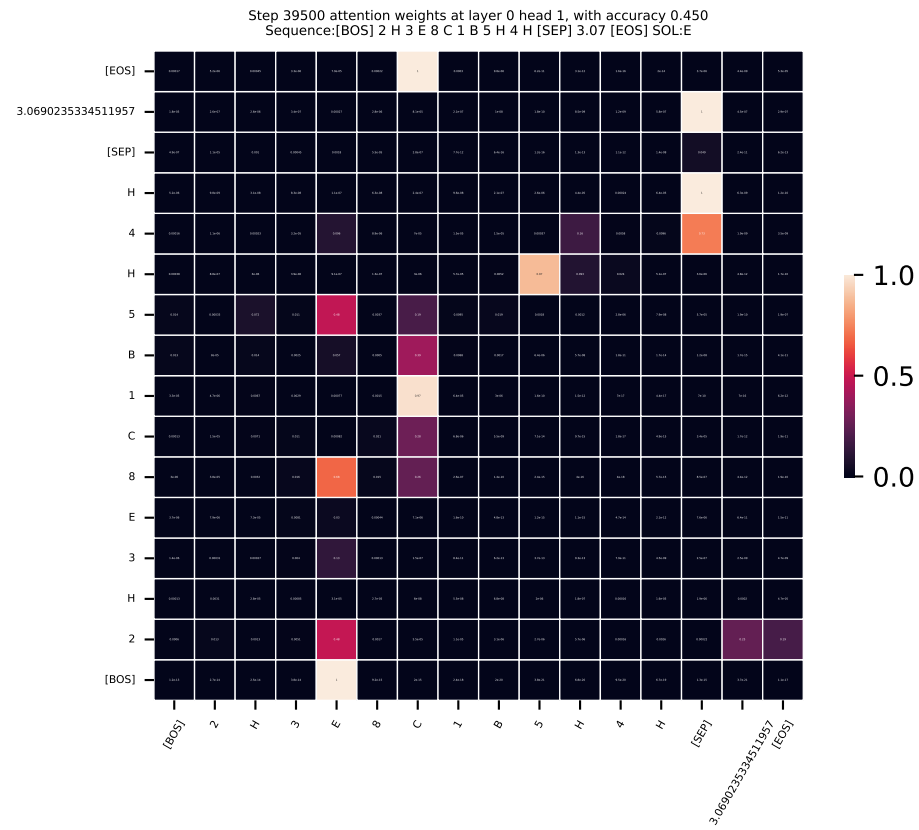

Step 39500 attention weights at layer 0 head 0, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

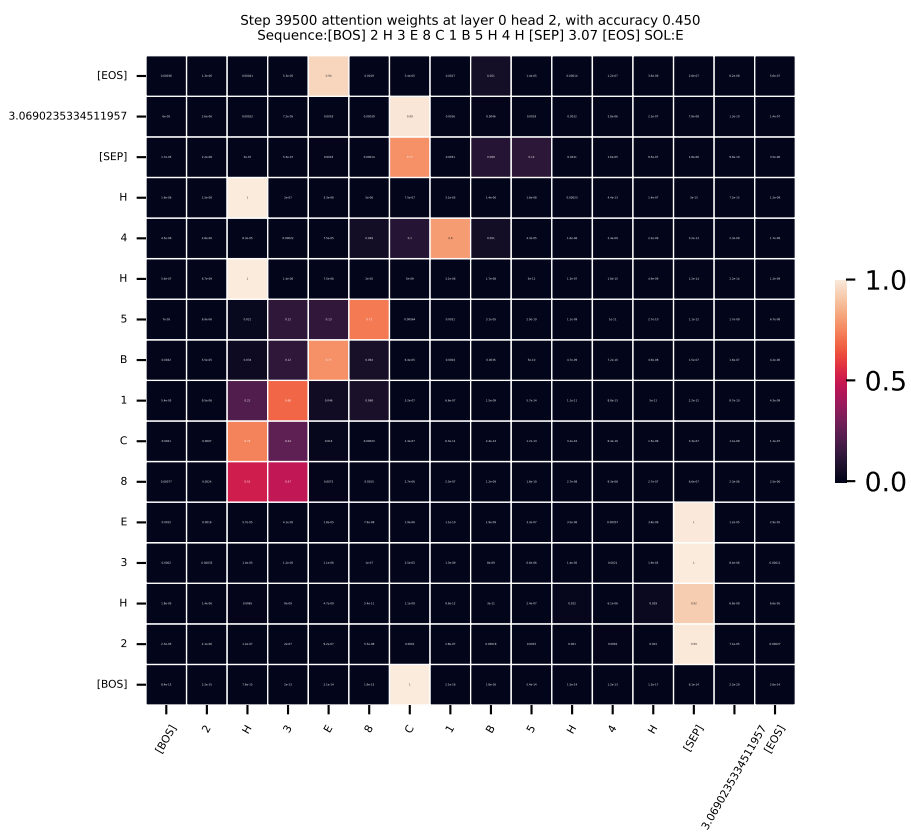

Step 39500 attention weights at layer 0 head 1, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

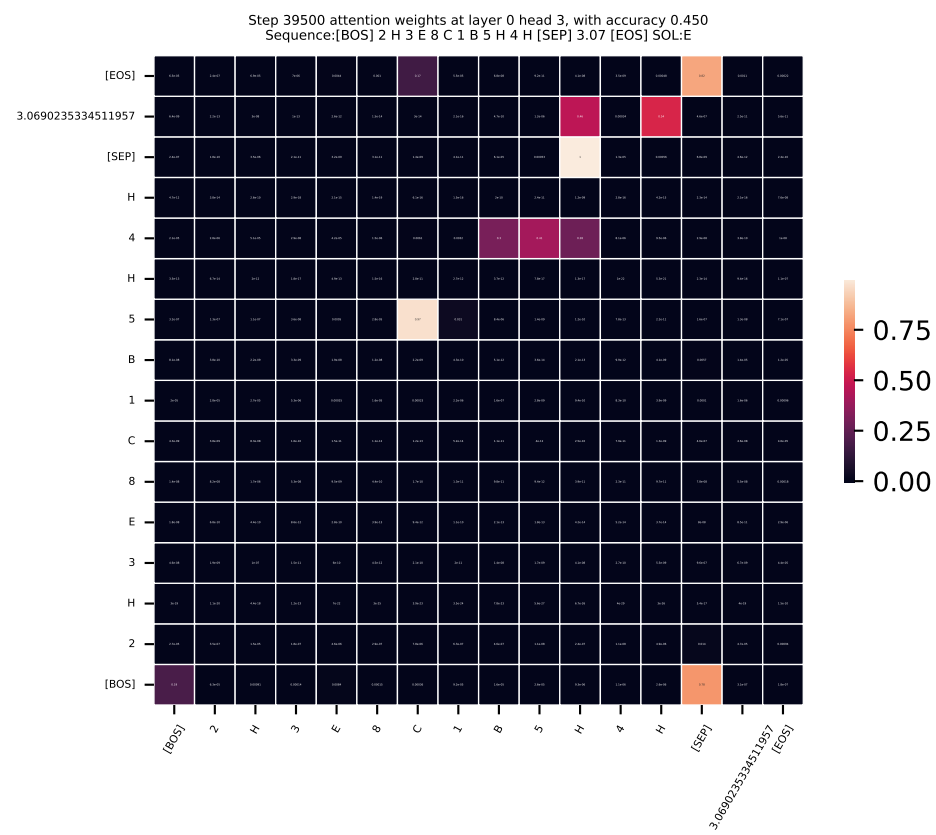

Step 39500 attention weights at layer 0 head 2, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

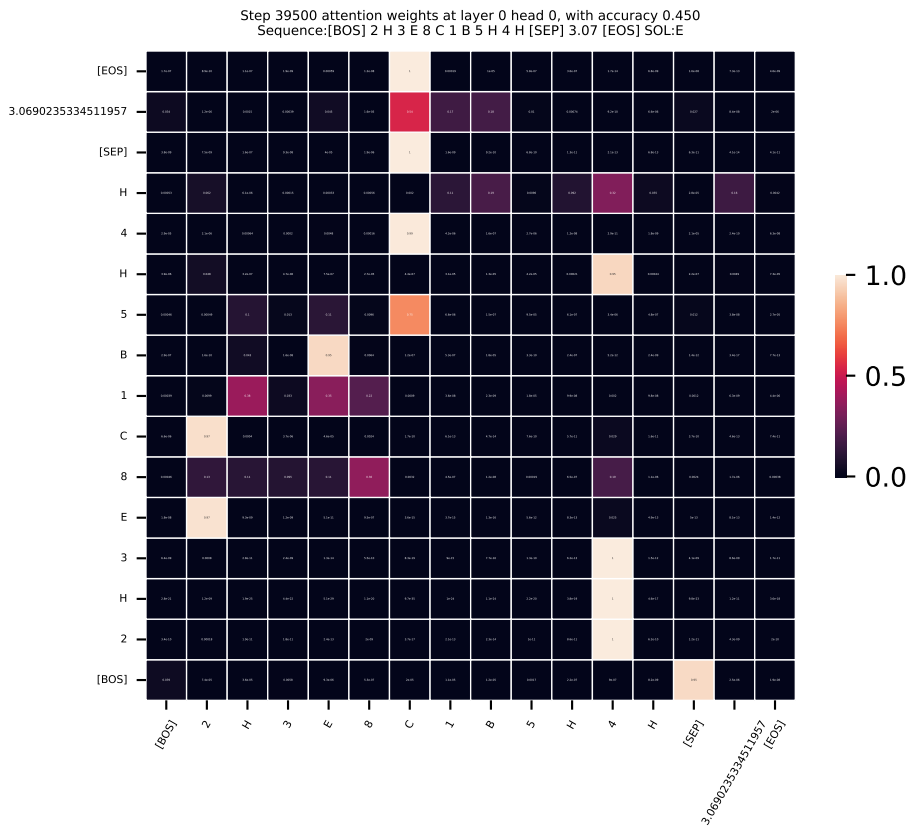

Step 39500 attention weights at layer 0 head 3, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

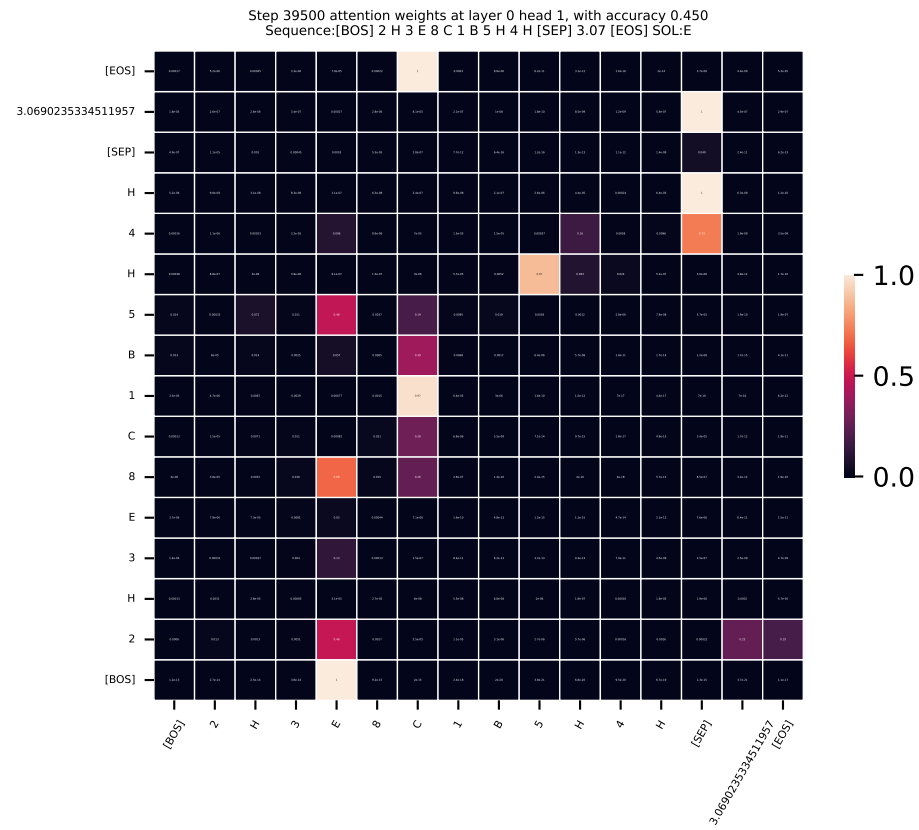

Step 39500 attention weights at layer 0 head 0, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

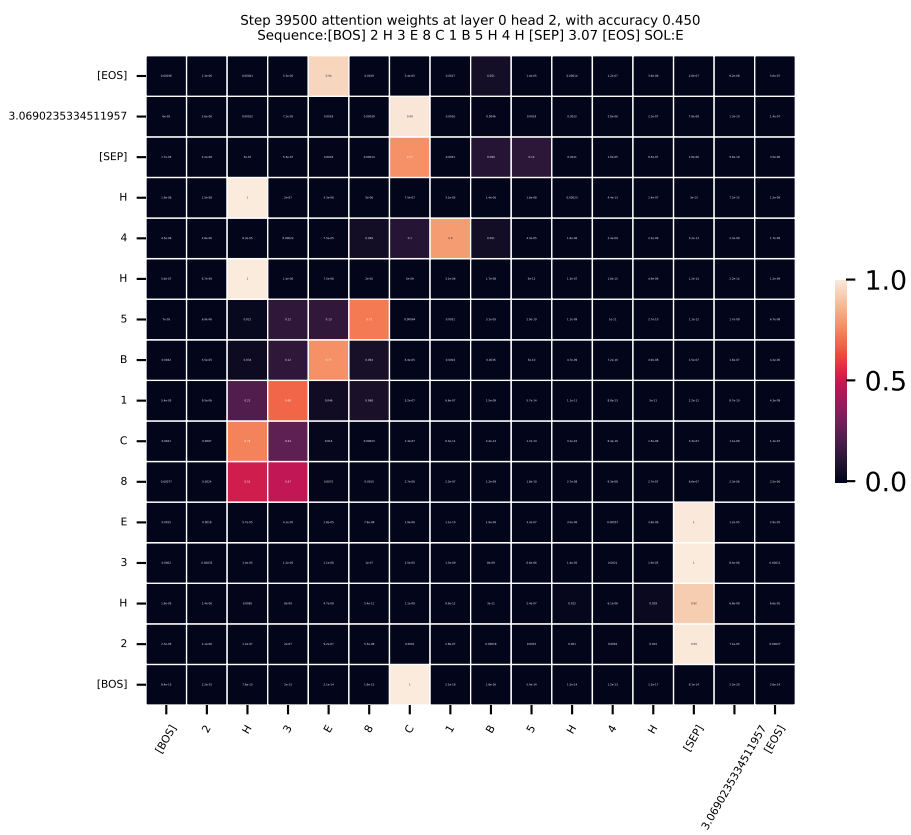

Step 39500 attention weights at layer 0 head 1, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

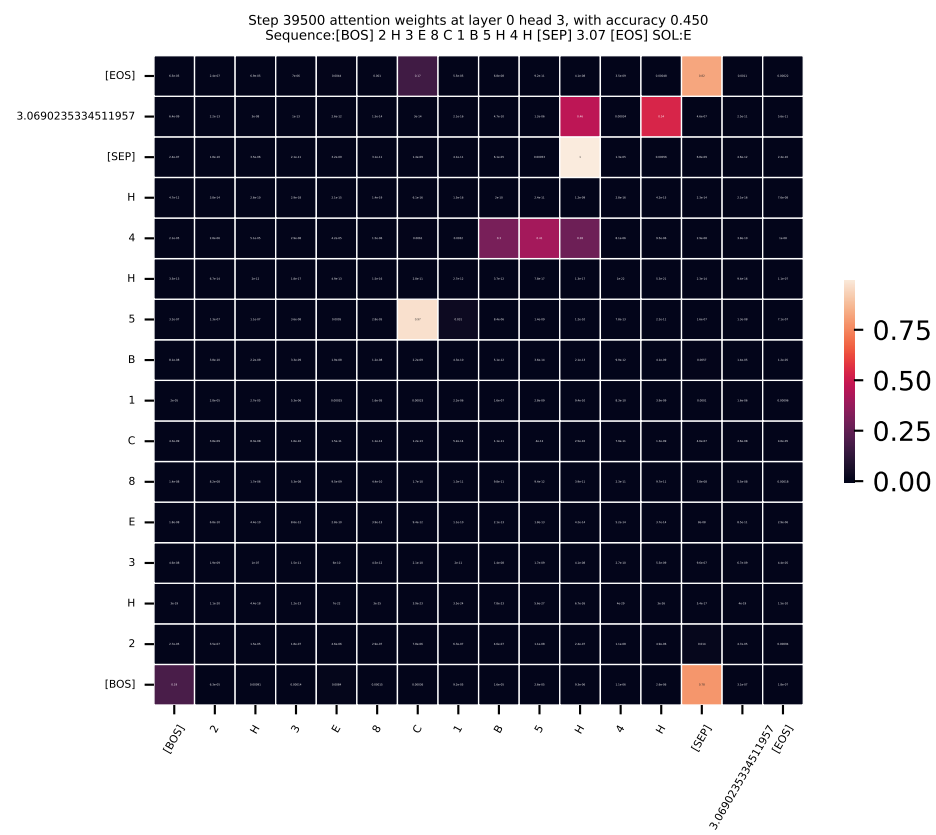

Step 39500 attention weights at layer 0 head 2, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

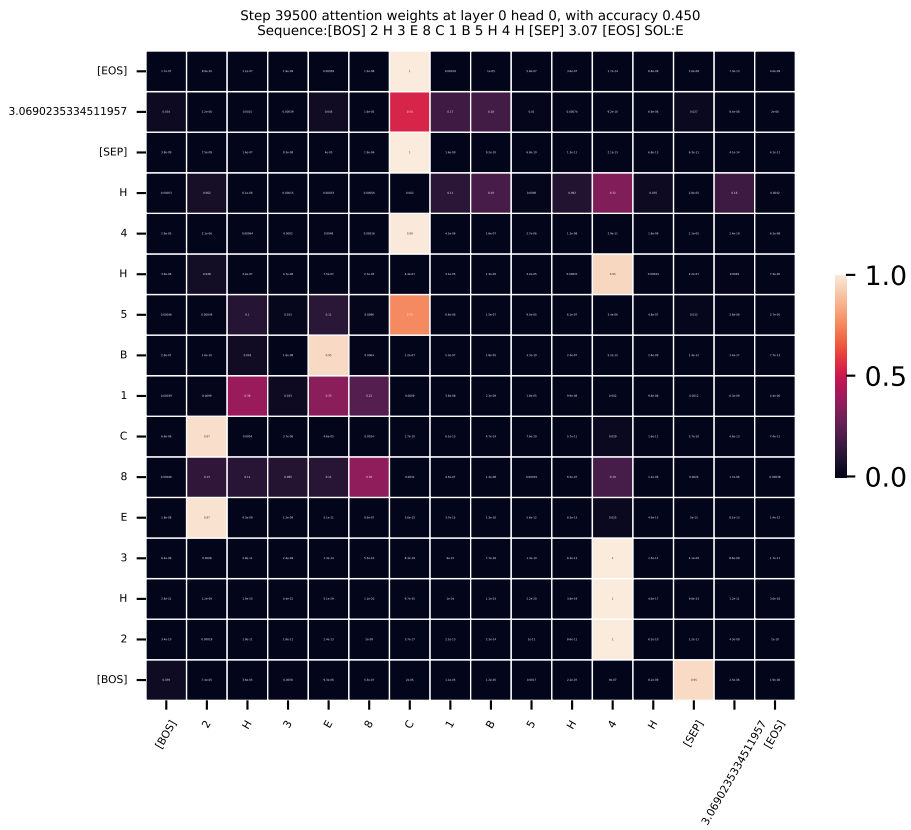

Step 39500 attention weights at layer 0 head 3, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

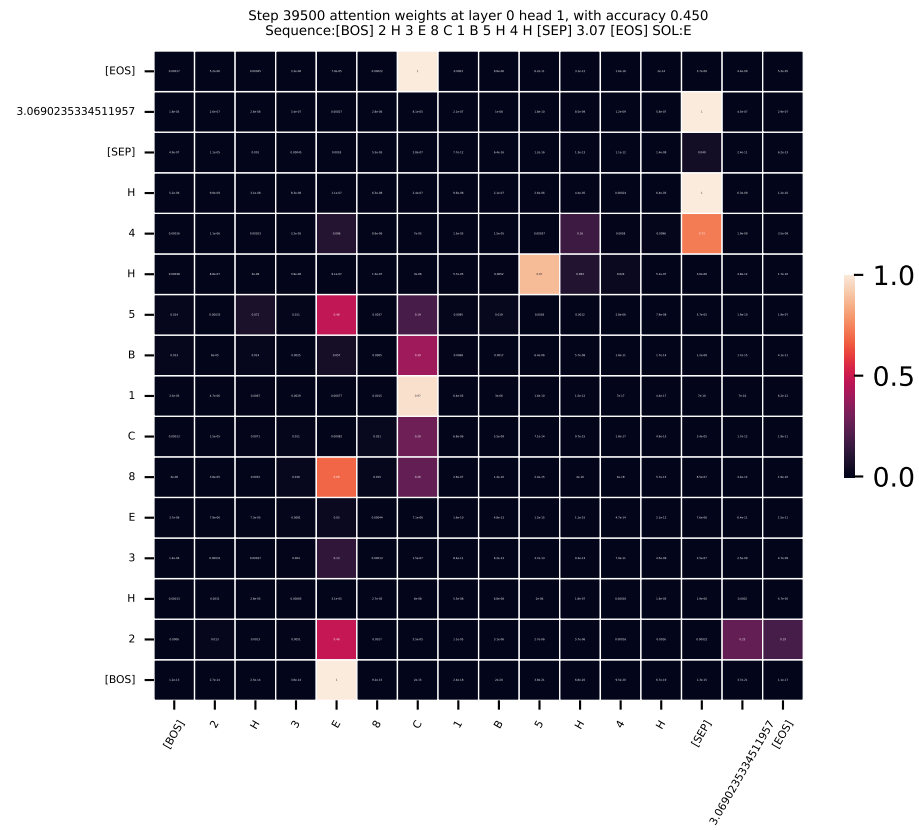

Step 39500 attention weights at layer 0 head 0, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

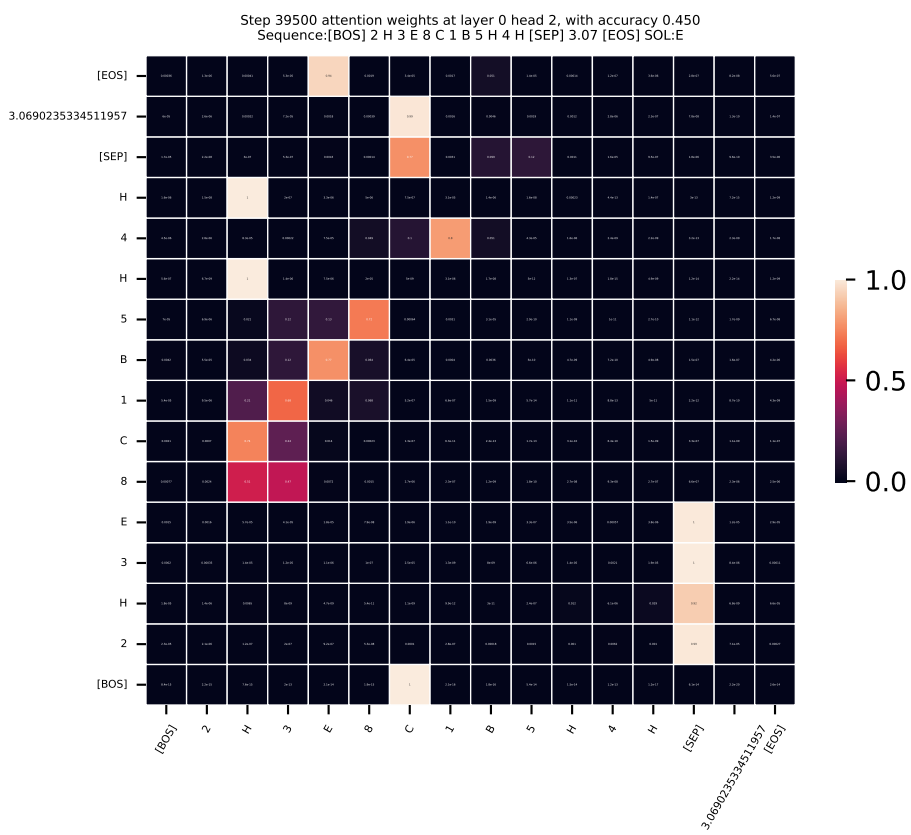

Step 39500 attention weights at layer 0 head 1, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

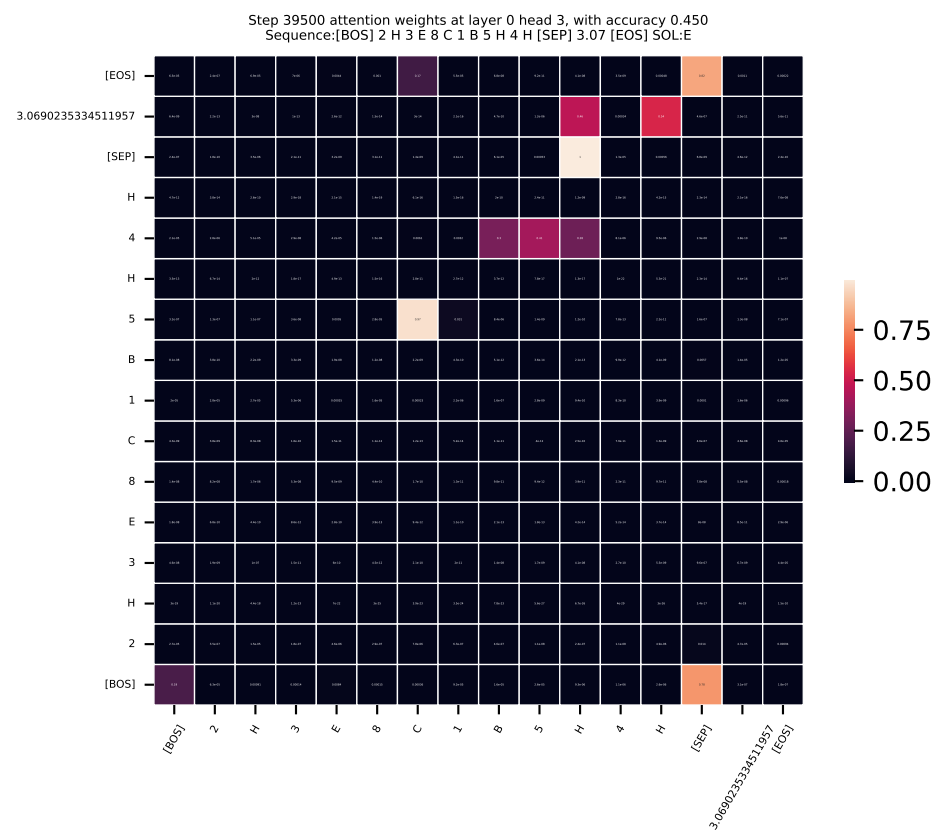

Step 39500 attention weights at layer 0 head 2, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

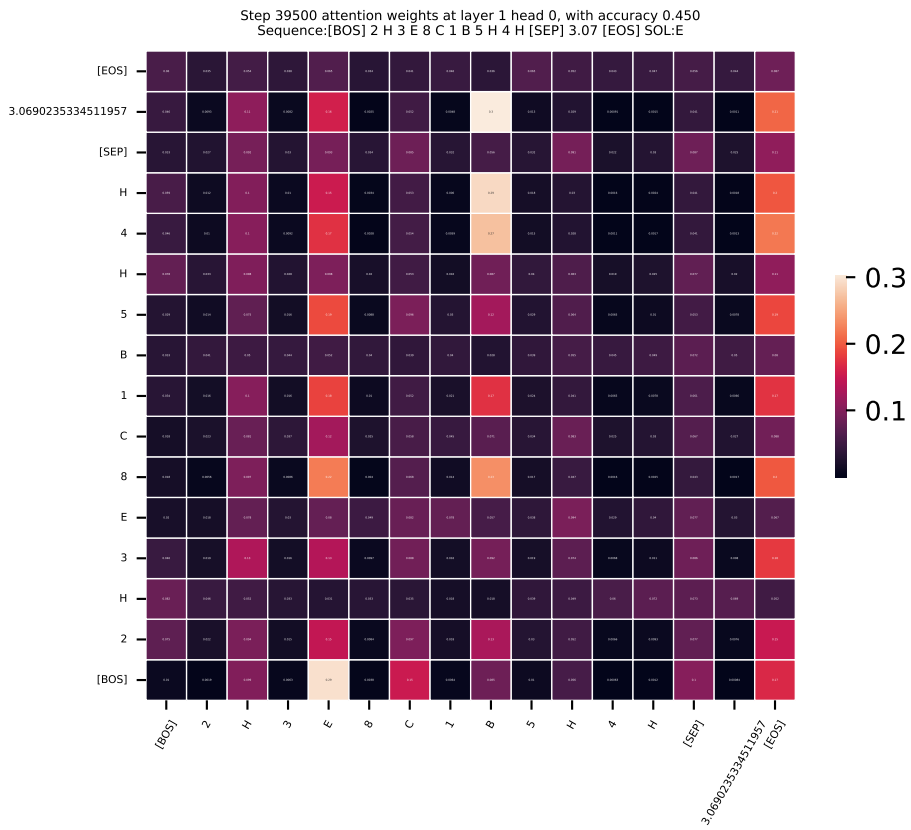

Step 39500 attention weights at layer 0 head 3, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

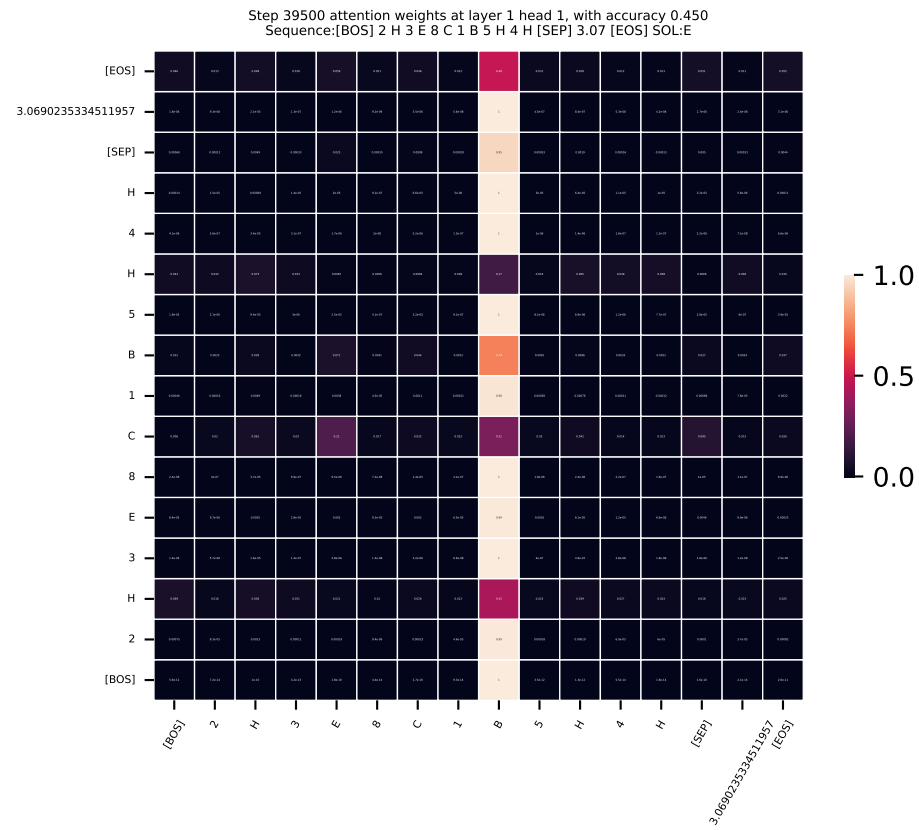

Step 39500 attention weights at layer 0 head 0, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

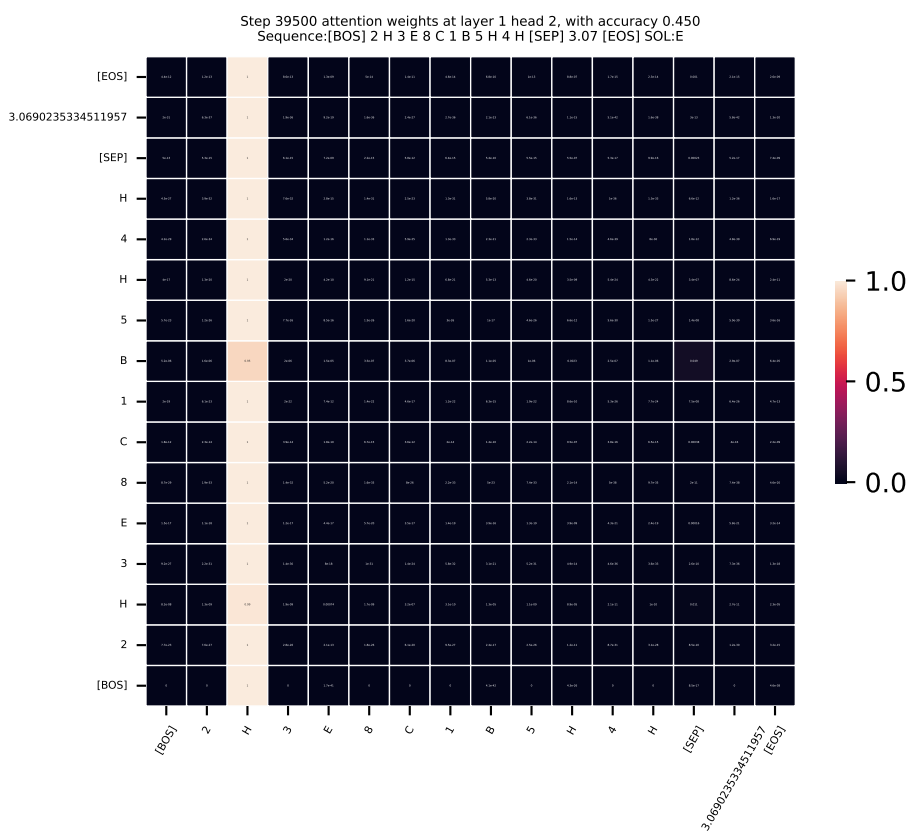

Step 39500 attention weights at layer 0 head 1, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

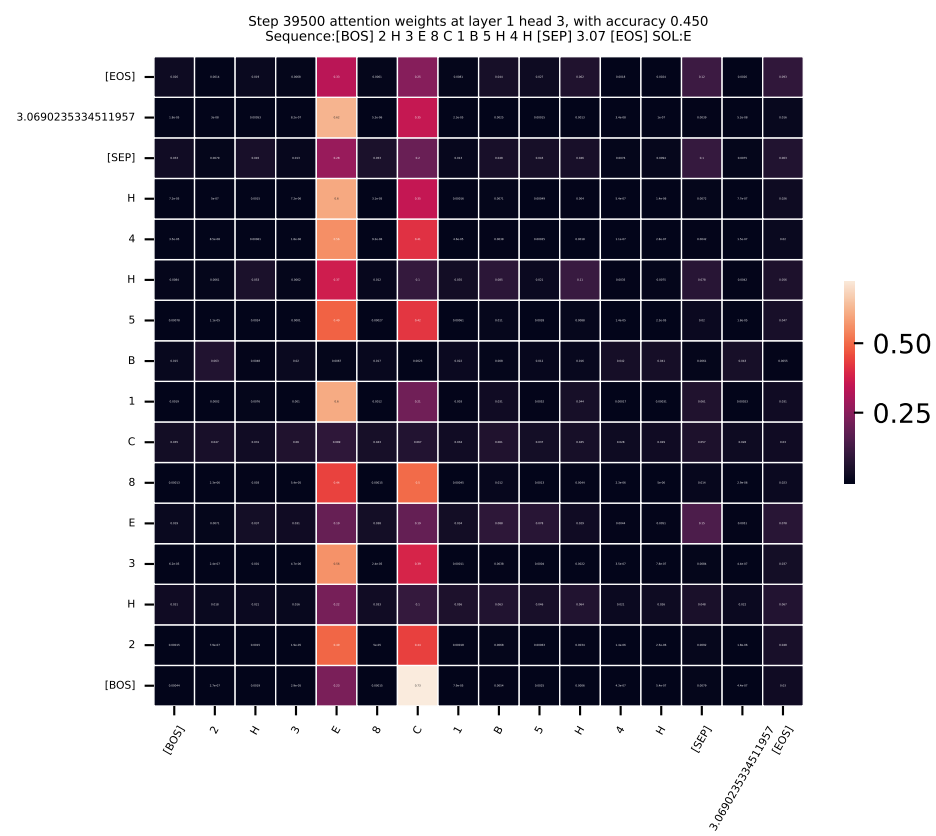

Step 39500 attention weights at layer 0 head 2, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

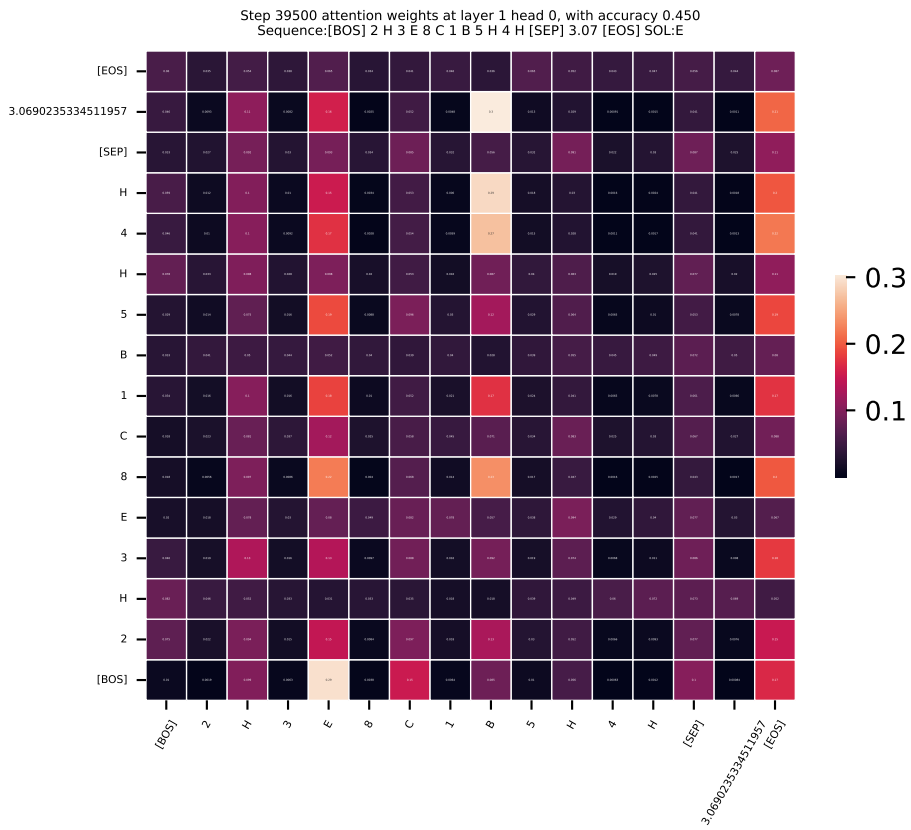

Step 39500 attention weights at layer 0 head 3, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

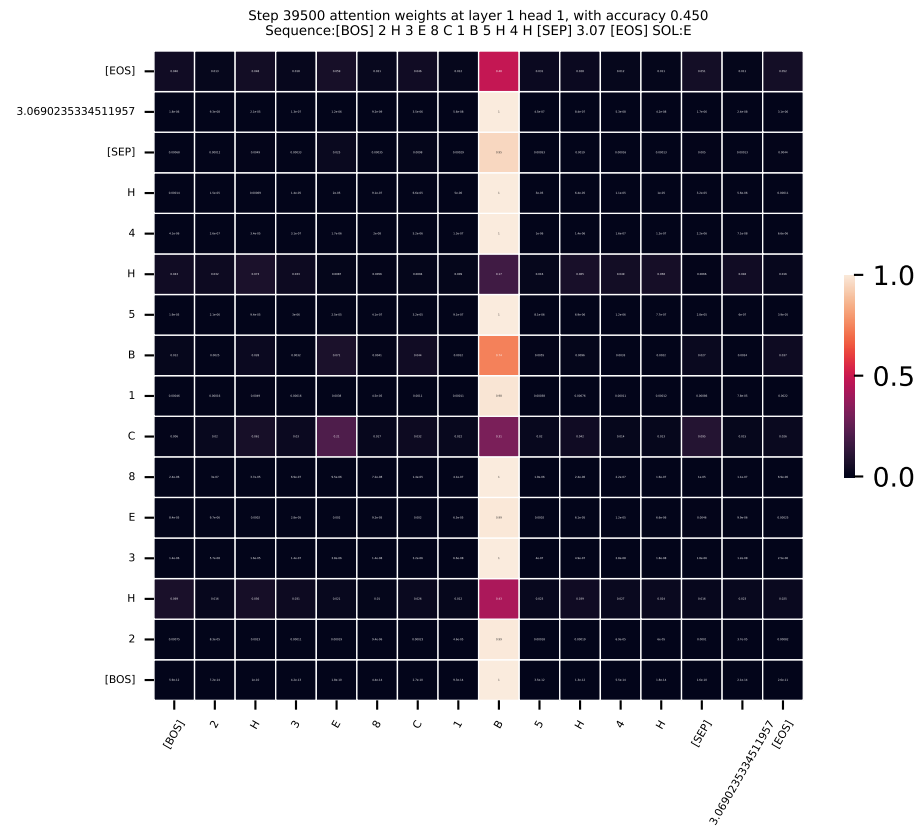

Step 39500 attention weights at layer 1 head 0, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.3
0.2
0.1

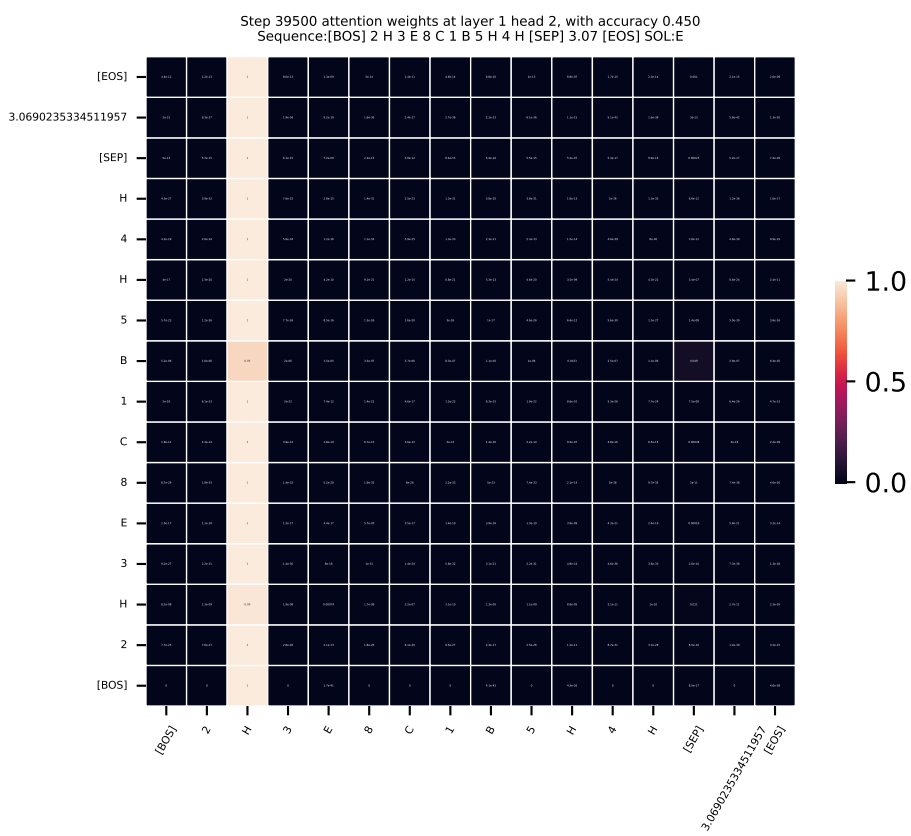

Step 39500 attention weights at layer 1 head 1, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

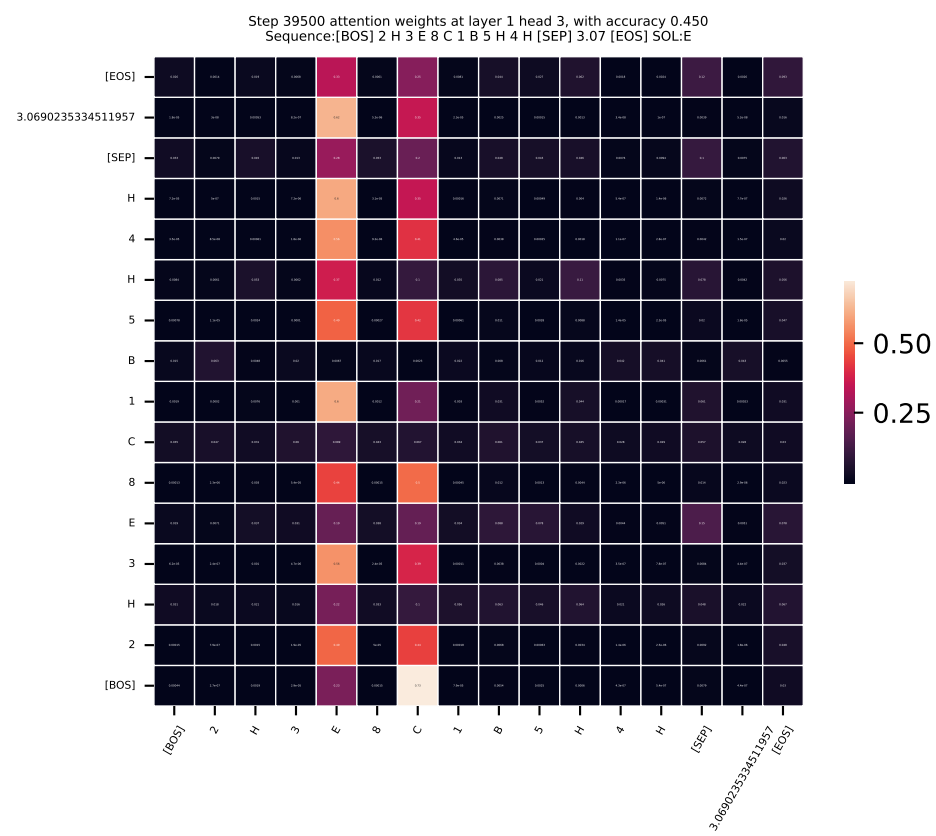

Step 39500 attention weights at layer 1 head 2, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

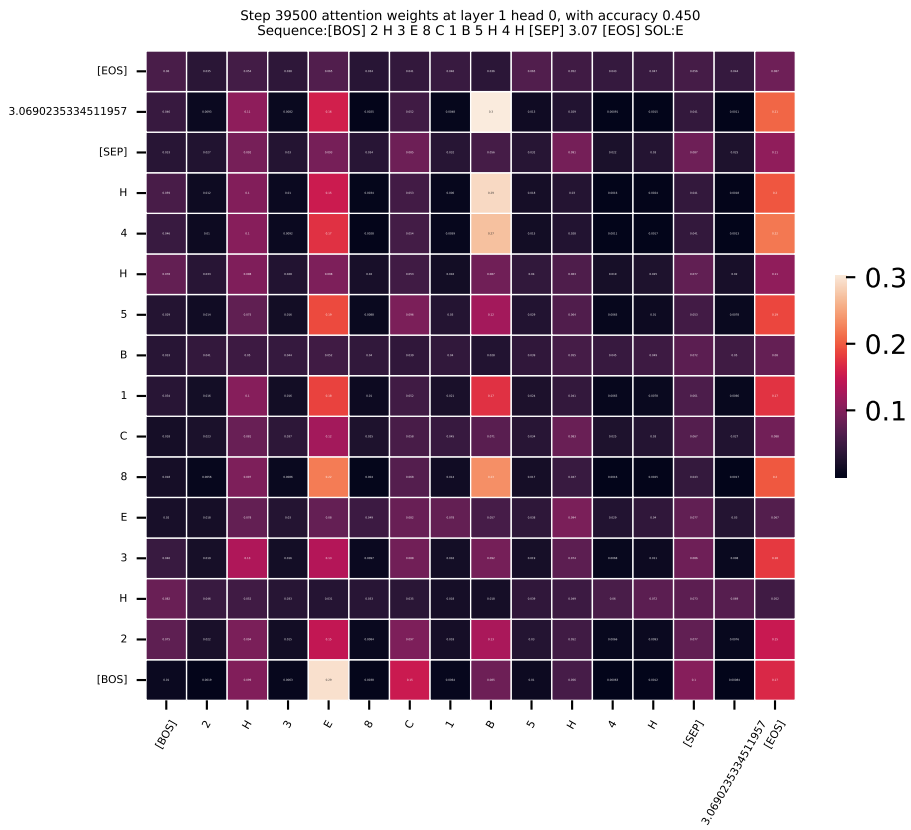

Step 39500 attention weights at layer 1 head 3, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.50
0.25

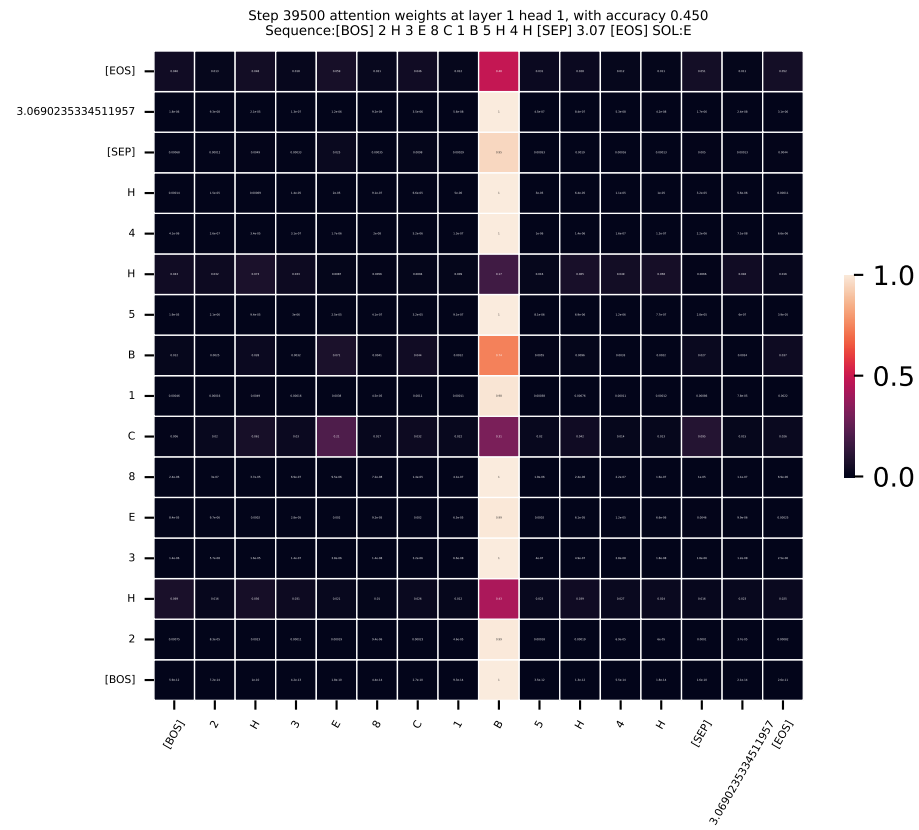

Step 39500 attention weights at layer 1 head 0, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.3
0.2
0.1

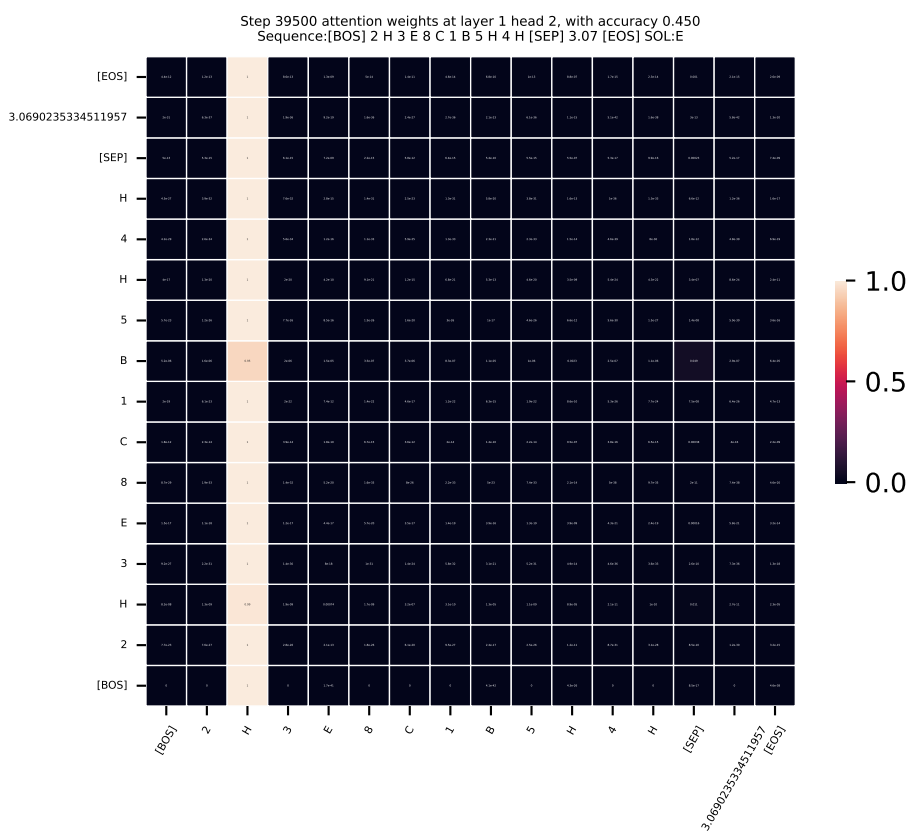

Step 39500 attention weights at layer 1 head 1, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

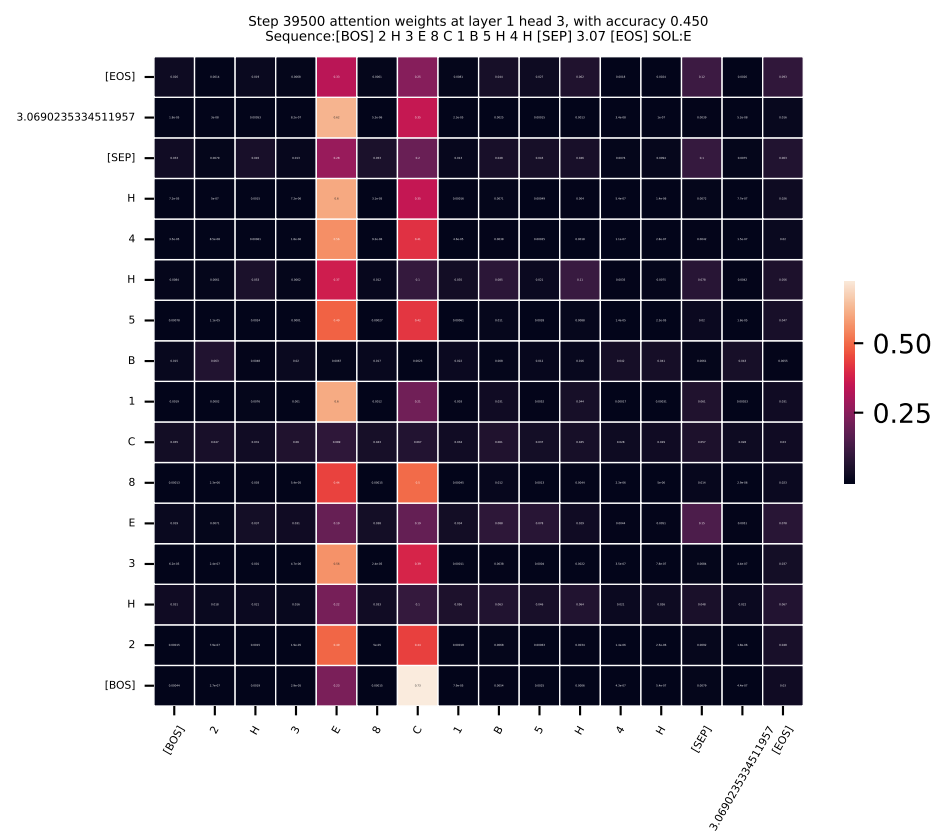

Step 39500 attention weights at layer 1 head 2, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

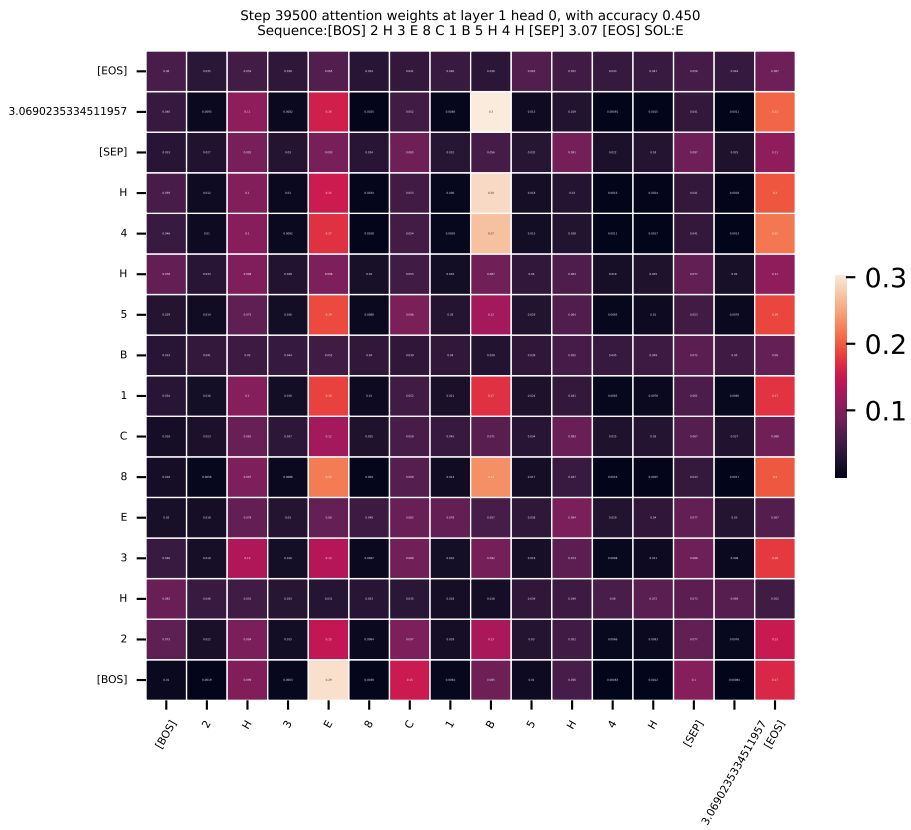

Step 39500 attention weights at layer 1 head 3, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.50
0.25

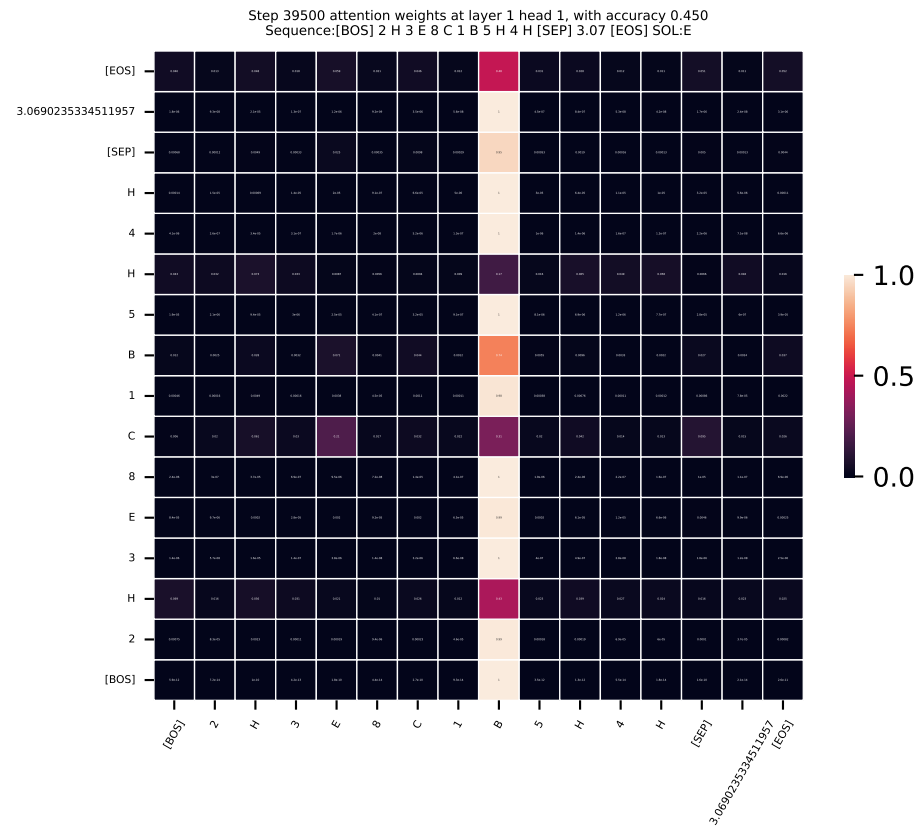

Step 39500 attention weights at layer 1 head 0, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.3
0.2
0.1

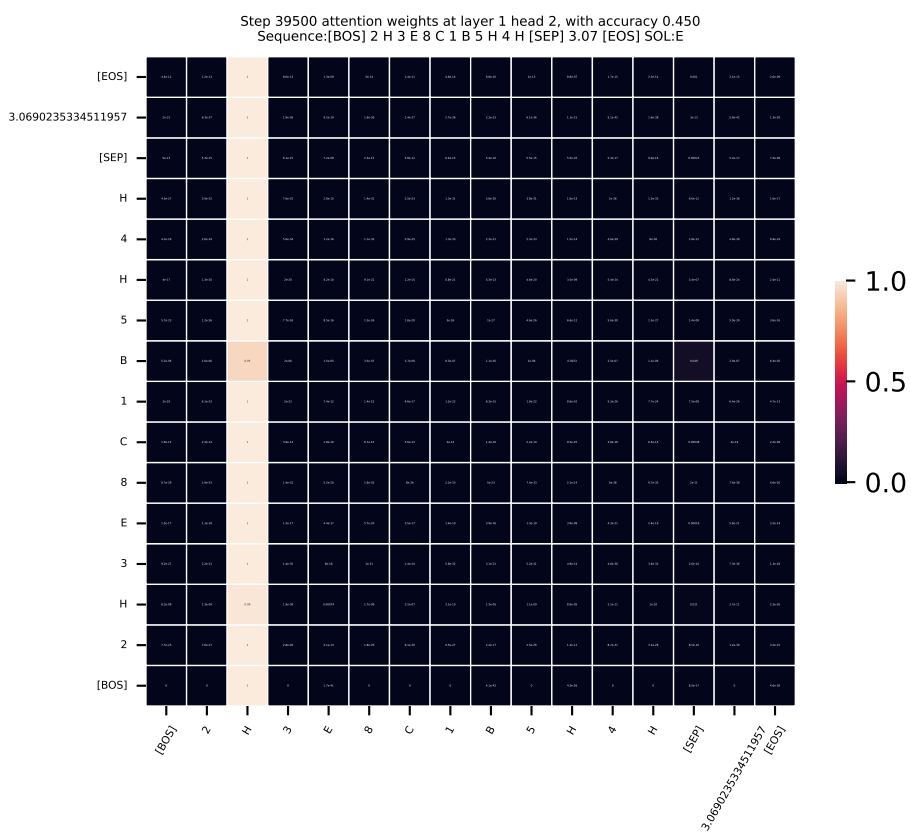

Step 39500 attention weights at layer 1 head 1, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

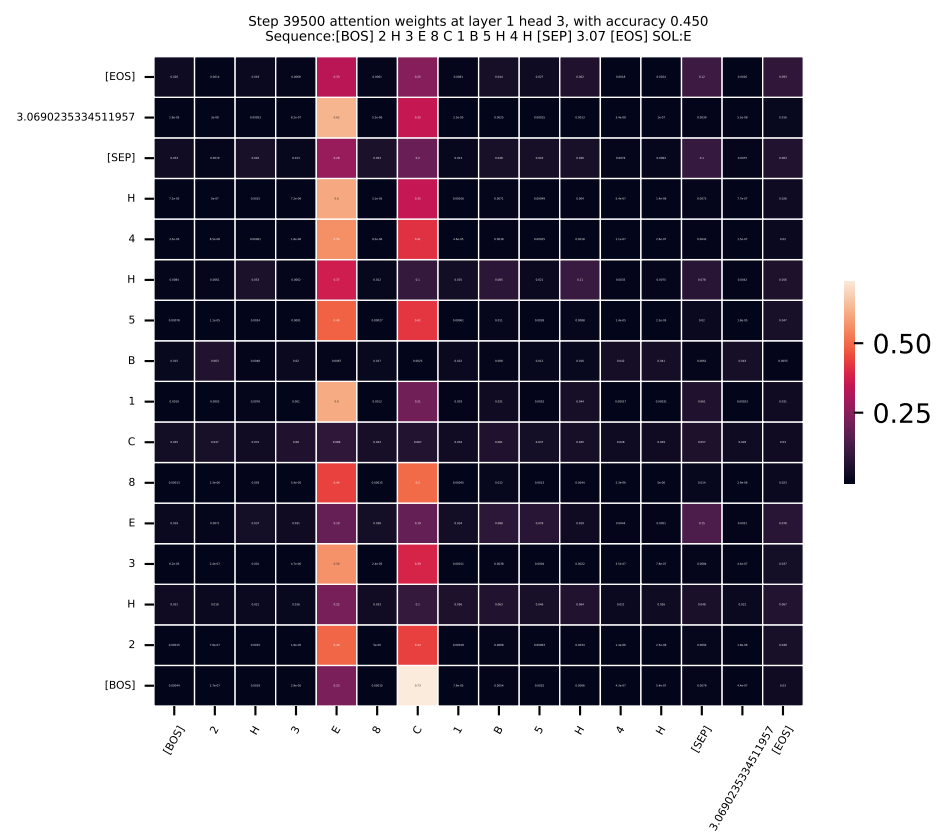

Step 39500 attention weights at layer 1 head 2, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

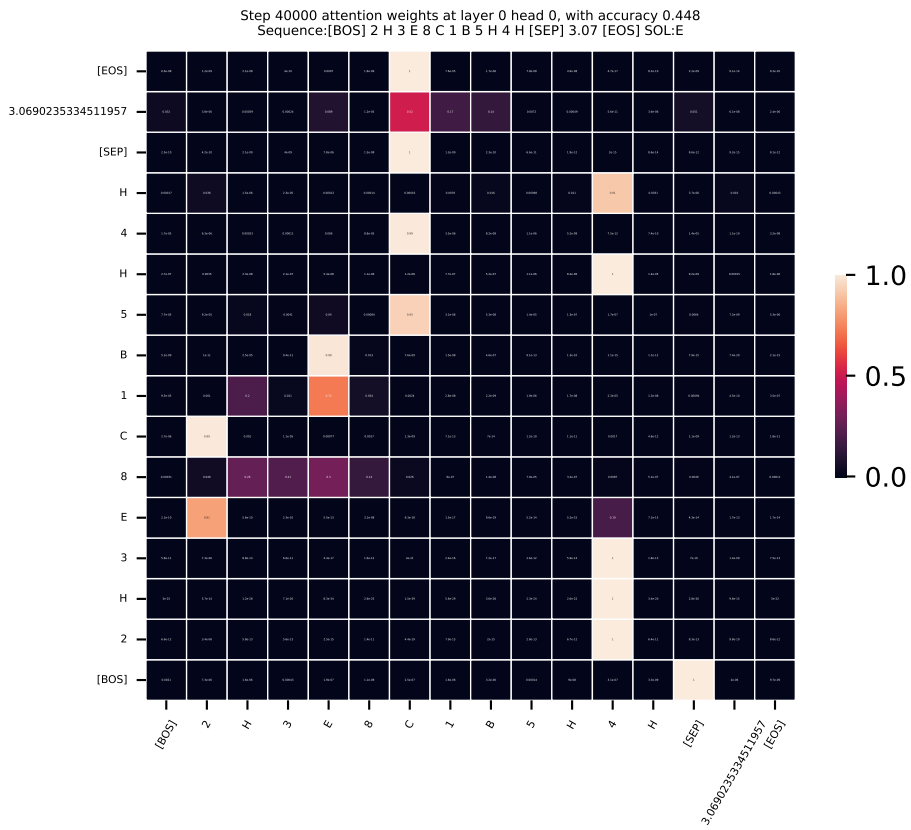

Step 39500 attention weights at layer 1 head 3, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.50
0.25

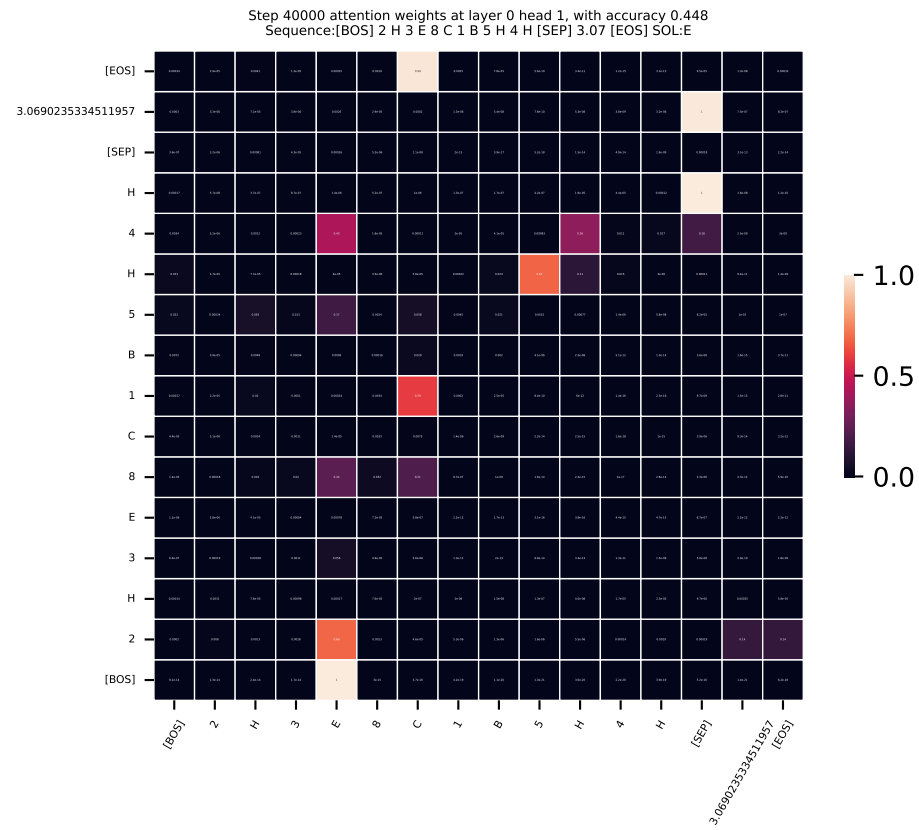

Step 39500 attention weights at layer 1 head 0, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.3
0.2
0.1

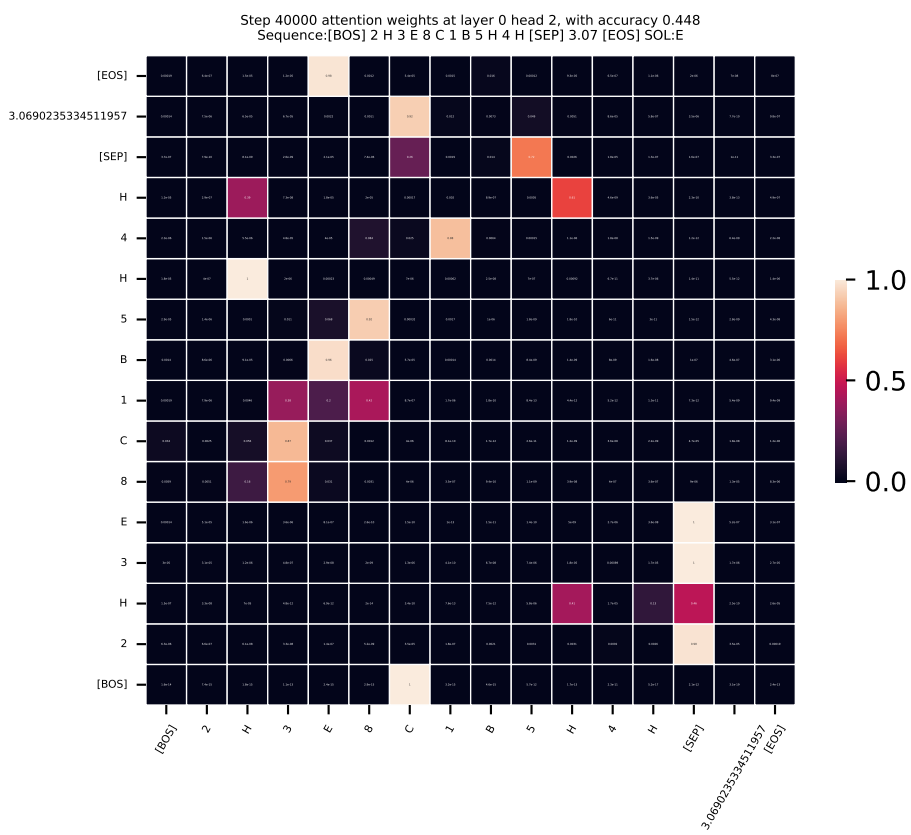

Step 39500 attention weights at layer 1 head 1, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

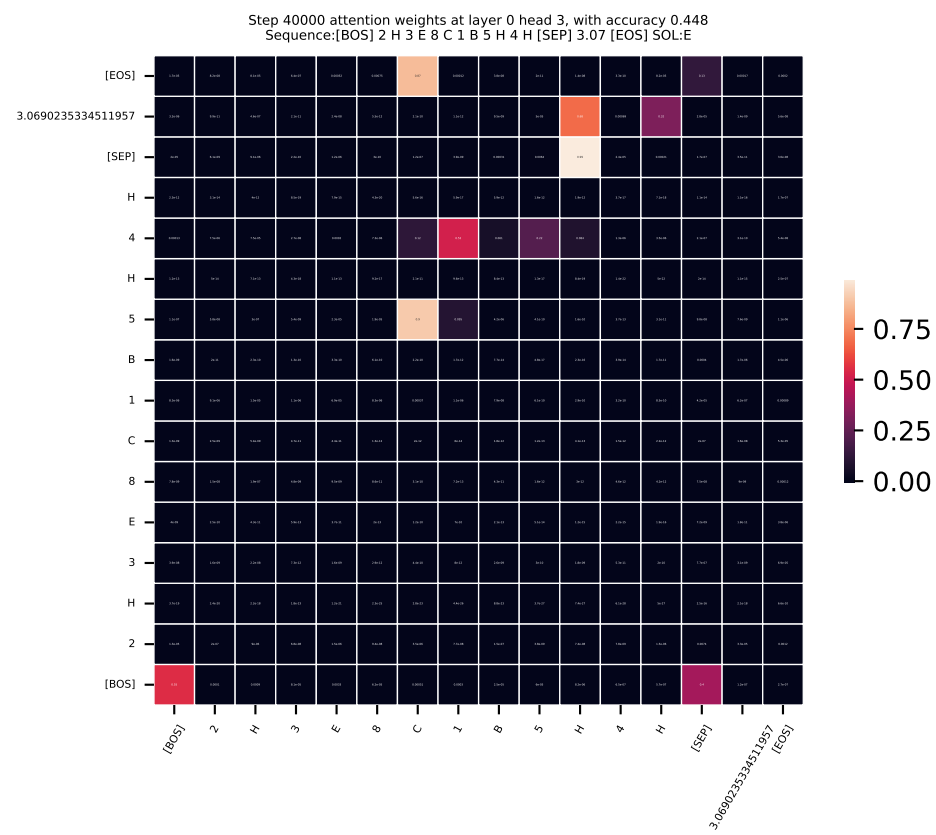

Step 39500 attention weights at layer 1 head 2, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

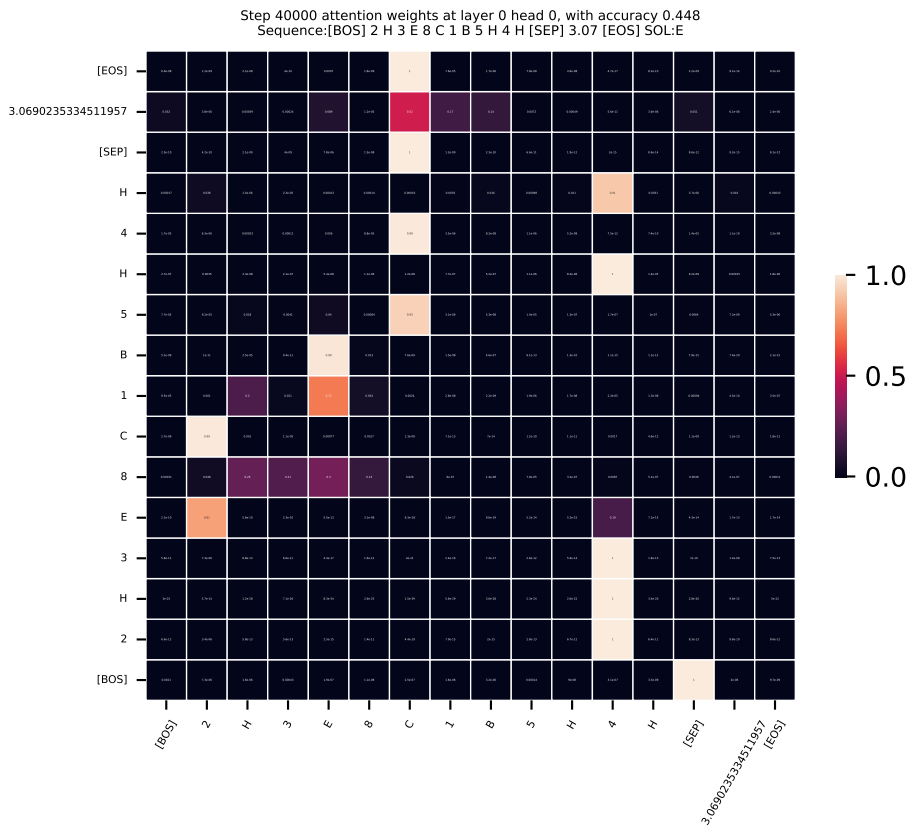

Step 39500 attention weights at layer 1 head 3, with accuracy 0.450
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.50
0.25

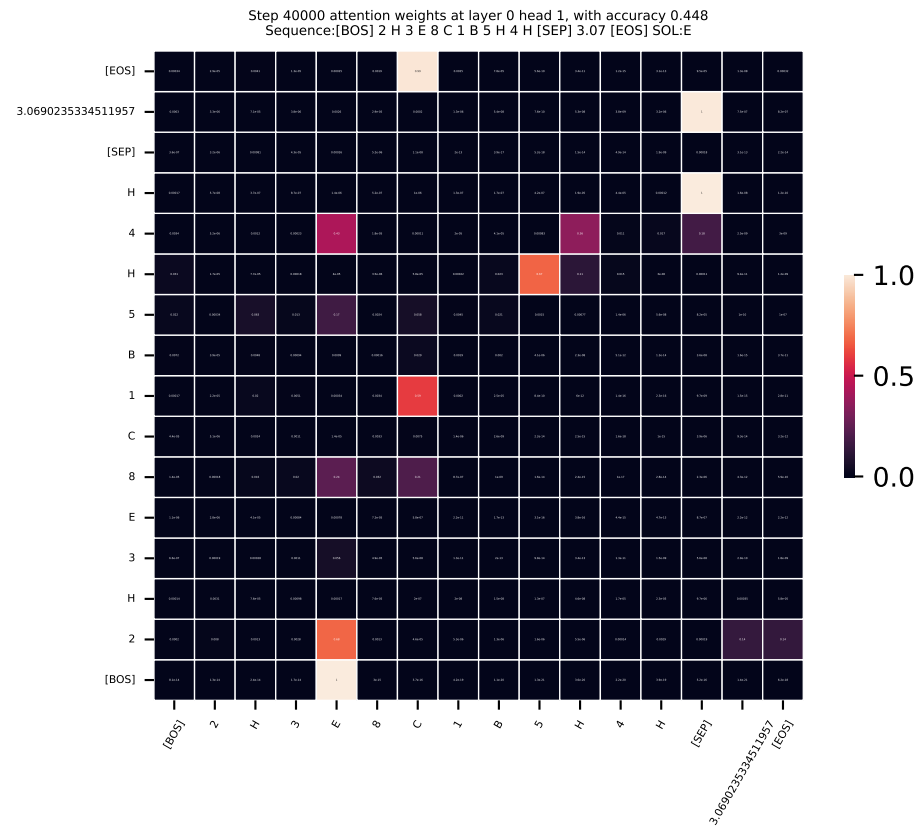

Step 40000 attention weights at layer 0 head 0, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

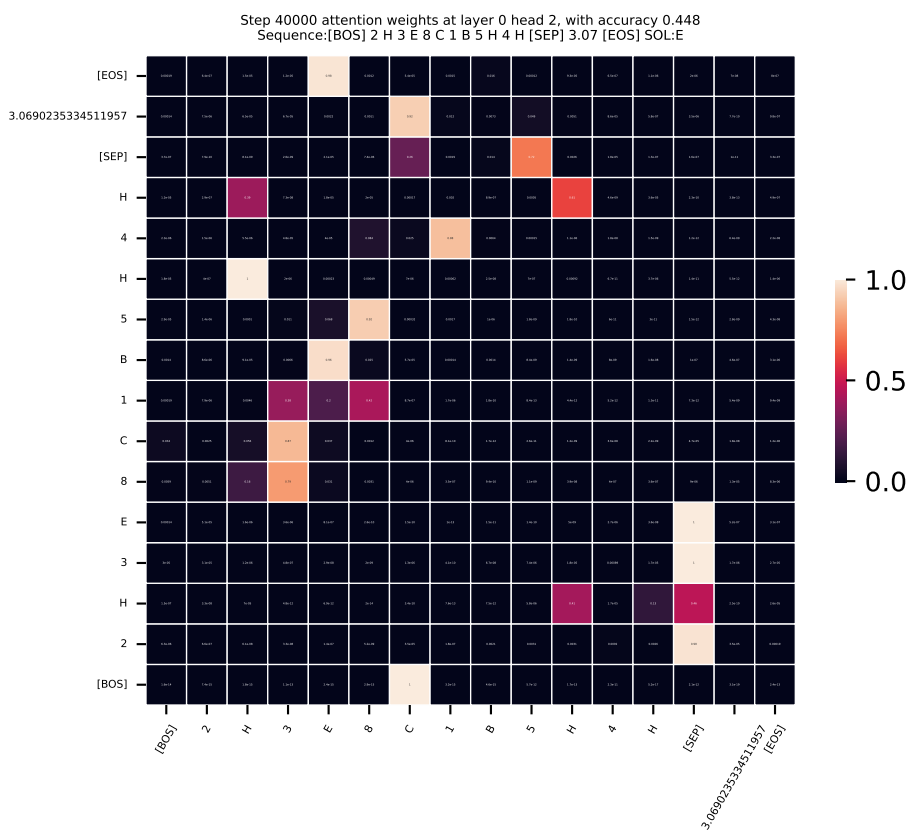

Step 40000 attention weights at layer 0 head 1, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

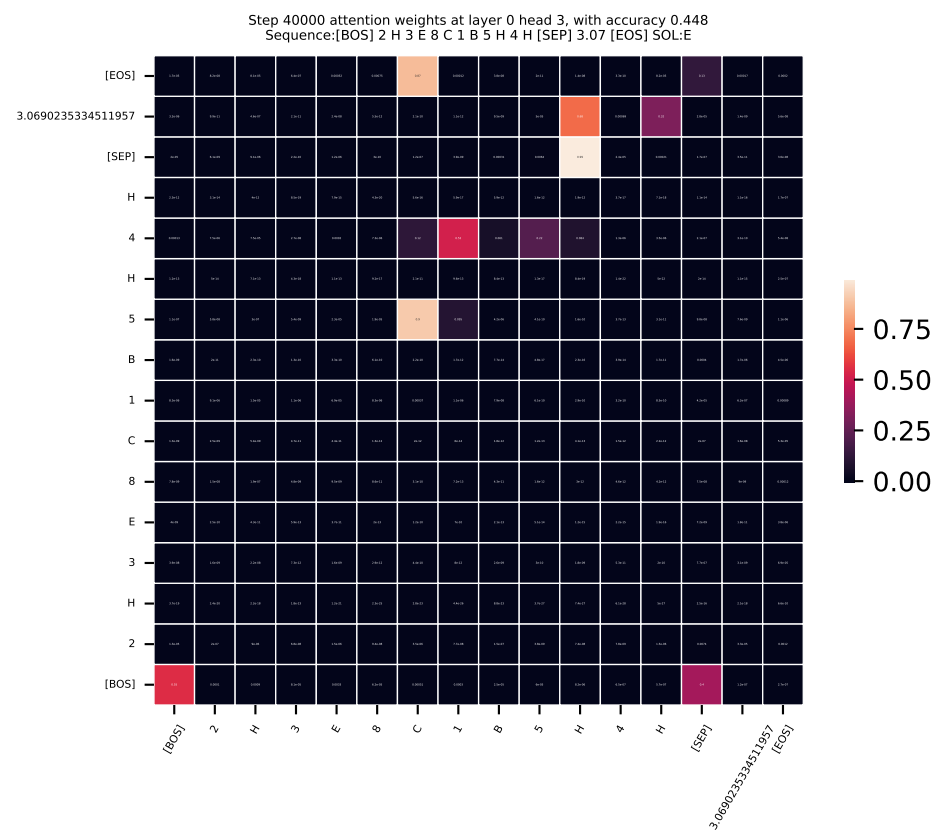

Step 40000 attention weights at layer 0 head 2, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

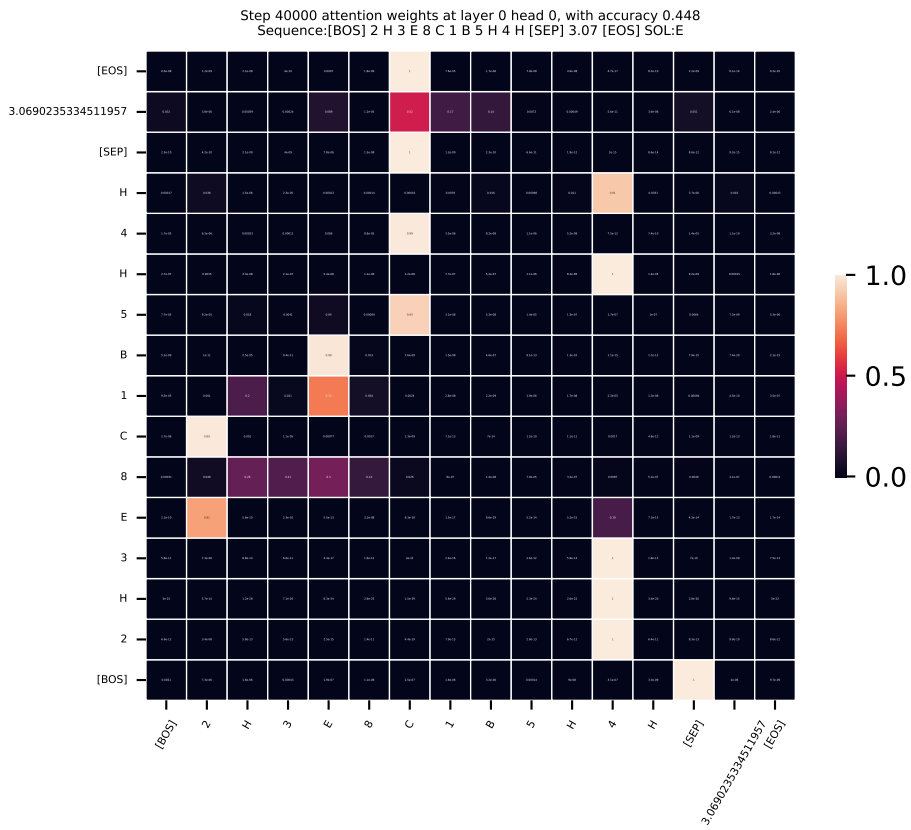

Step 40000 attention weights at layer 0 head 3, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

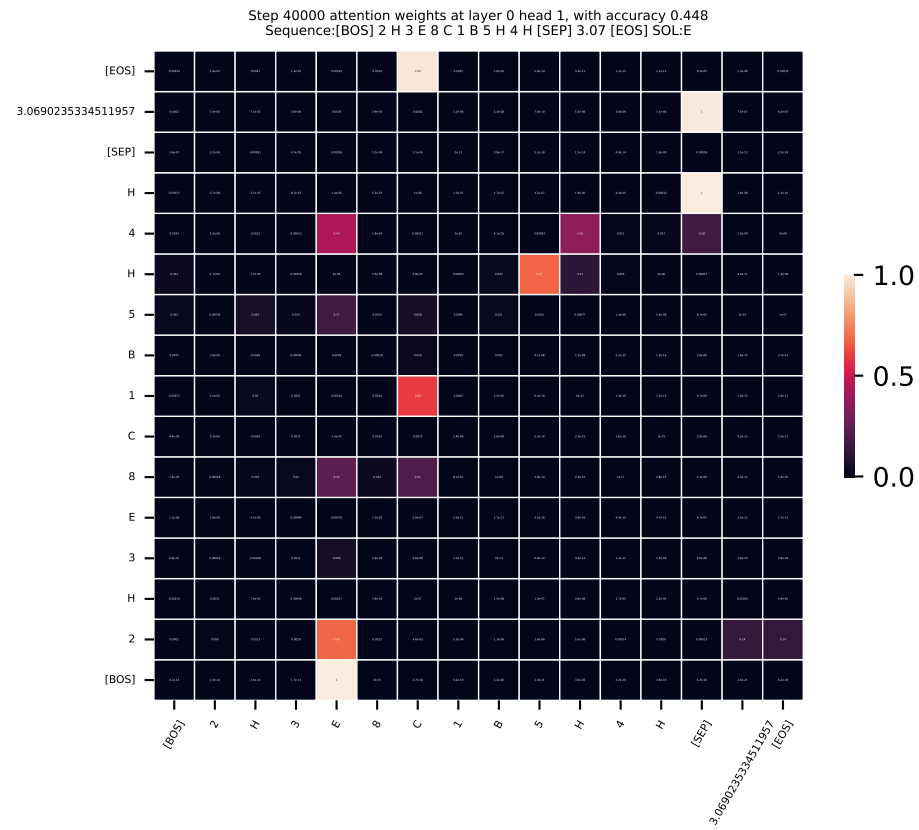

Step 40000 attention weights at layer 0 head 0, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

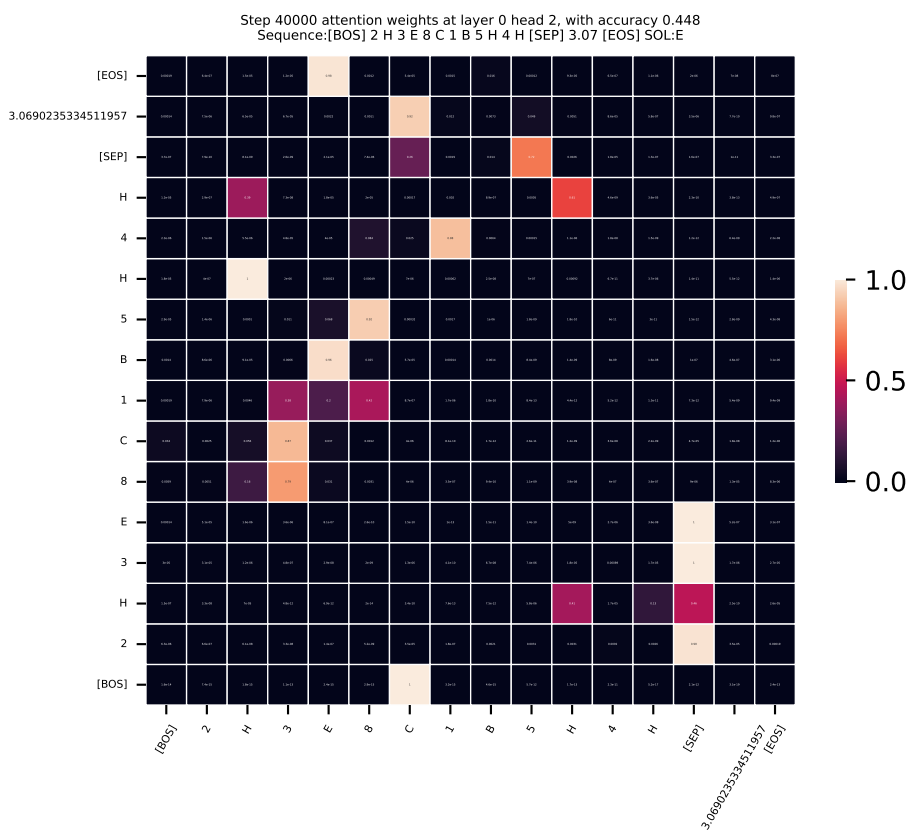

Step 40000 attention weights at layer 0 head 1, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

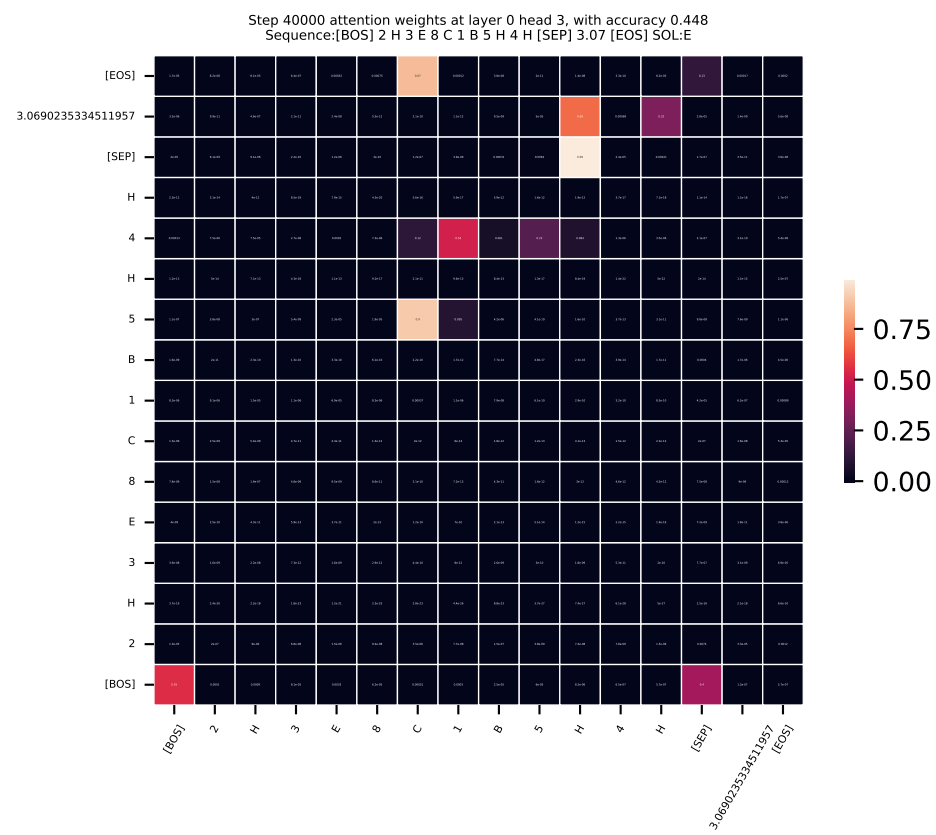

Step 40000 attention weights at layer 0 head 2, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

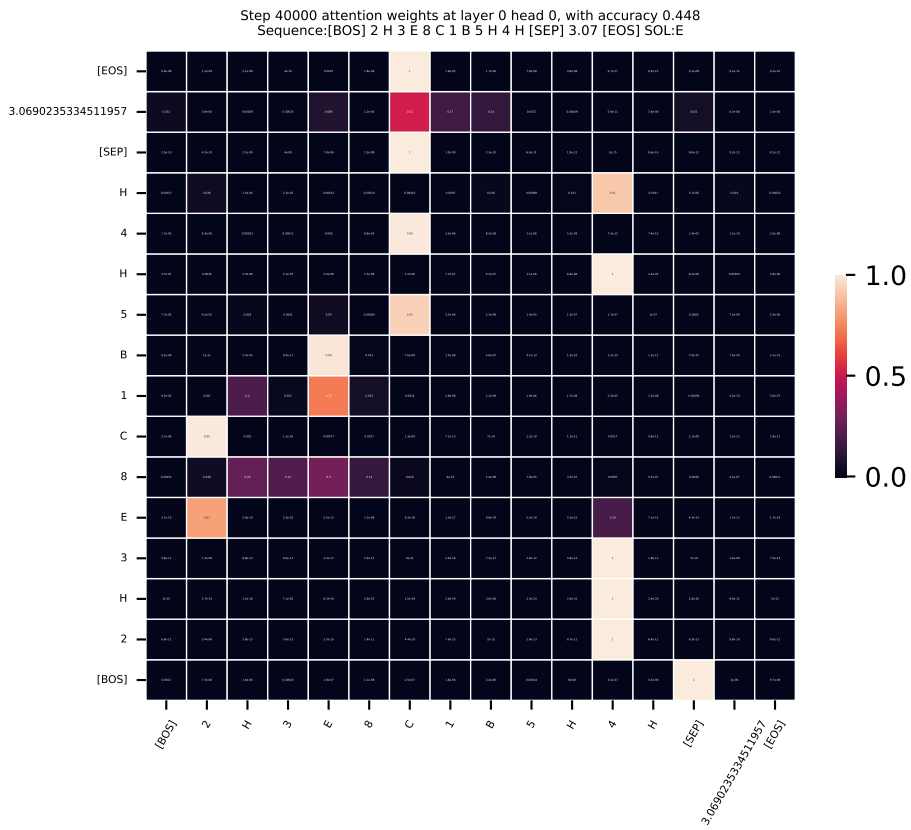

Step 40000 attention weights at layer 0 head 3, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

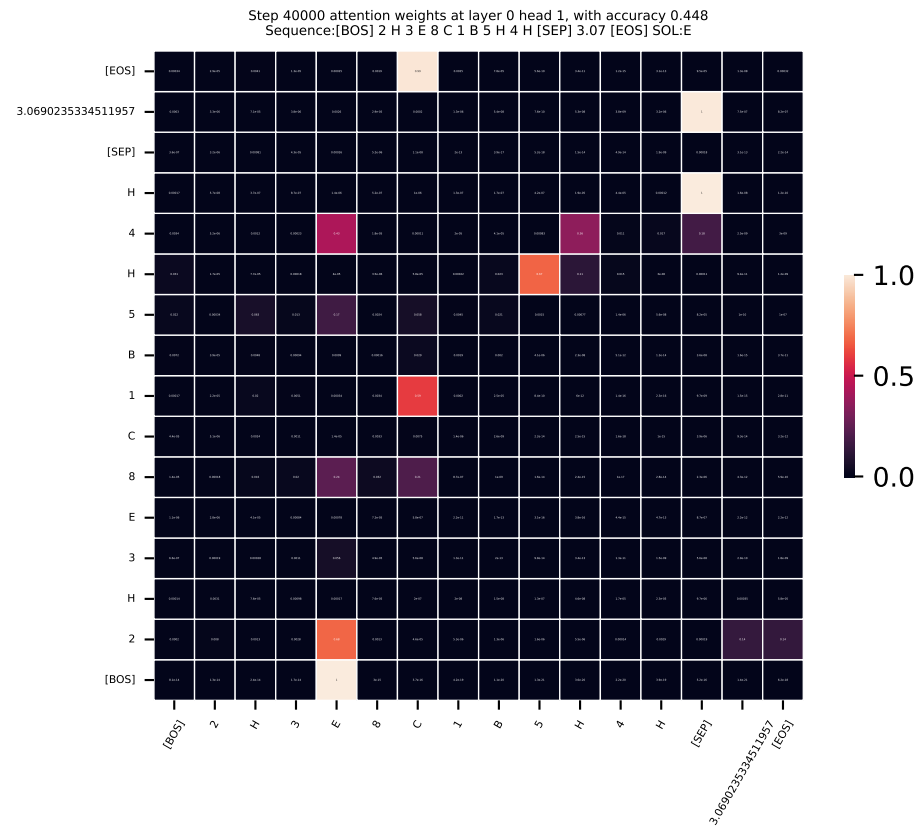

Step 40000 attention weights at layer 0 head 0, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

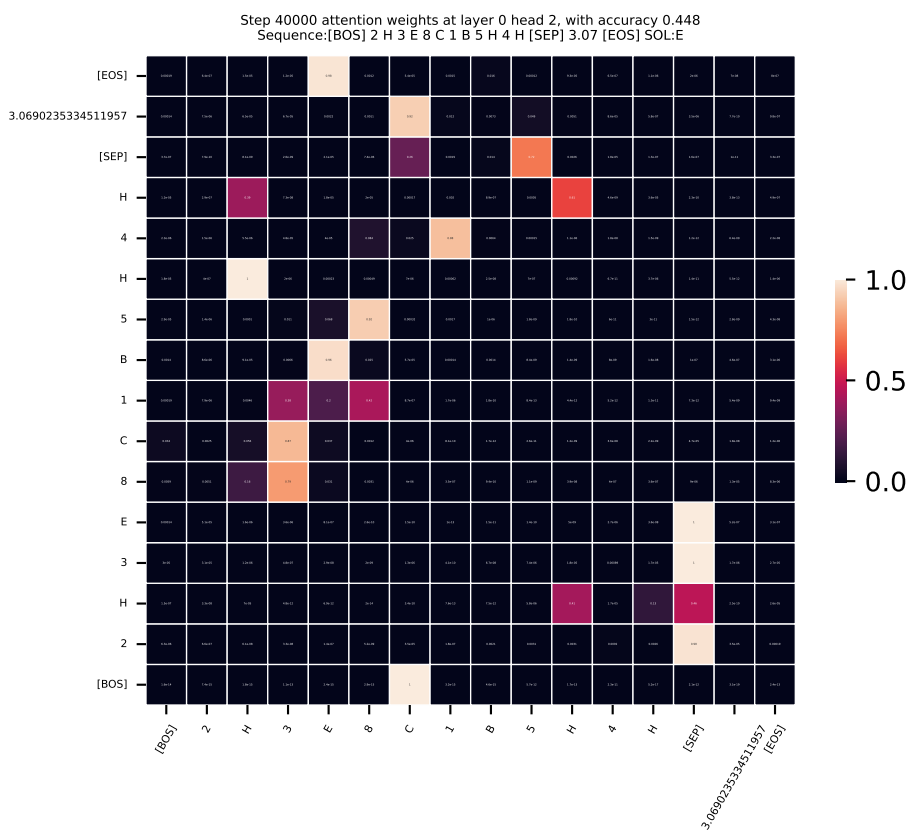

Step 40000 attention weights at layer 0 head 1, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

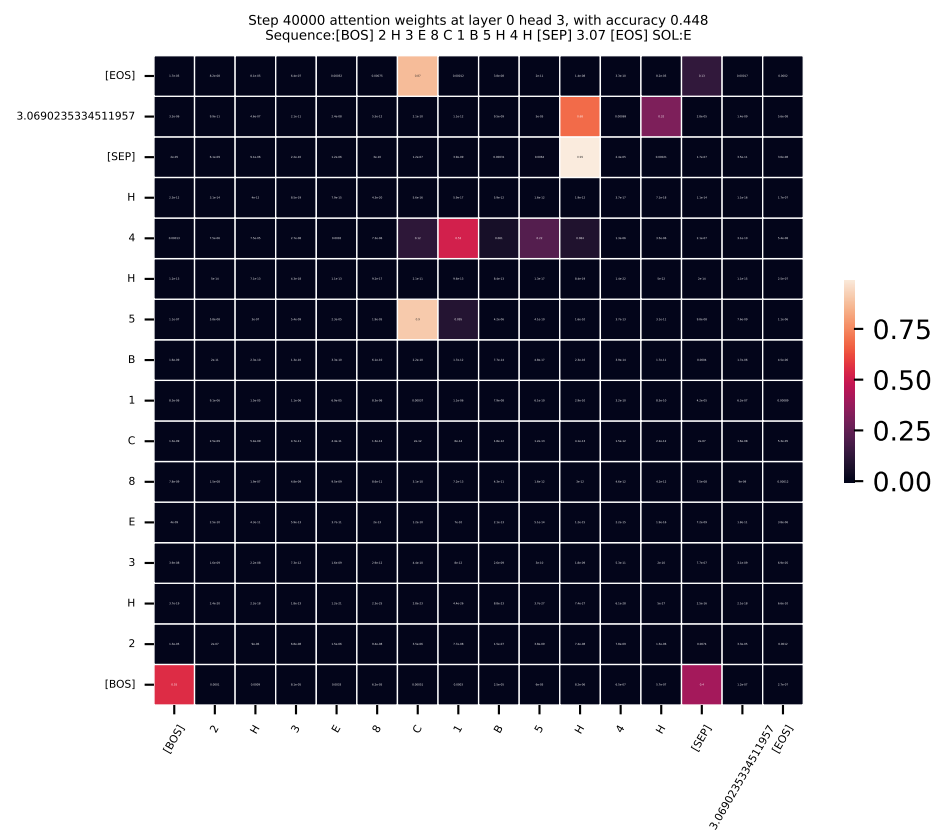

Step 40000 attention weights at layer 0 head 2, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

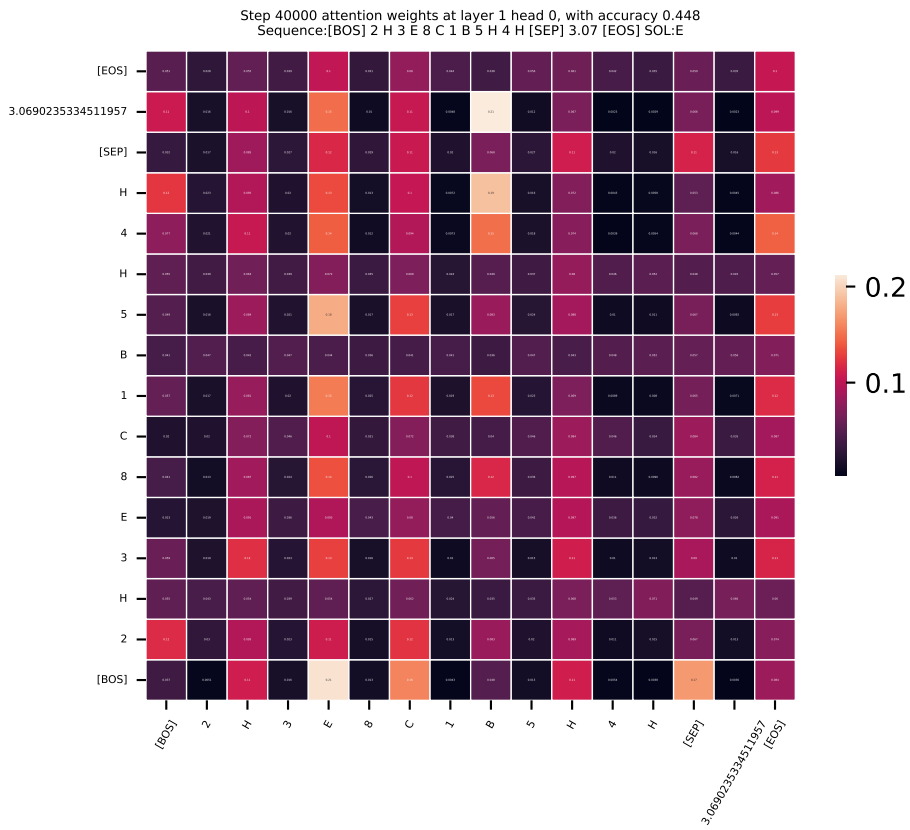

Step 40000 attention weights at layer 0 head 3, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

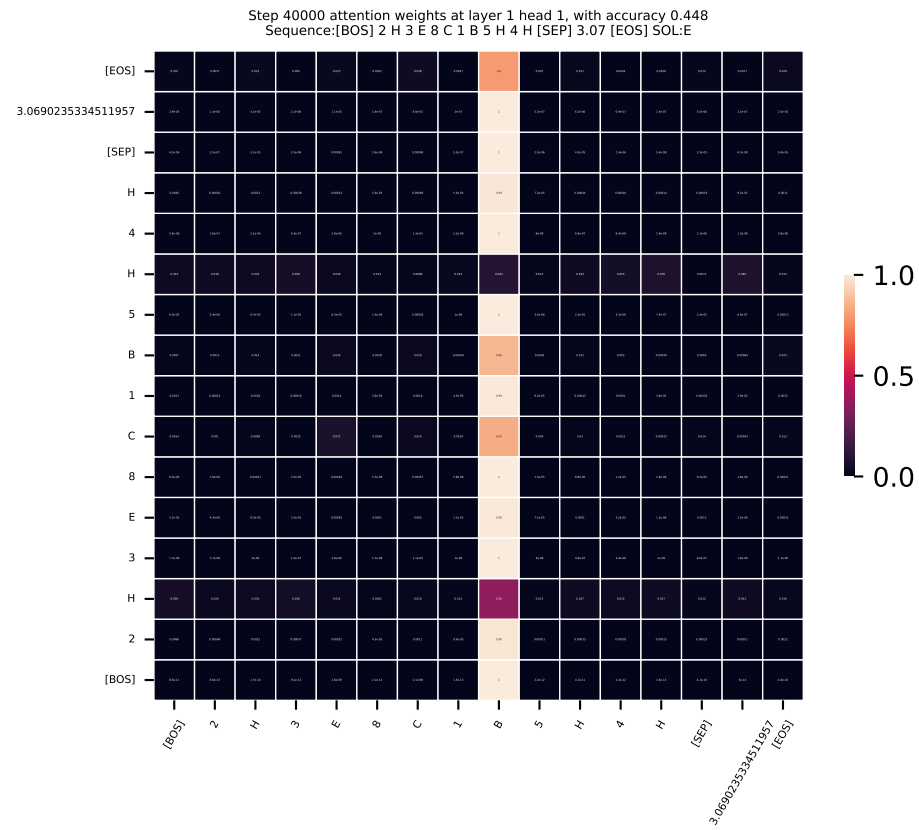

Step 40000 attention weights at layer 0 head 0, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

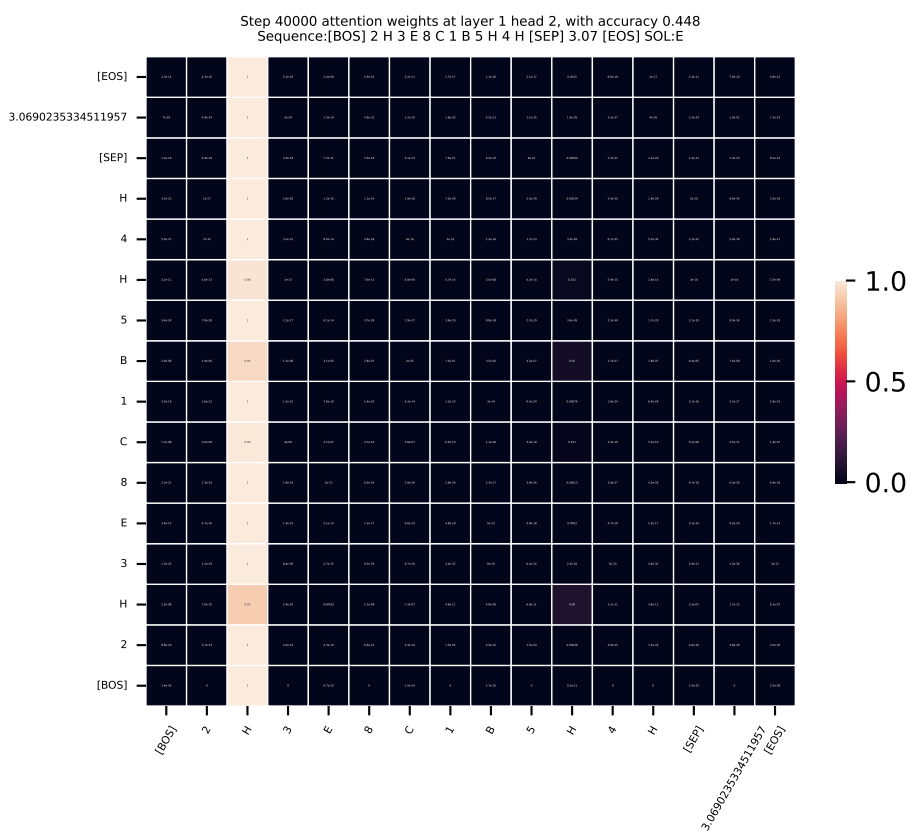

Step 40000 attention weights at layer 0 head 1, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

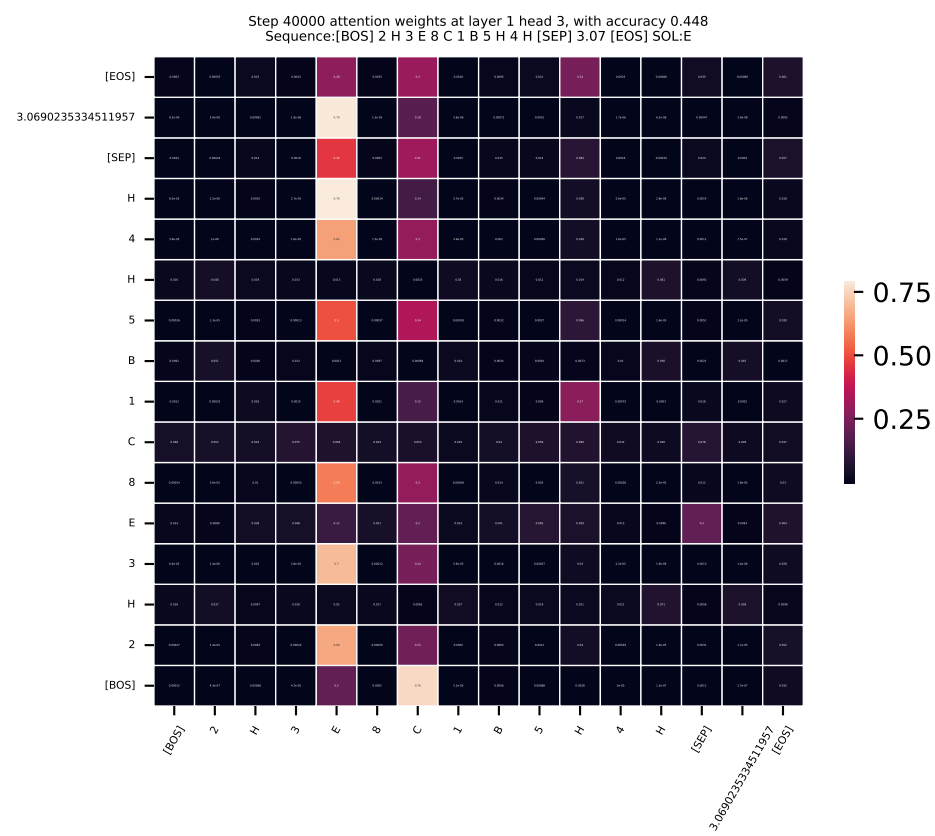

Step 40000 attention weights at layer 0 head 2, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

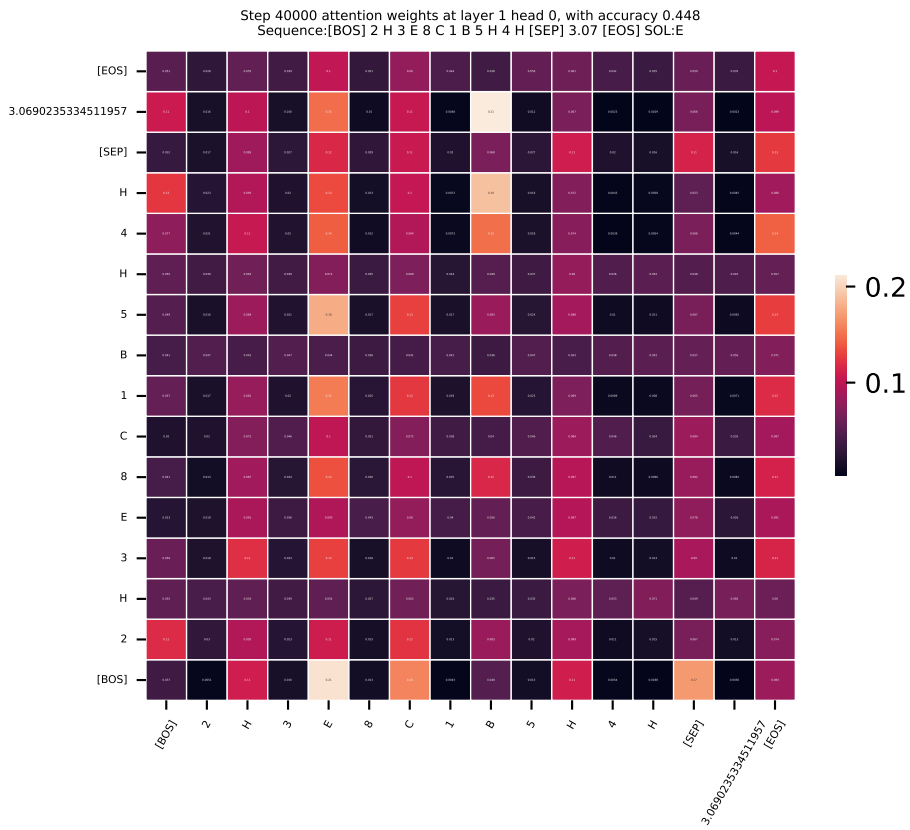

Step 40000 attention weights at layer 0 head 3, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

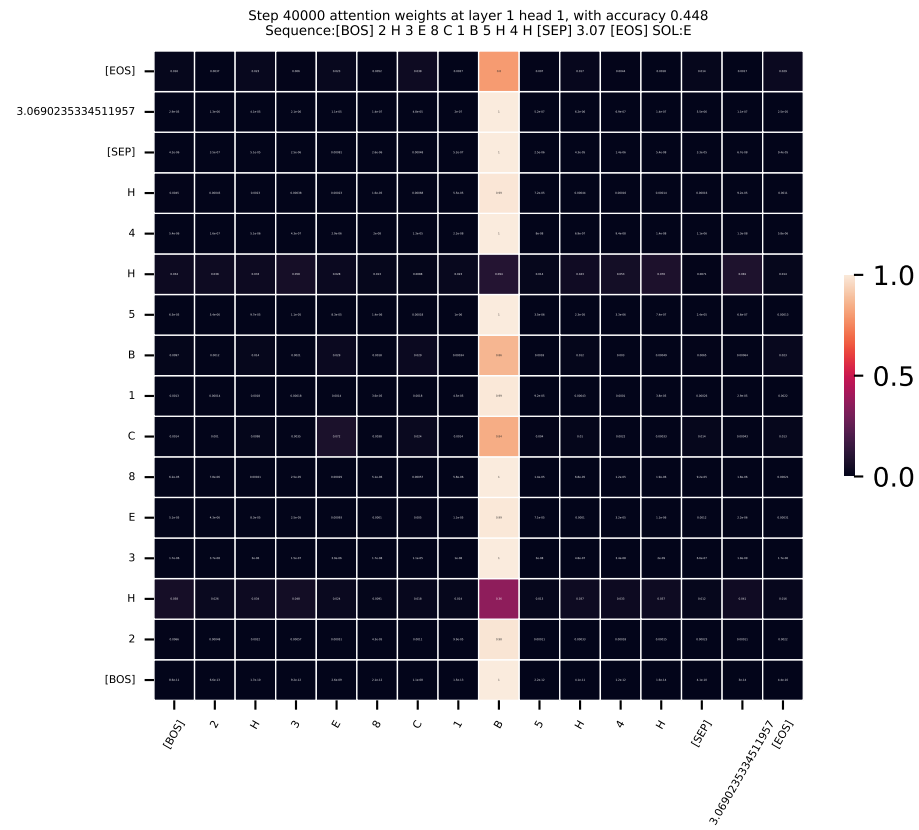

Step 40000 attention weights at layer 1 head 0, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.2
0.1

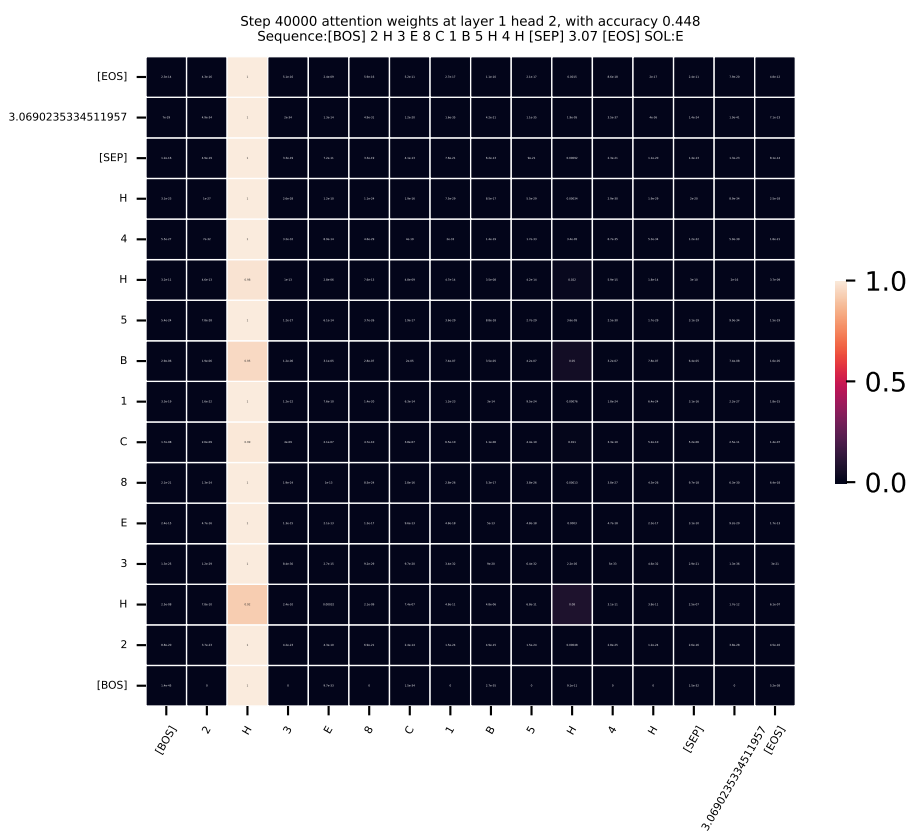

Step 40000 attention weights at layer 1 head 1, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

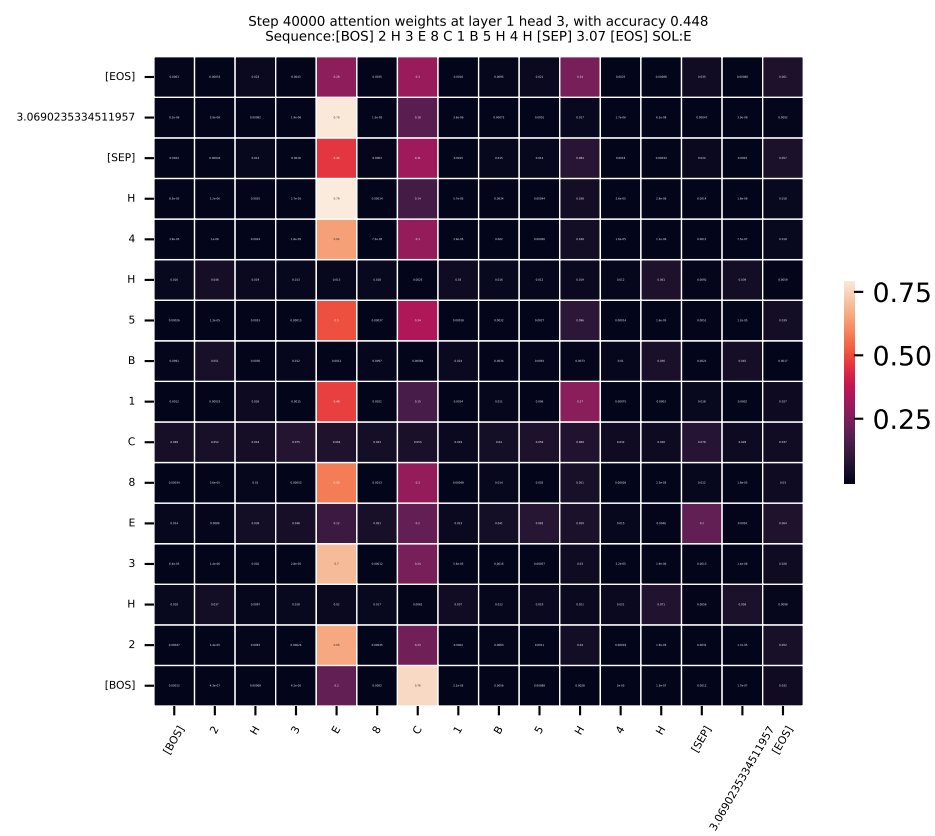

Step 40000 attention weights at layer 1 head 2, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

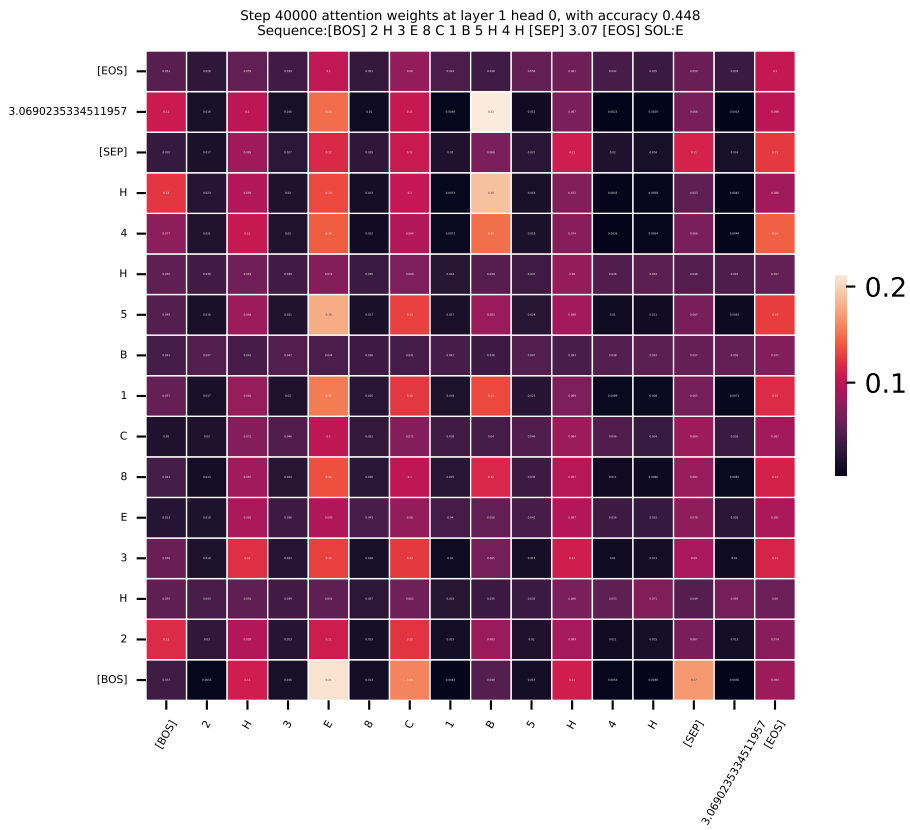

Step 40000 attention weights at layer 1 head 3, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

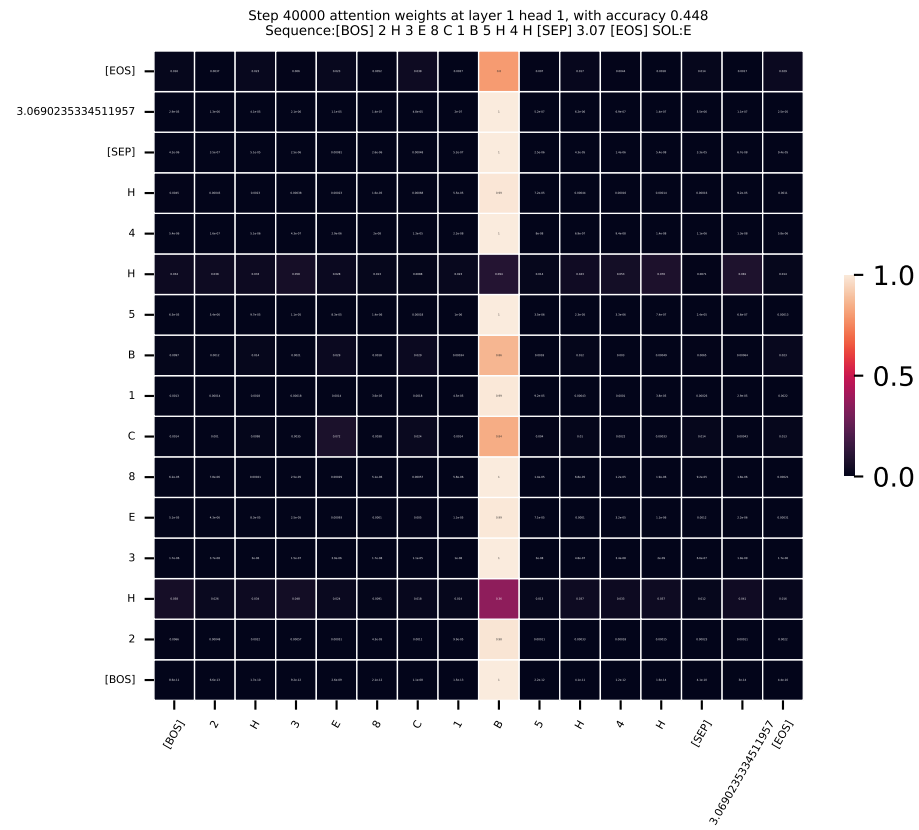

Step 40000 attention weights at layer 1 head 0, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.2
0.1

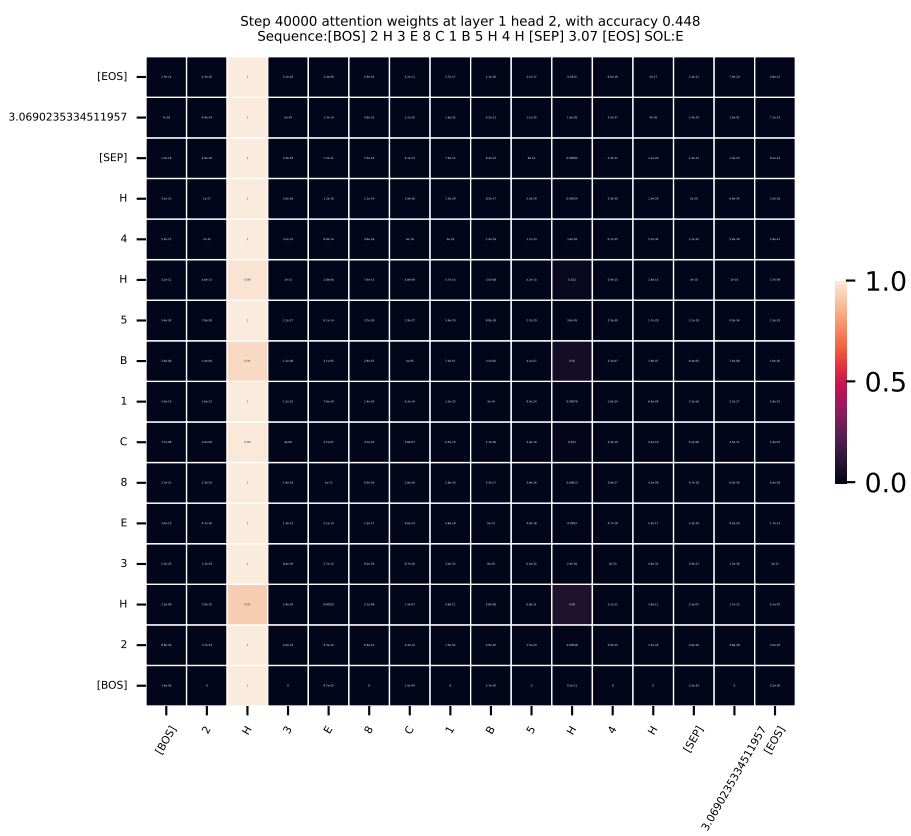

Step 40000 attention weights at layer 1 head 1, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

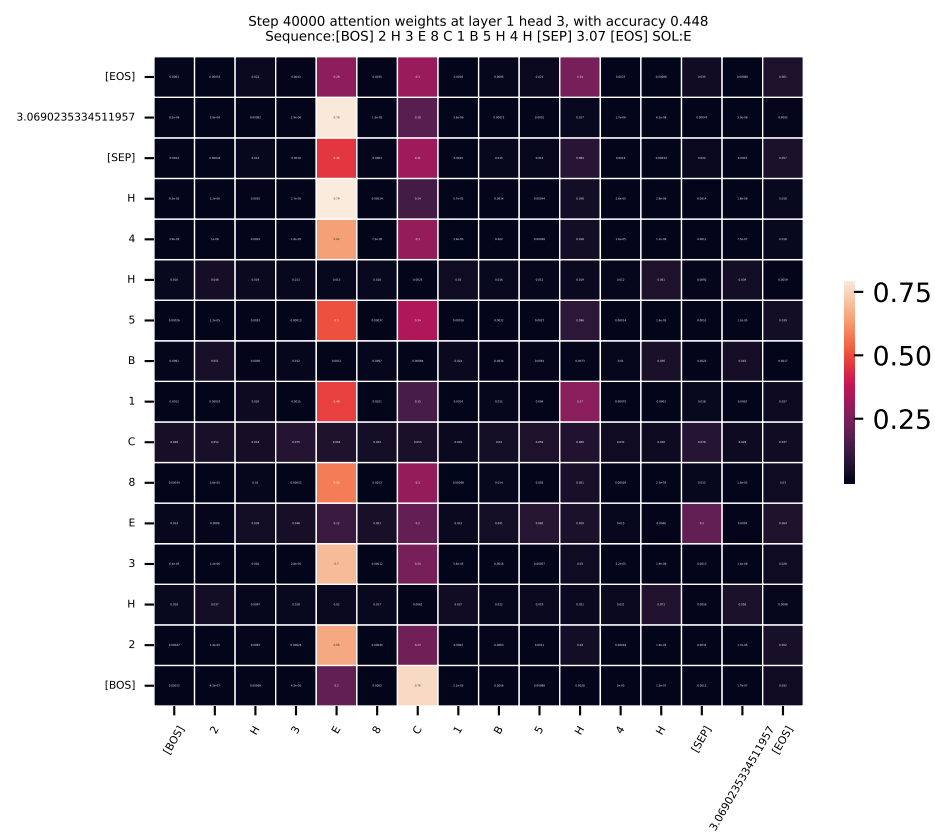

Step 40000 attention weights at layer 1 head 2, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

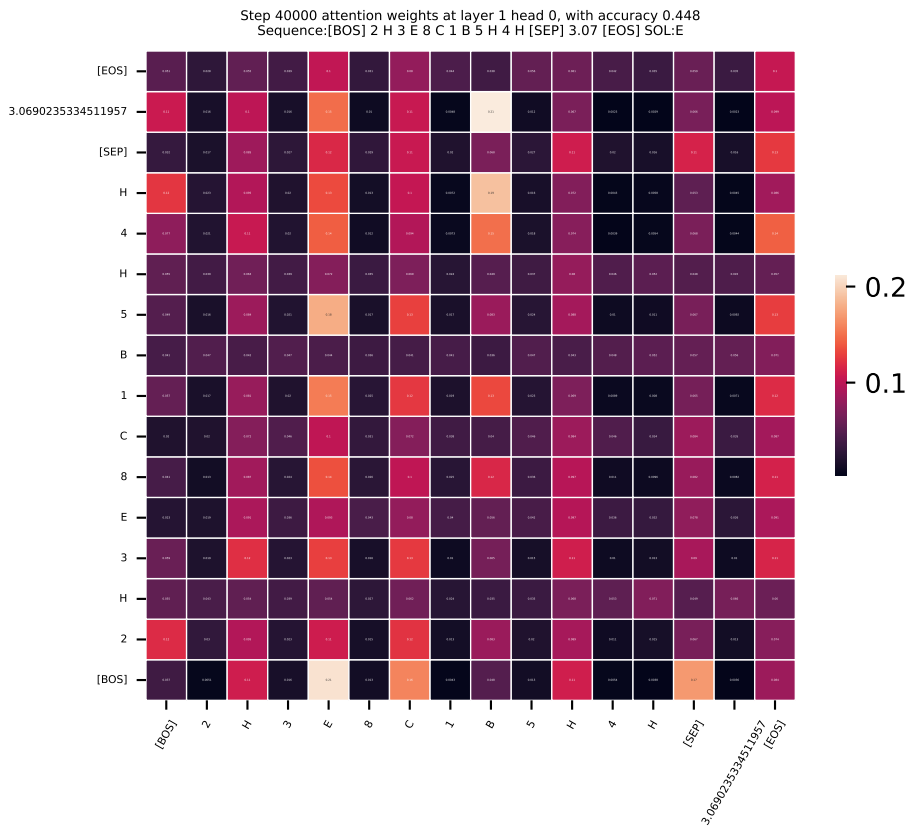

Step 40000 attention weights at layer 1 head 3, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

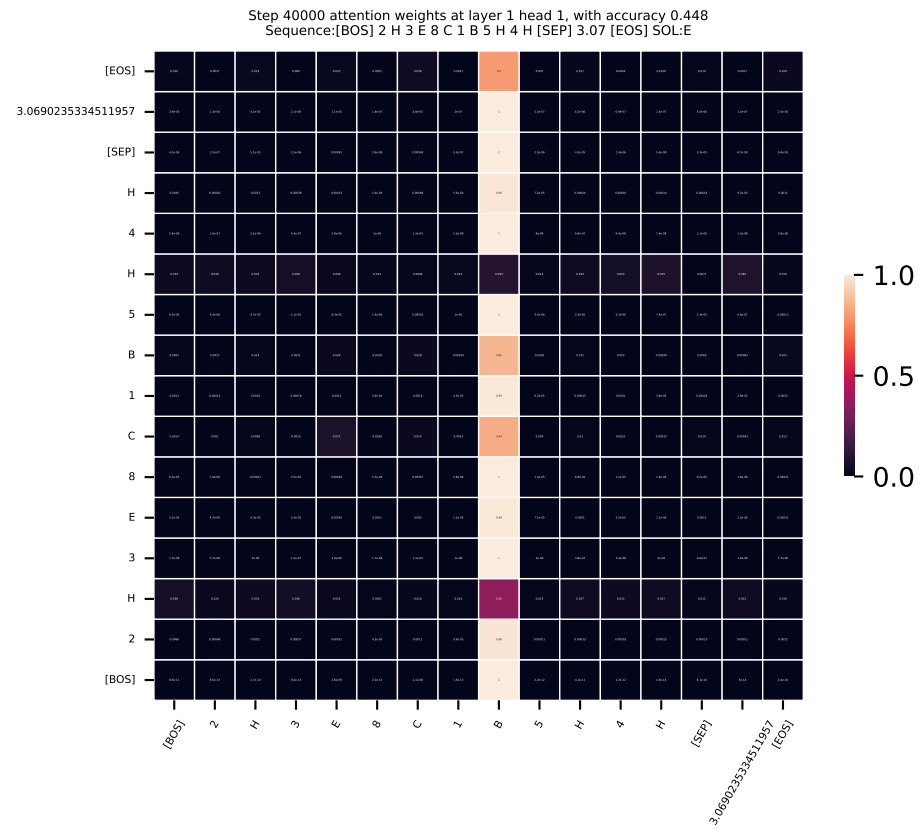

Step 40000 attention weights at layer 1 head 0, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.2
0.1

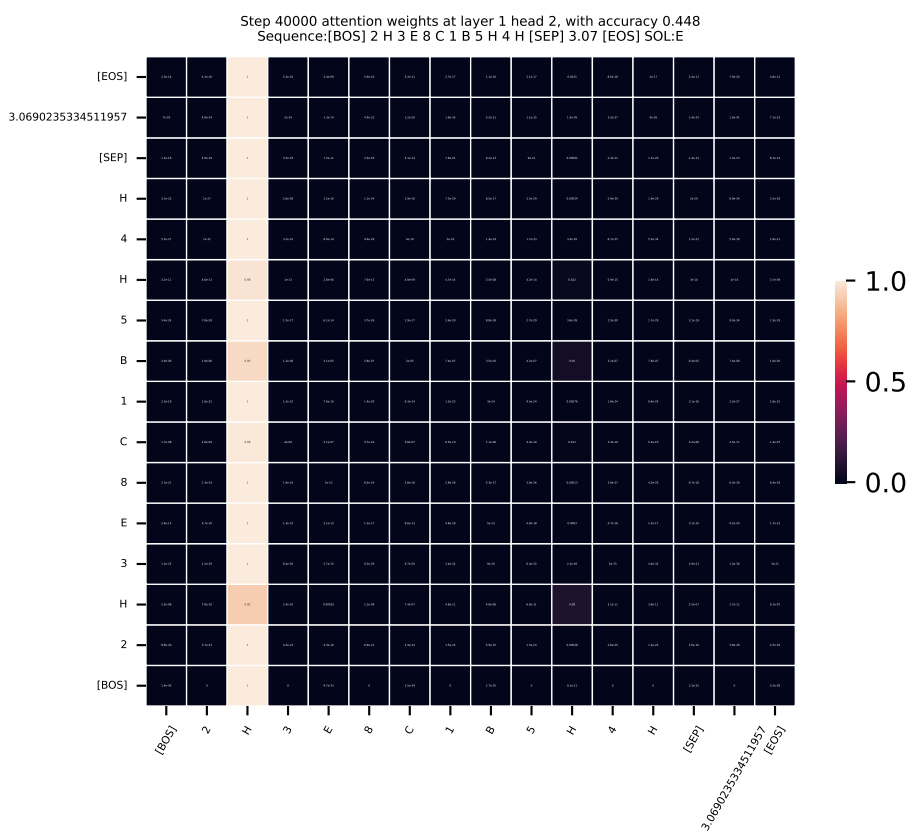

Step 40000 attention weights at layer 1 head 1, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

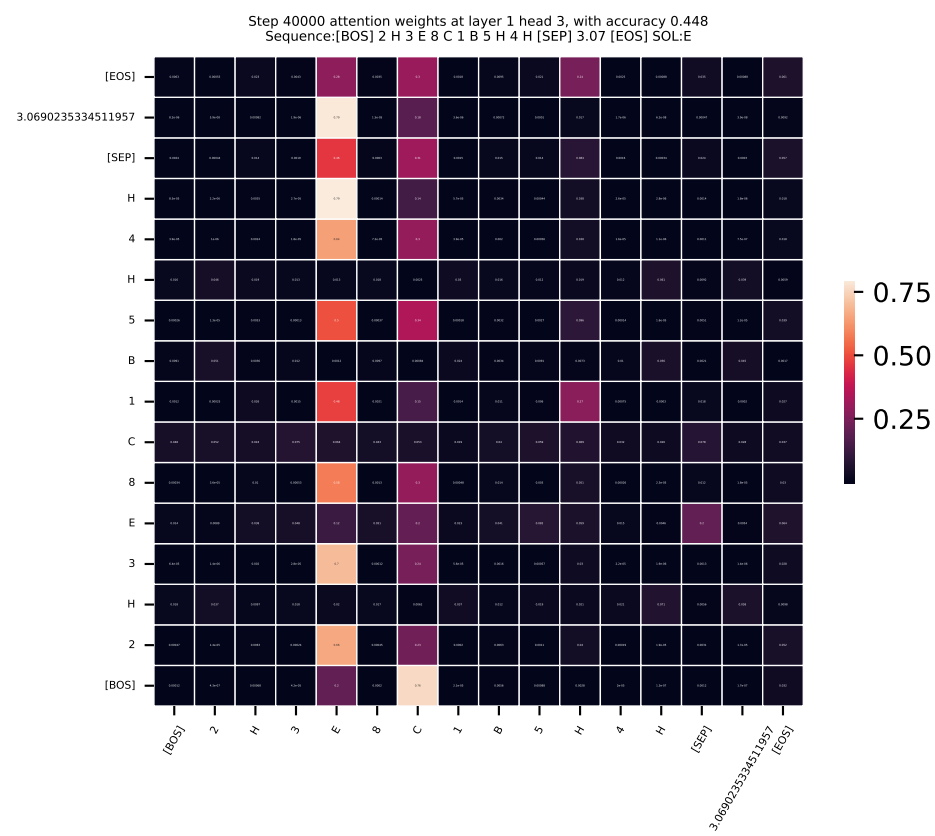

Step 40000 attention weights at layer 1 head 2, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

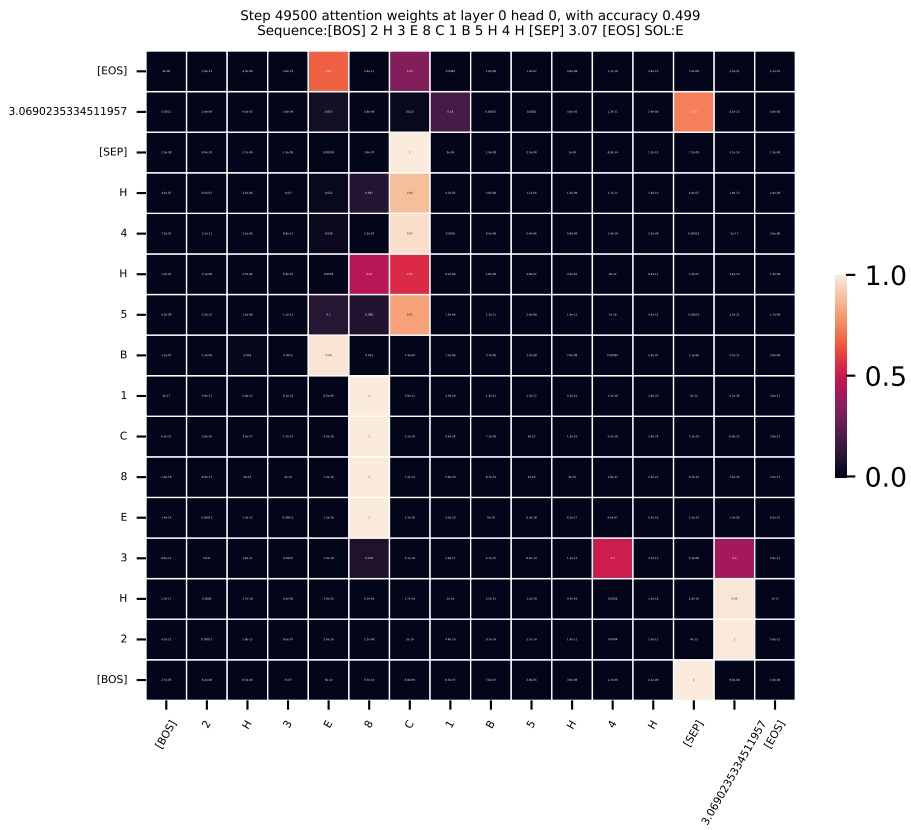

Step 40000 attention weights at layer 1 head 3, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

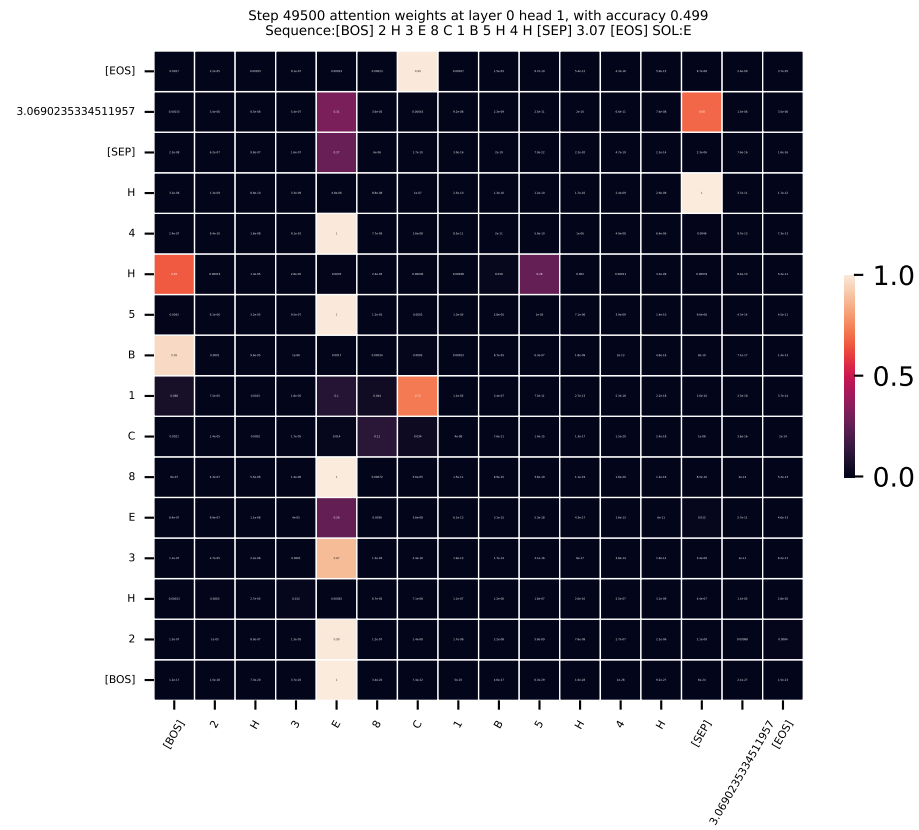

Step 40000 attention weights at layer 1 head 0, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.2
0.1

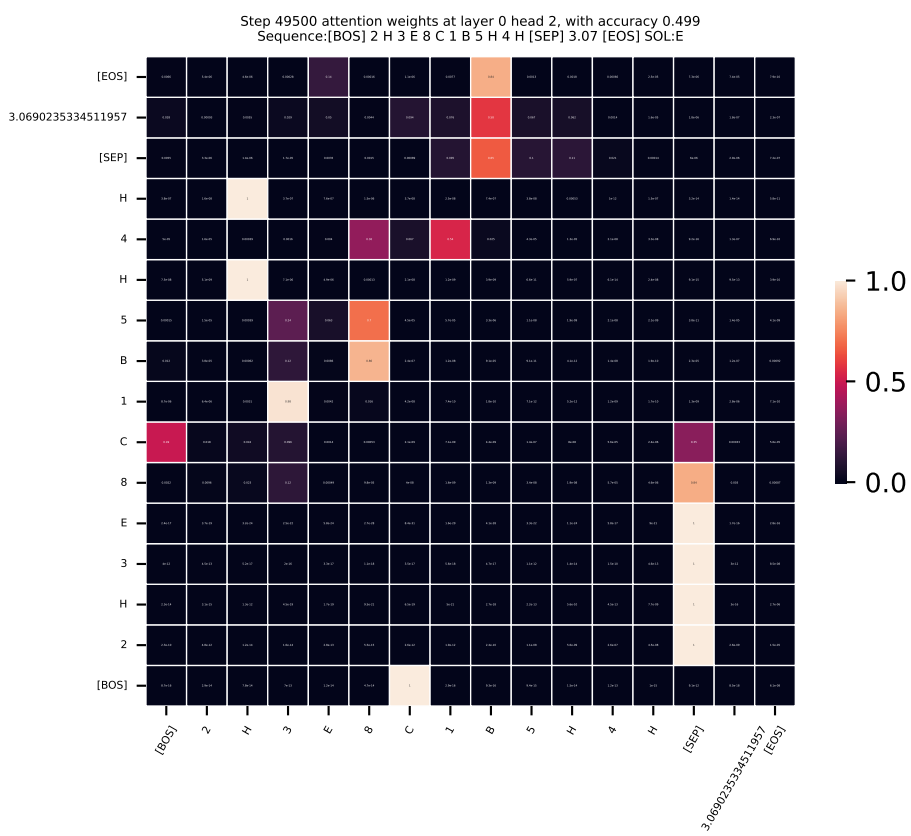

Step 40000 attention weights at layer 1 head 1, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

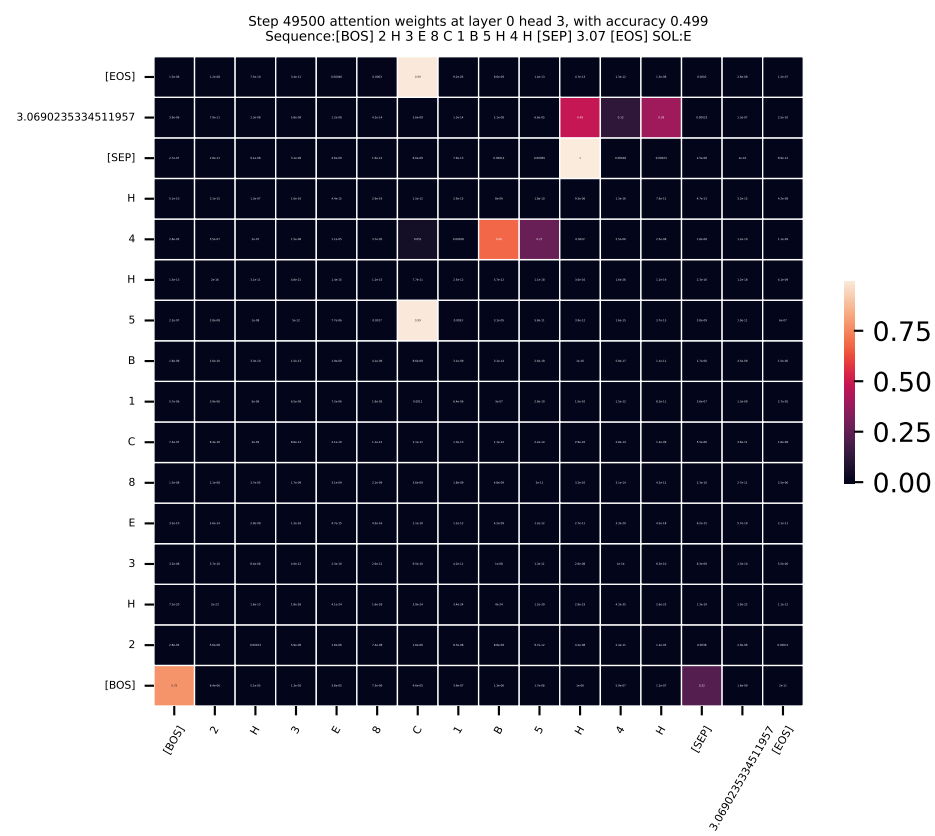

Step 40000 attention weights at layer 1 head 2, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

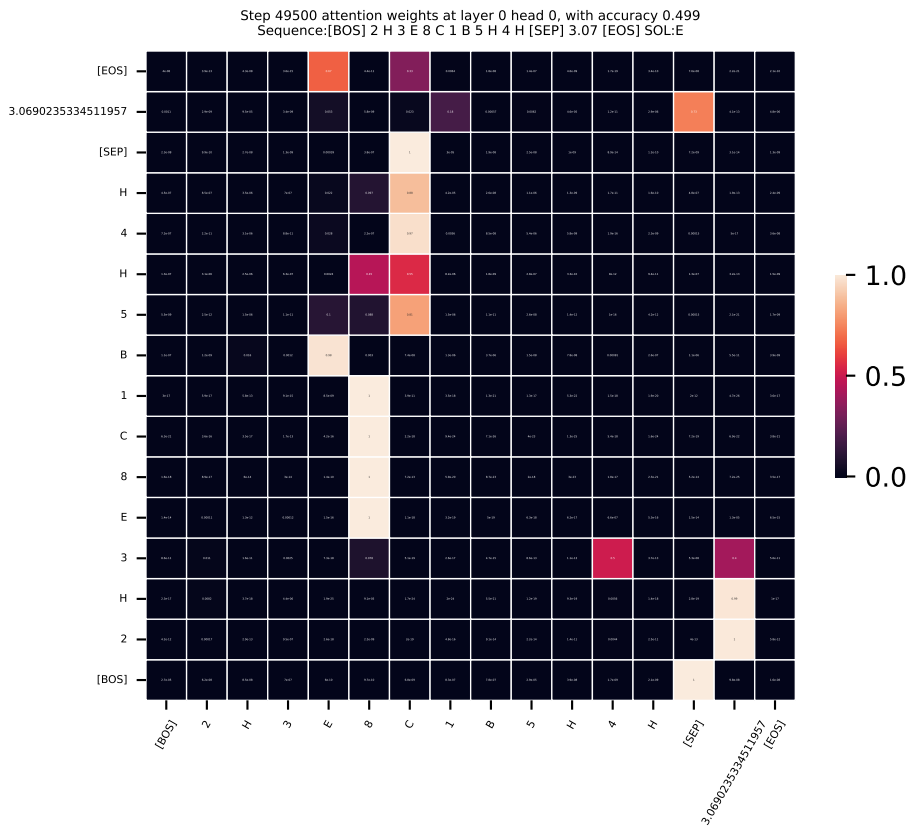

Step 40000 attention weights at layer 1 head 3, with accuracy 0.448
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

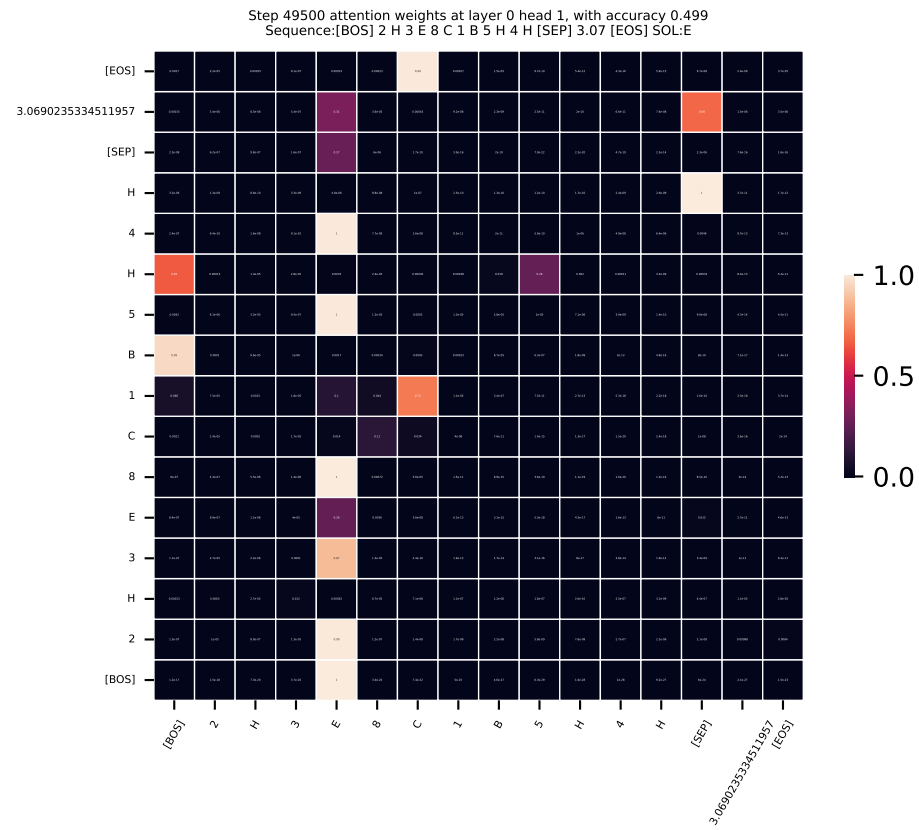

Step 49500 attention weights at layer 0 head 0, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

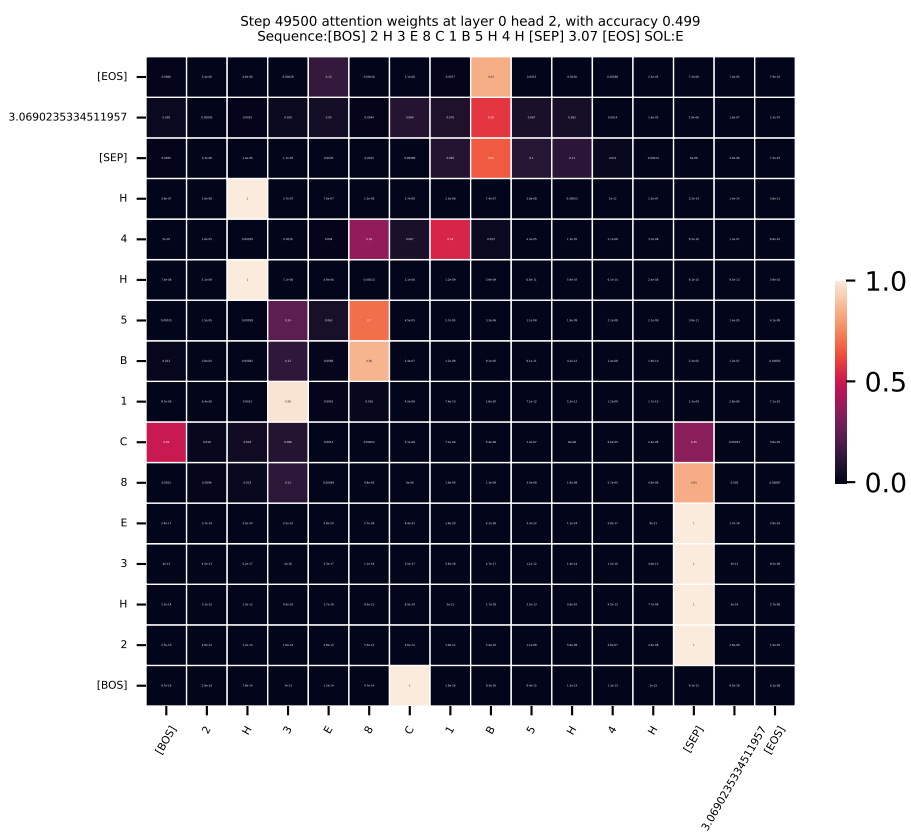

Step 49500 attention weights at layer 0 head 1, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

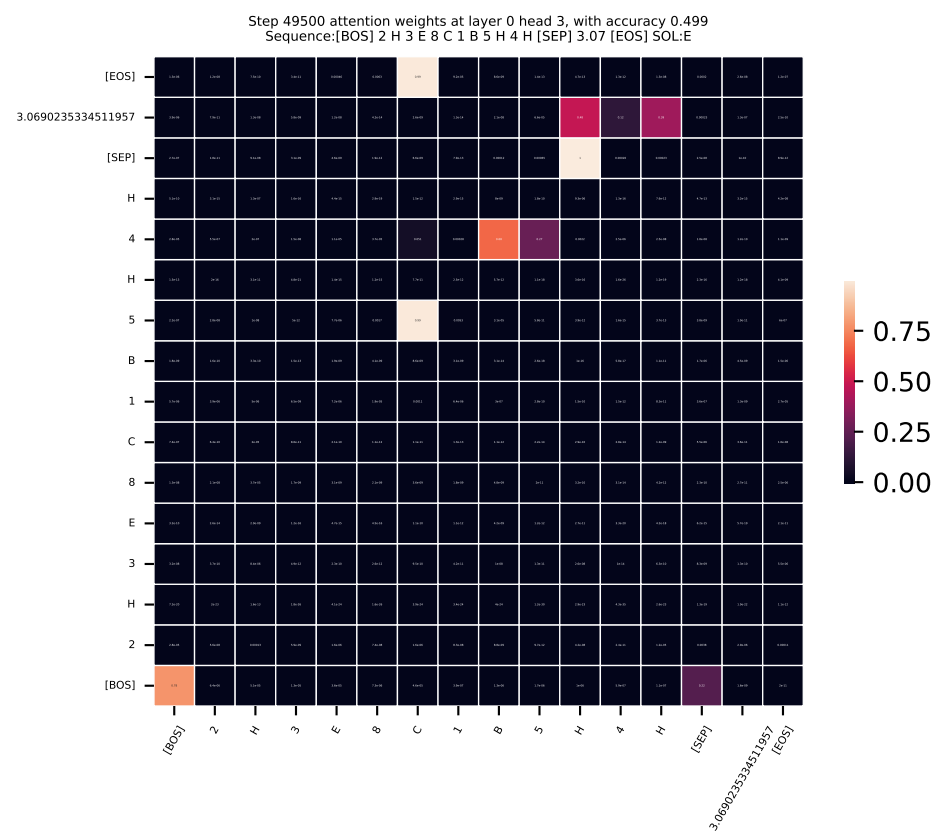

Step 49500 attention weights at layer 0 head 2, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

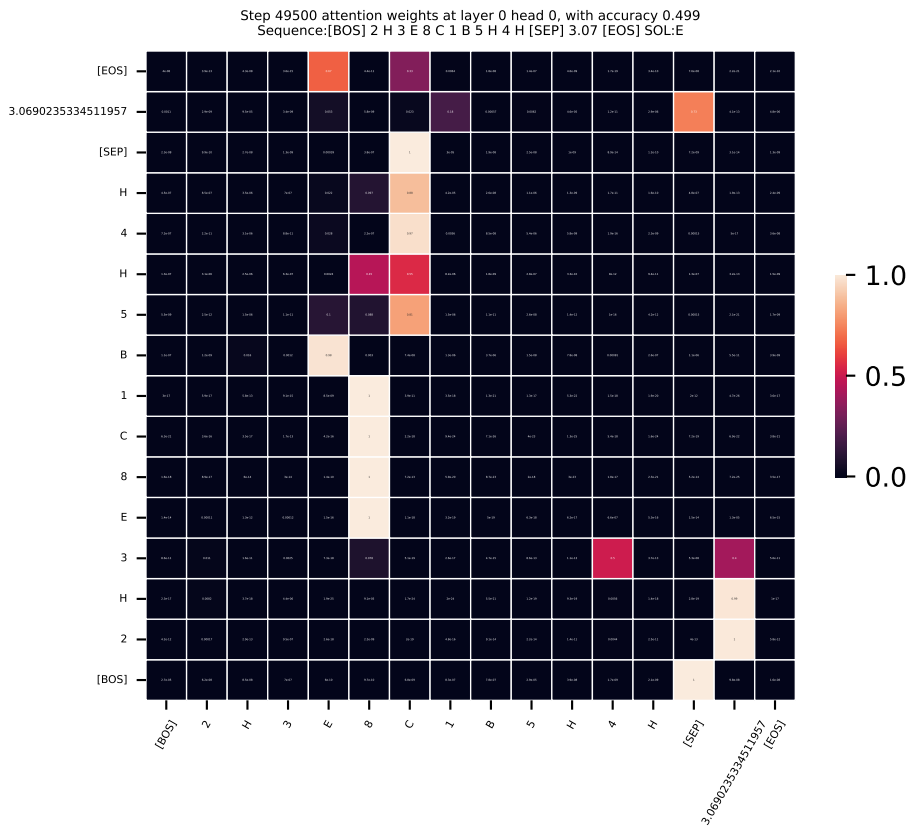

Step 49500 attention weights at layer 0 head 3, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.75
0.50
0.25
0.00

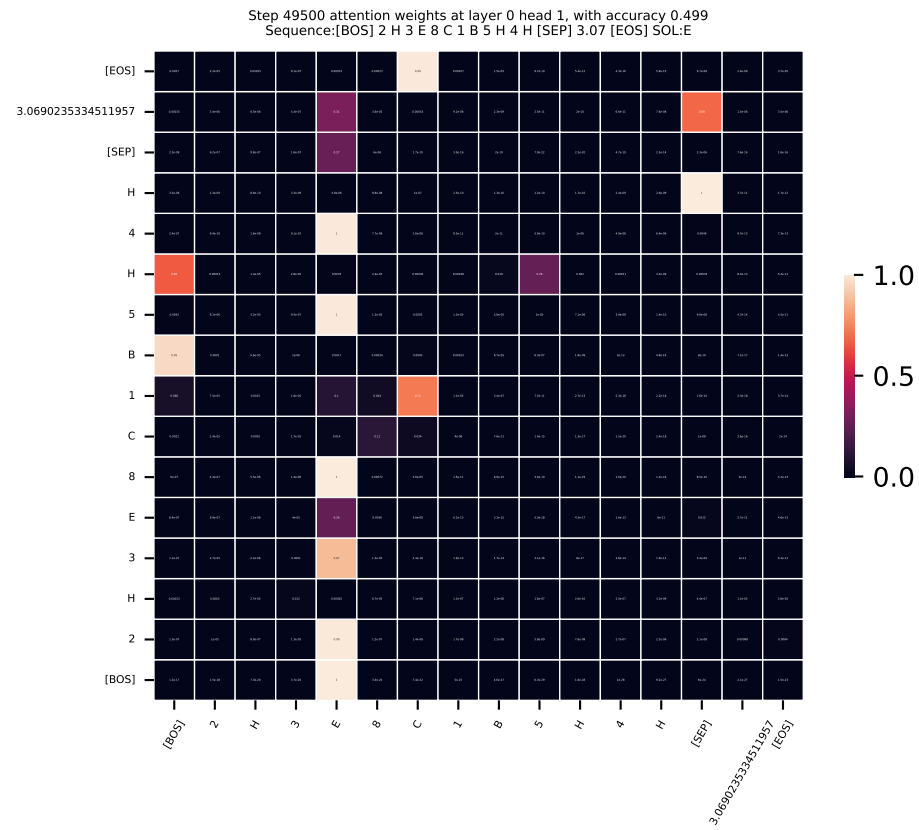

Step 49500 attention weights at layer 0 head 0, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

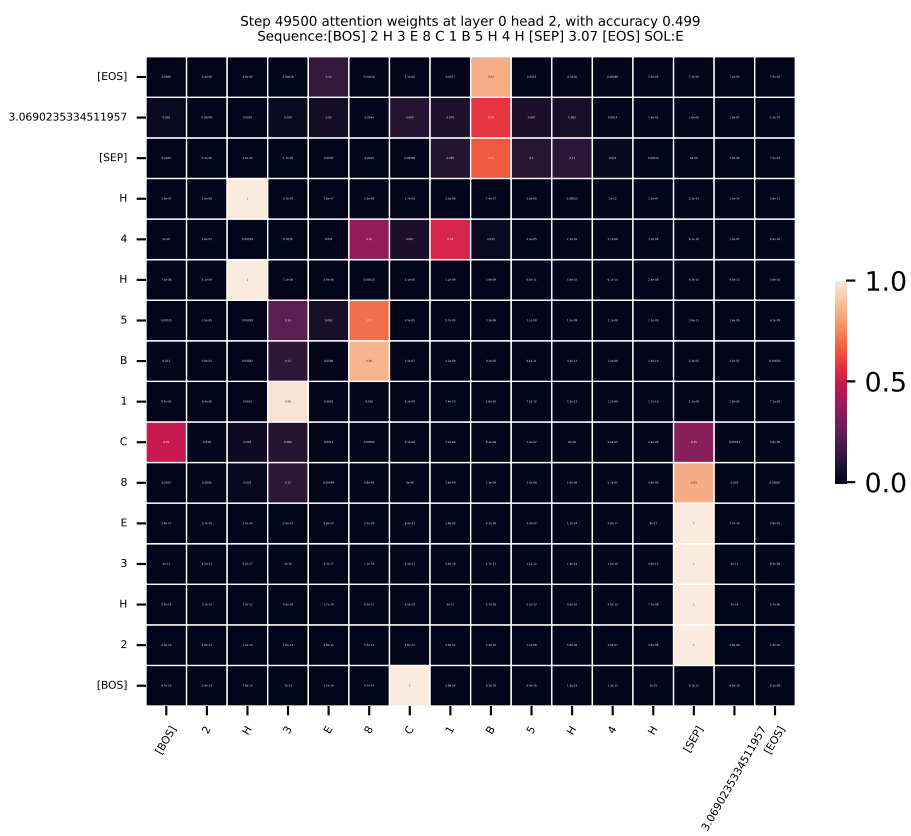

Step 49500 attention weights at layer 0 head 1, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

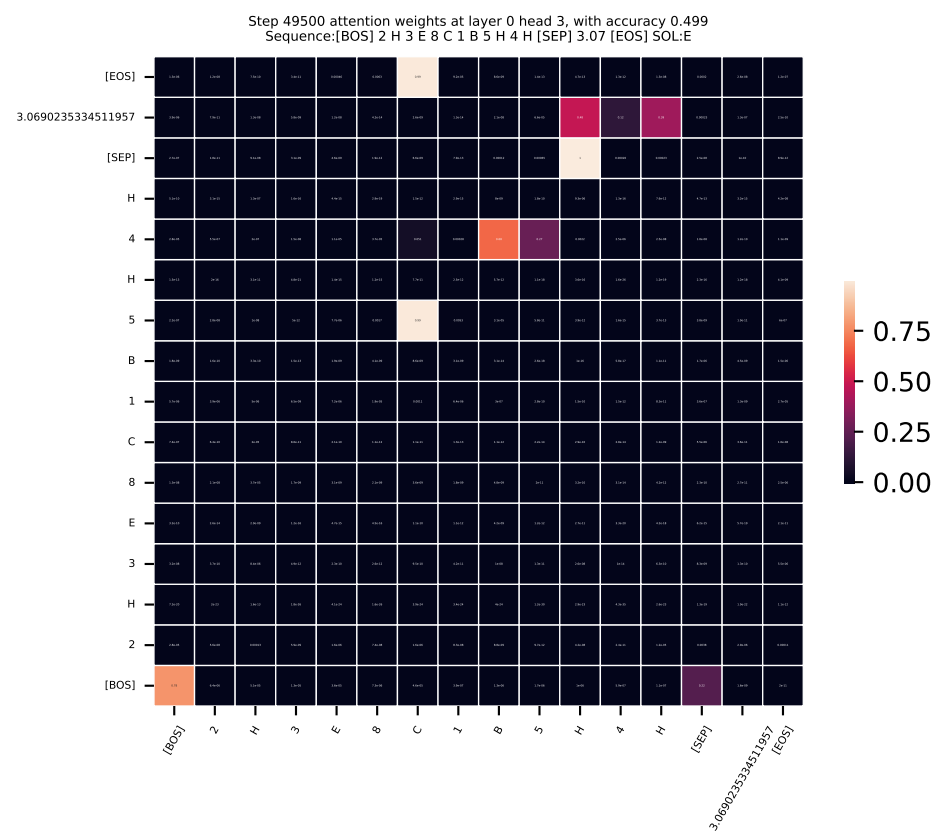

Step 49500 attention weights at layer 0 head 2, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

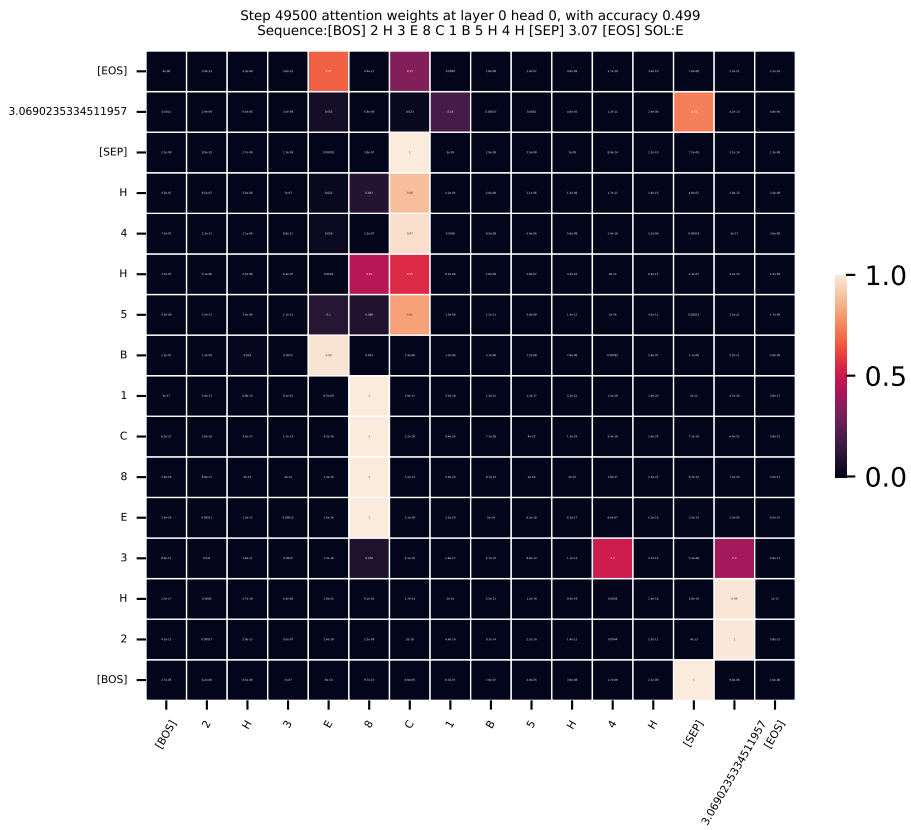

Step 49500 attention weights at layer 0 head 3, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.75
0.50
0.25
0.00

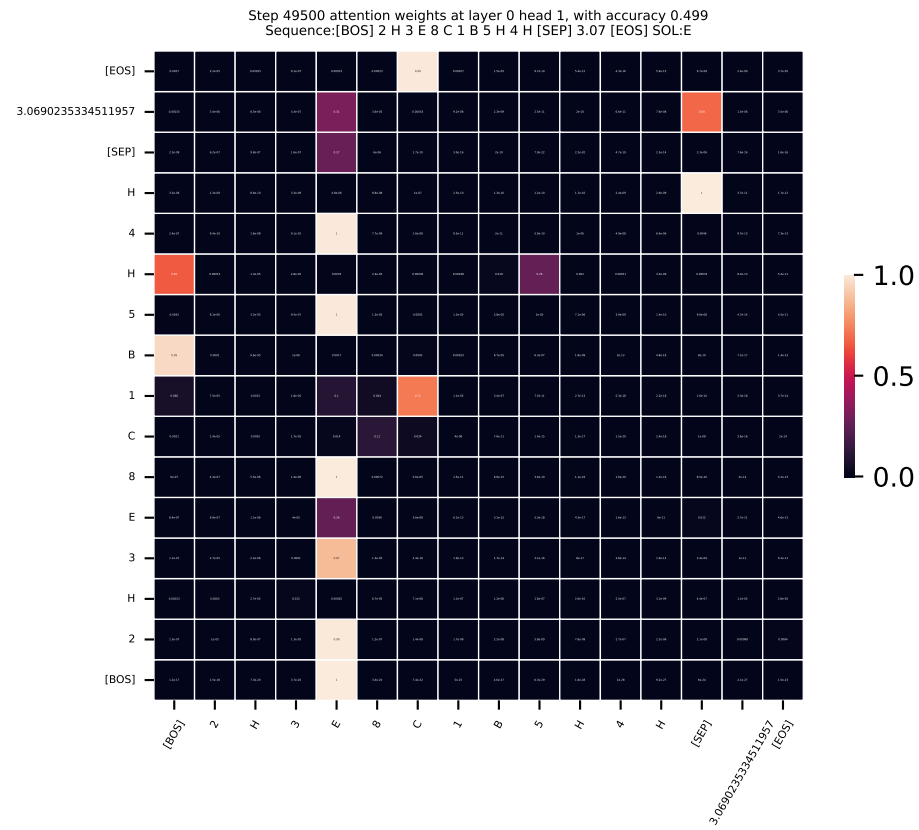

Step 49500 attention weights at layer 0 head 0, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

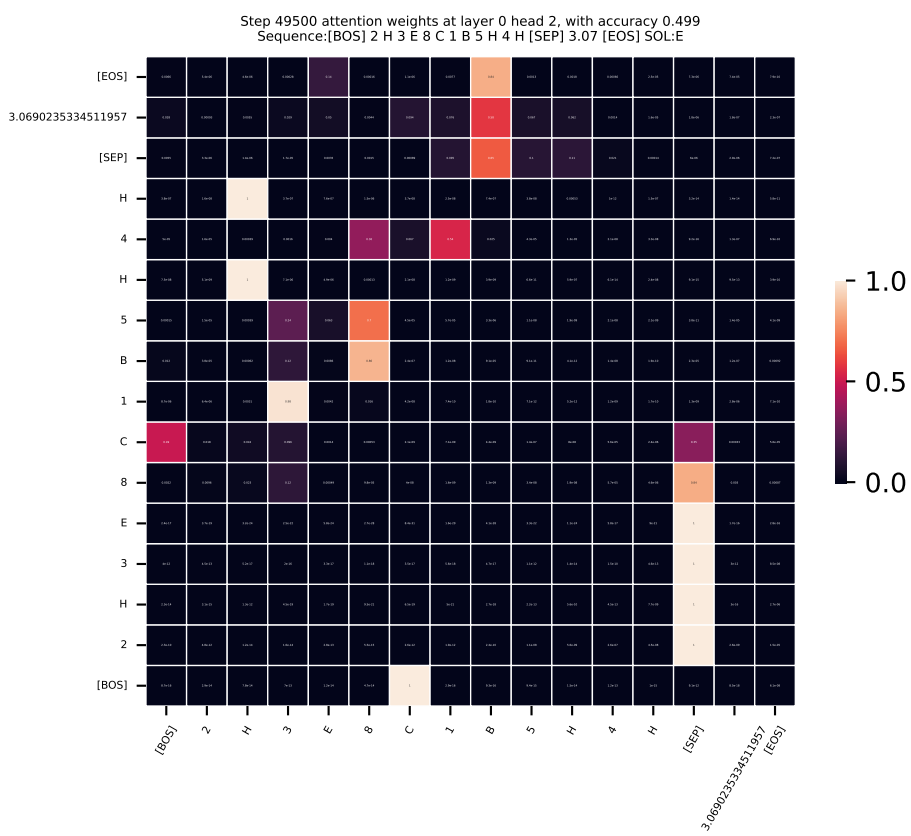

Step 49500 attention weights at layer 0 head 1, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

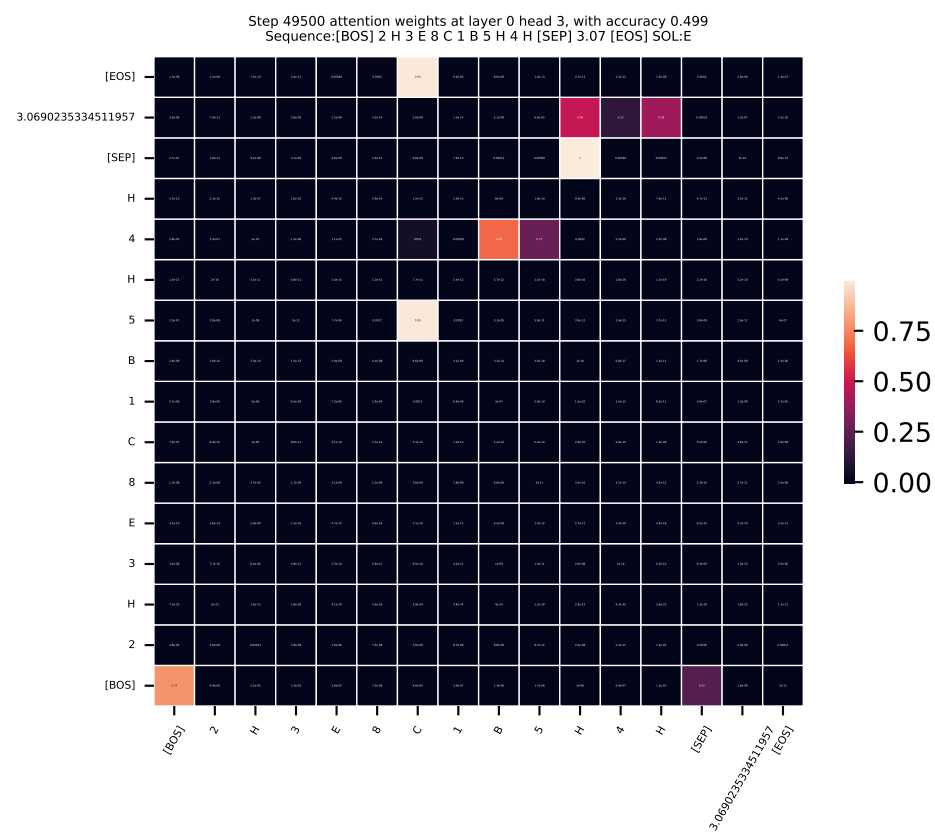

Step 49500 attention weights at layer 0 head 2, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

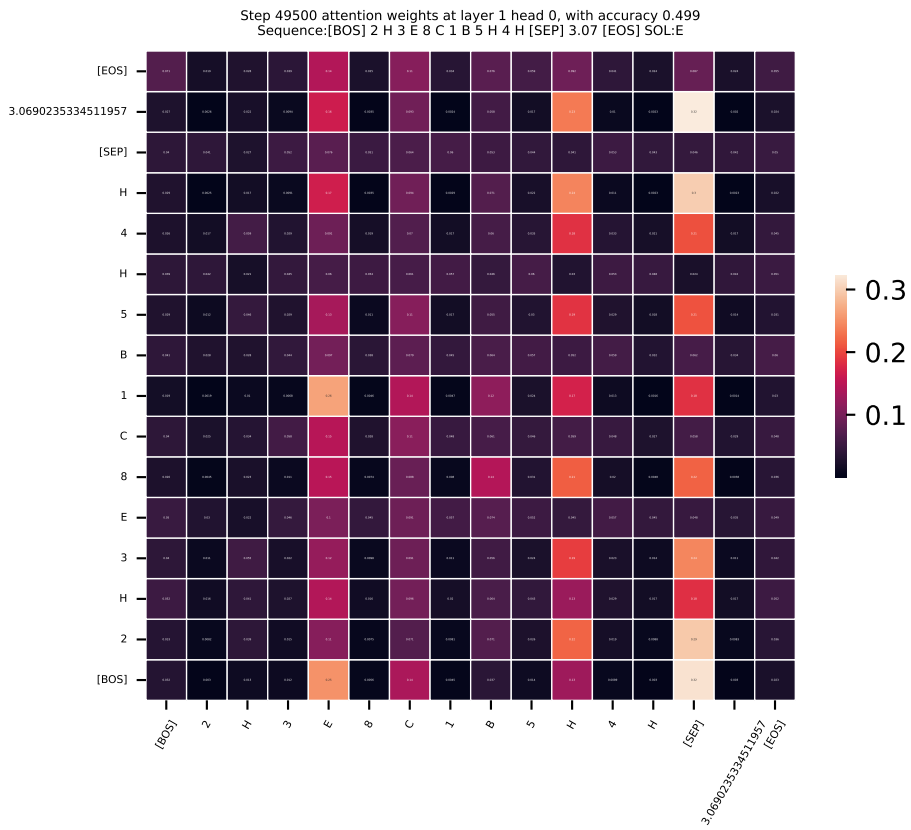

Step 49500 attention weights at layer 0 head 3, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.75
0.50
0.25
0.00

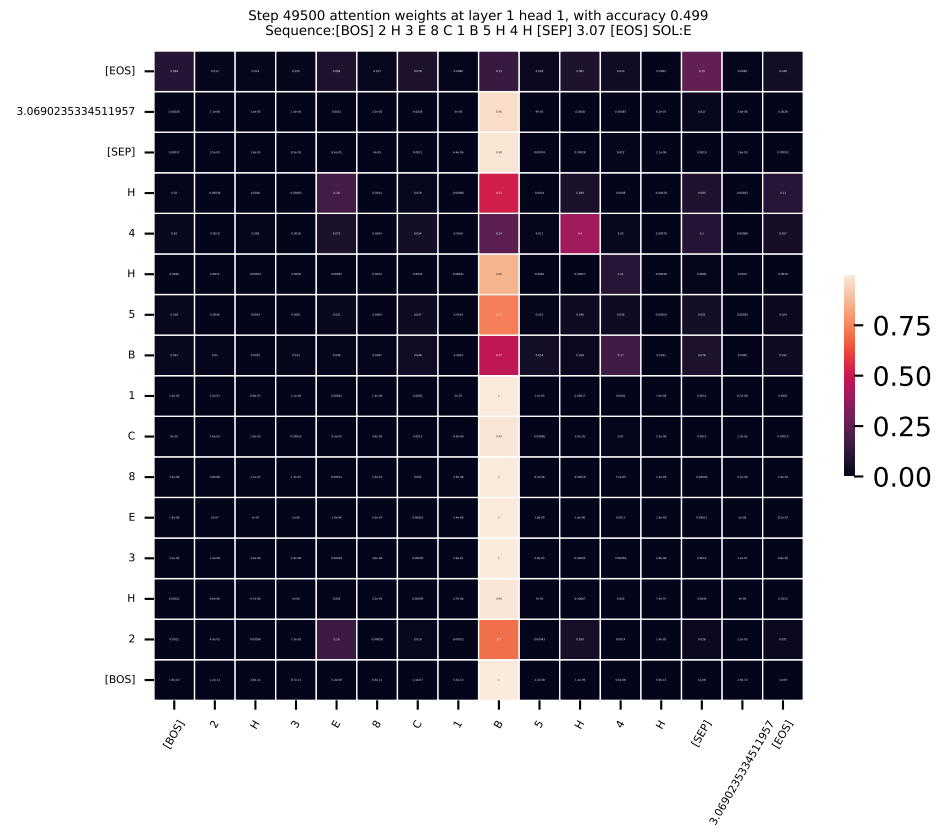

Step 49500 attention weights at layer 0 head 0, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

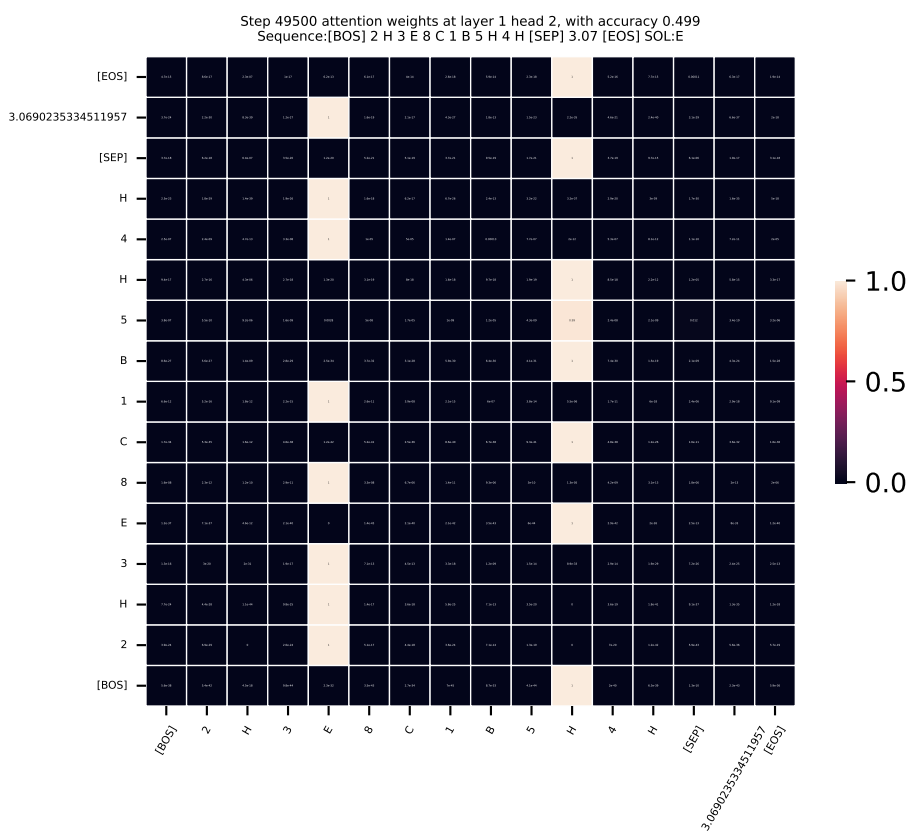

Step 49500 attention weights at layer 0 head 1, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

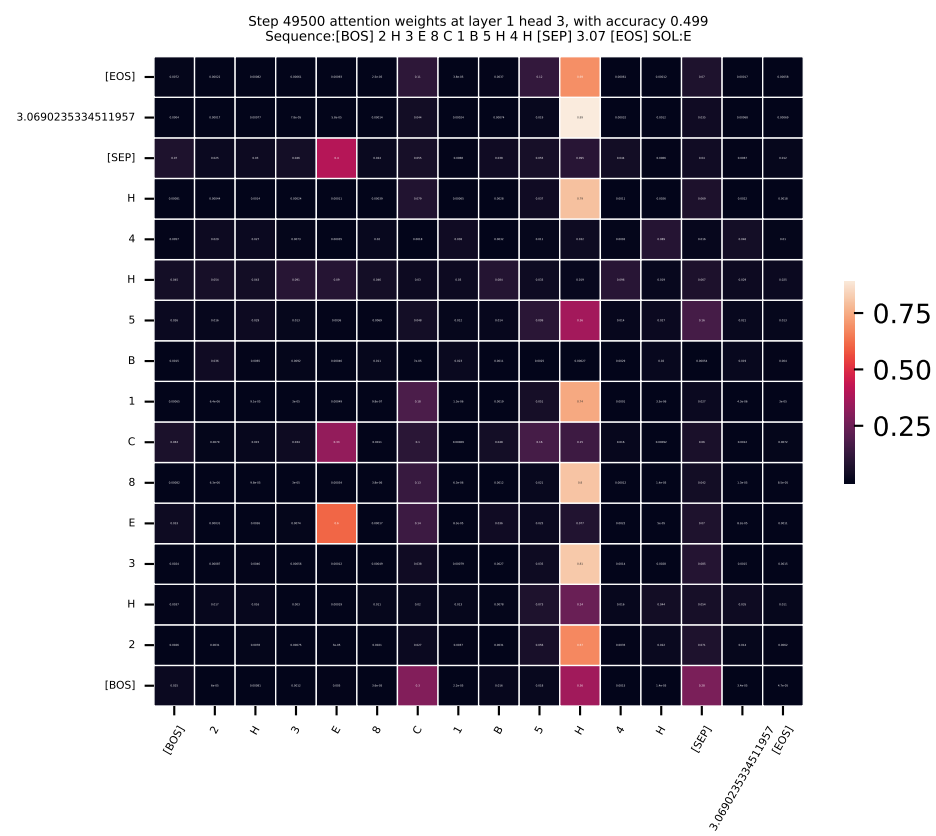

Step 49500 attention weights at layer 0 head 2, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

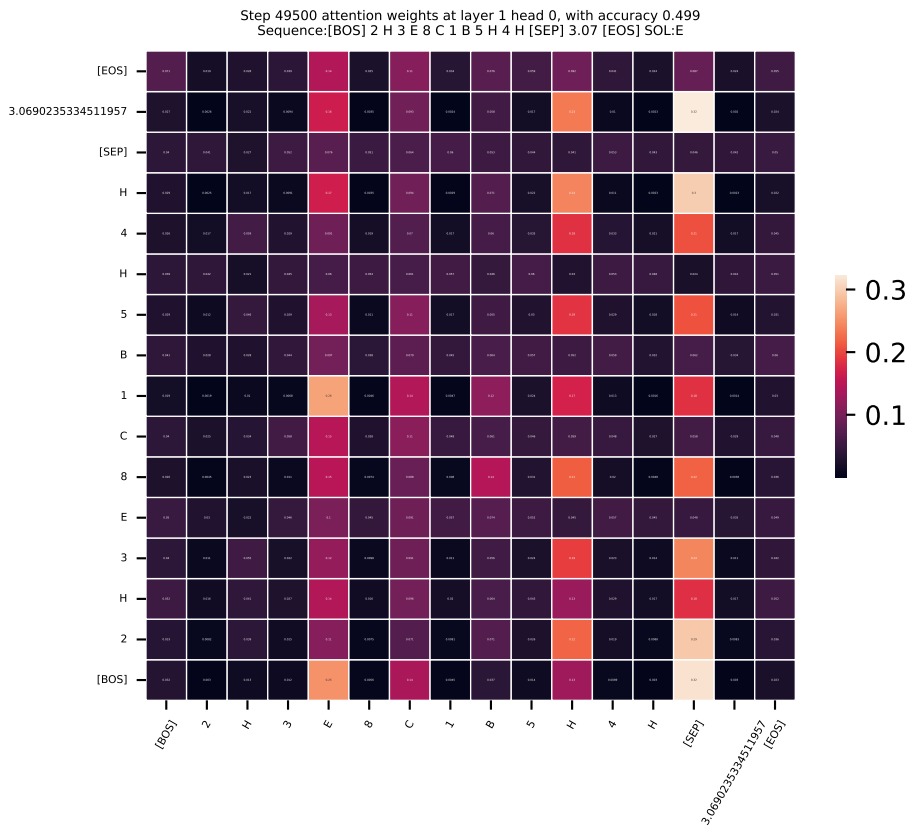

Step 49500 attention weights at layer 0 head 3, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.75
0.50
0.25
0.00

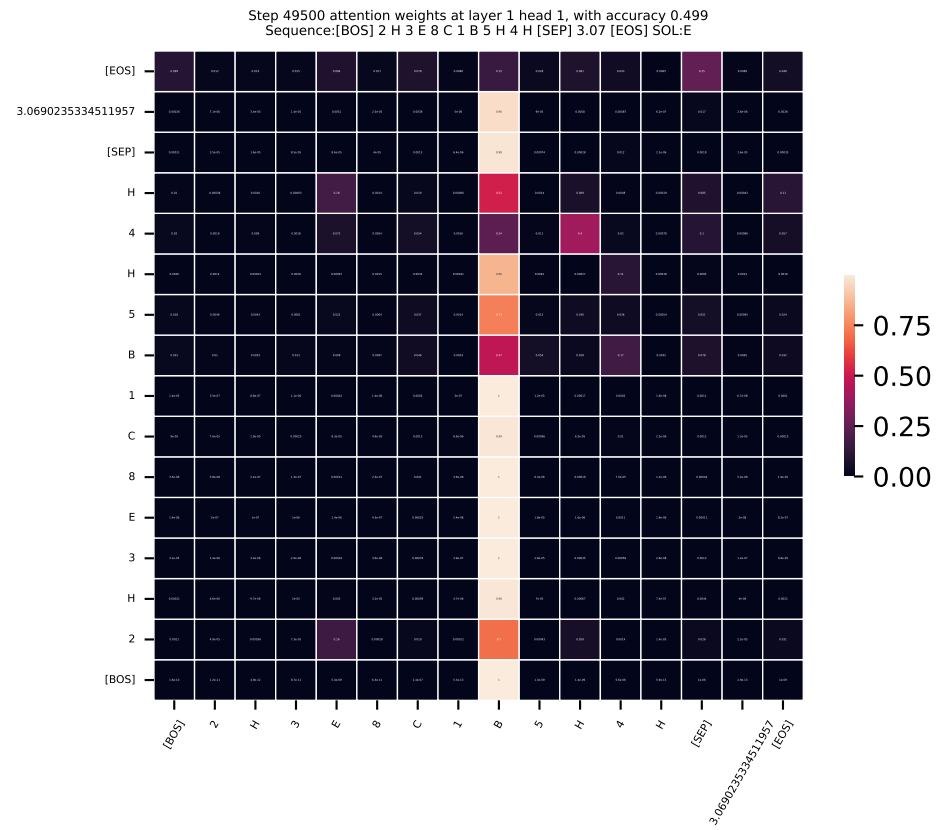

Step 49500 attention weights at layer 1 head 0, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.3
0.2
0.1

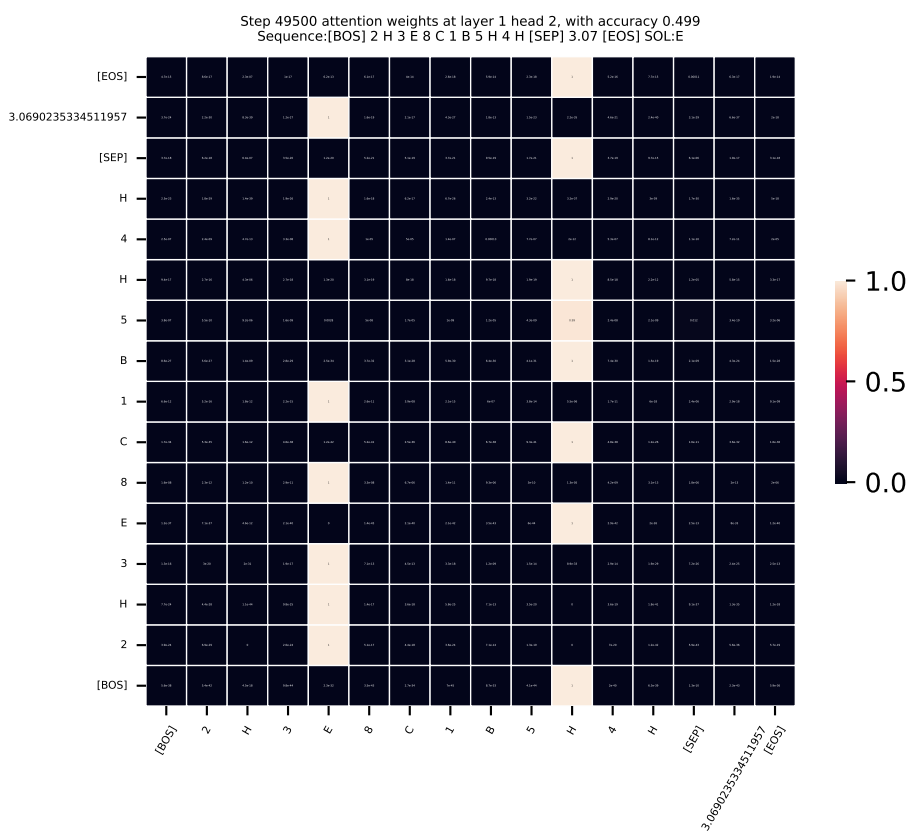

Step 49500 attention weights at layer 1 head 1, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

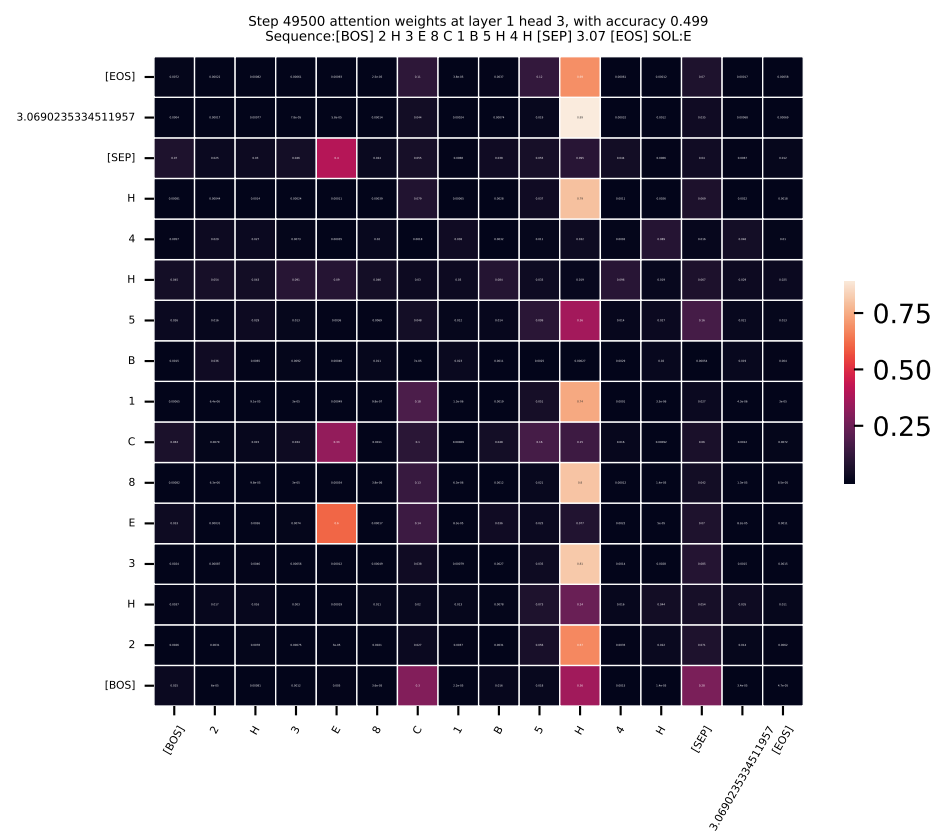

Step 49500 attention weights at layer 1 head 2, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

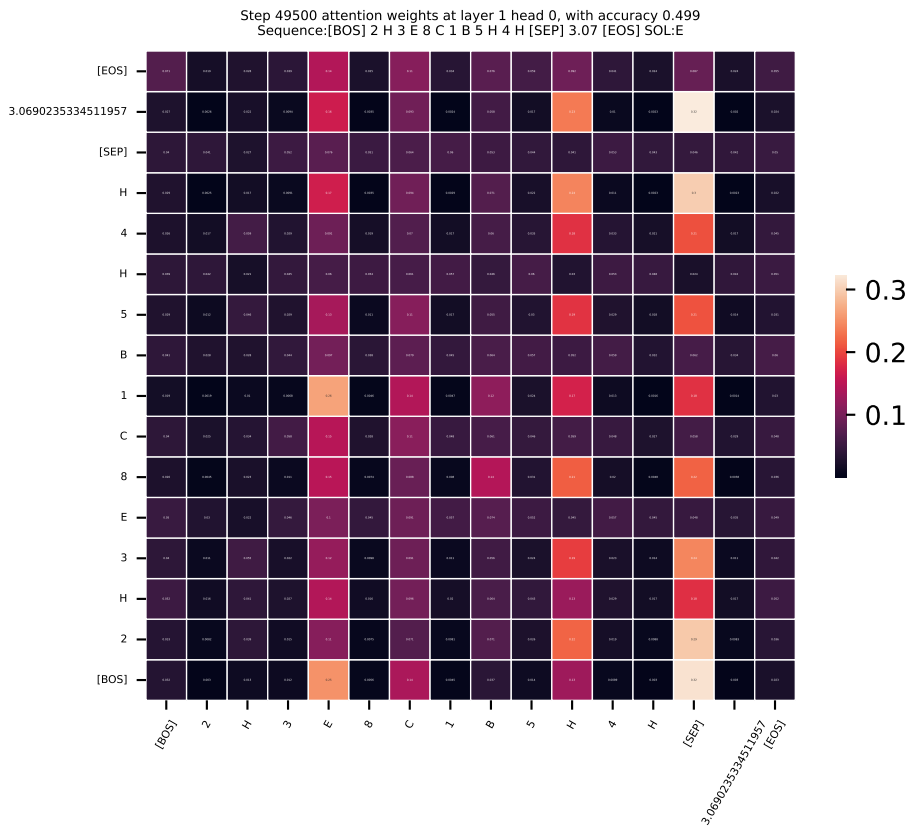

Step 49500 attention weights at layer 1 head 3, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

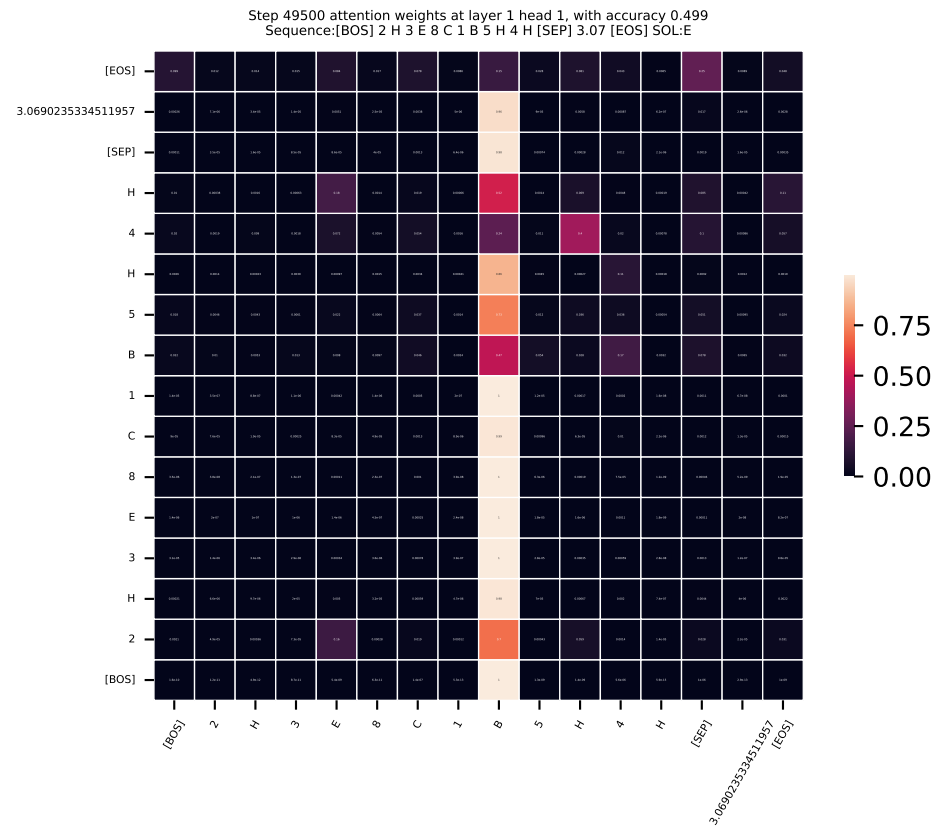

Step 49500 attention weights at layer 1 head 0, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.3
0.2
0.1

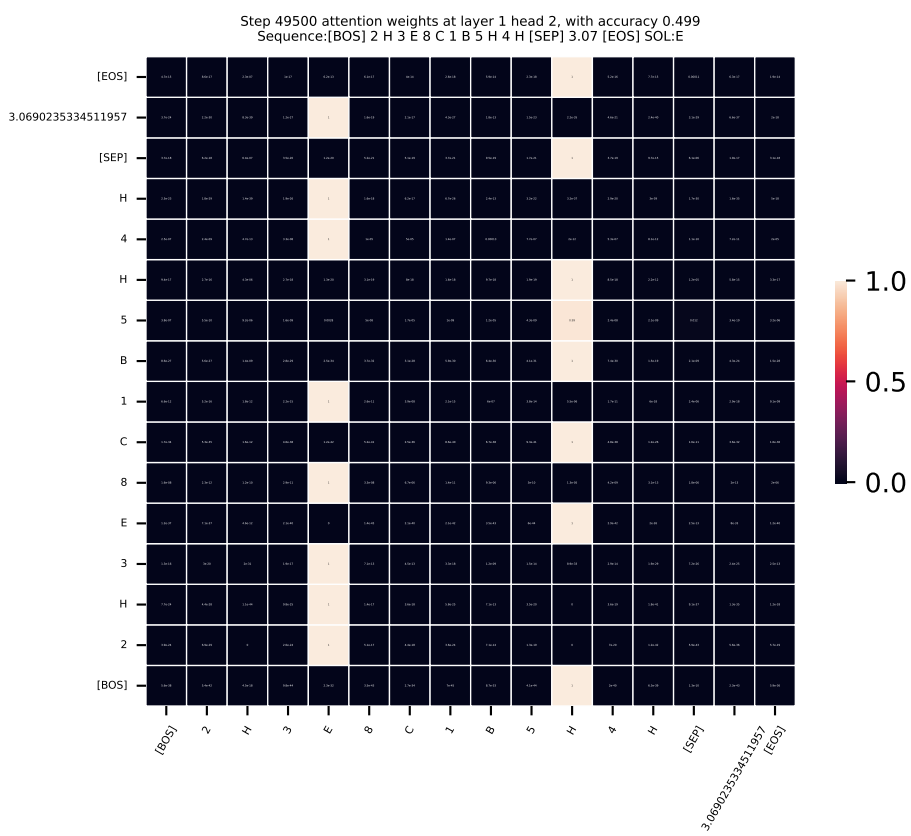

Step 49500 attention weights at layer 1 head 1, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

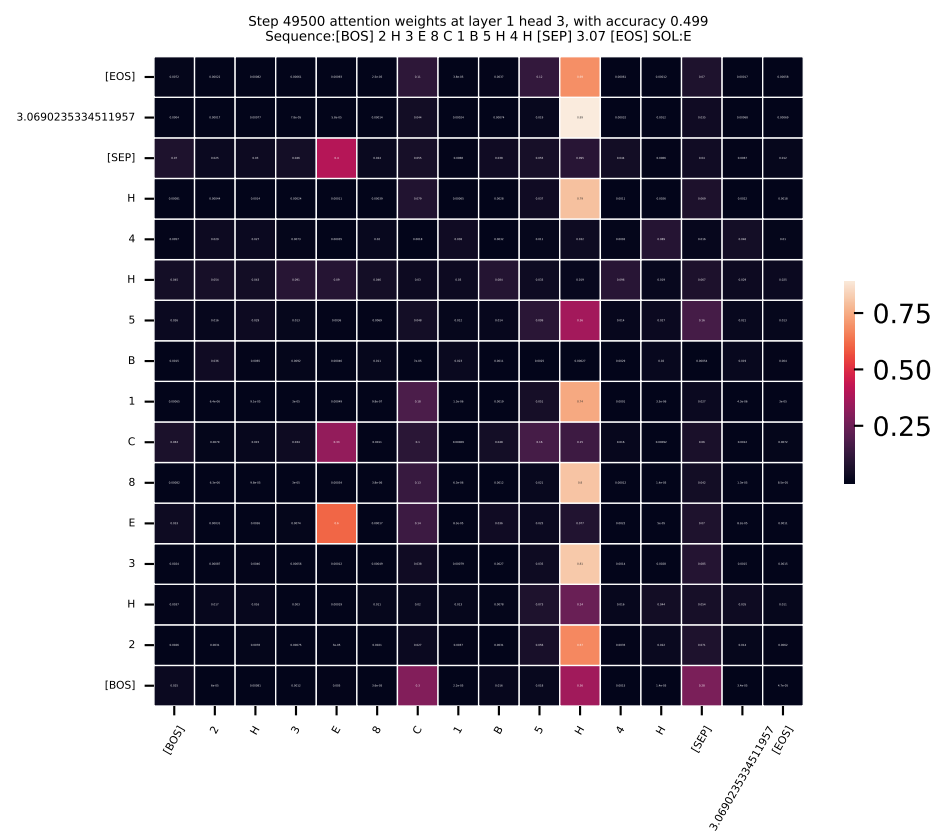

Step 49500 attention weights at layer 1 head 2, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

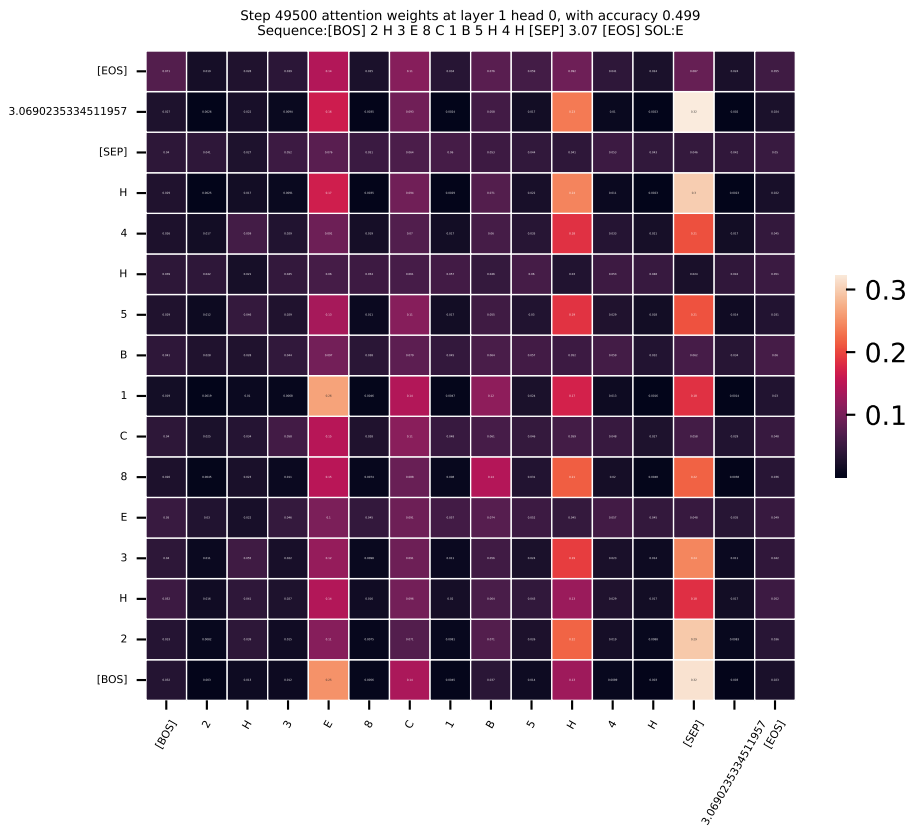

Step 49500 attention weights at layer 1 head 3, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

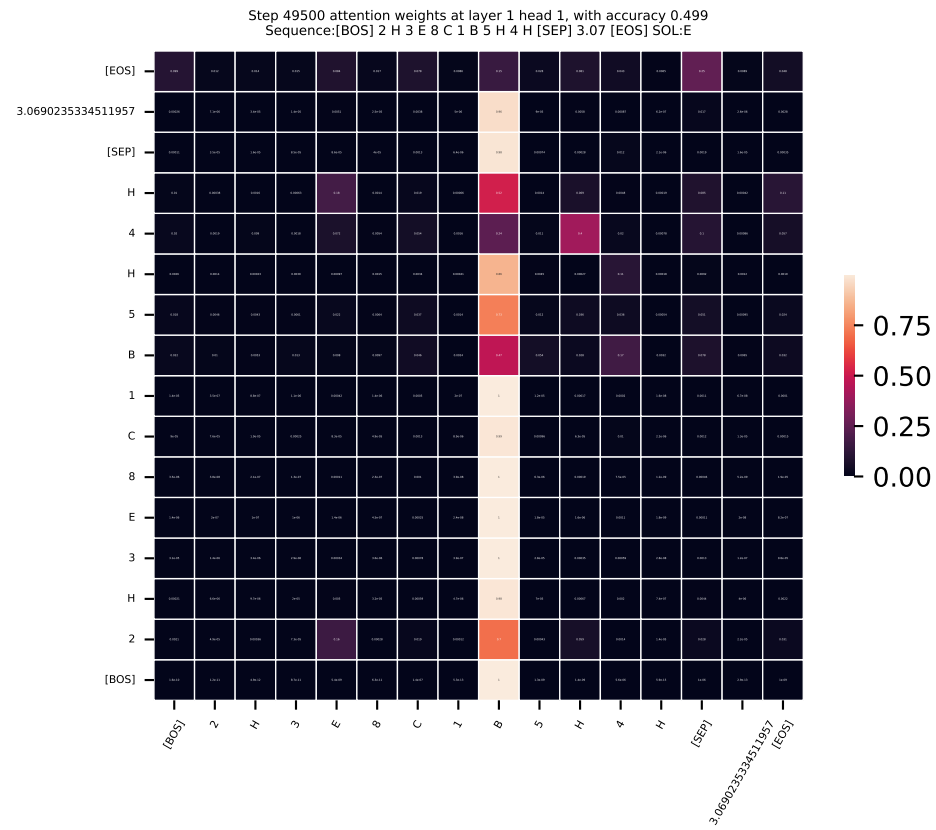

Step 49500 attention weights at layer 1 head 0, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.3
0.2
0.1

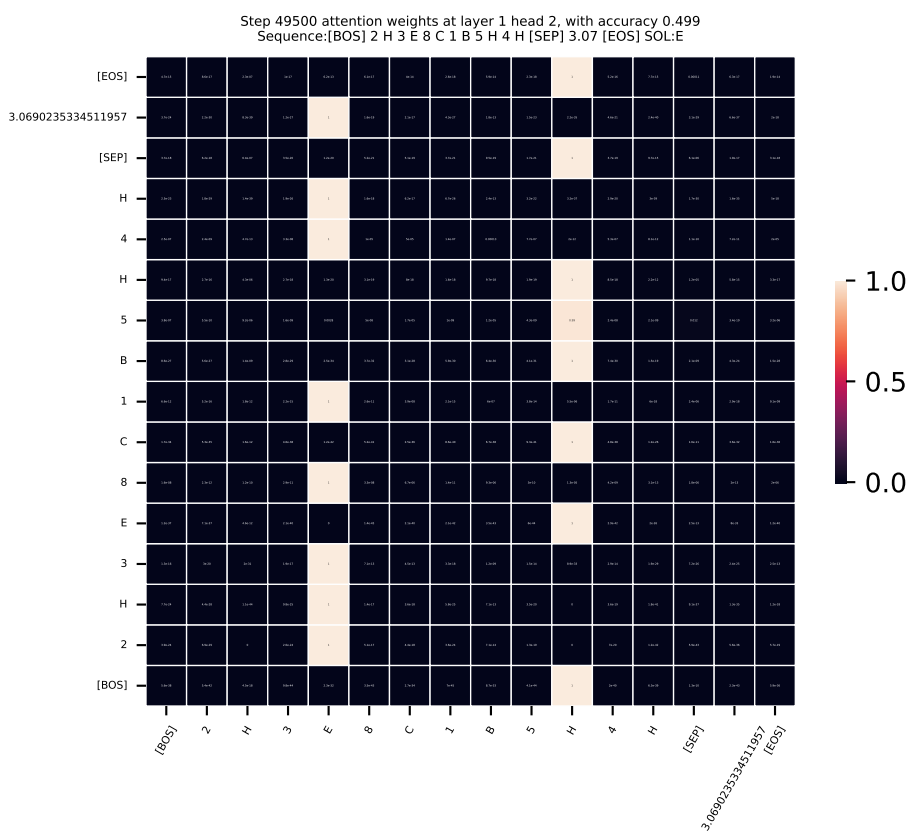

Step 49500 attention weights at layer 1 head 1, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

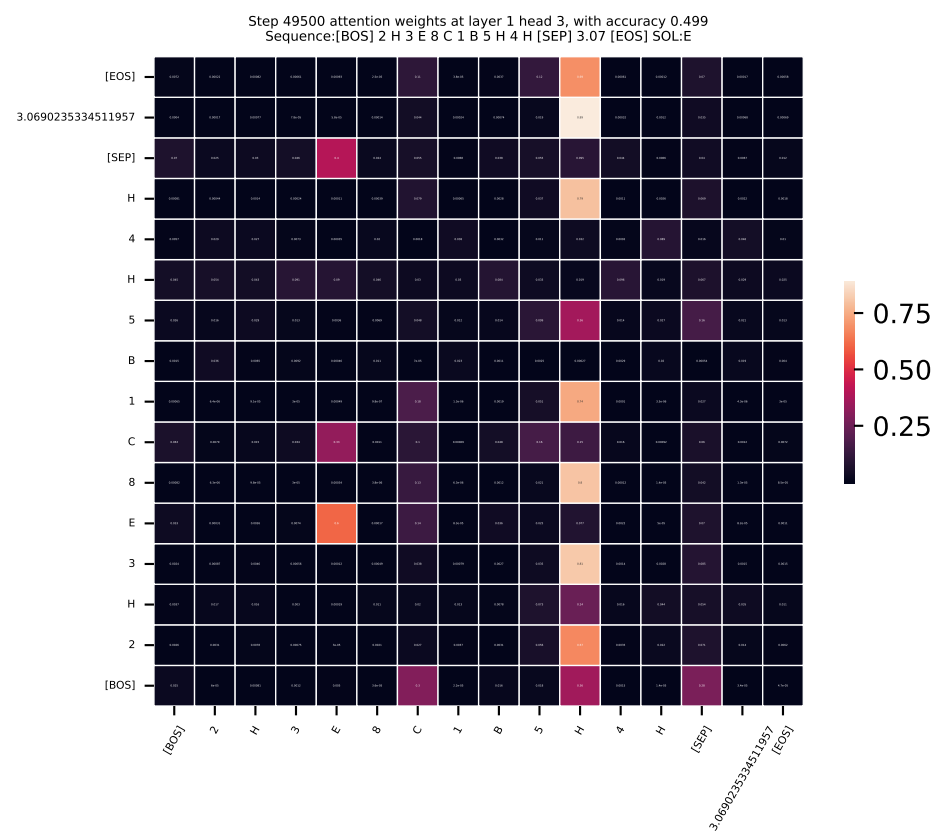

Step 49500 attention weights at layer 1 head 2, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

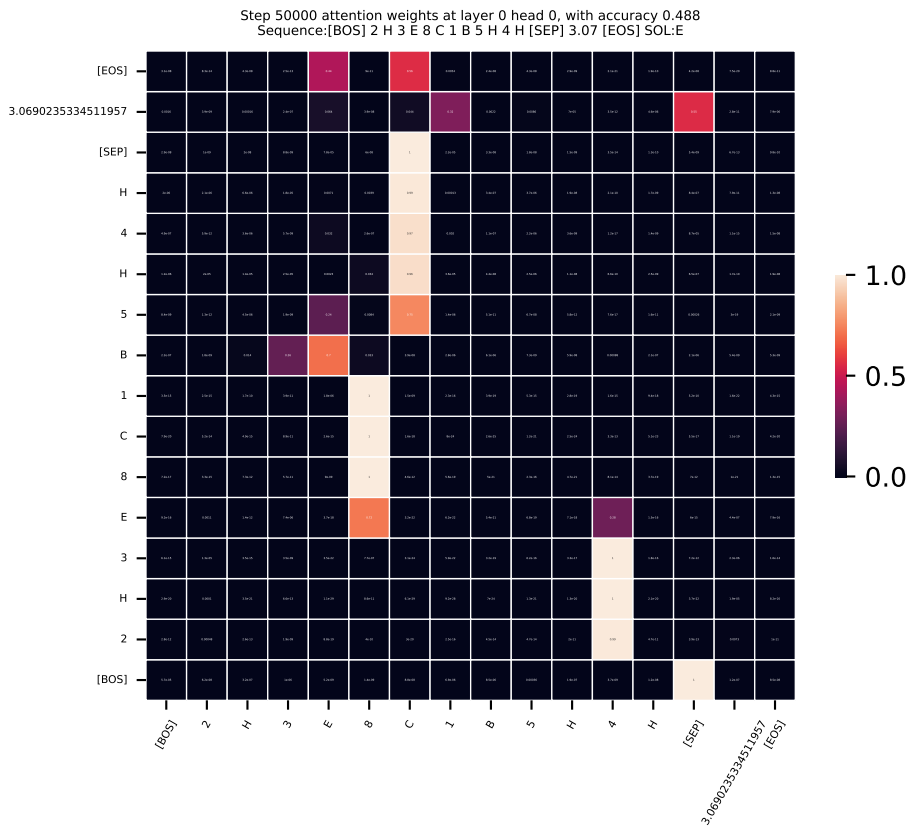

Step 49500 attention weights at layer 1 head 3, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

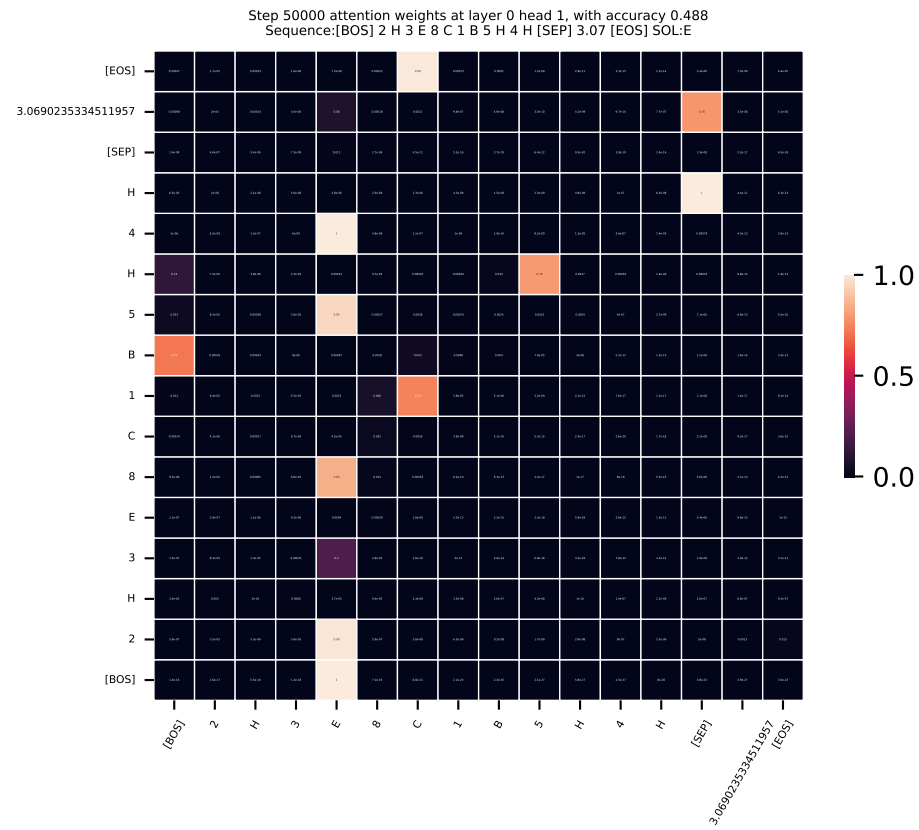

Step 49500 attention weights at layer 1 head 0, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.3
0.2
0.1

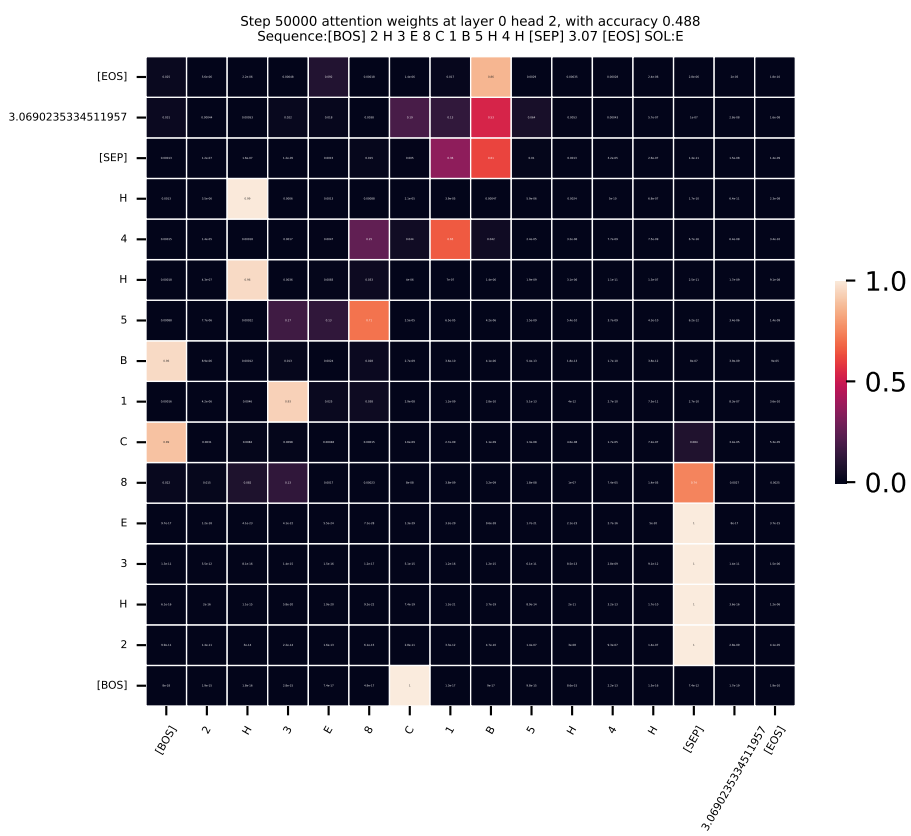

Step 49500 attention weights at layer 1 head 1, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

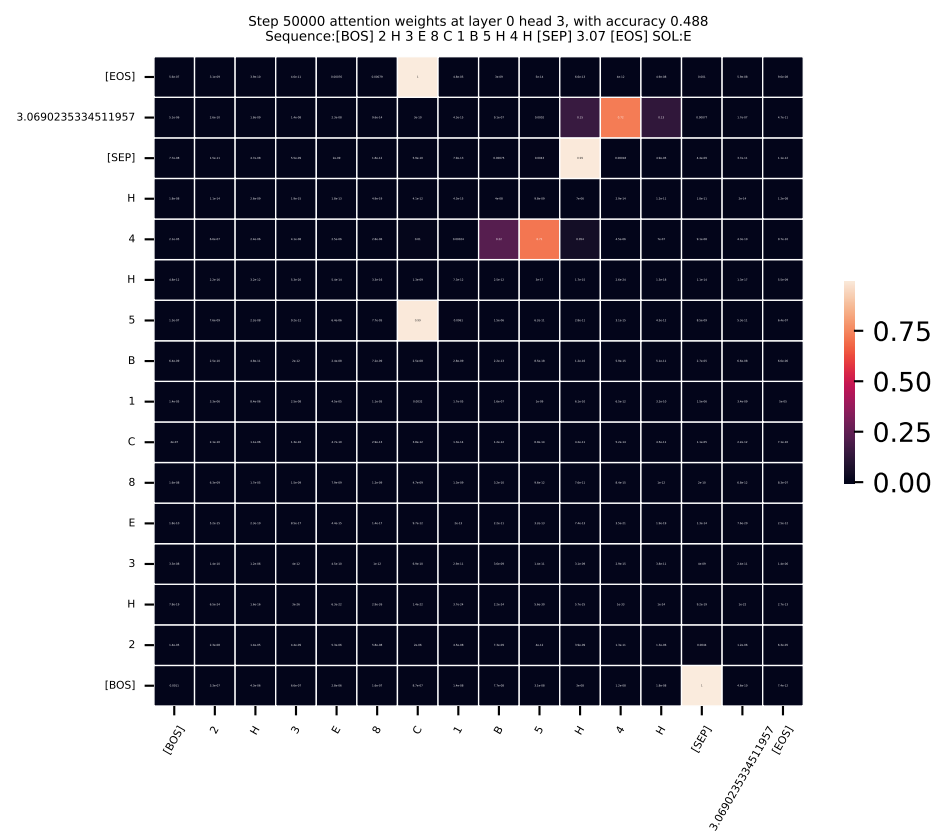

Step 49500 attention weights at layer 1 head 2, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

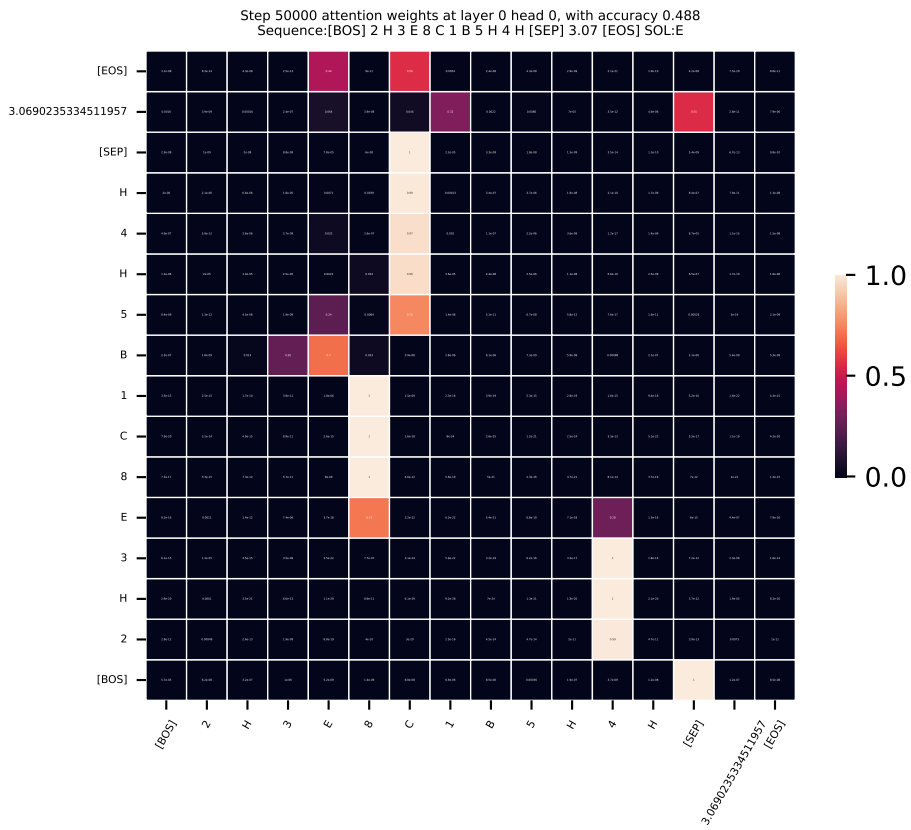

Step 49500 attention weights at layer 1 head 3, with accuracy 0.499
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

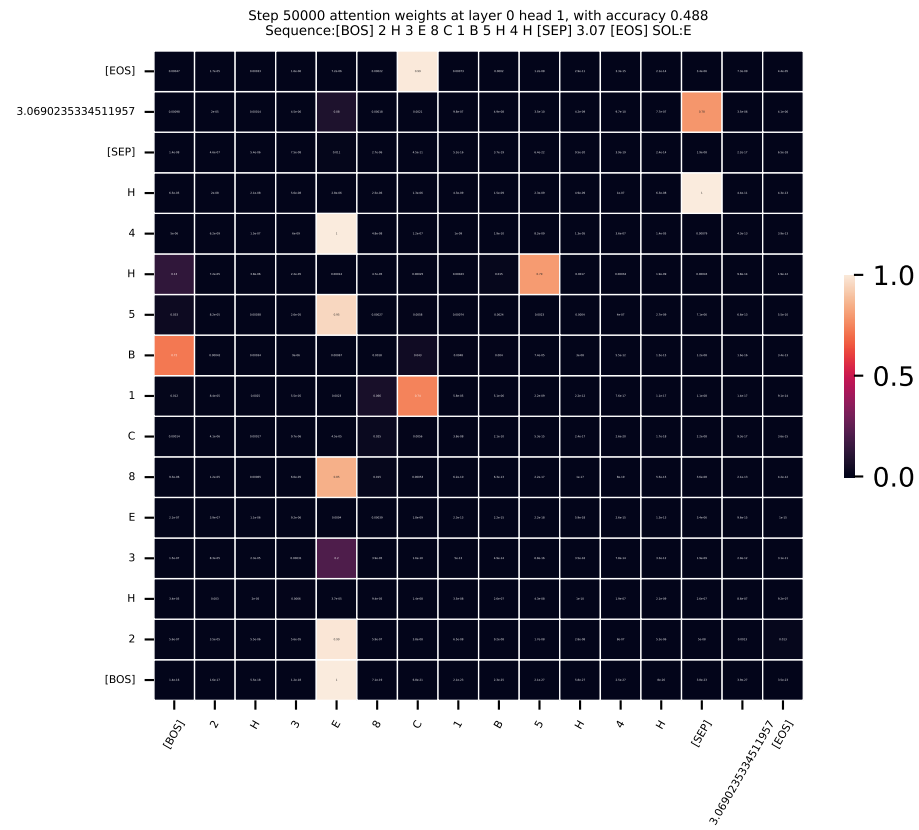

Step 50000 attention weights at layer 0 head 0, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

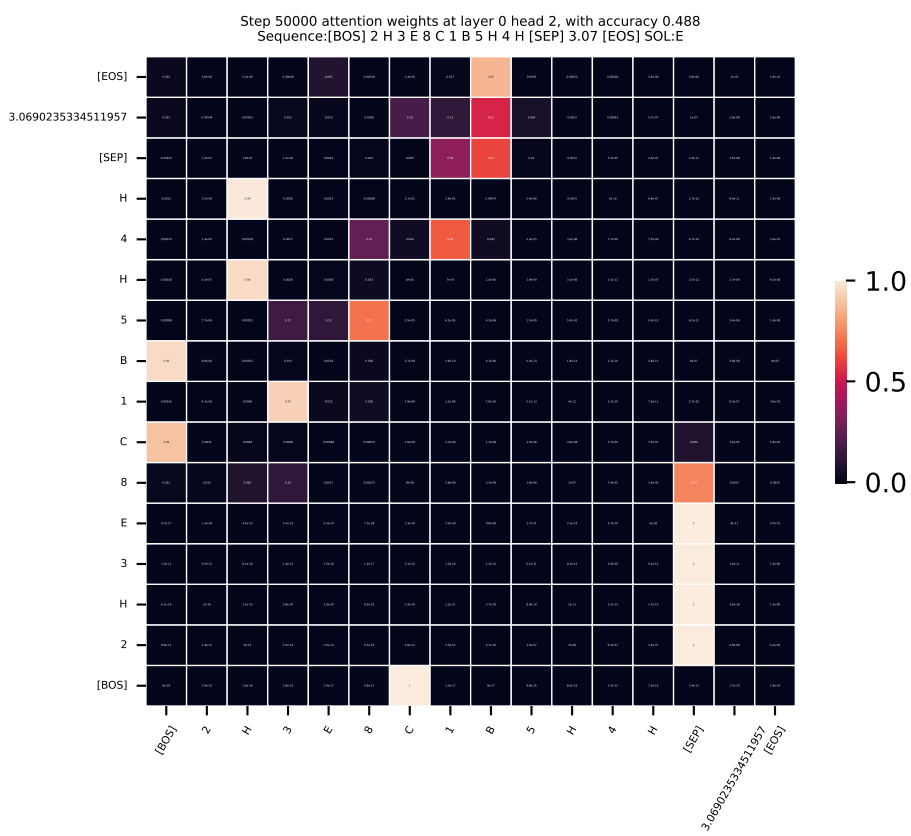

Step 50000 attention weights at layer 0 head 1, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

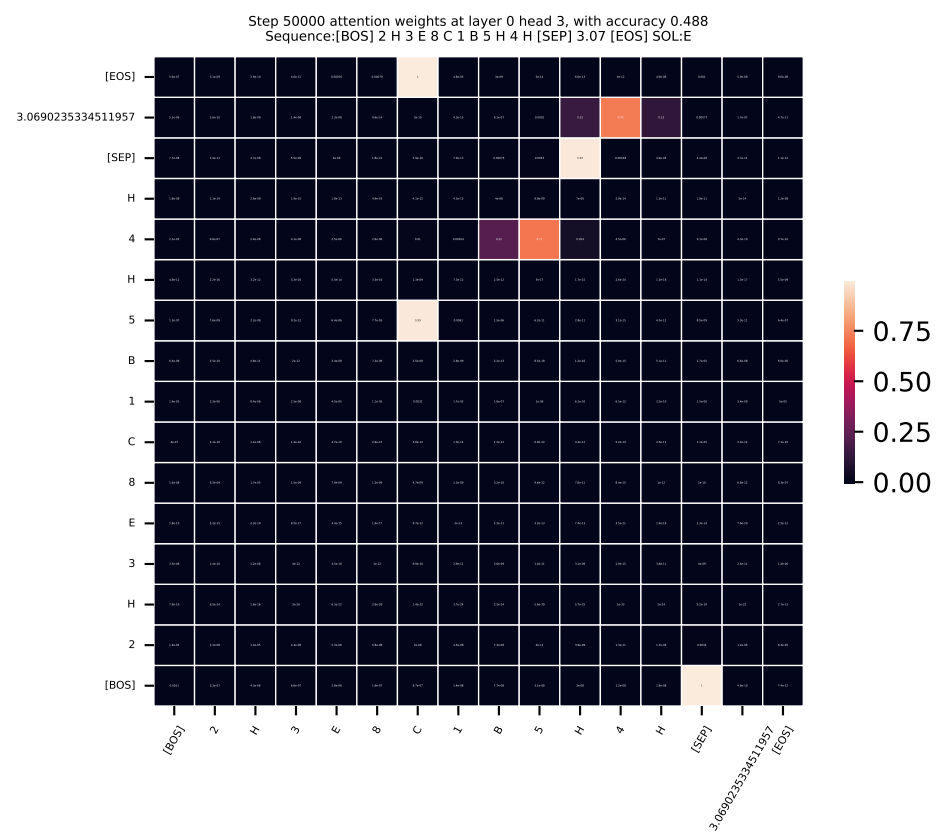

Step 50000 attention weights at layer 0 head 2, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

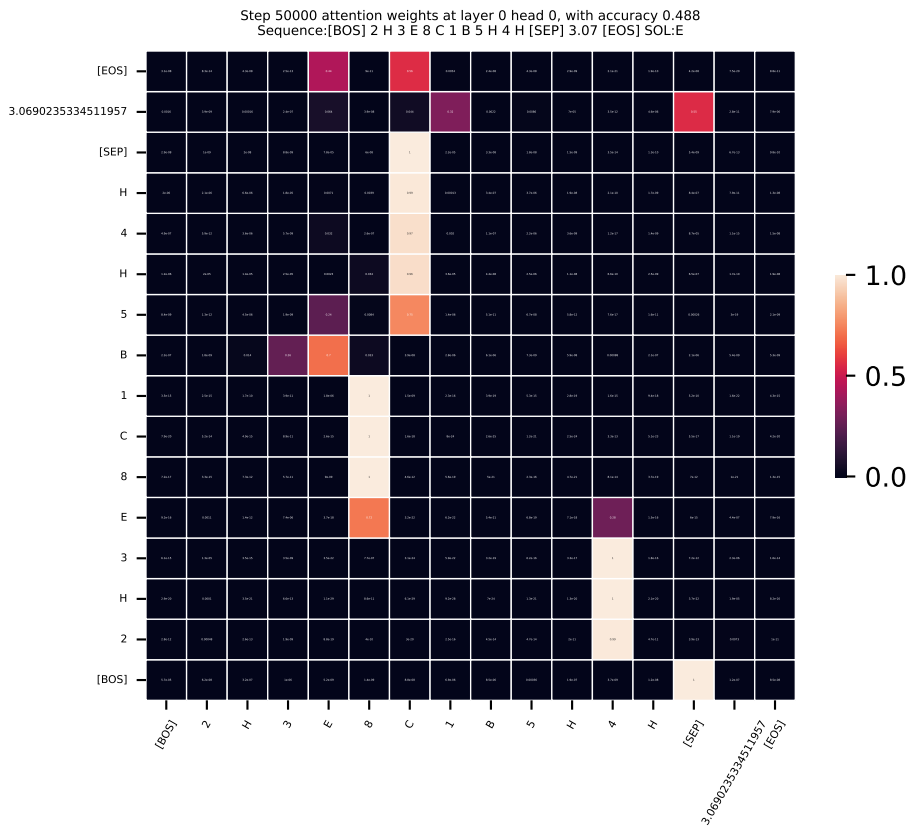

Step 50000 attention weights at layer 0 head 3, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

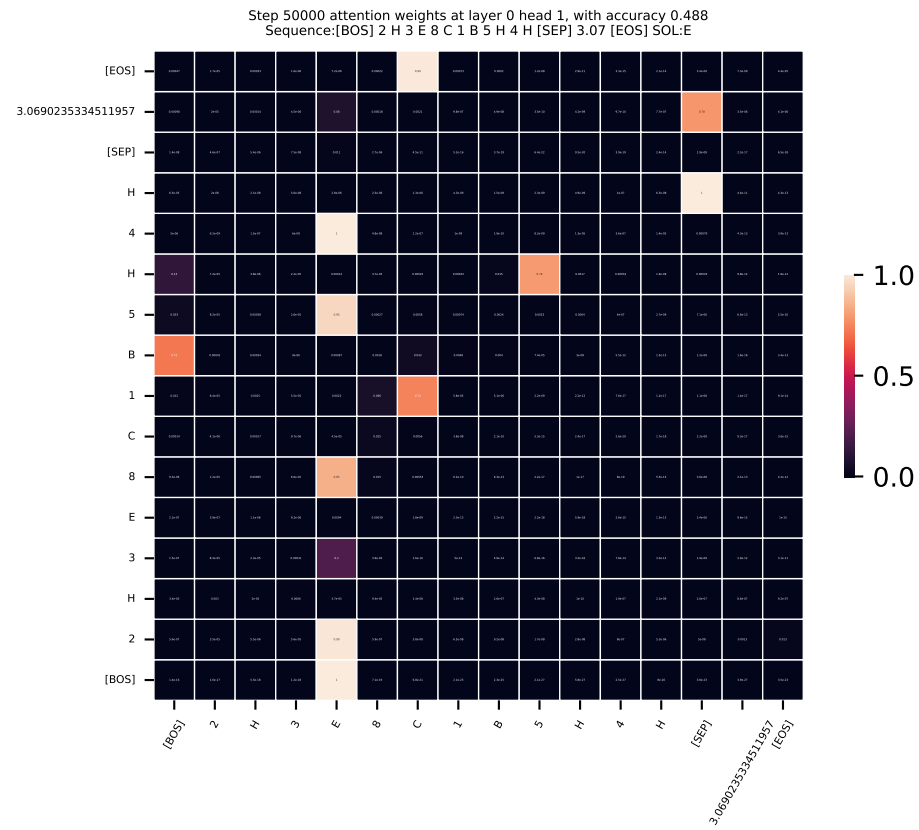

Step 50000 attention weights at layer 0 head 0, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

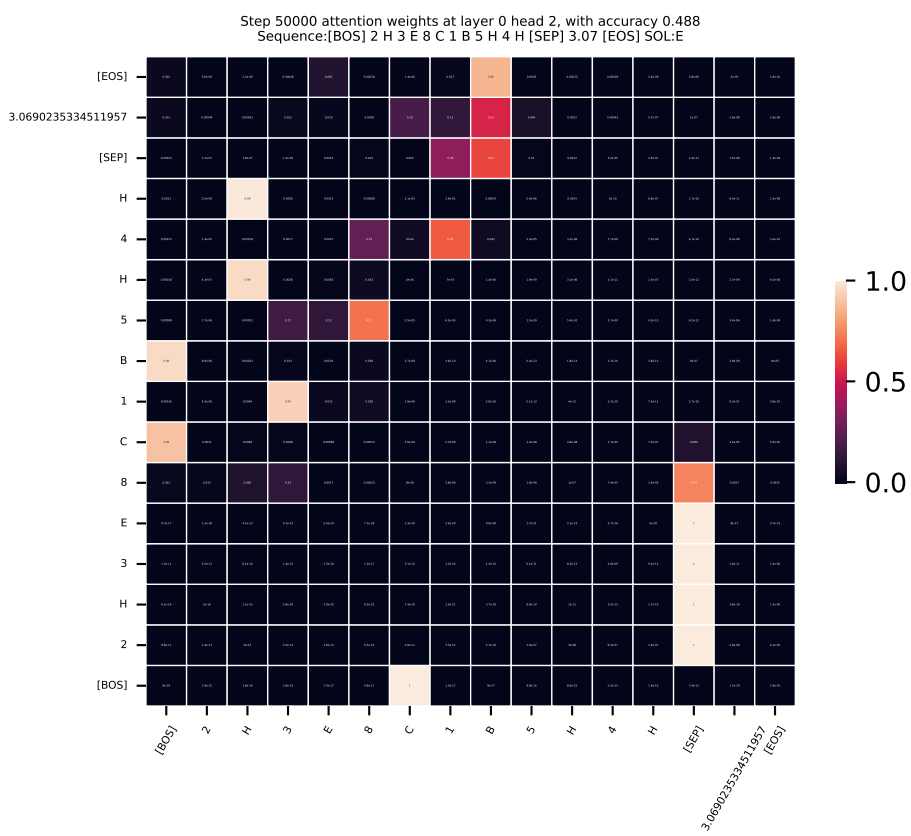

Step 50000 attention weights at layer 0 head 1, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

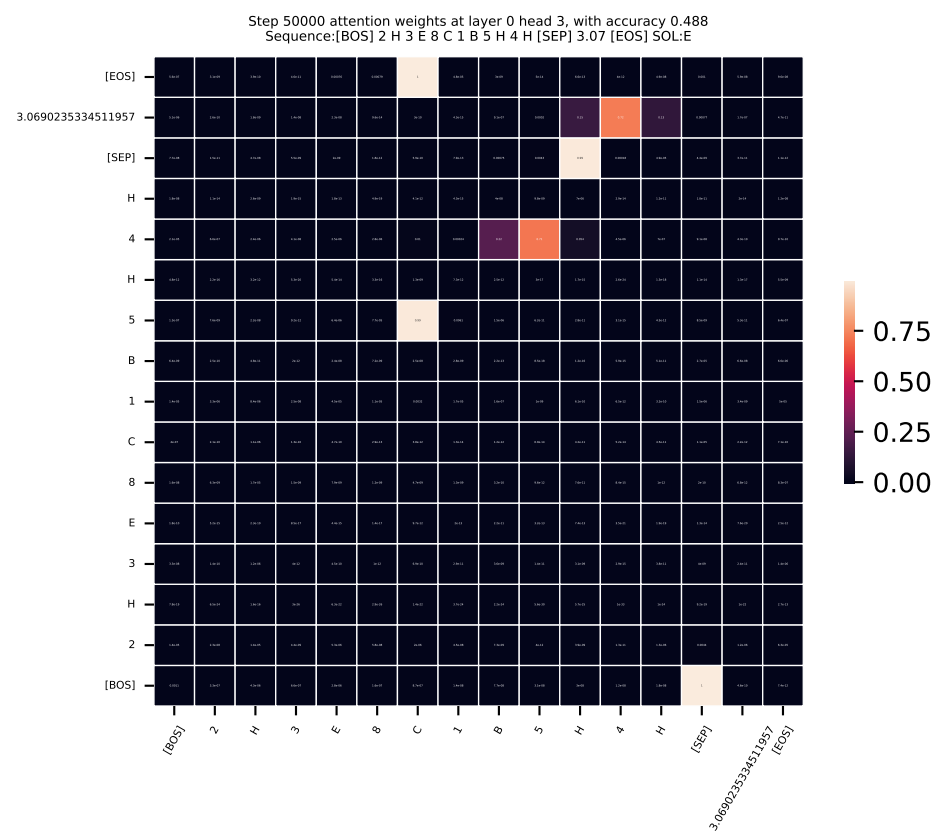

Step 50000 attention weights at layer 0 head 2, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

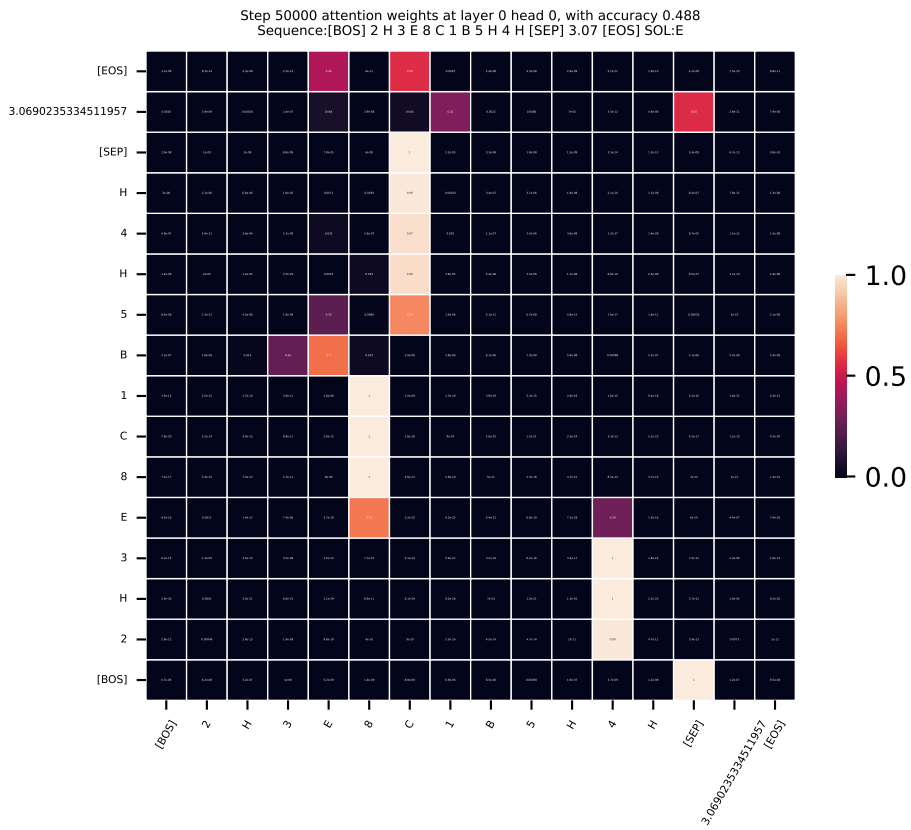

Step 50000 attention weights at layer 0 head 3, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

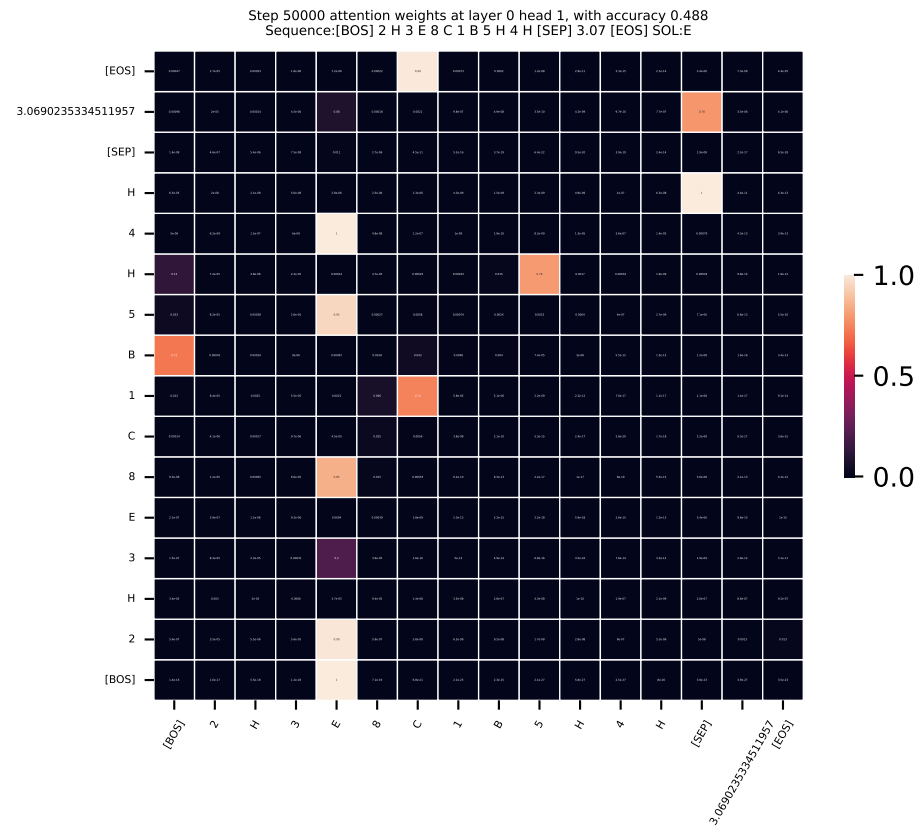

Step 50000 attention weights at layer 0 head 0, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

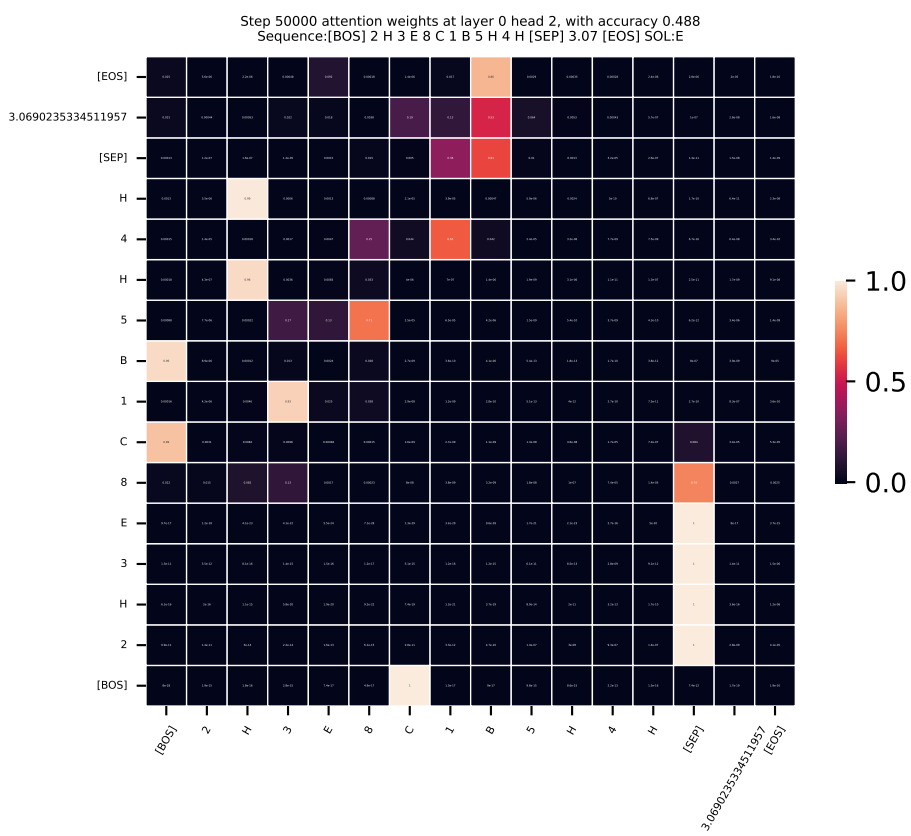

Step 50000 attention weights at layer 0 head 1, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

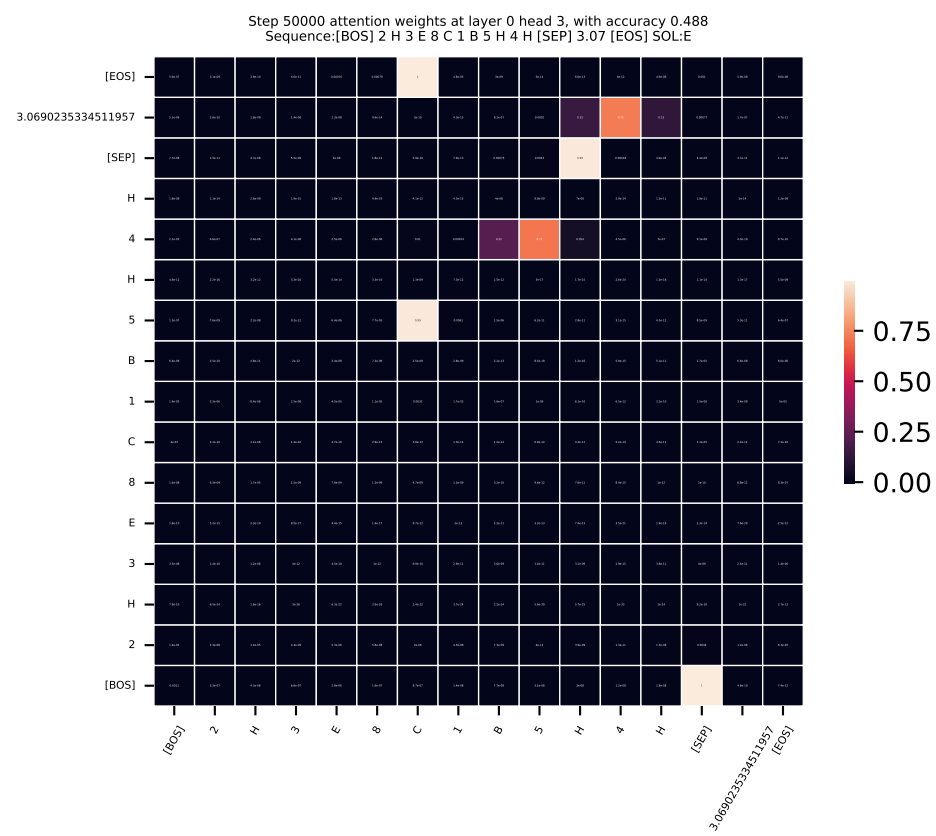

Step 50000 attention weights at layer 0 head 2, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

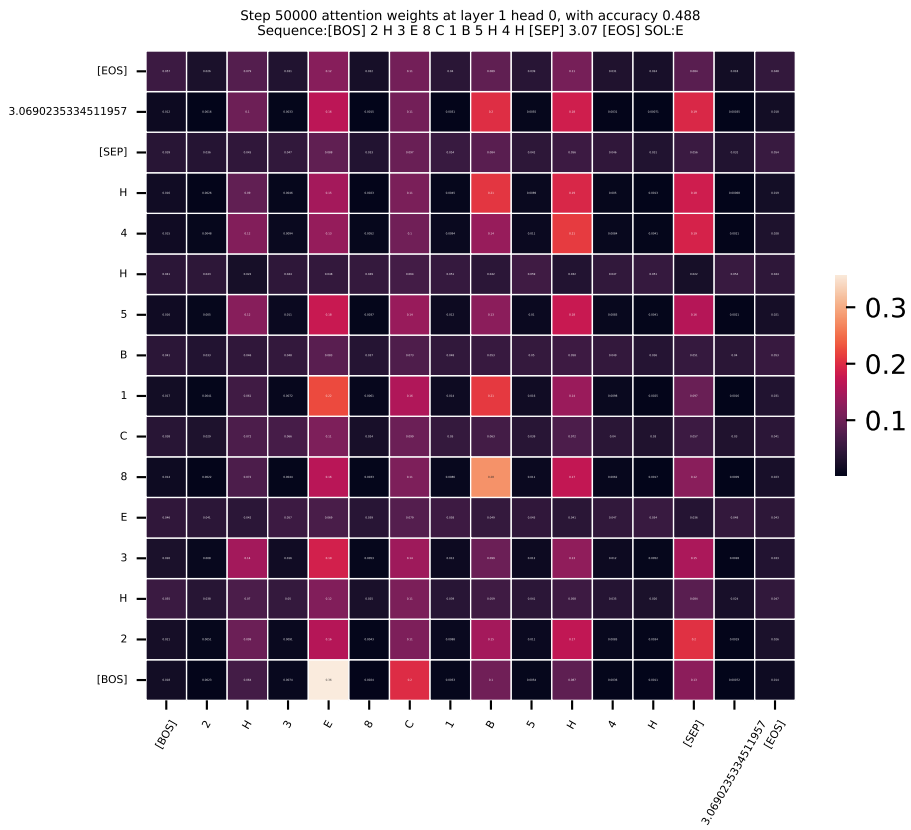

Step 50000 attention weights at layer 0 head 3, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

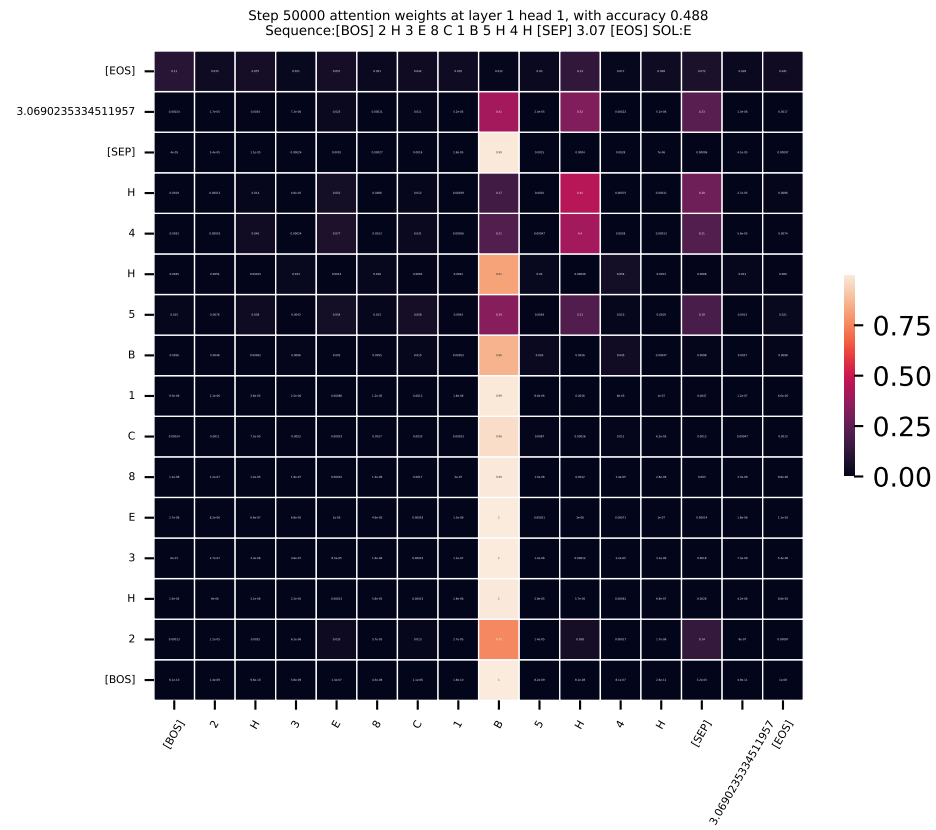

Step 50000 attention weights at layer 0 head 0, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

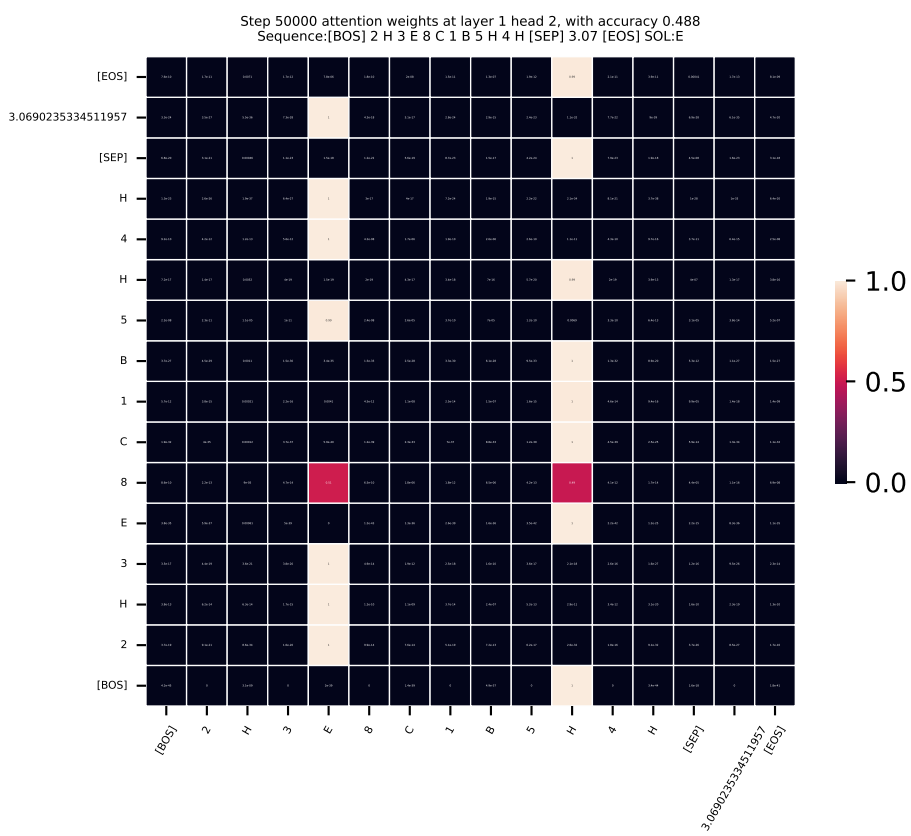

Step 50000 attention weights at layer 0 head 1, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

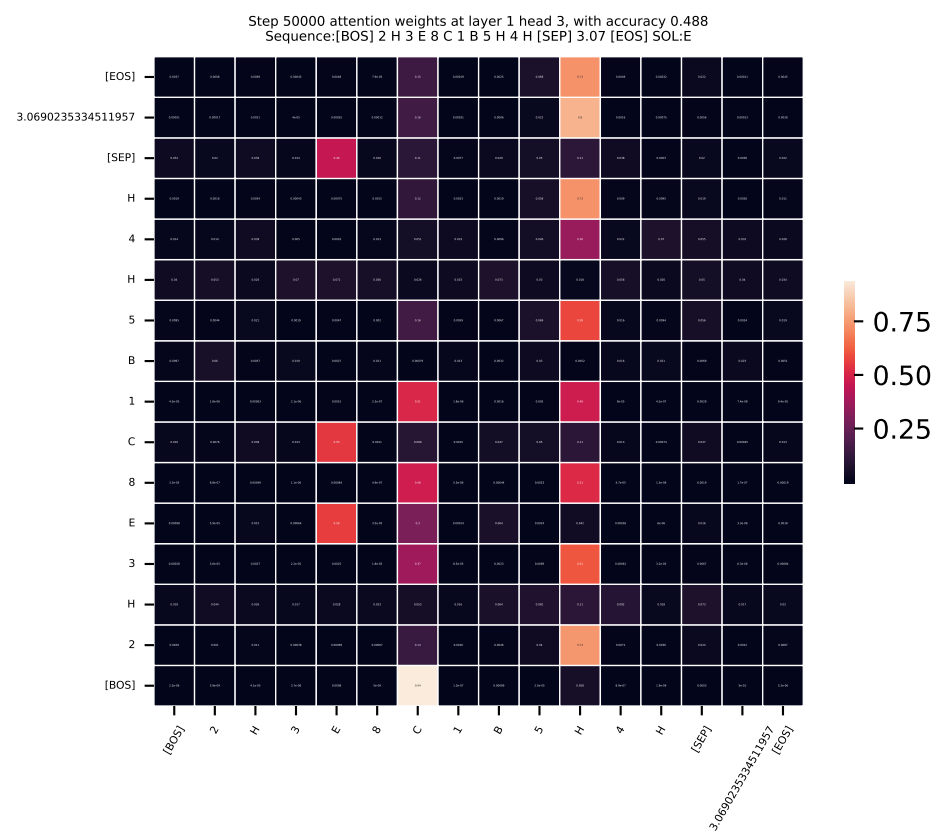

Step 50000 attention weights at layer 0 head 2, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

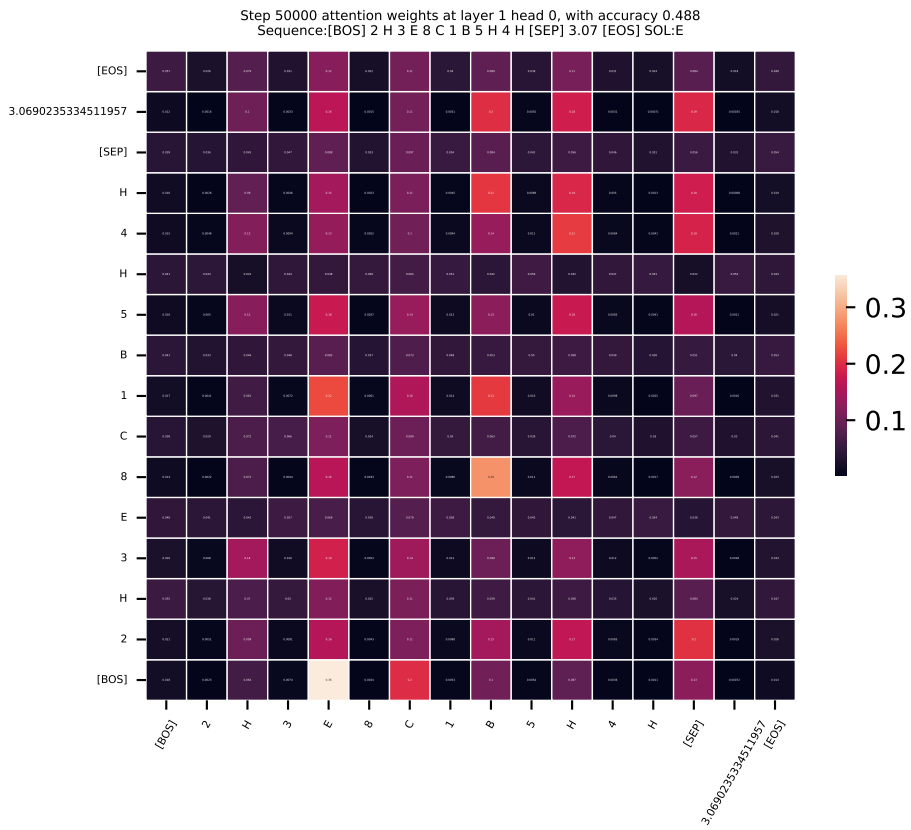

Step 50000 attention weights at layer 0 head 3, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

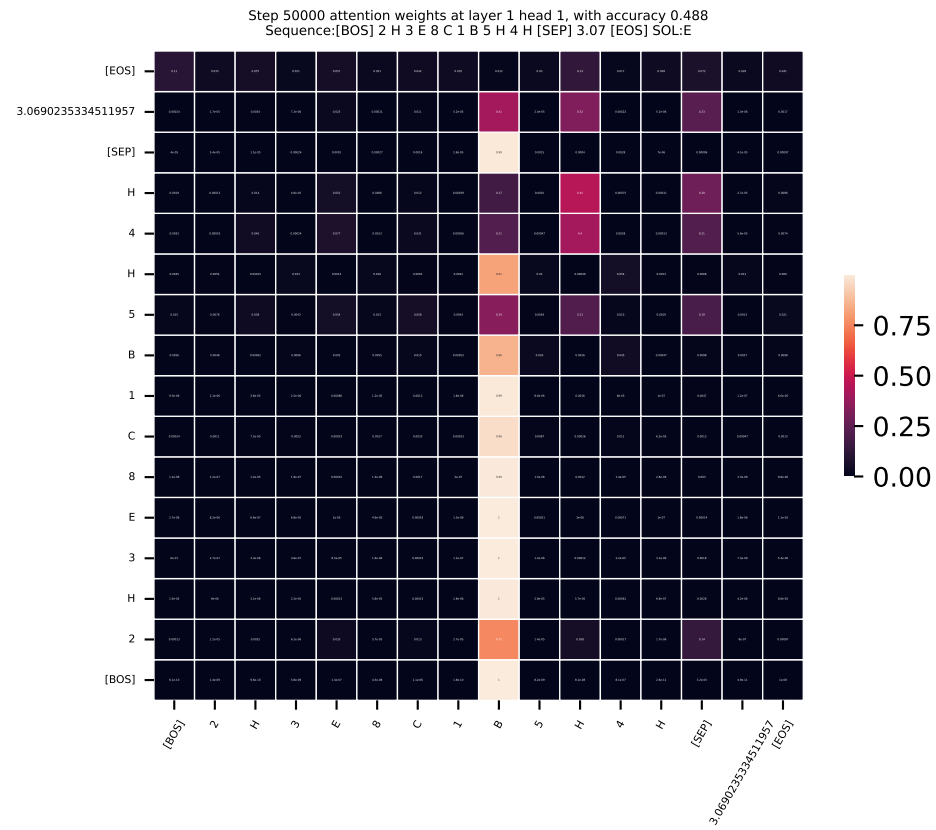

Step 50000 attention weights at layer 1 head 0, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.3
0.2
0.1

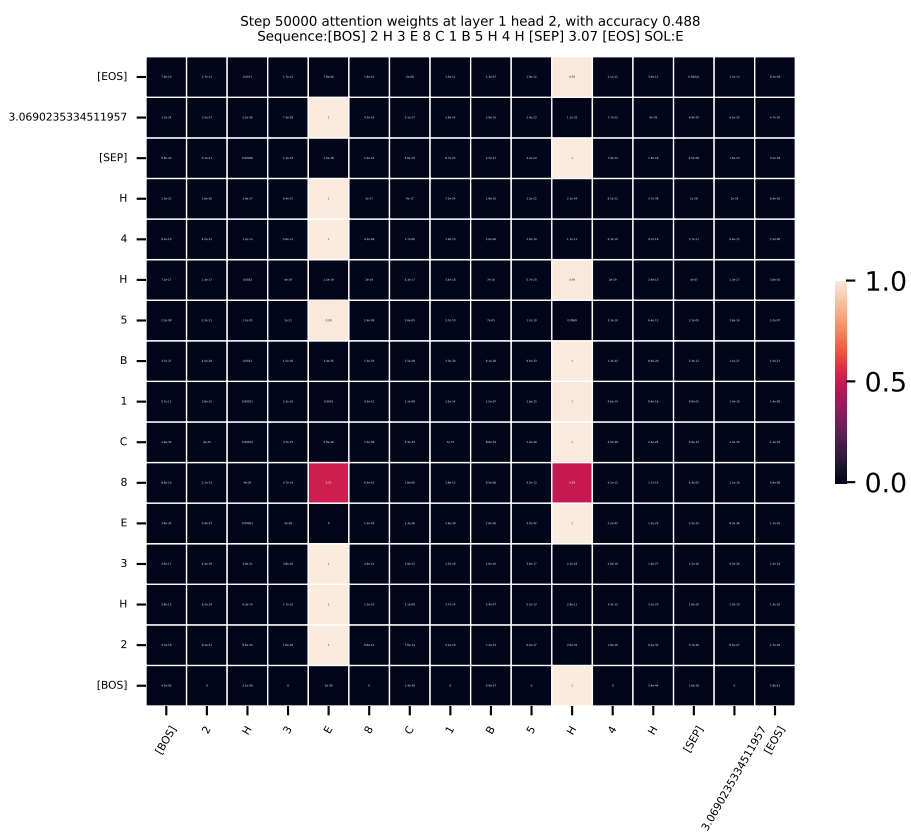

Step 50000 attention weights at layer 1 head 1, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

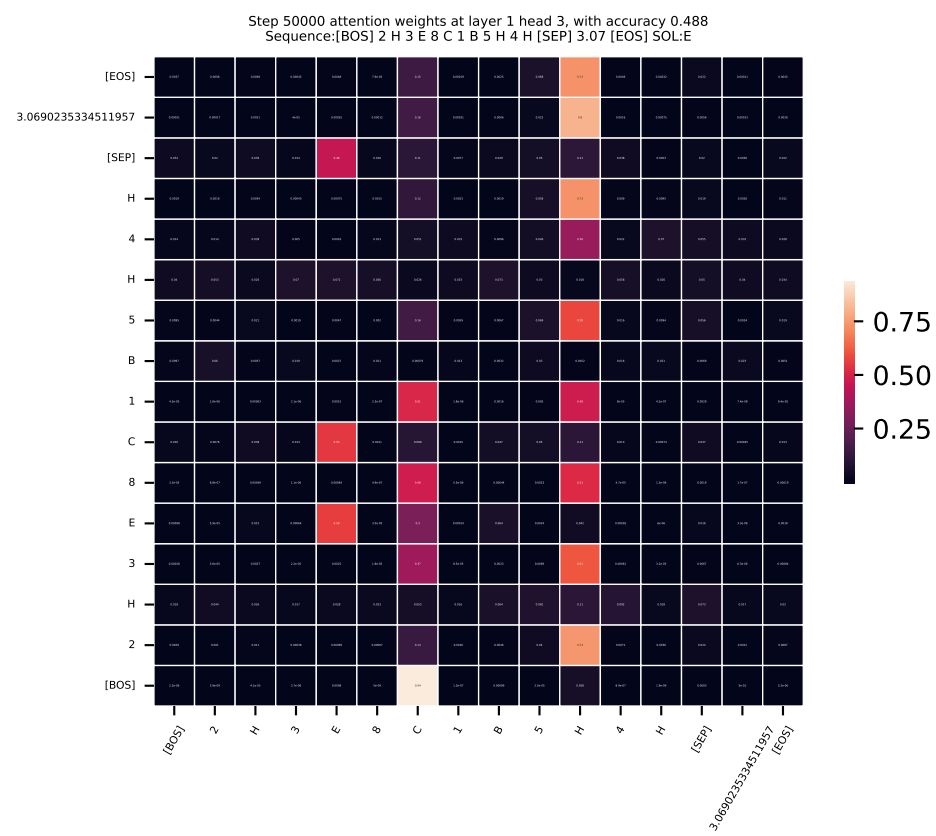

Step 50000 attention weights at layer 1 head 2, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

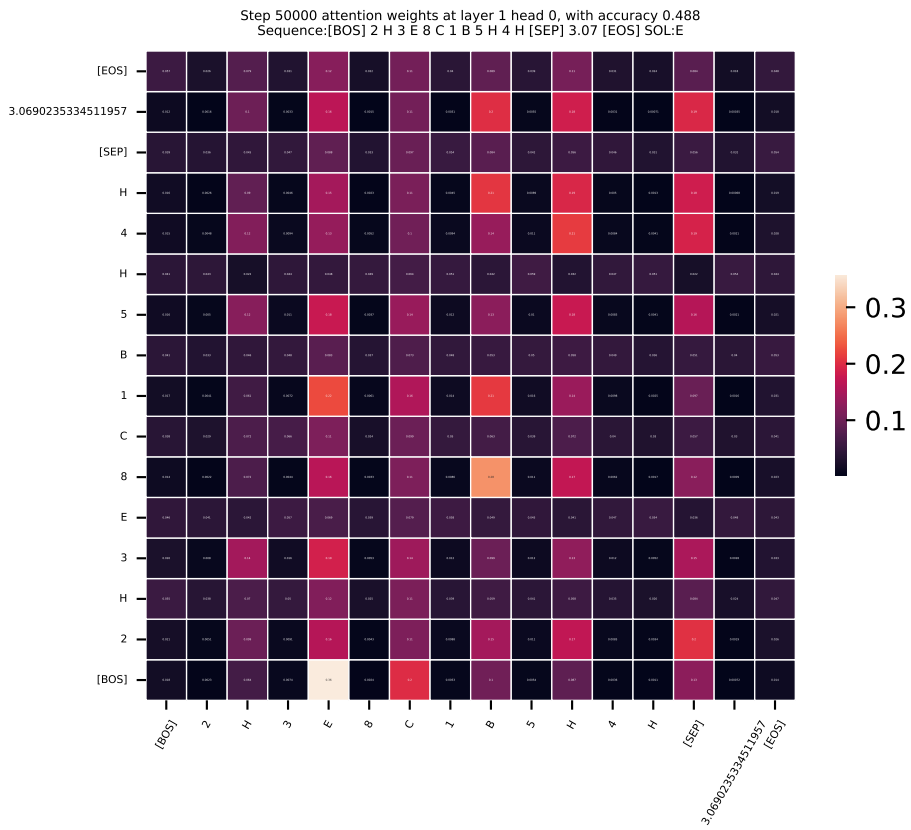

Step 50000 attention weights at layer 1 head 3, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

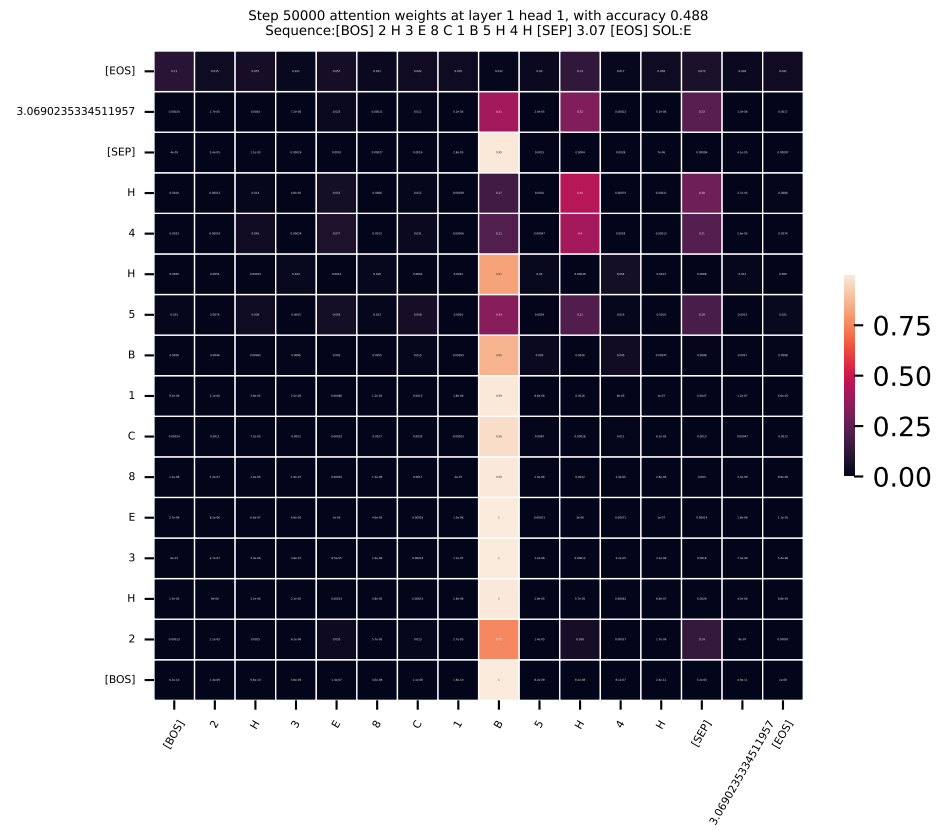

Step 50000 attention weights at layer 1 head 0, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.3
0.2
0.1

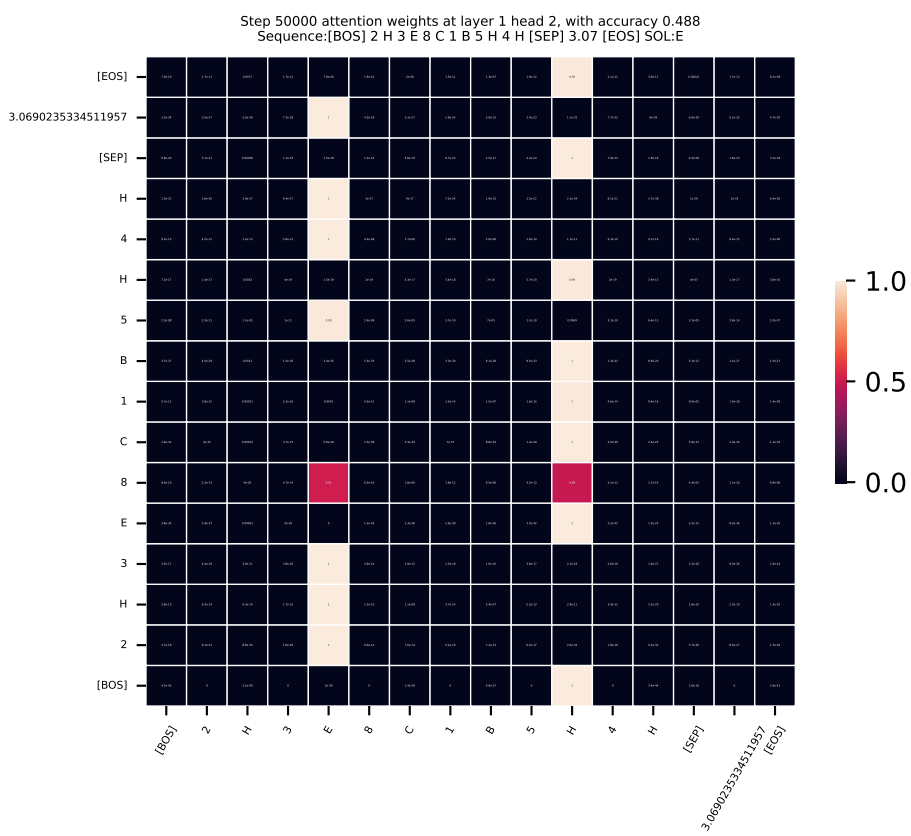

Step 50000 attention weights at layer 1 head 1, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

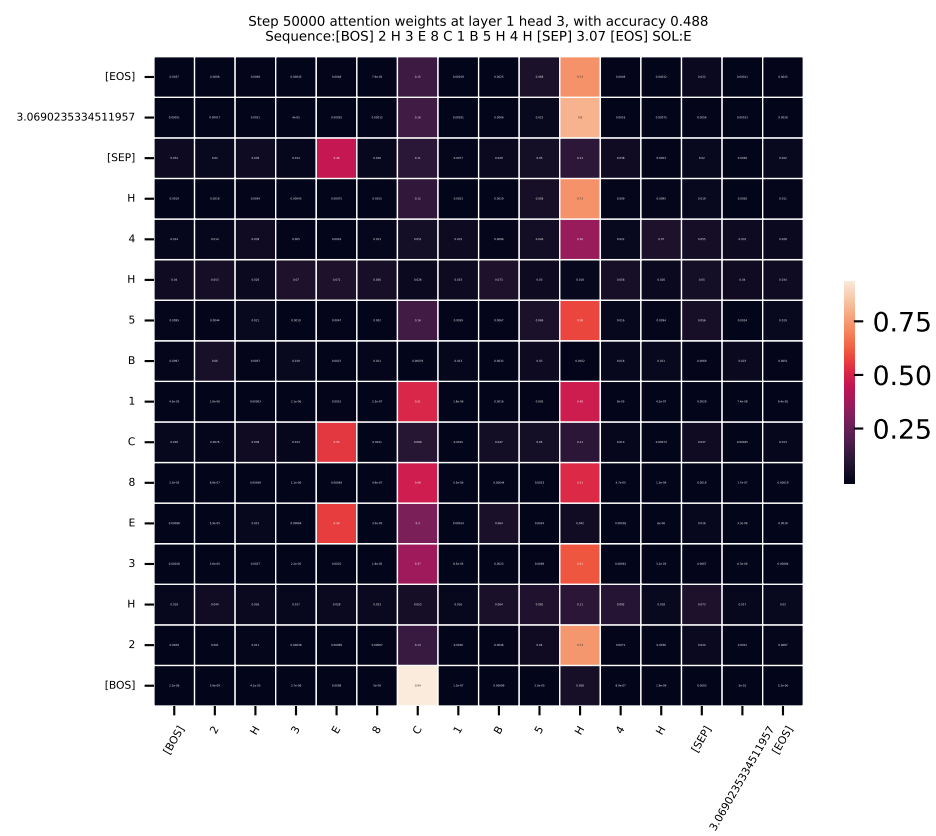

Step 50000 attention weights at layer 1 head 2, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

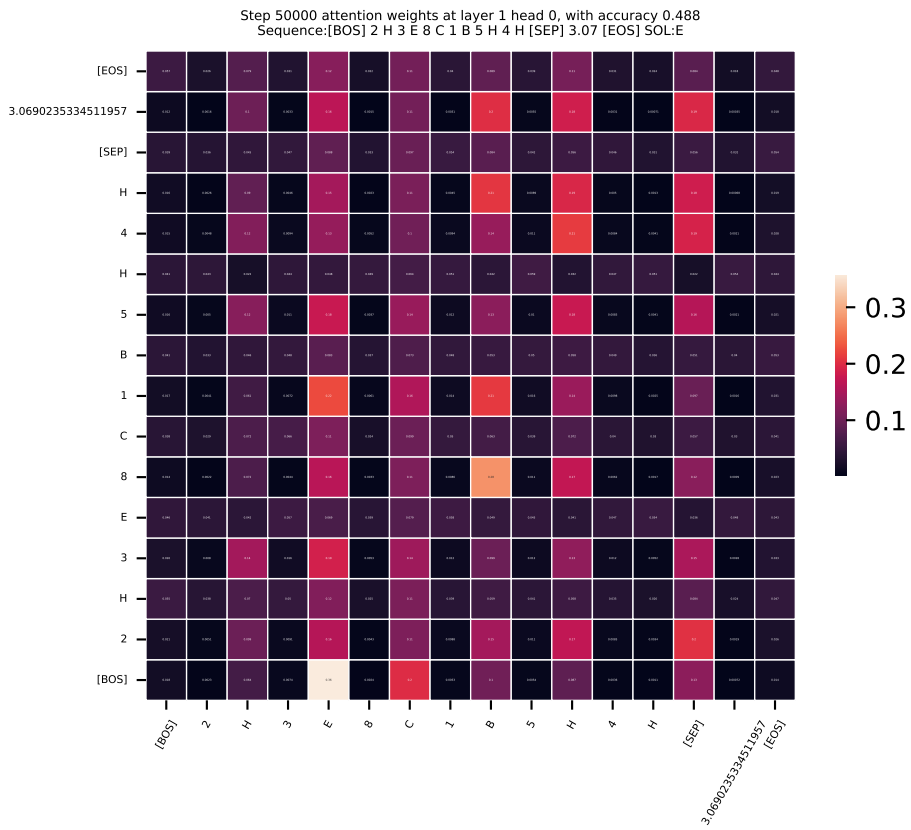

Step 50000 attention weights at layer 1 head 3, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

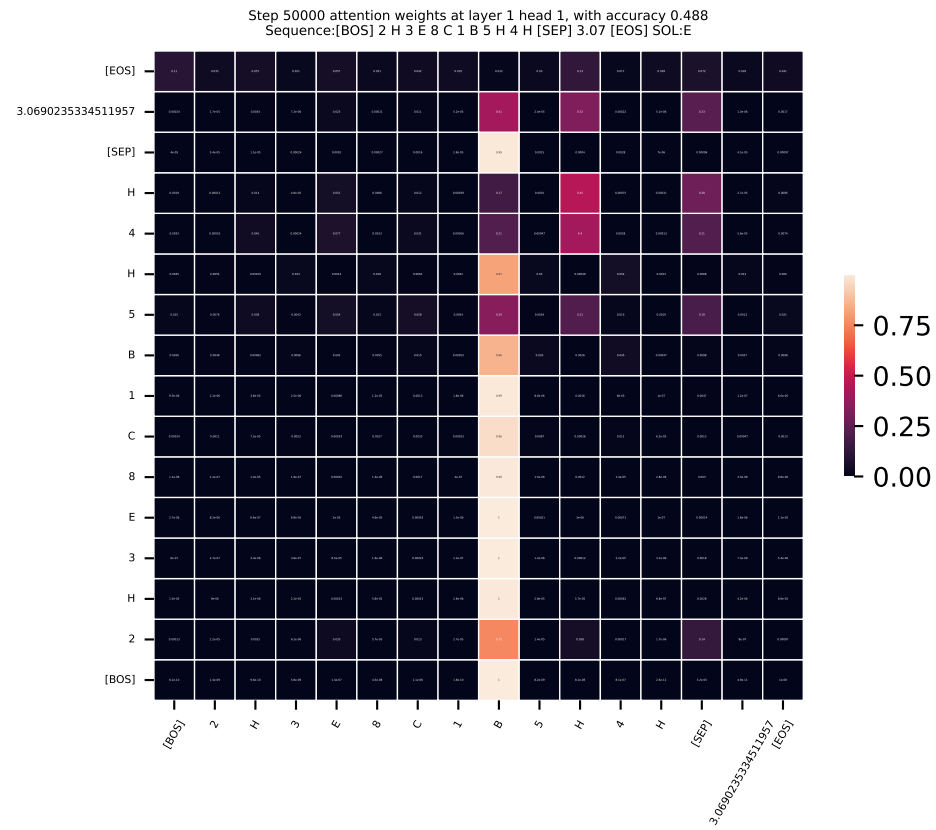

Step 50000 attention weights at layer 1 head 0, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.3
0.2
0.1

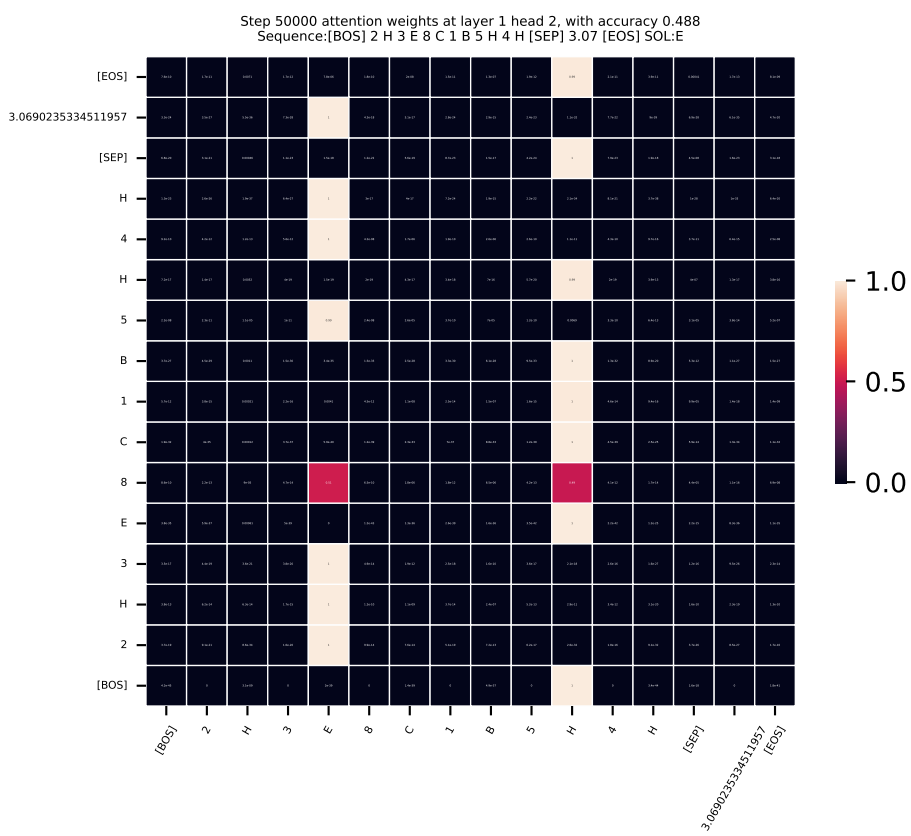

Step 50000 attention weights at layer 1 head 1, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

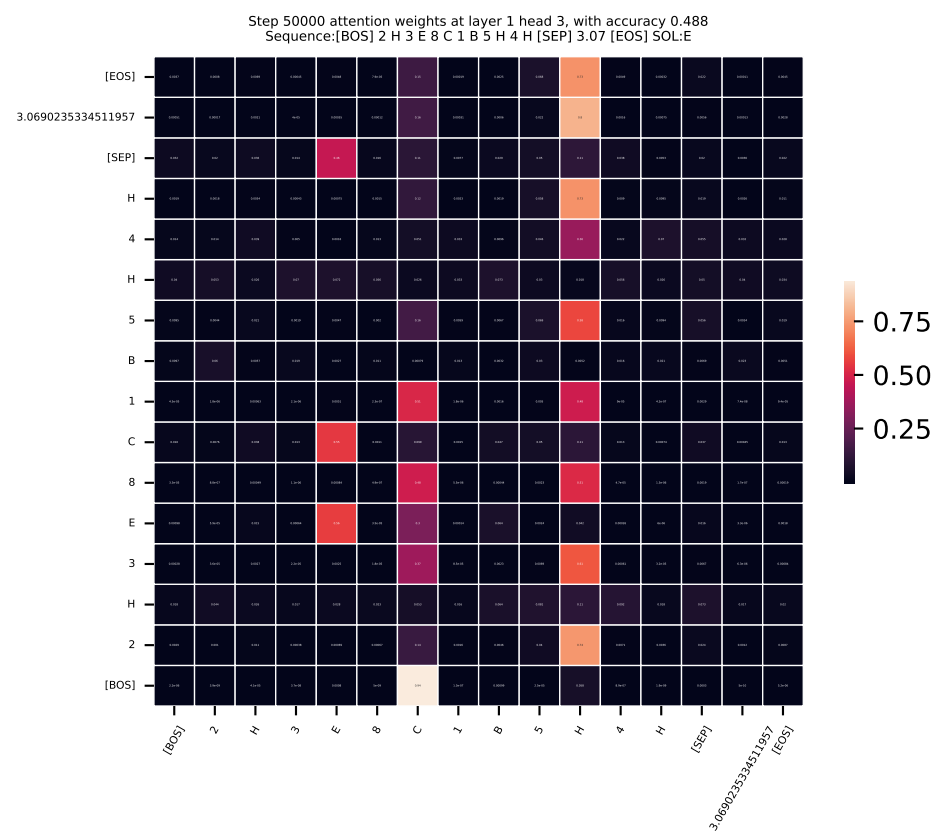

Step 50000 attention weights at layer 1 head 2, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

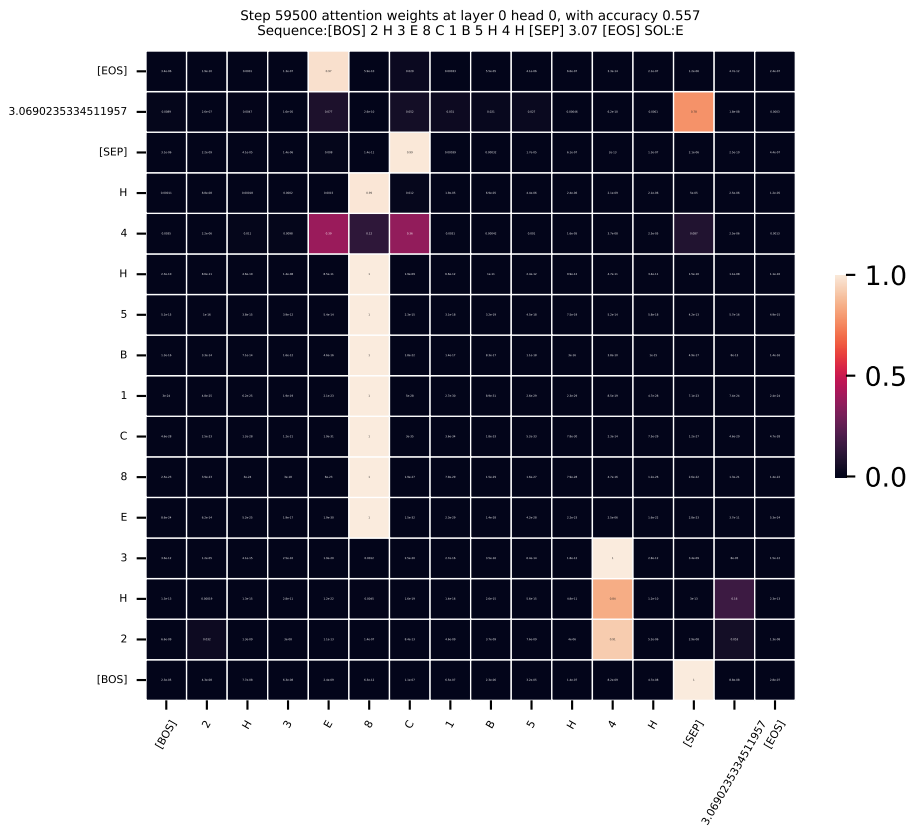

Step 50000 attention weights at layer 1 head 3, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

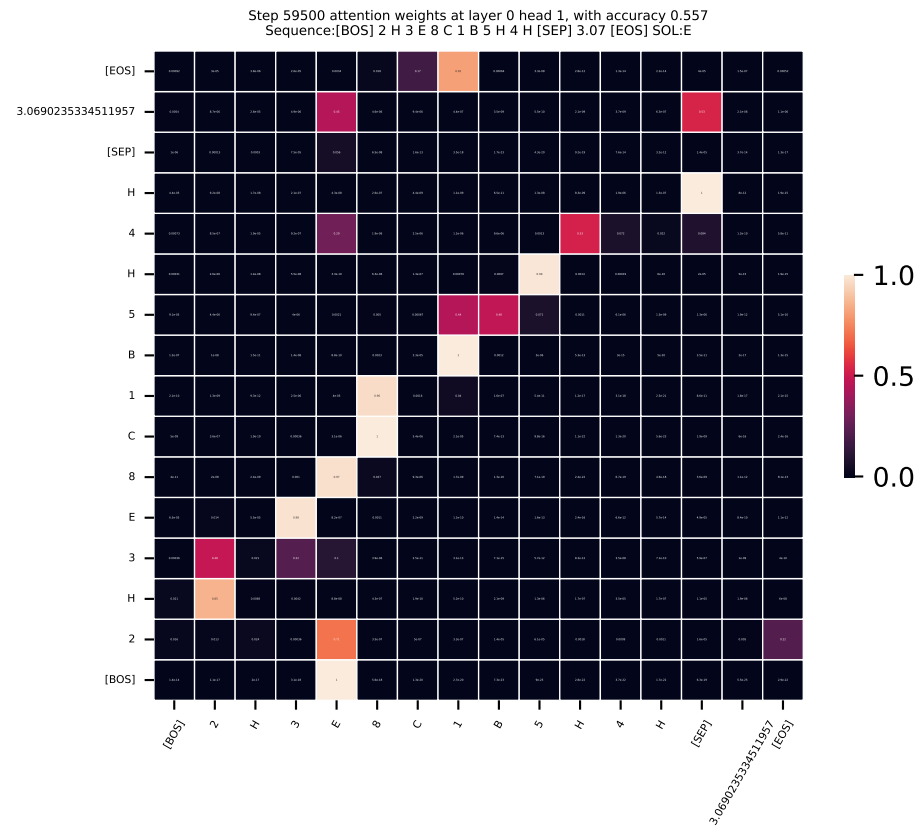

Step 50000 attention weights at layer 1 head 0, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.3
0.2
0.1

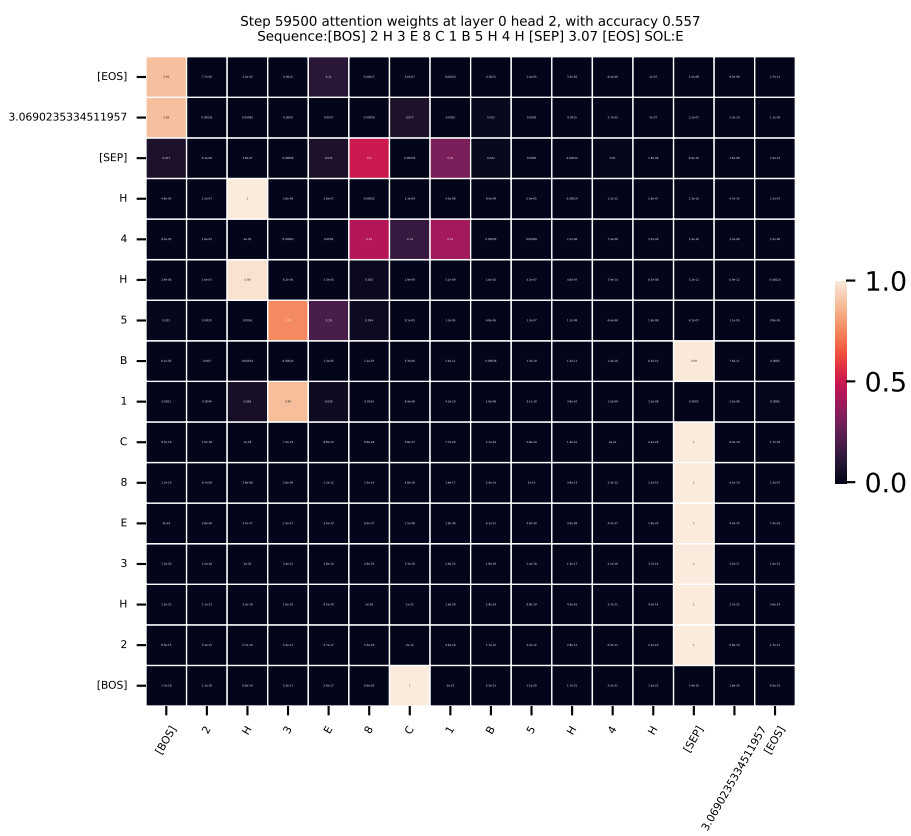

Step 50000 attention weights at layer 1 head 1, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

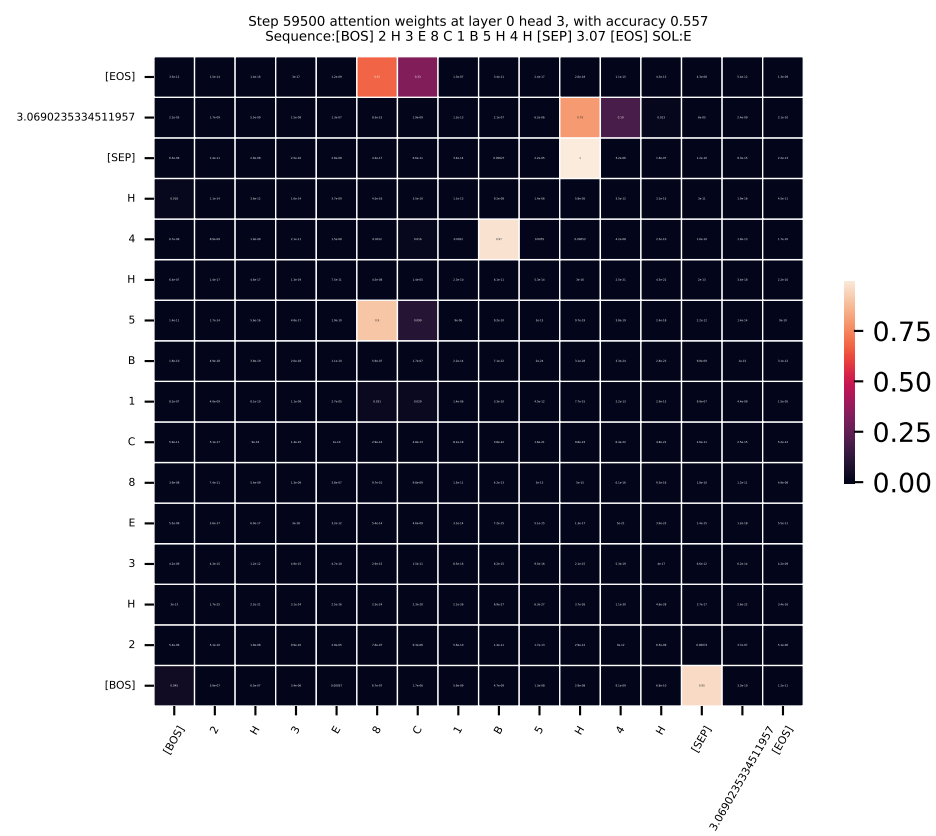

Step 50000 attention weights at layer 1 head 2, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

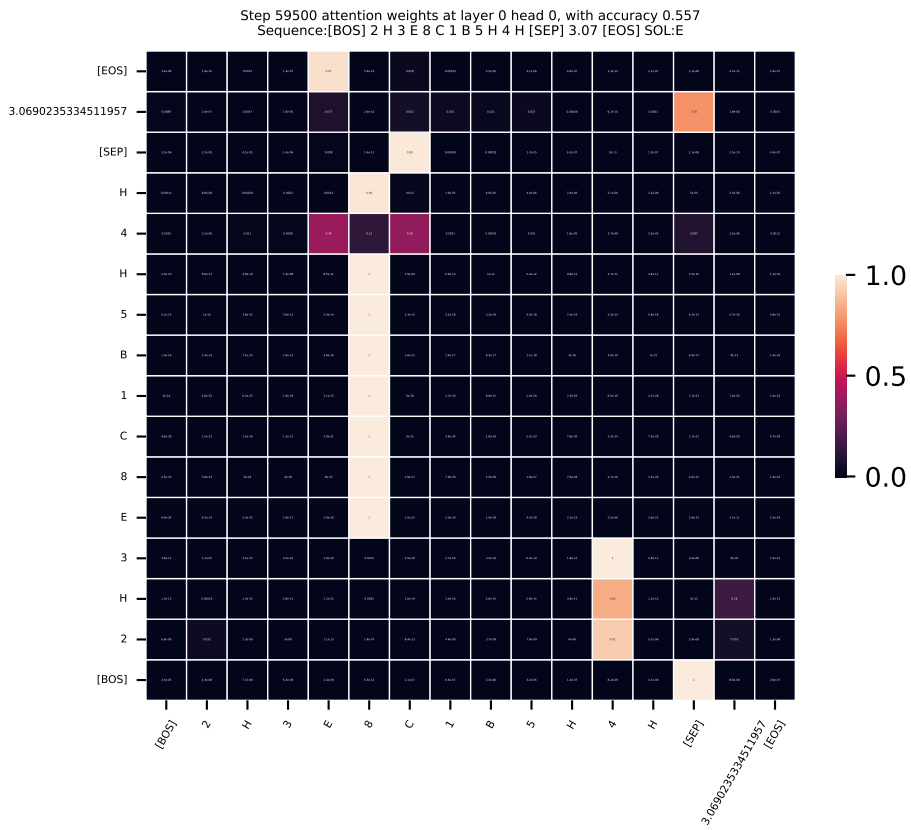

Step 50000 attention weights at layer 1 head 3, with accuracy 0.488
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

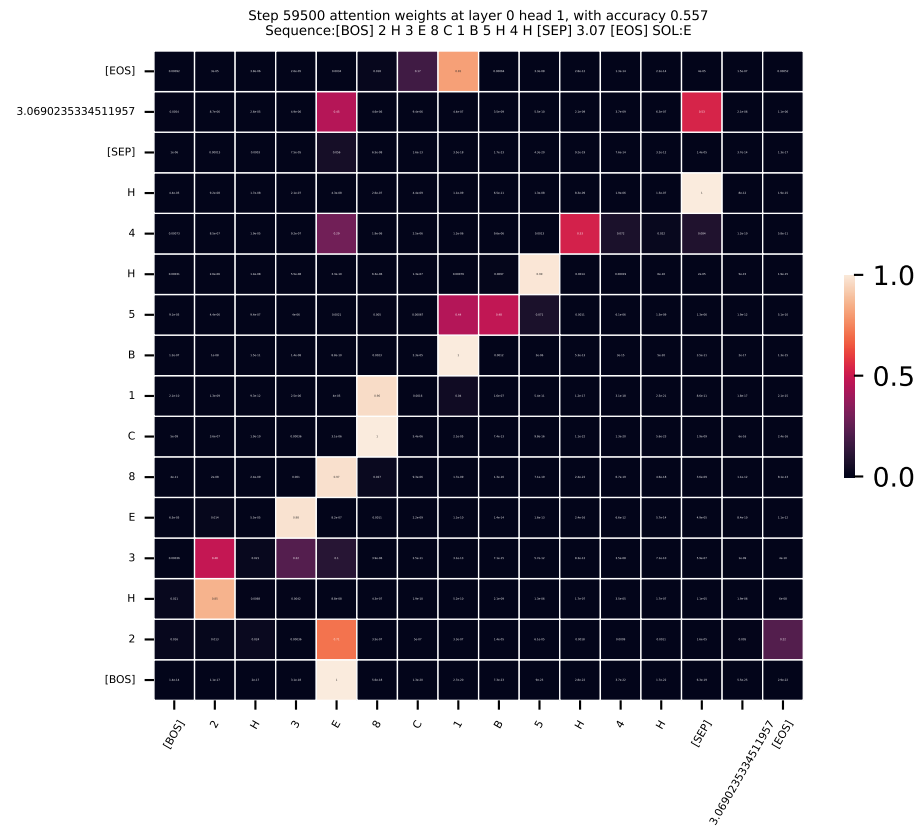

Step 59500 attention weights at layer 0 head 0, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

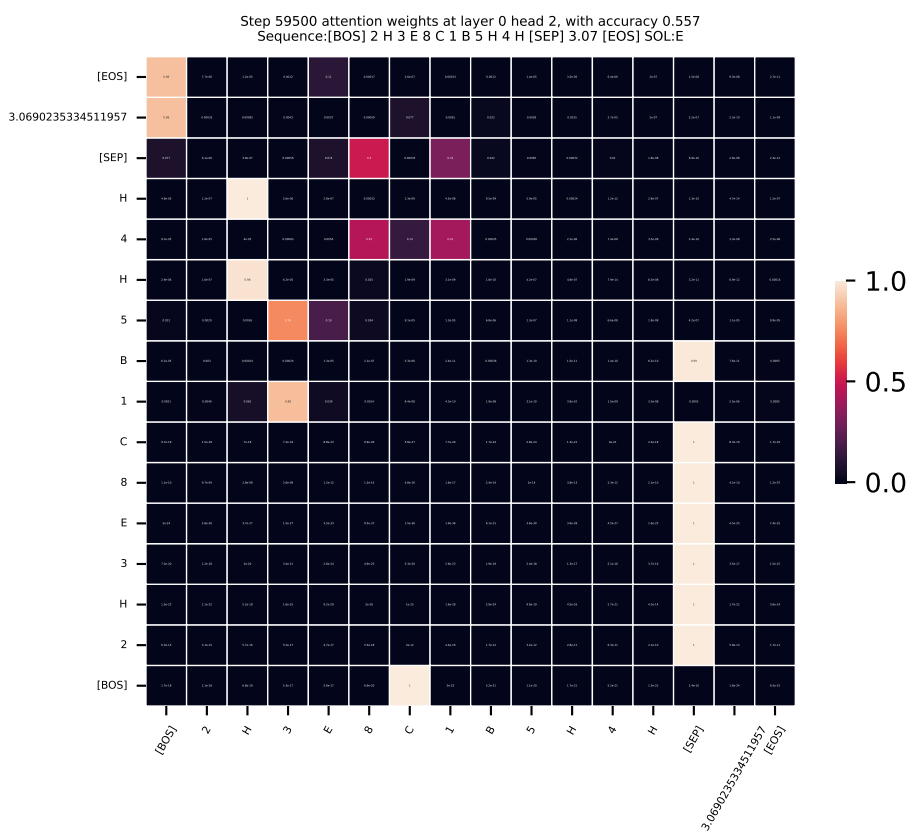

Step 59500 attention weights at layer 0 head 1, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

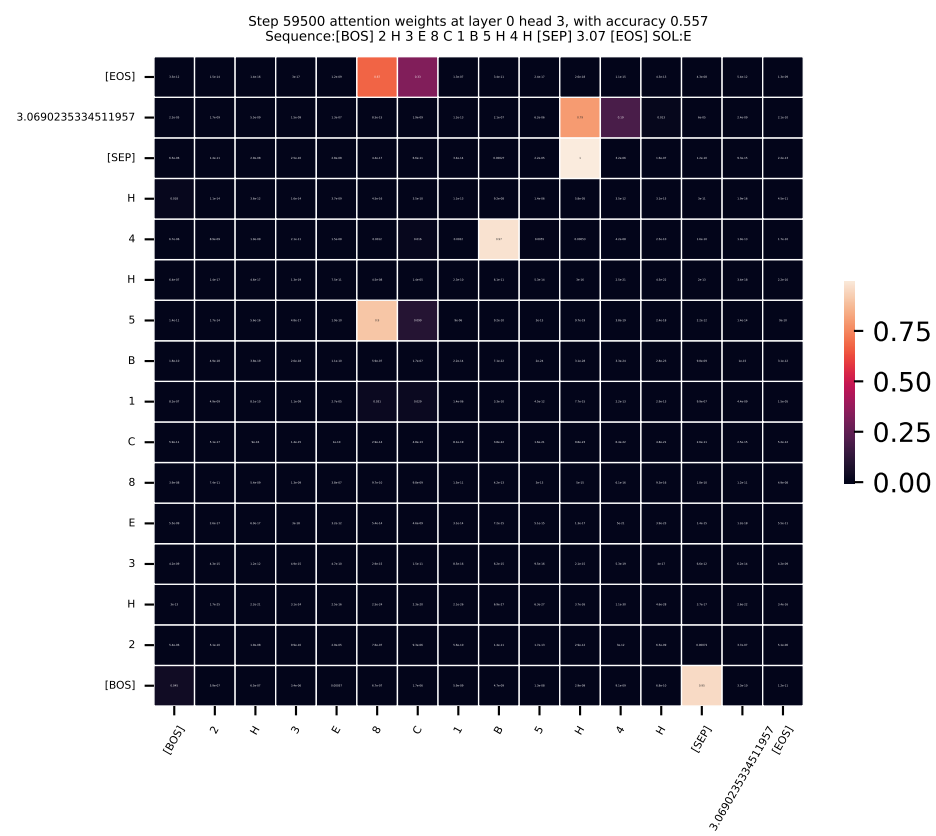

Step 59500 attention weights at layer 0 head 2, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

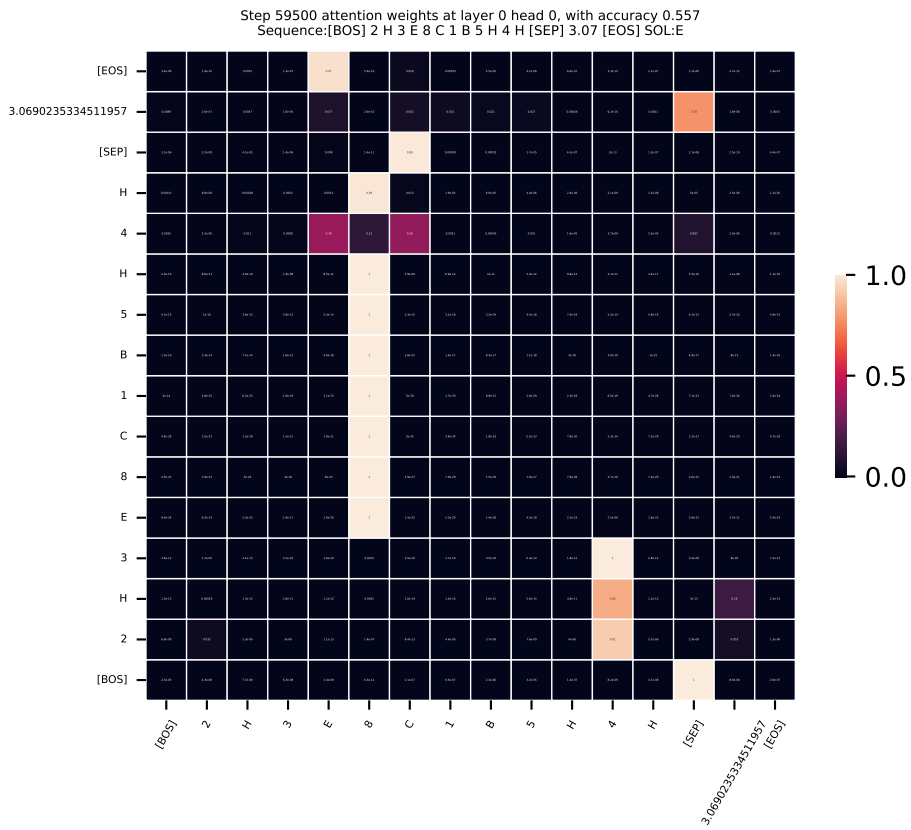

Step 59500 attention weights at layer 0 head 3, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

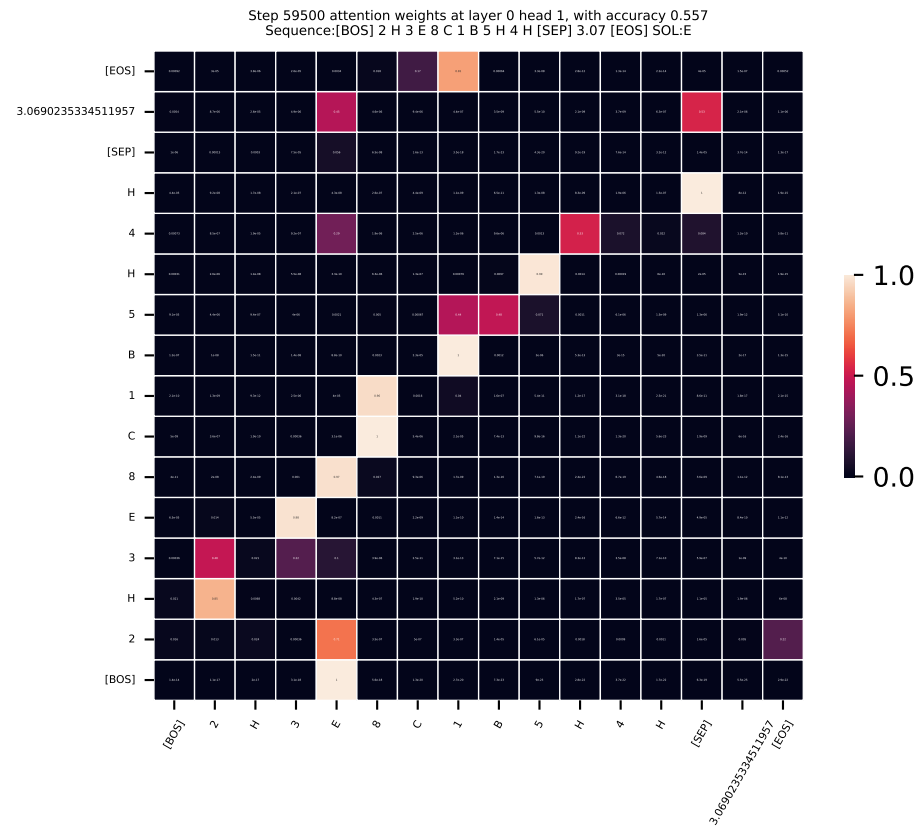

Step 59500 attention weights at layer 0 head 0, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

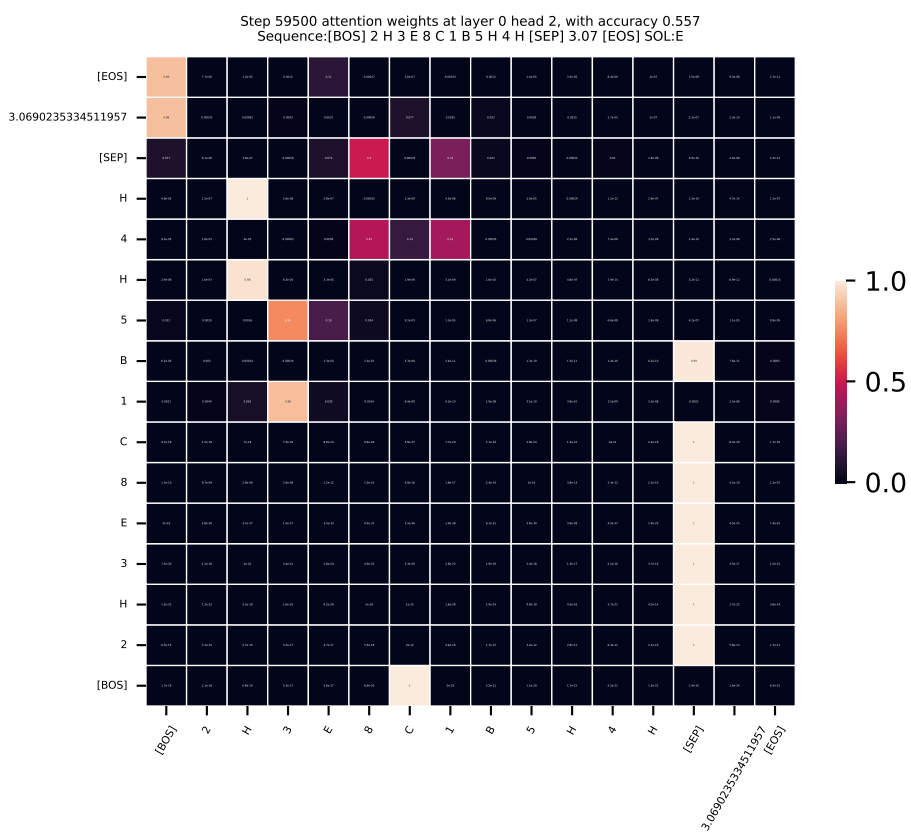

Step 59500 attention weights at layer 0 head 1, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

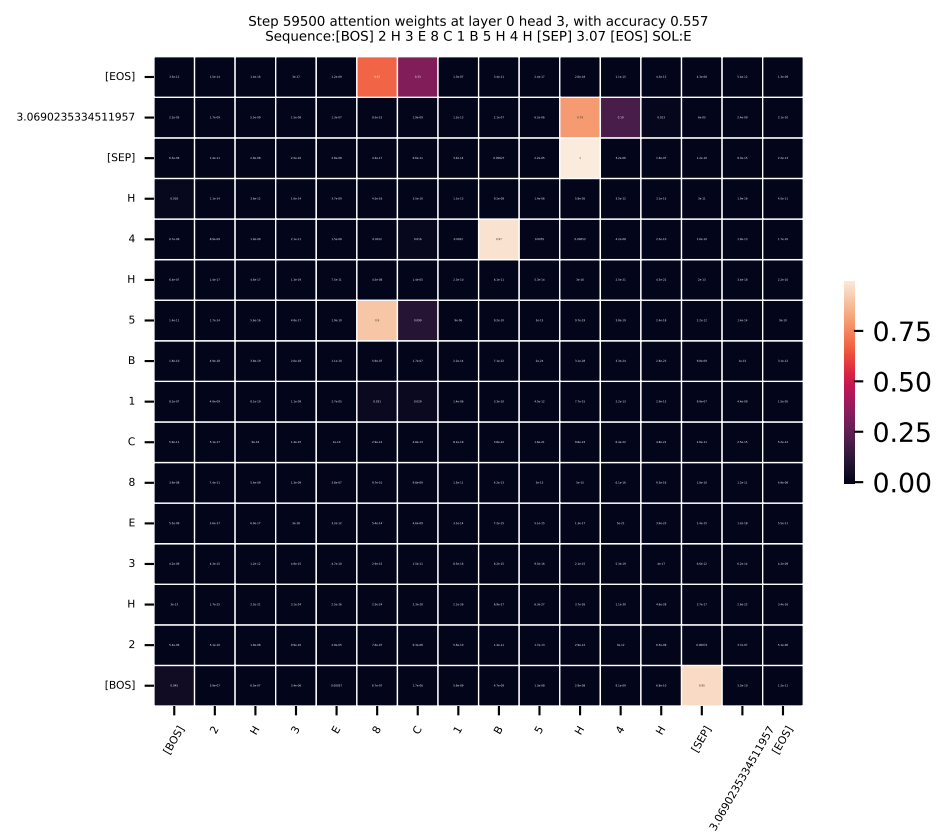

Step 59500 attention weights at layer 0 head 2, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

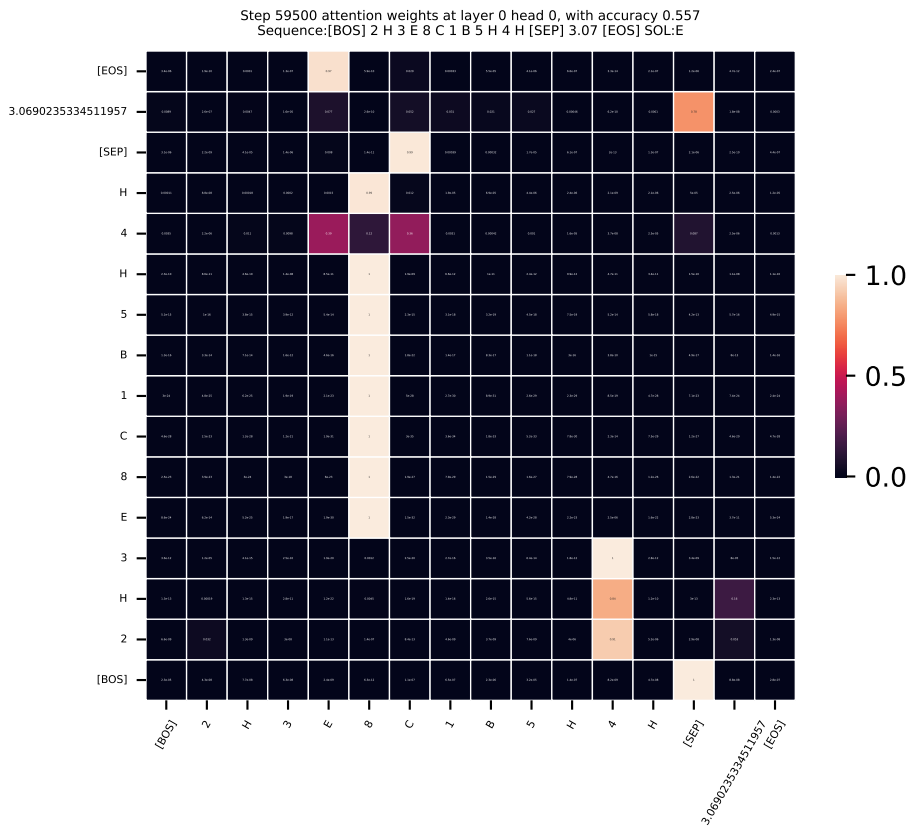

Step 59500 attention weights at layer 0 head 3, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

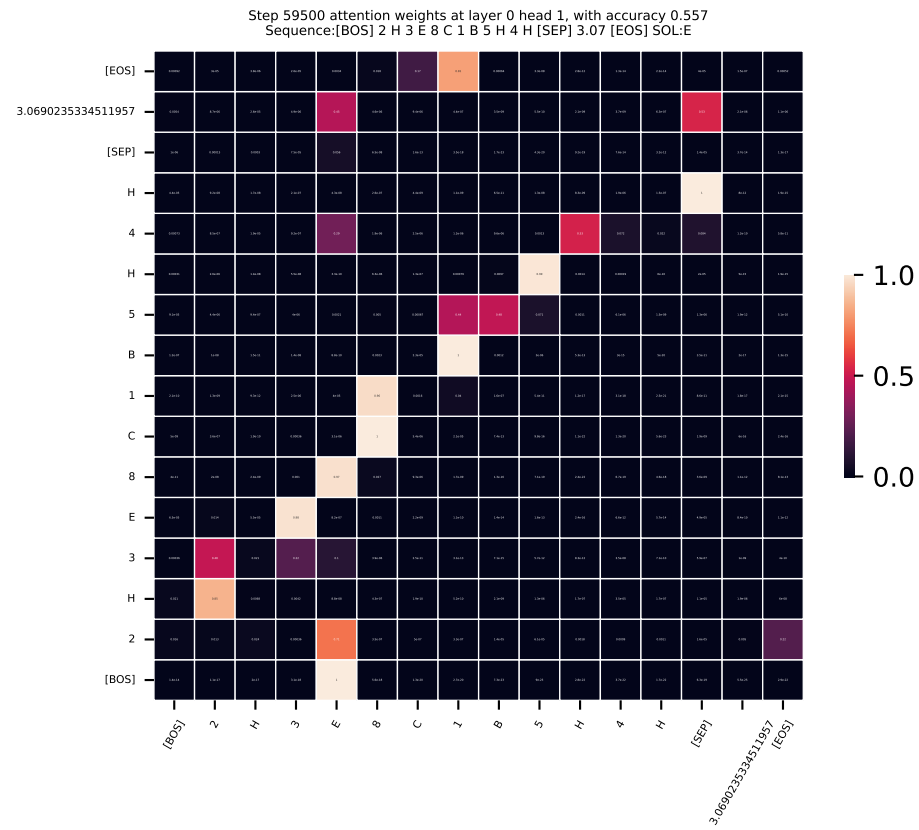

Step 59500 attention weights at layer 0 head 0, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

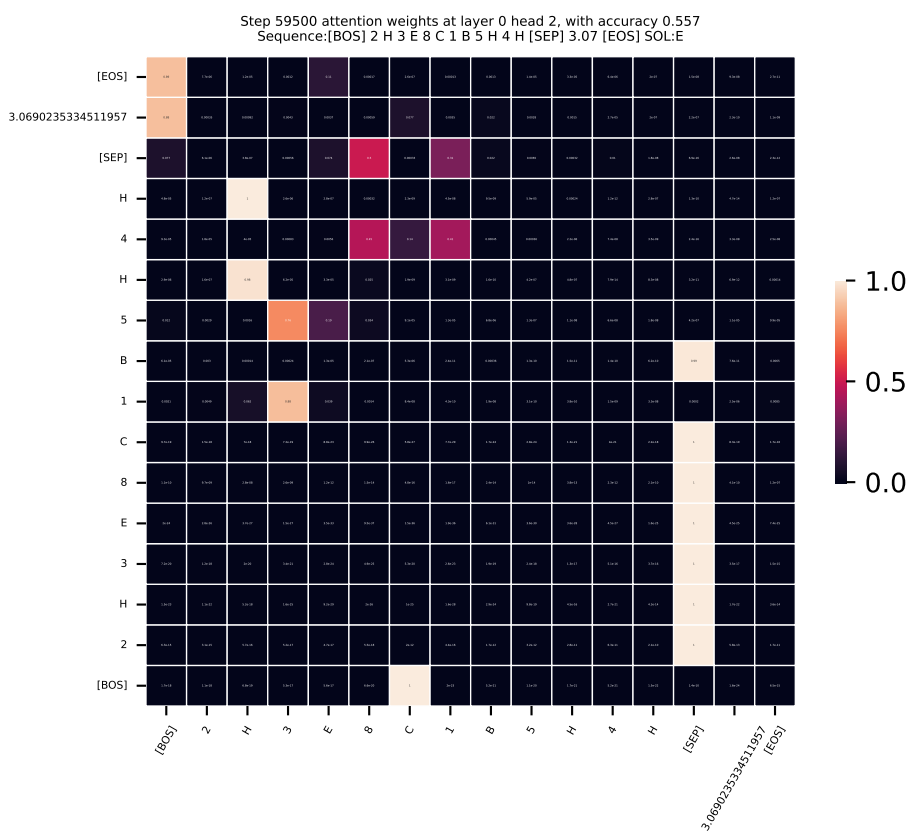

Step 59500 attention weights at layer 0 head 1, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

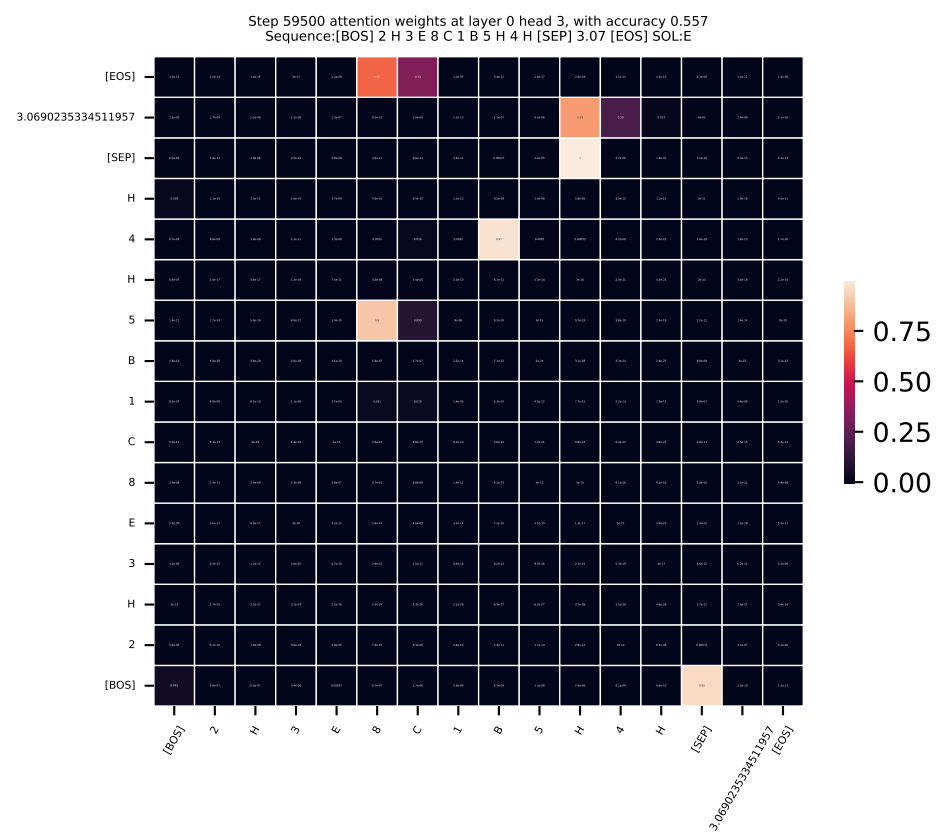

Step 59500 attention weights at layer 0 head 2, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

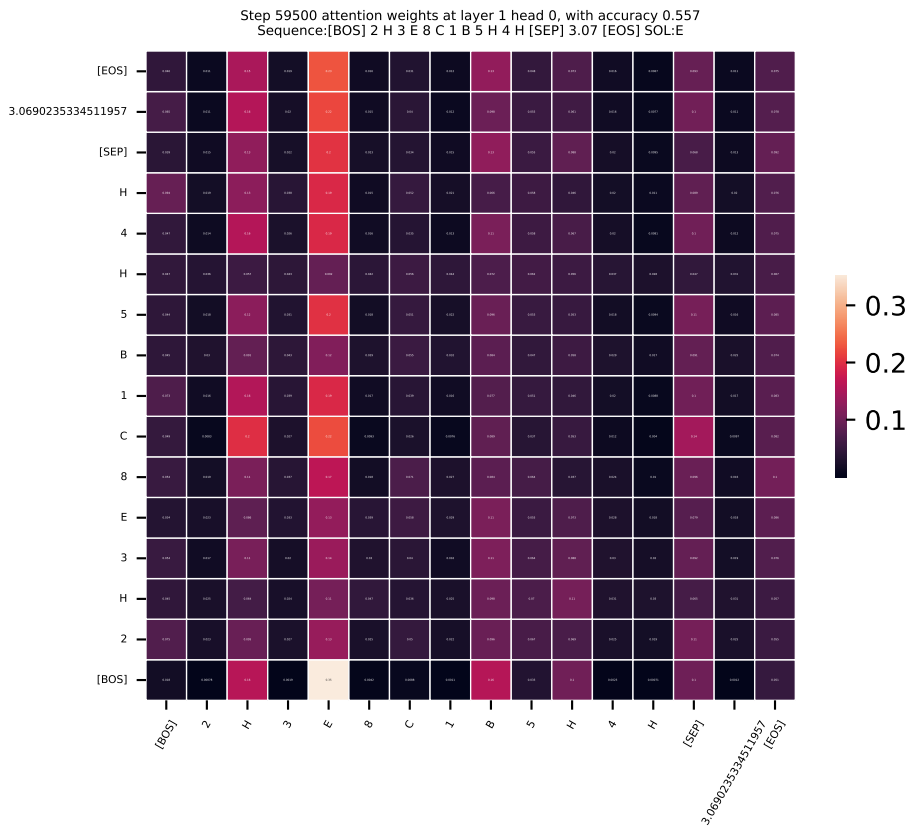

Step 59500 attention weights at layer 0 head 3, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

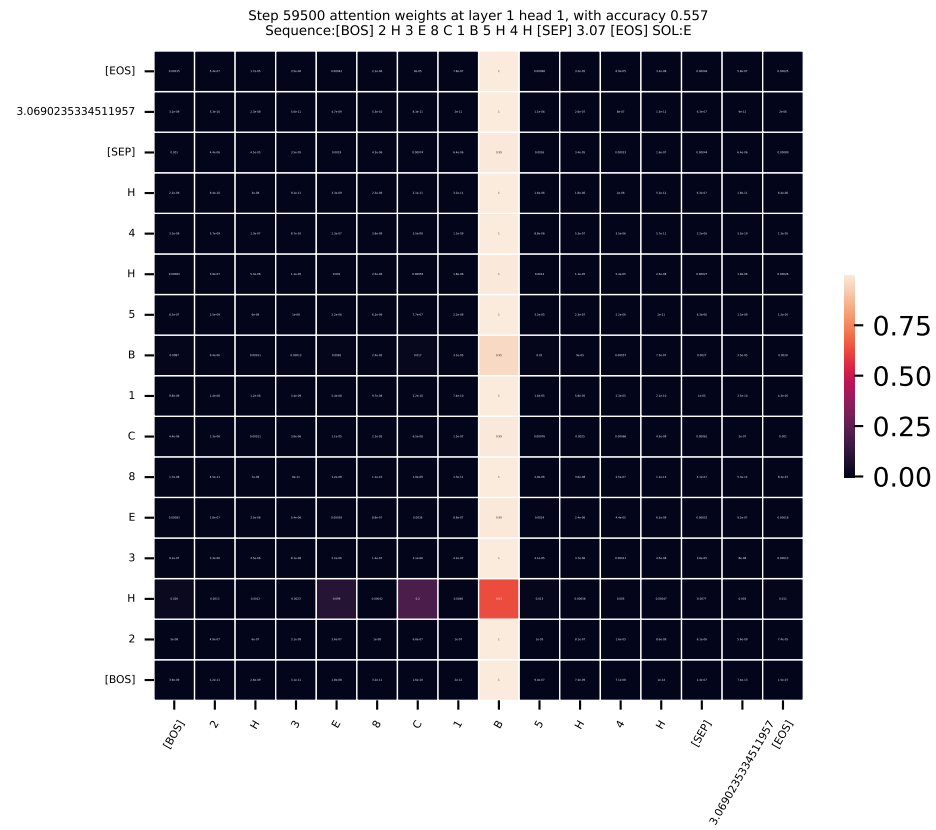

Step 59500 attention weights at layer 0 head 0, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

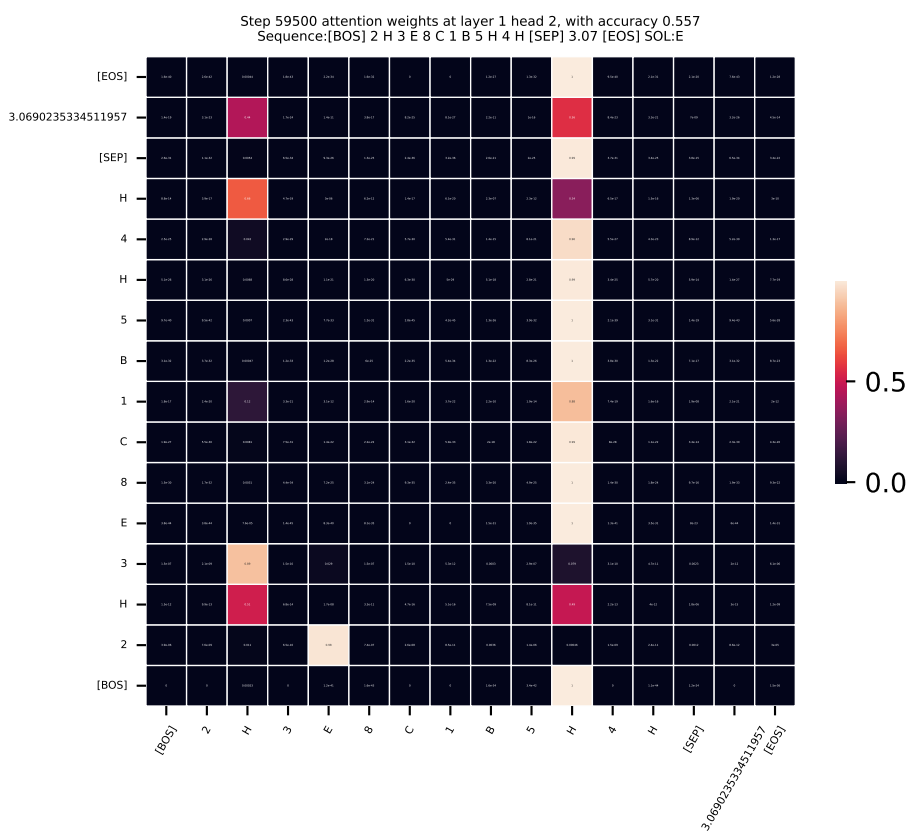

Step 59500 attention weights at layer 0 head 1, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

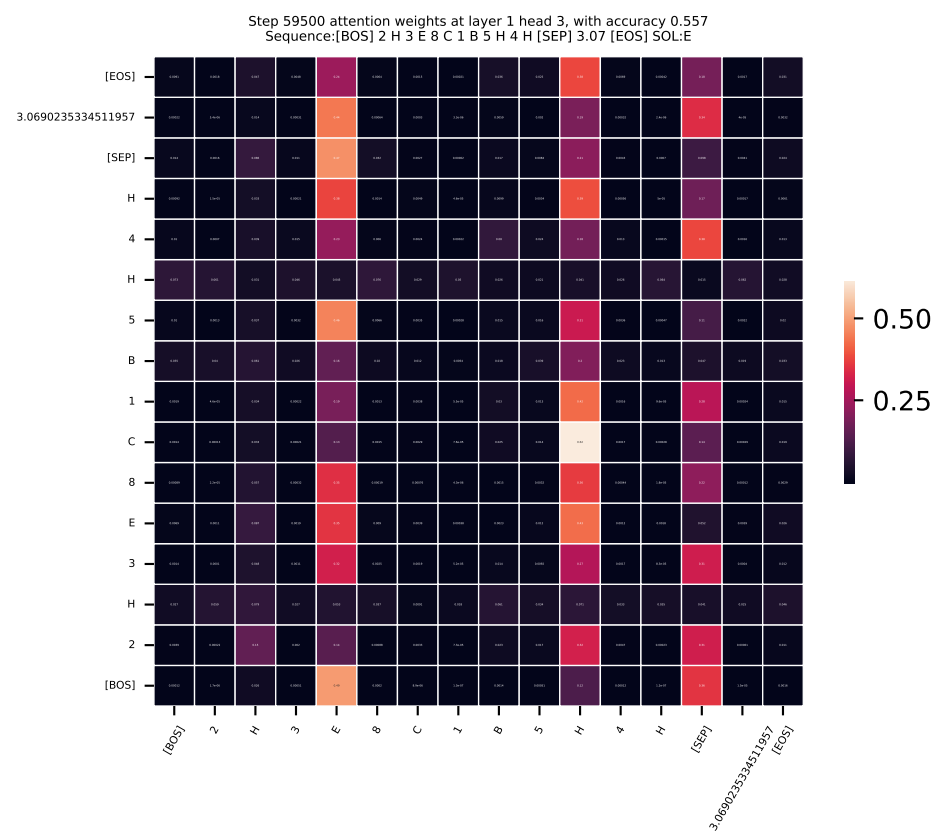

Step 59500 attention weights at layer 0 head 2, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

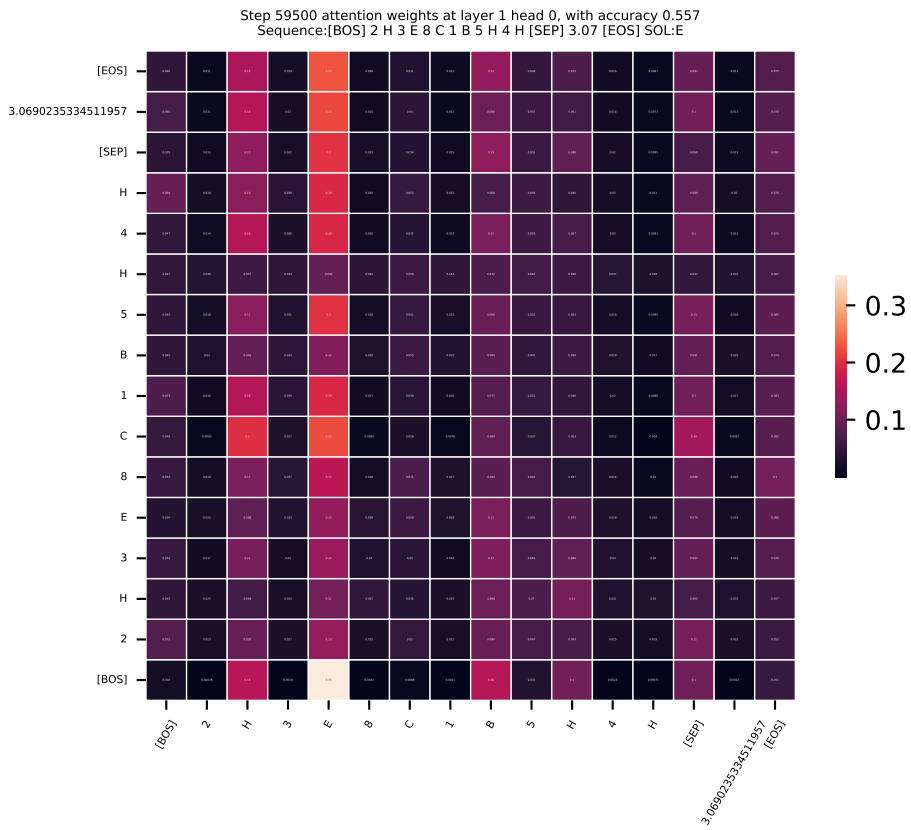

Step 59500 attention weights at layer 0 head 3, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

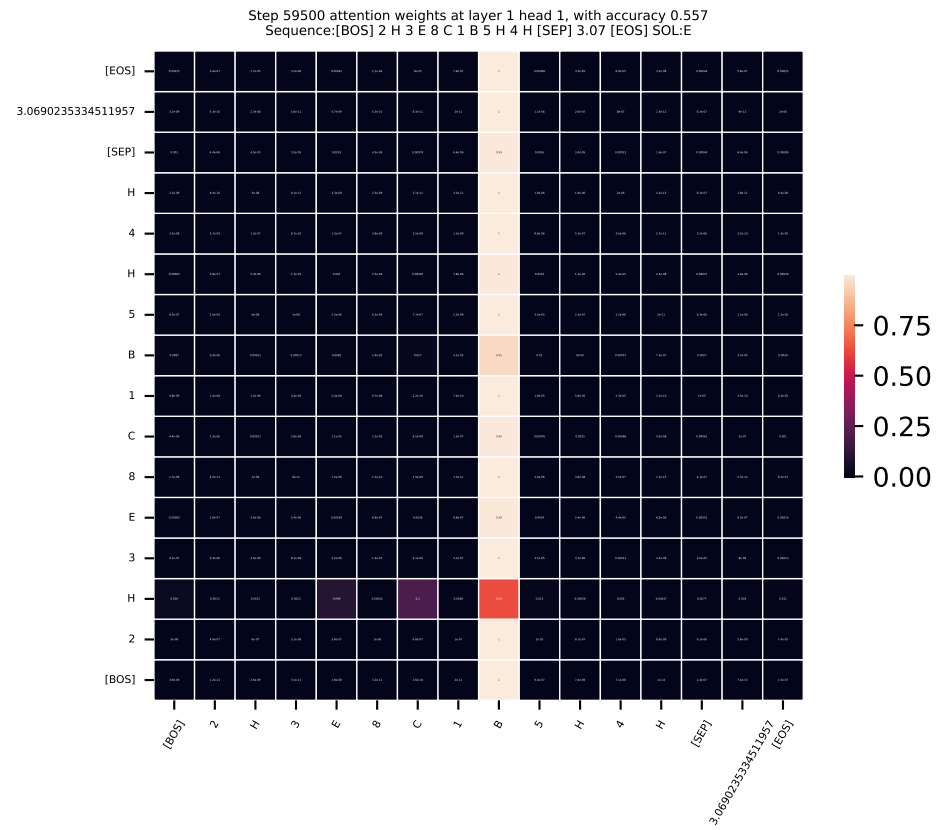

Step 59500 attention weights at layer 1 head 0, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.3
0.2
0.1

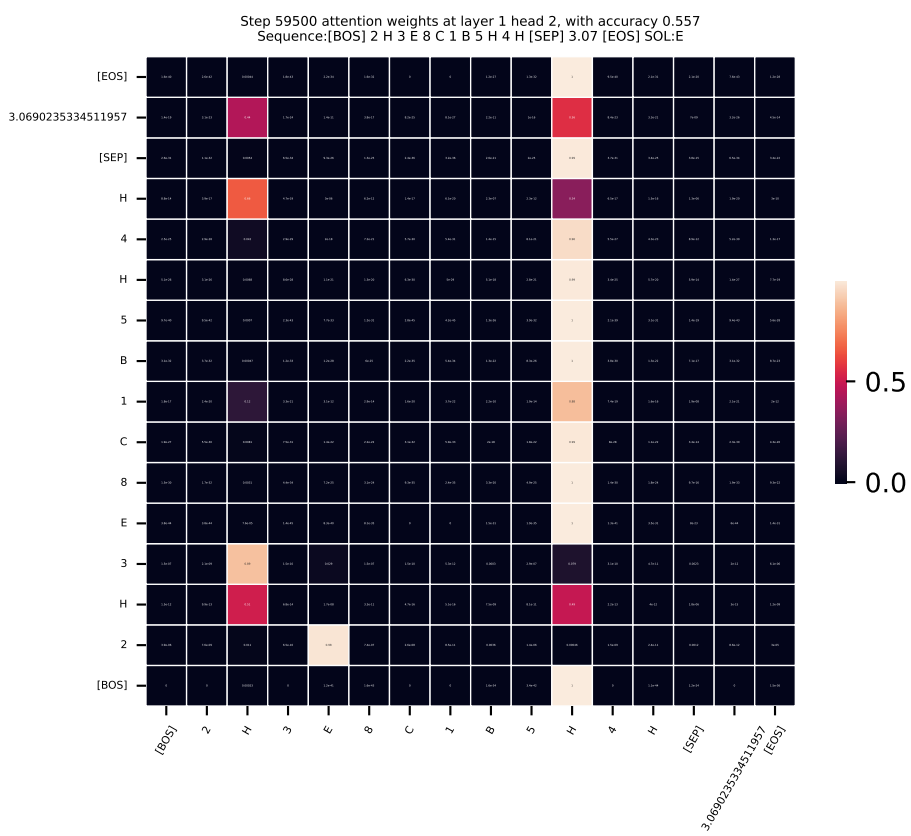

Step 59500 attention weights at layer 1 head 1, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

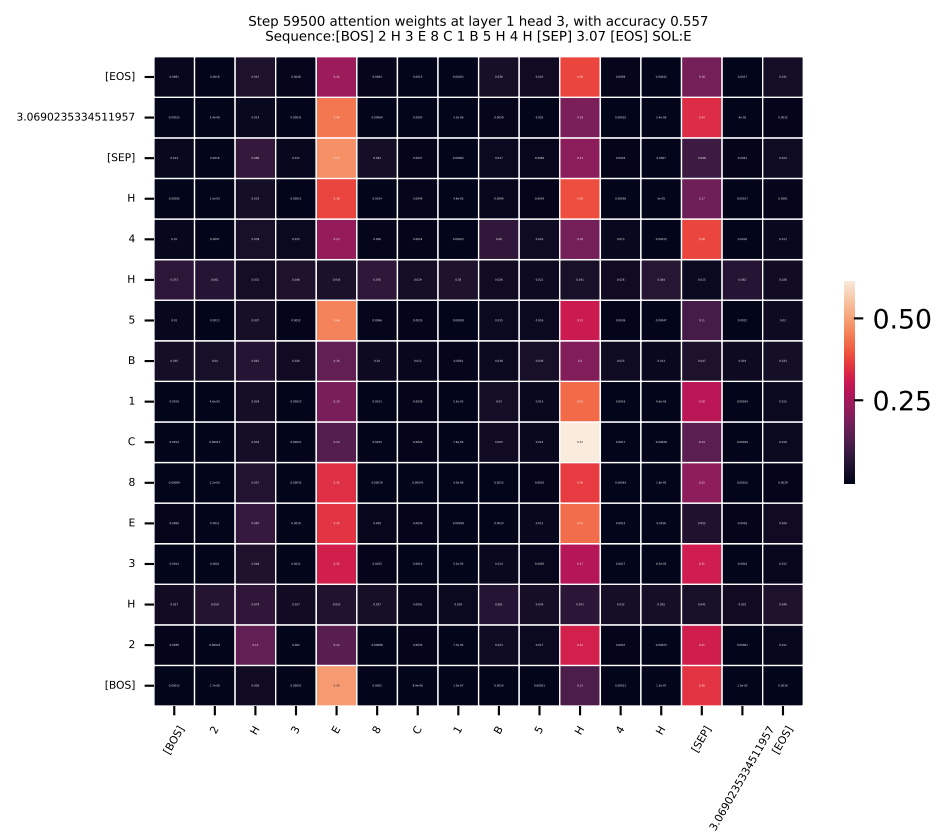

Step 59500 attention weights at layer 1 head 2, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.5
0.0

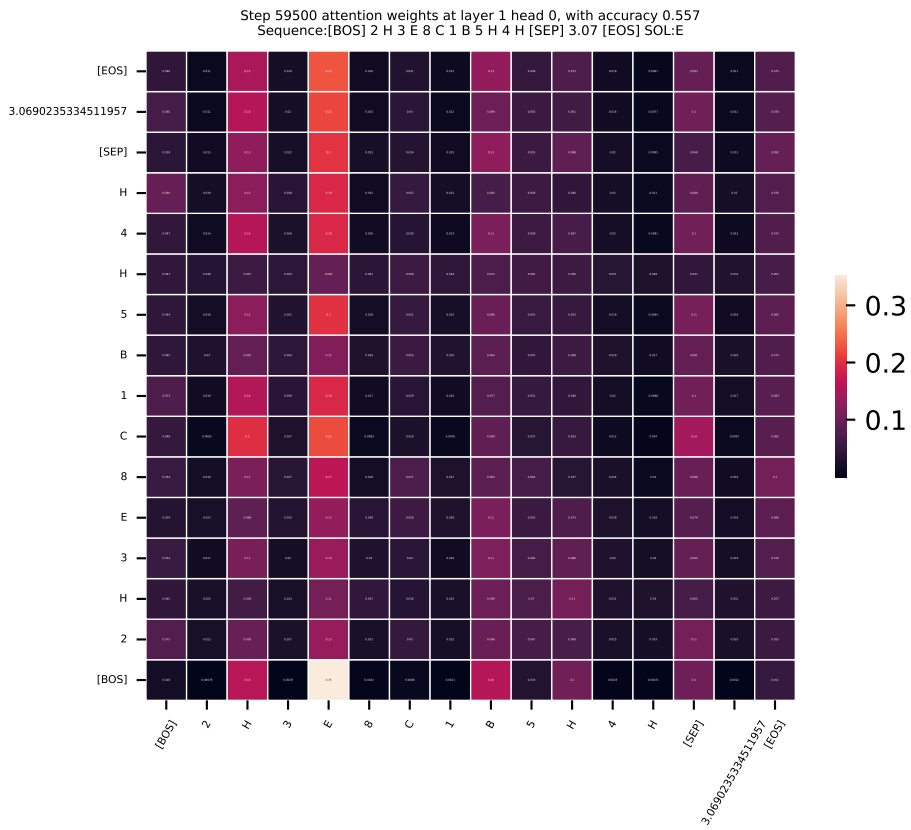

Step 59500 attention weights at layer 1 head 3, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.50
0.25

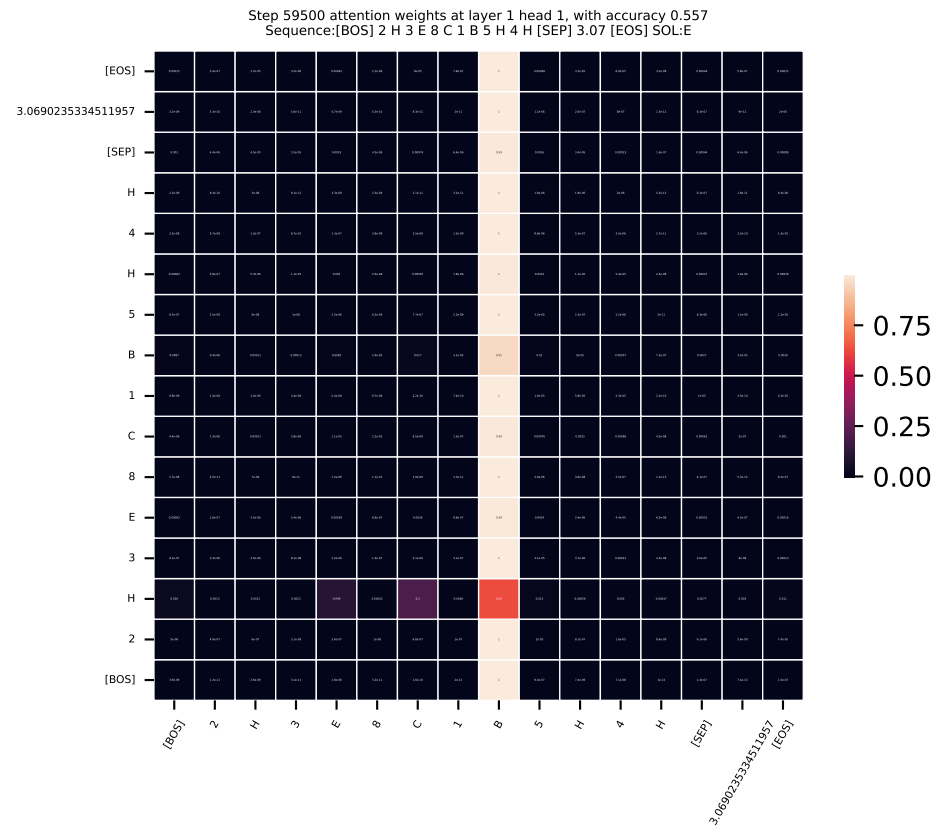

Step 59500 attention weights at layer 1 head 0, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.3
0.2
0.1

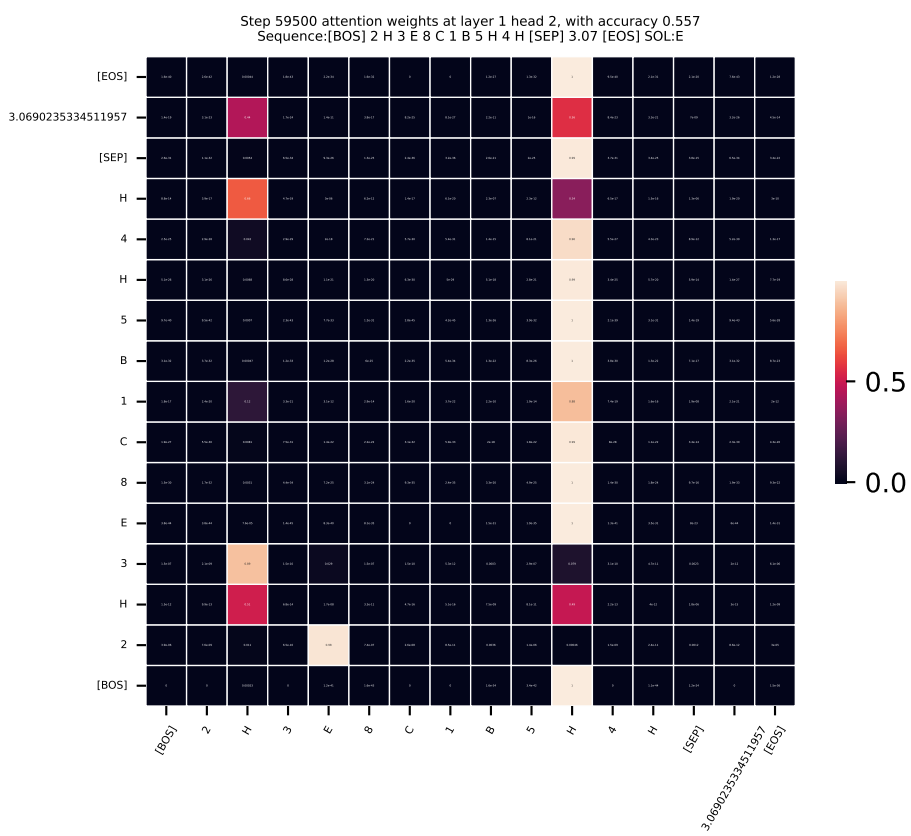

Step 59500 attention weights at layer 1 head 1, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

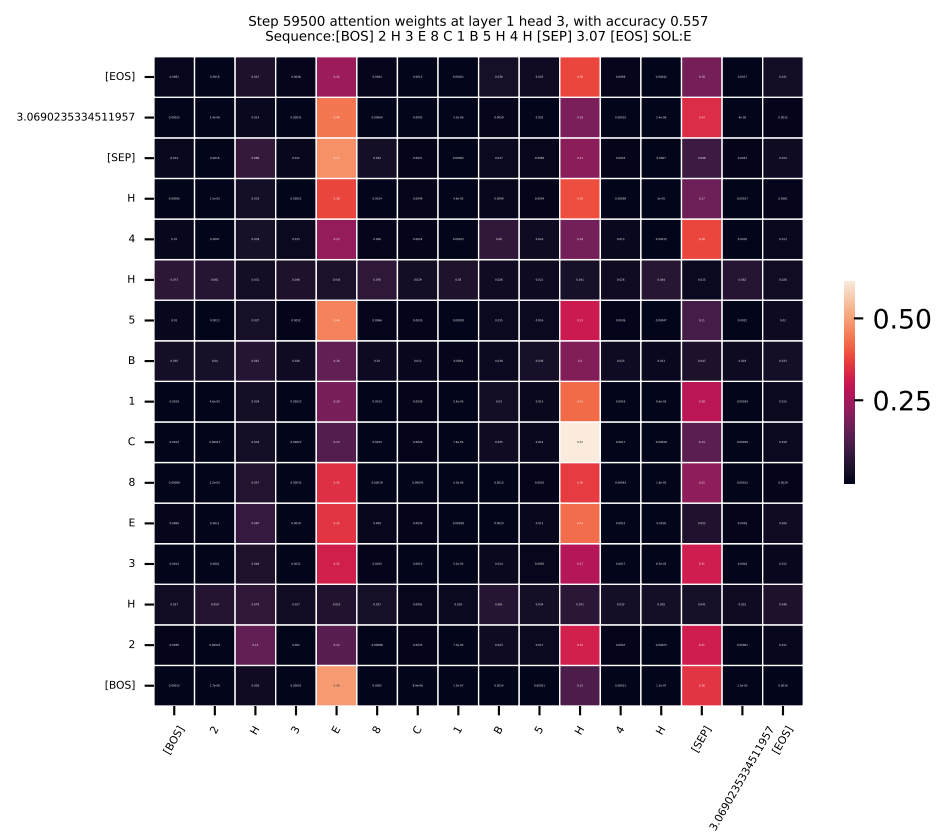

Step 59500 attention weights at layer 1 head 2, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.5
0.0

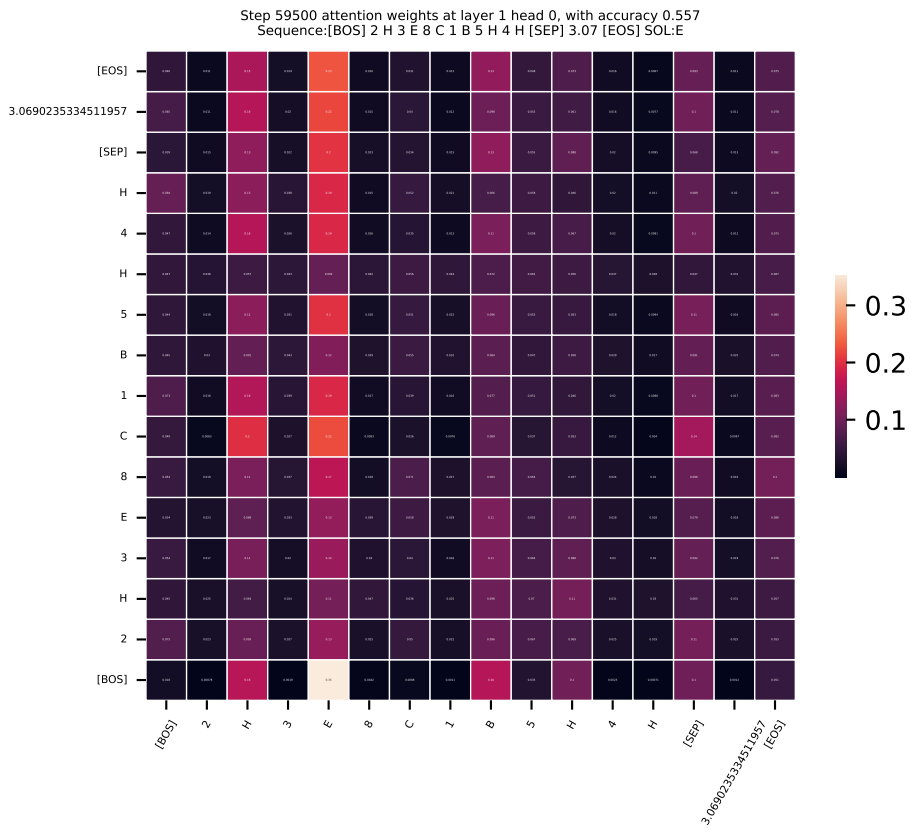

Step 59500 attention weights at layer 1 head 3, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.50
0.25

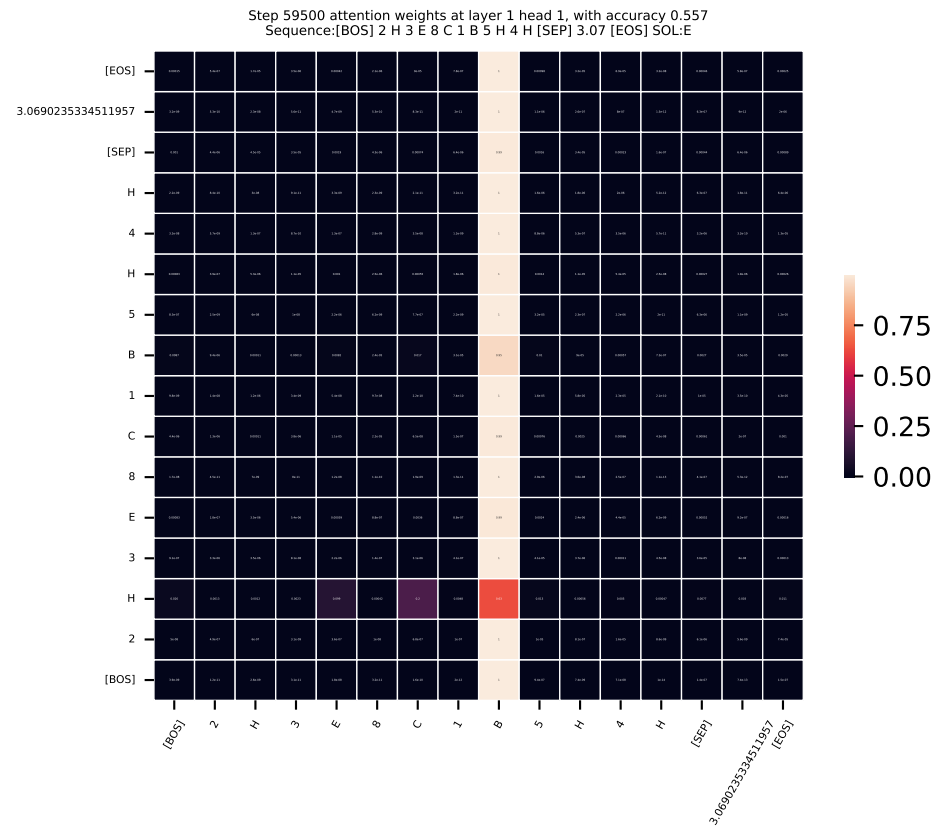

Step 59500 attention weights at layer 1 head 0, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.3
0.2
0.1

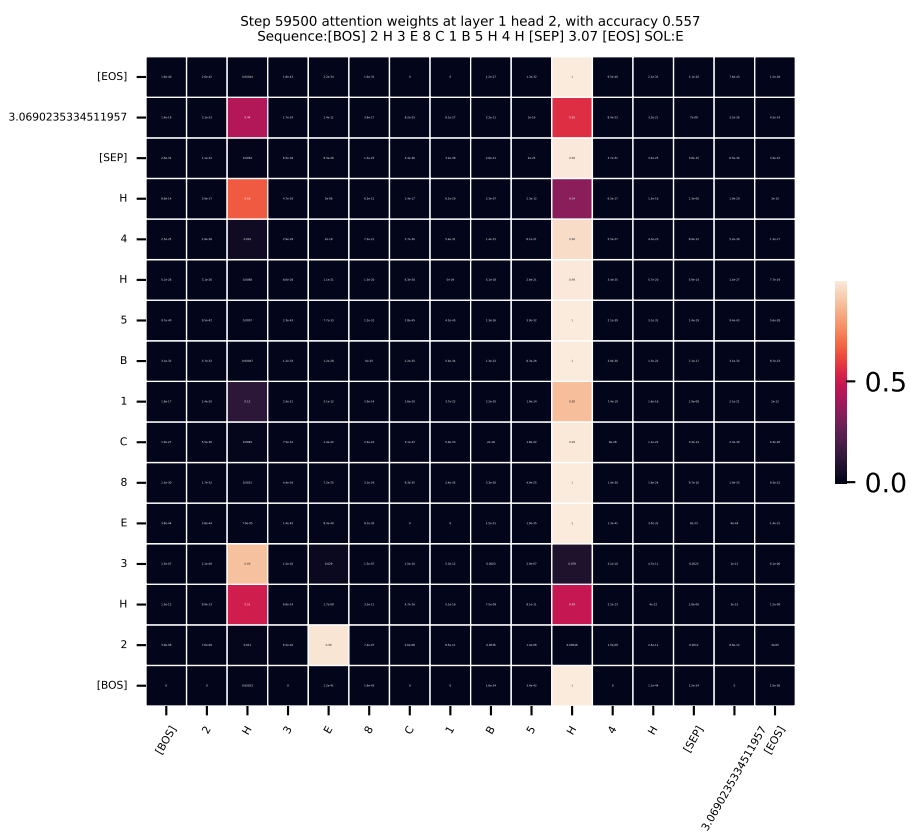

Step 59500 attention weights at layer 1 head 1, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

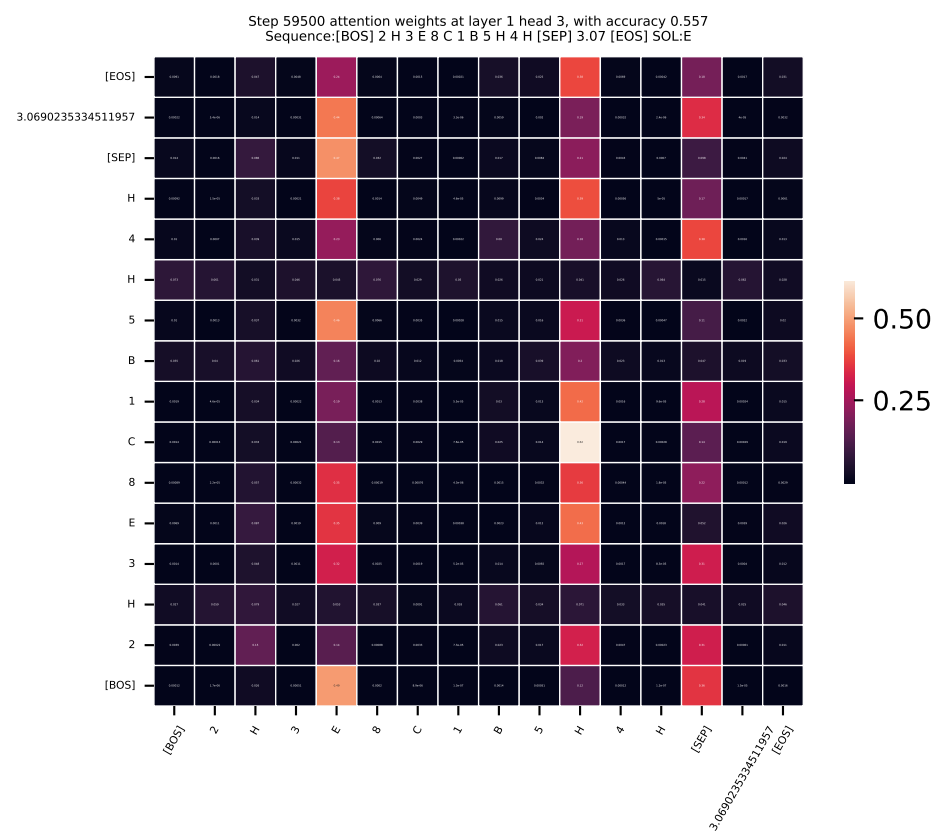

Step 59500 attention weights at layer 1 head 2, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.5
0.0

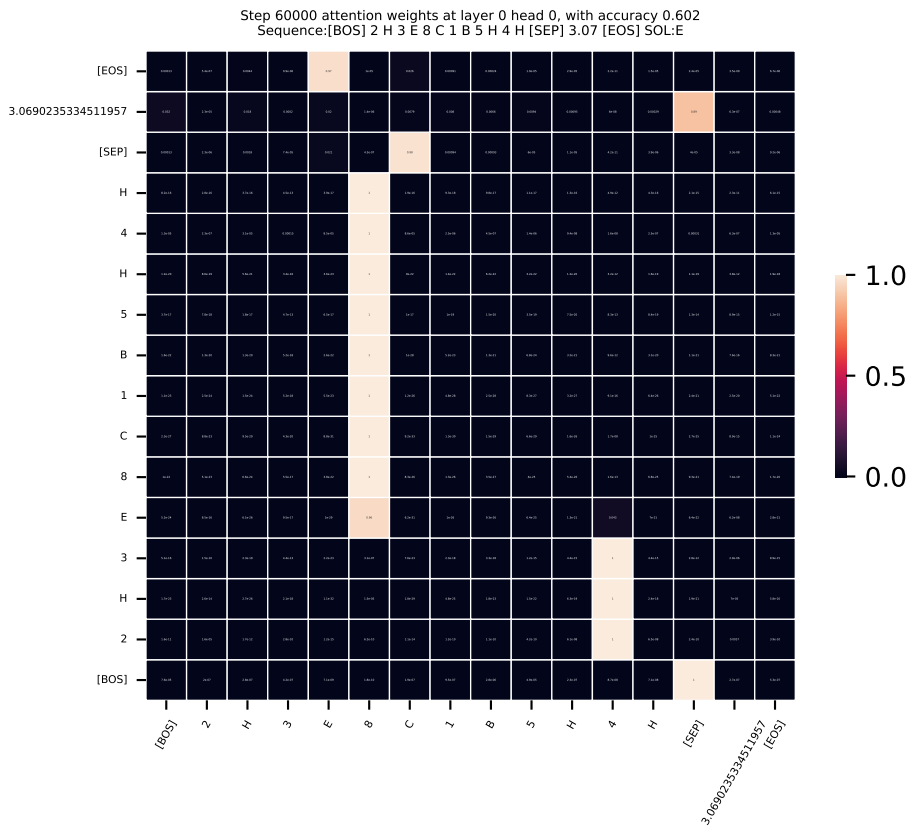

Step 59500 attention weights at layer 1 head 3, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.50
0.25

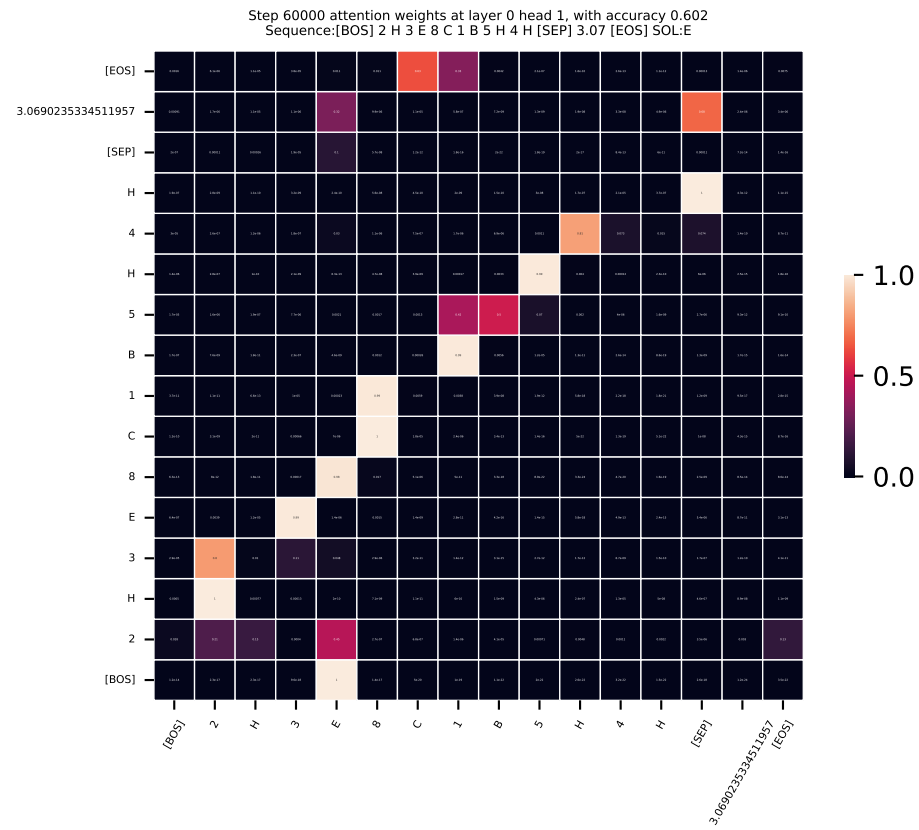

Step 59500 attention weights at layer 1 head 0, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.3
0.2
0.1

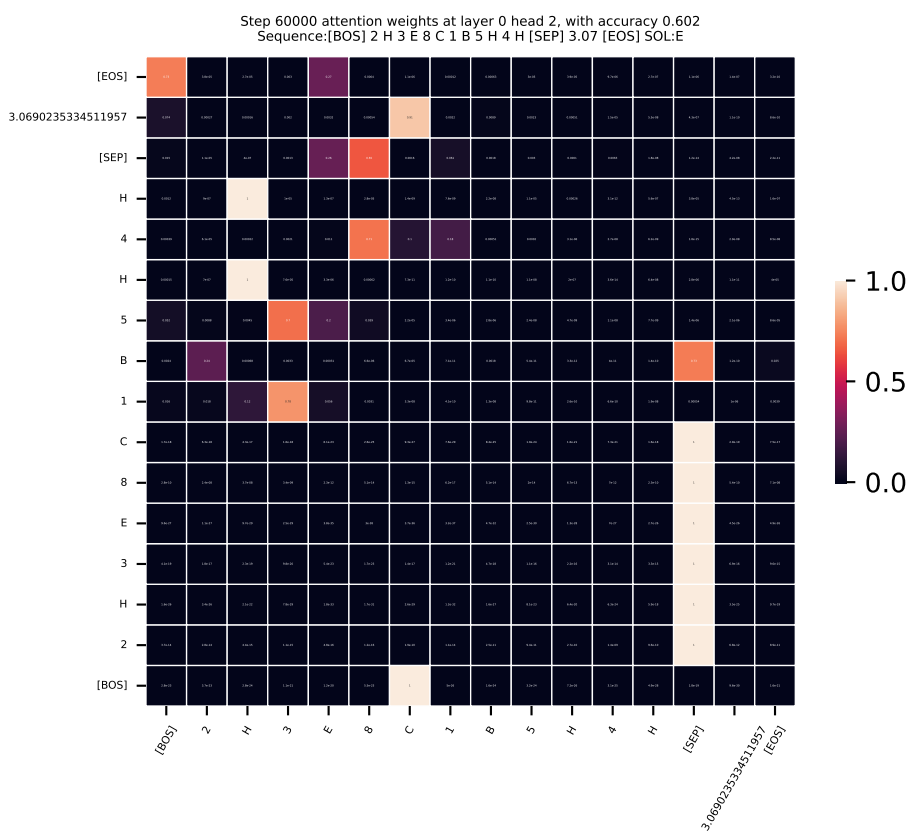

Step 59500 attention weights at layer 1 head 1, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

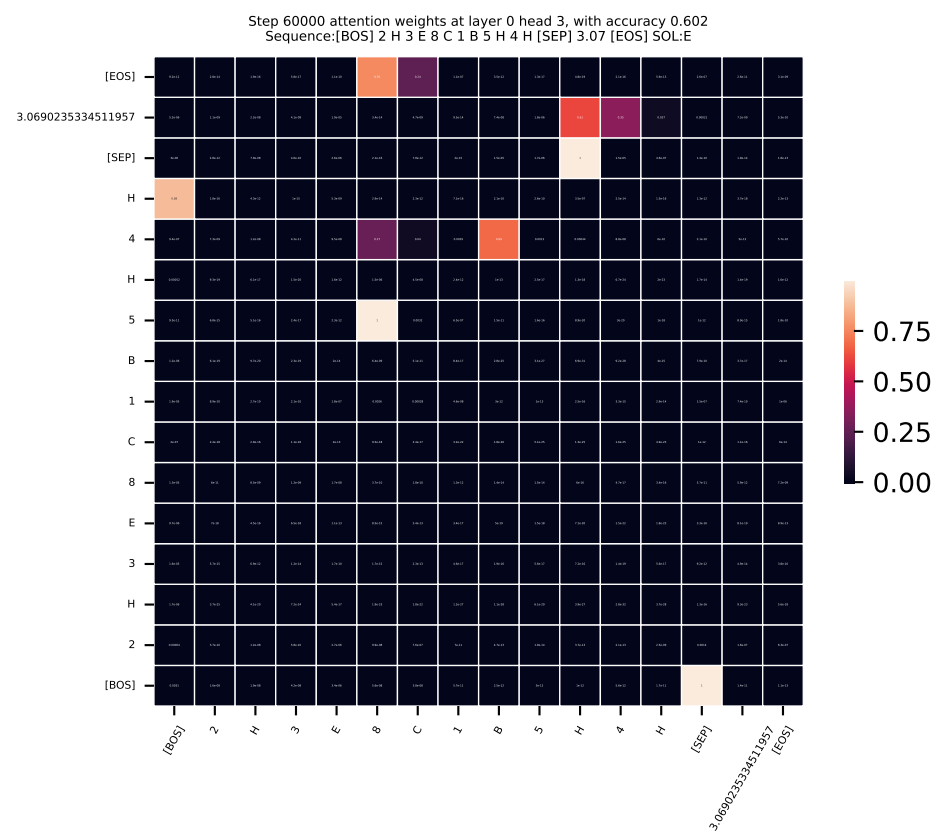

Step 59500 attention weights at layer 1 head 2, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.5
0.0

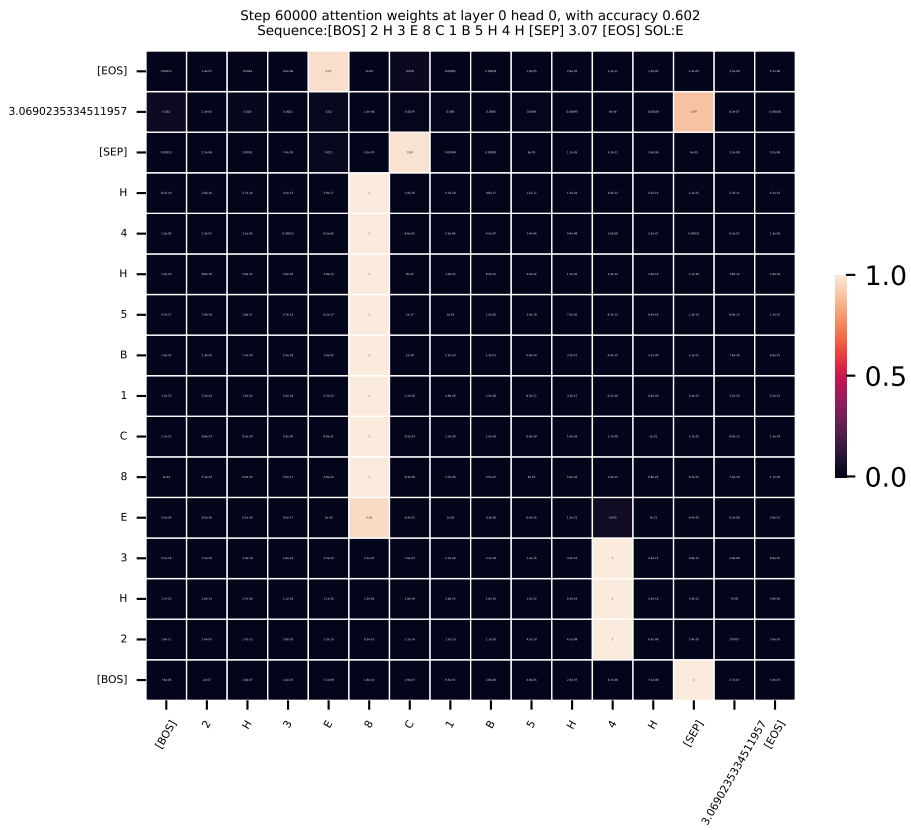

Step 59500 attention weights at layer 1 head 3, with accuracy 0.557
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.50
0.25

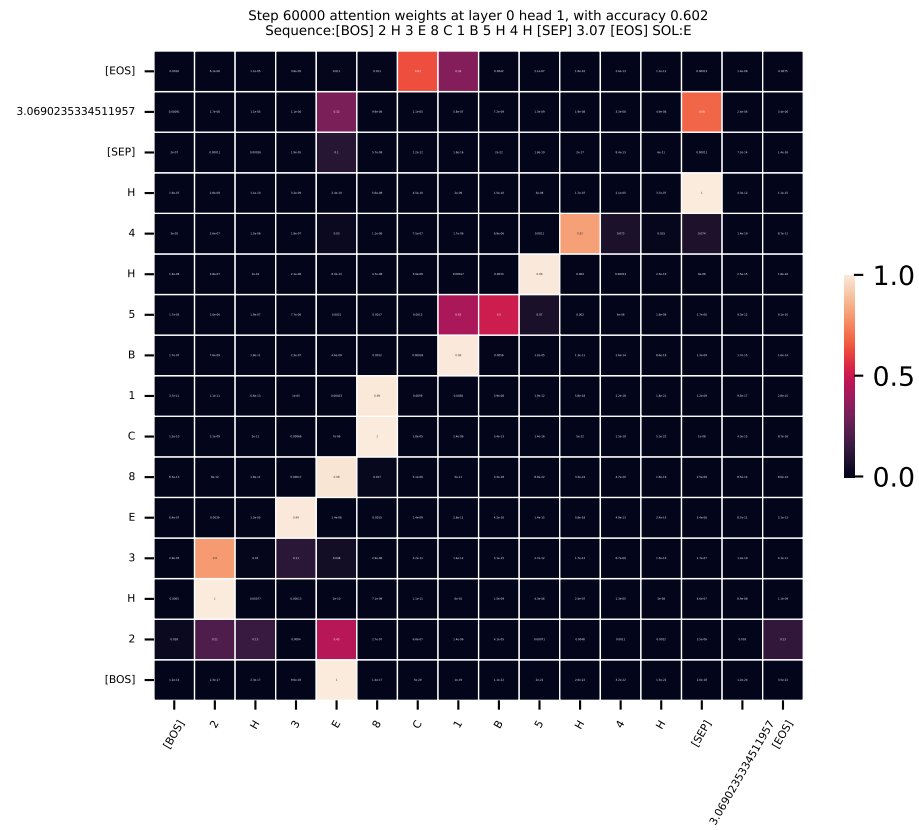

Step 60000 attention weights at layer 0 head 0, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

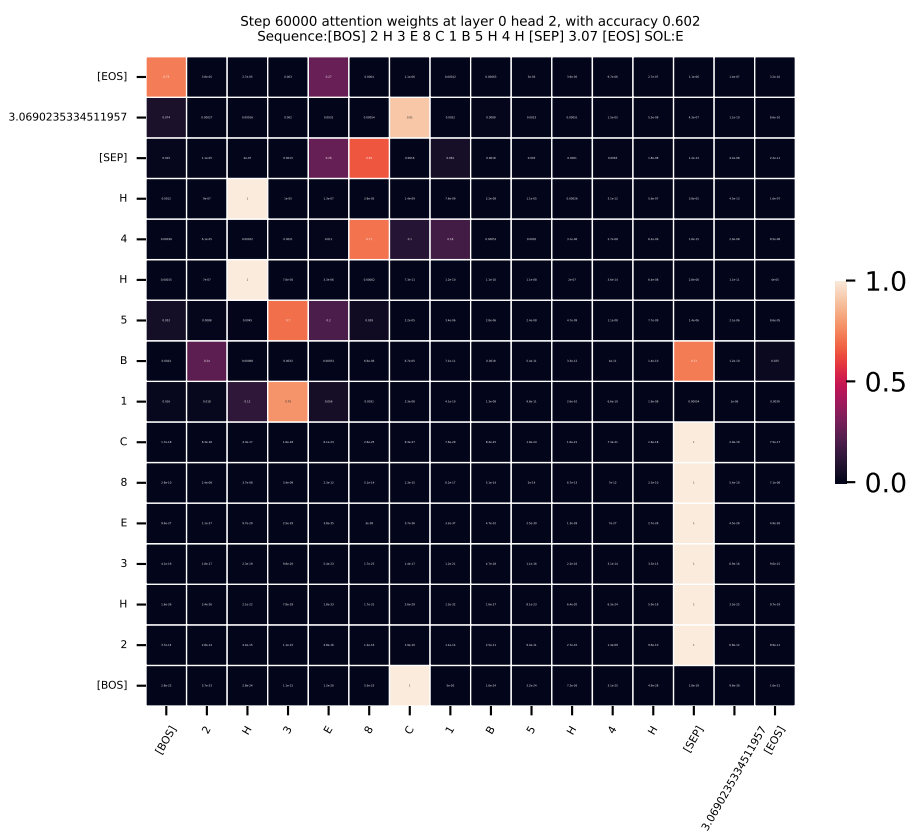

Step 60000 attention weights at layer 0 head 1, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

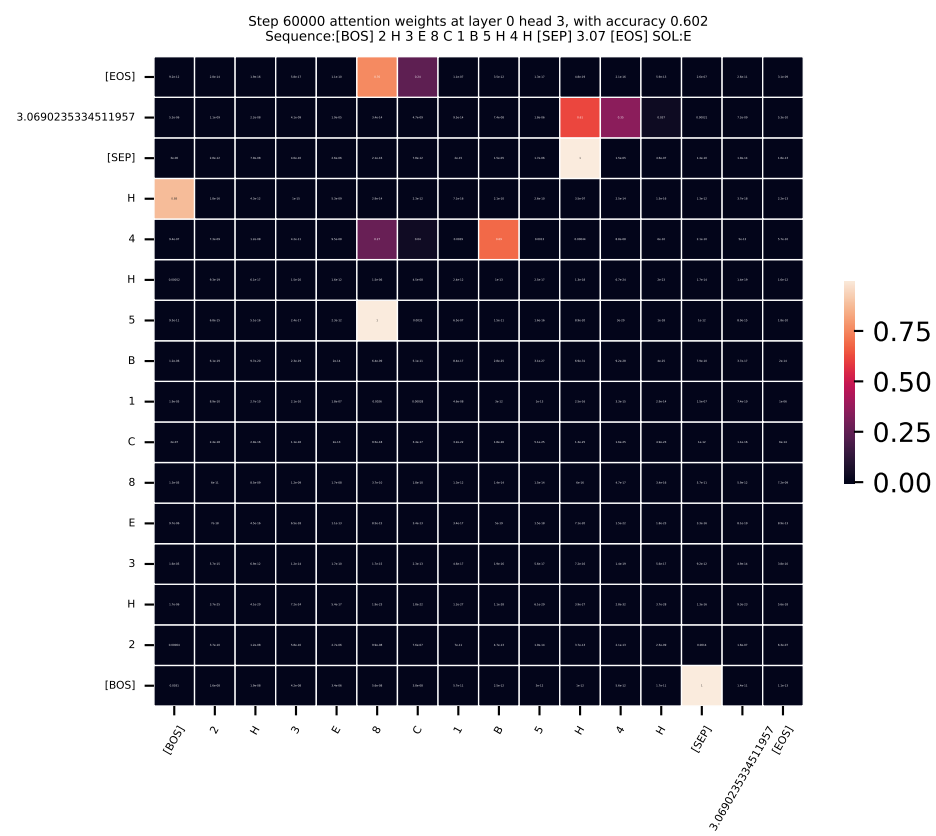

Step 60000 attention weights at layer 0 head 2, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

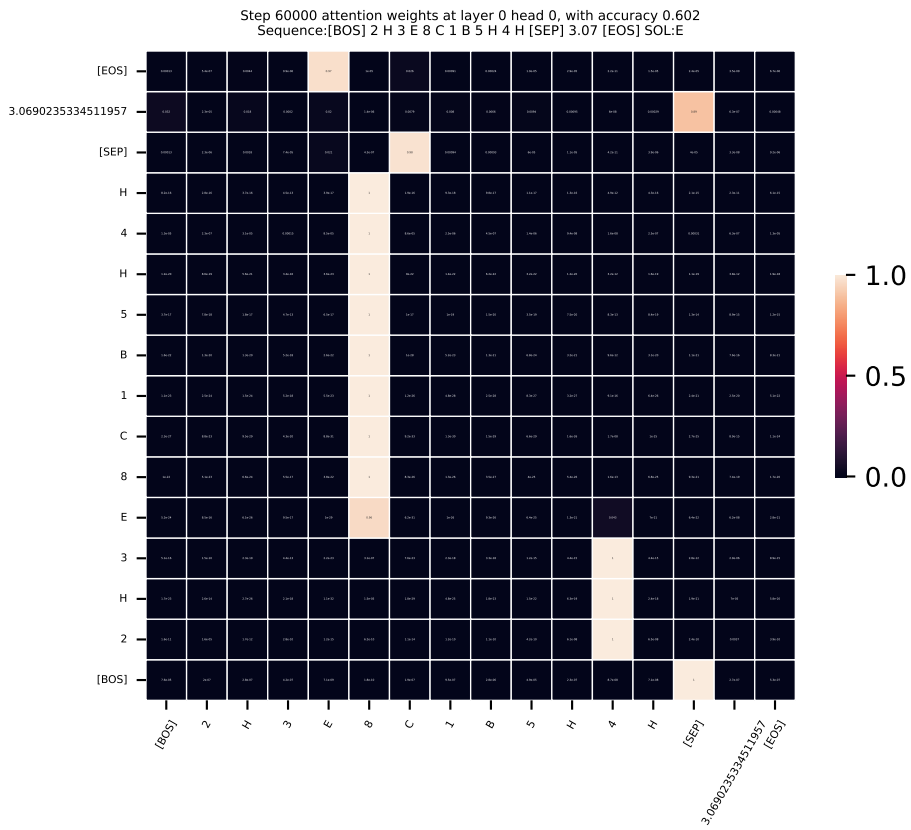

Step 60000 attention weights at layer 0 head 3, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

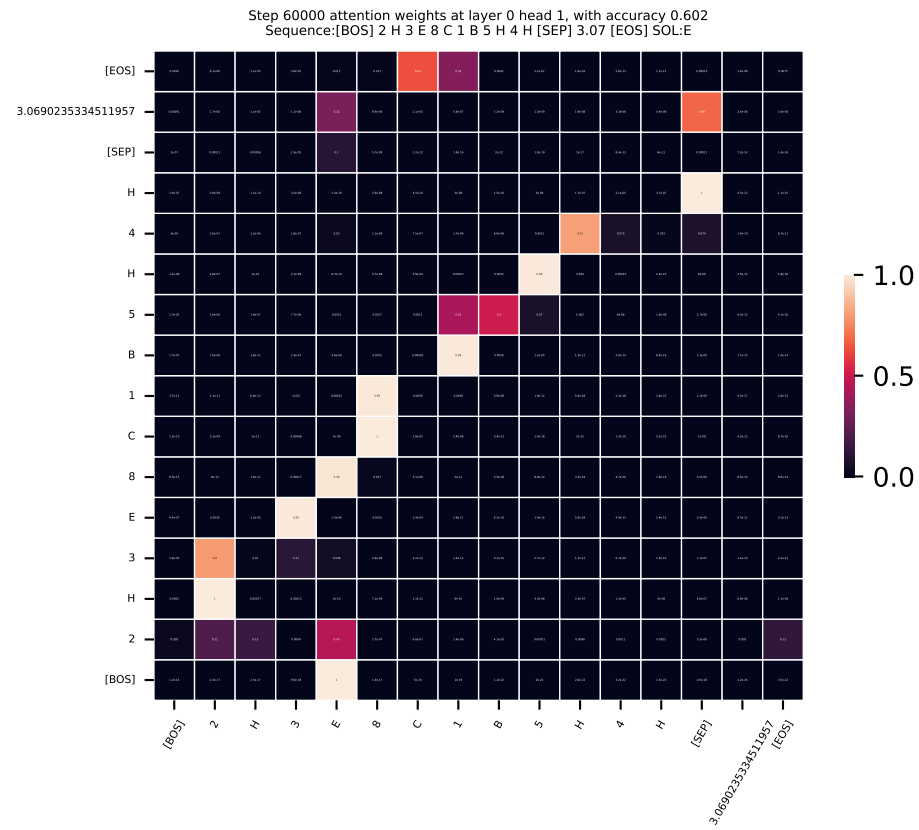

Step 60000 attention weights at layer 0 head 0, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

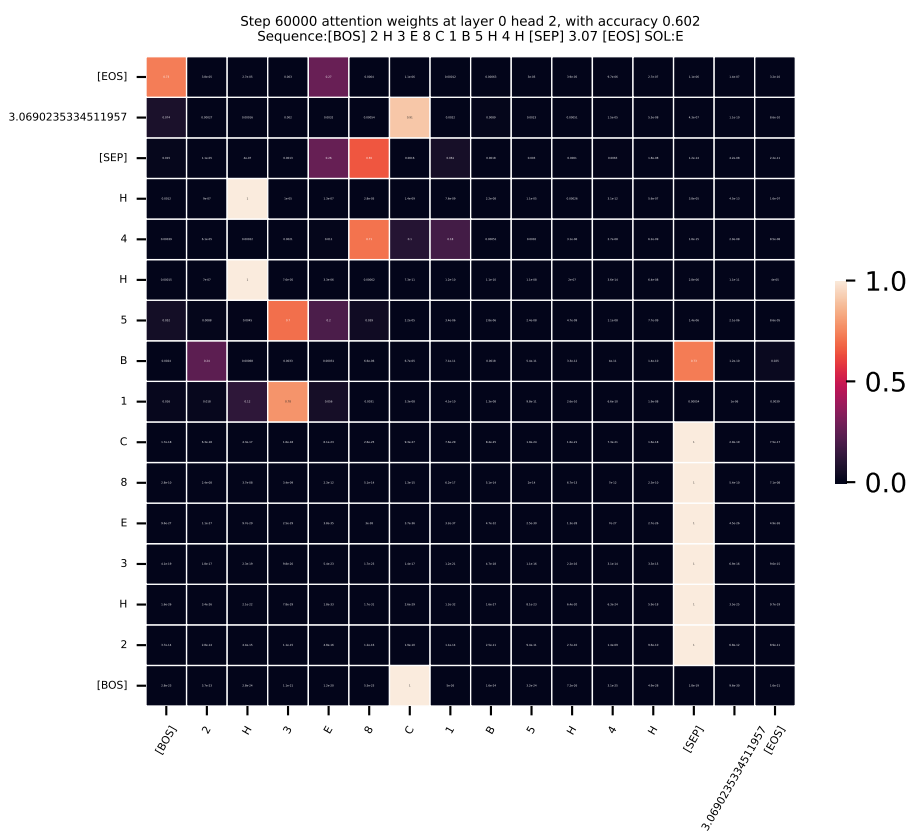

Step 60000 attention weights at layer 0 head 1, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

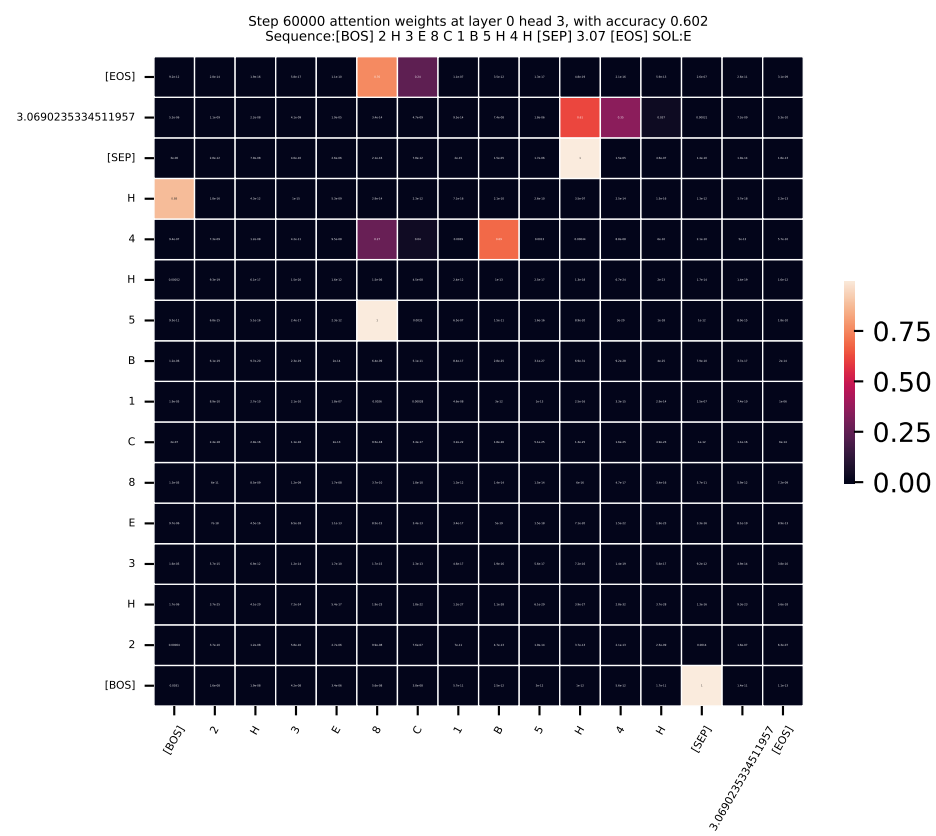

Step 60000 attention weights at layer 0 head 2, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

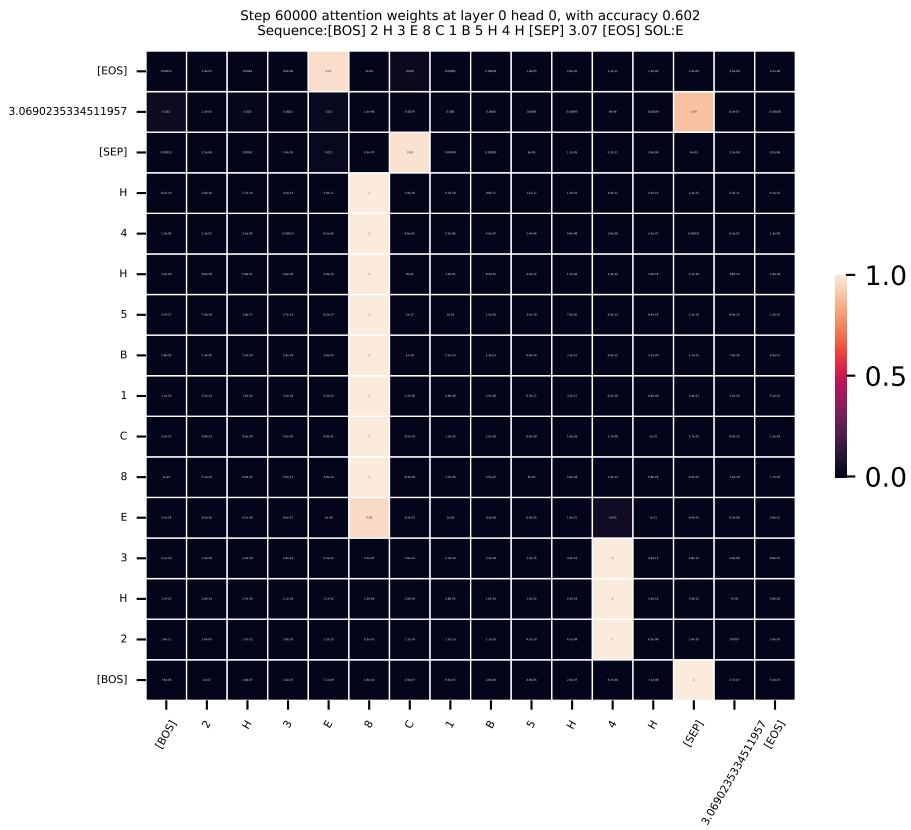

Step 60000 attention weights at layer 0 head 3, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

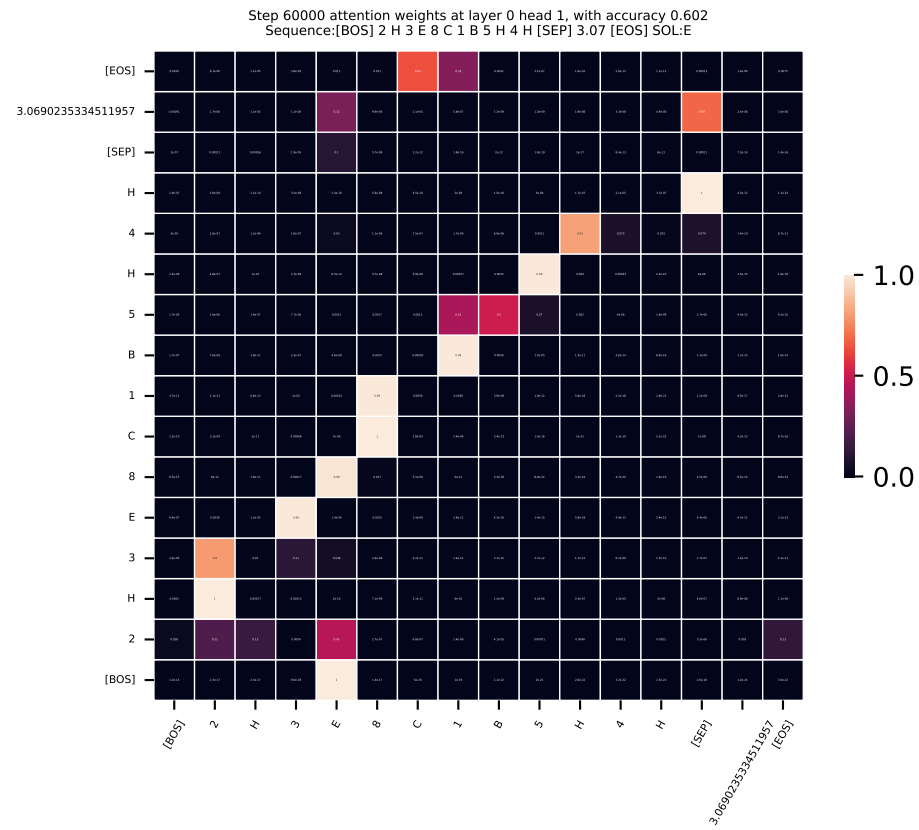

Step 60000 attention weights at layer 0 head 0, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

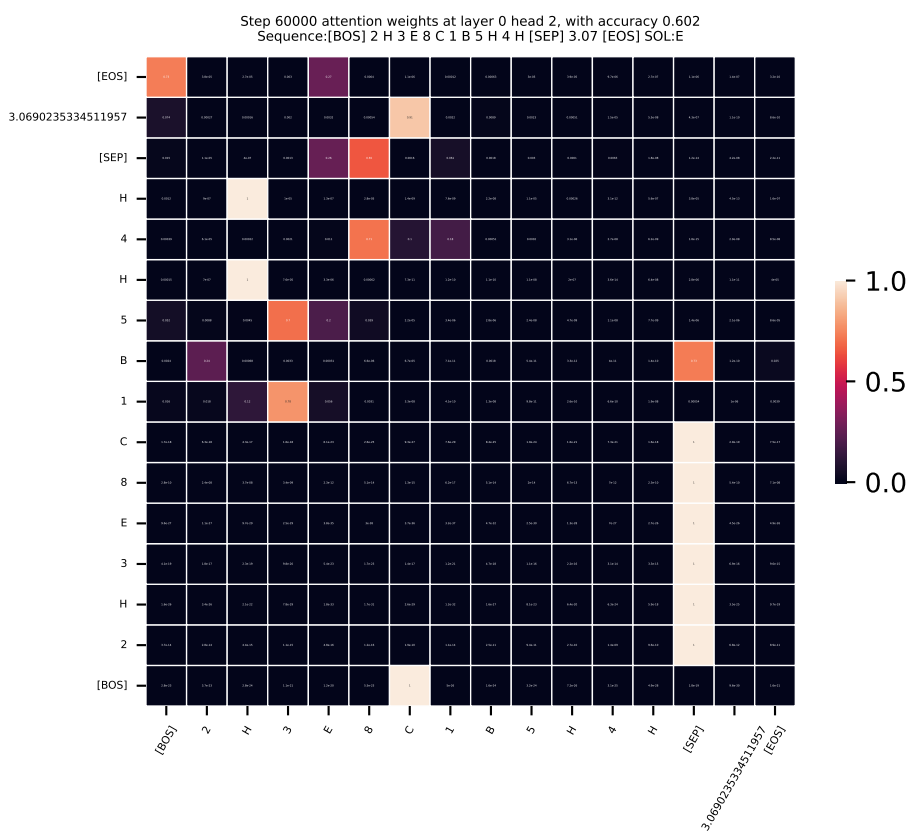

Step 60000 attention weights at layer 0 head 1, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

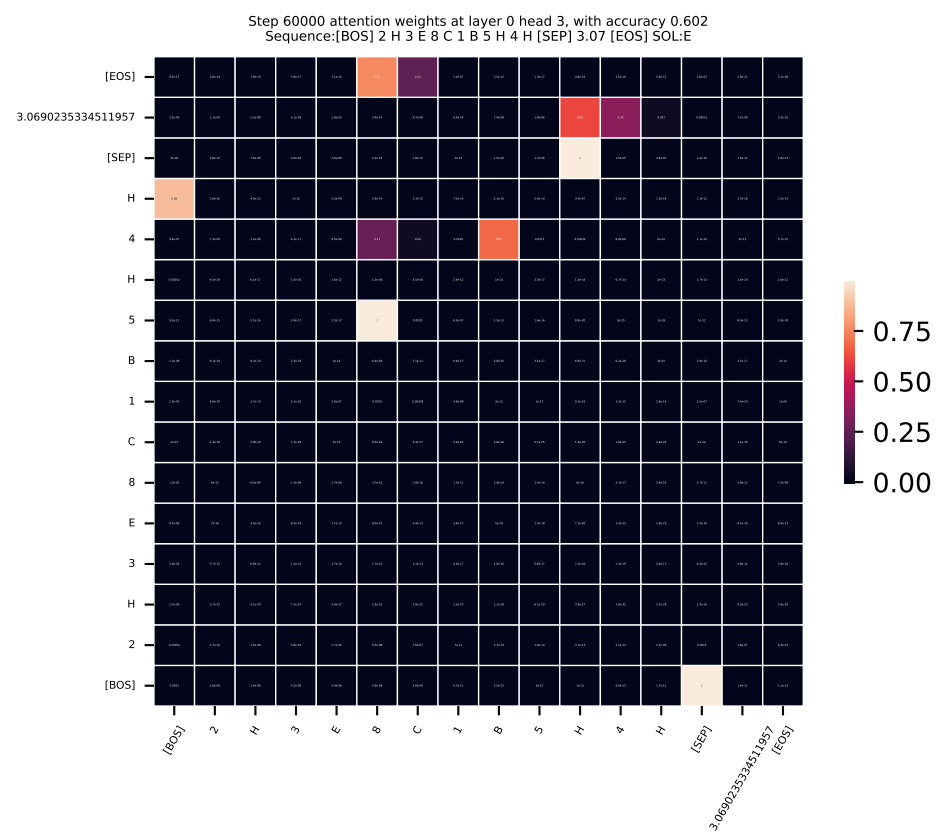

Step 60000 attention weights at layer 0 head 2, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

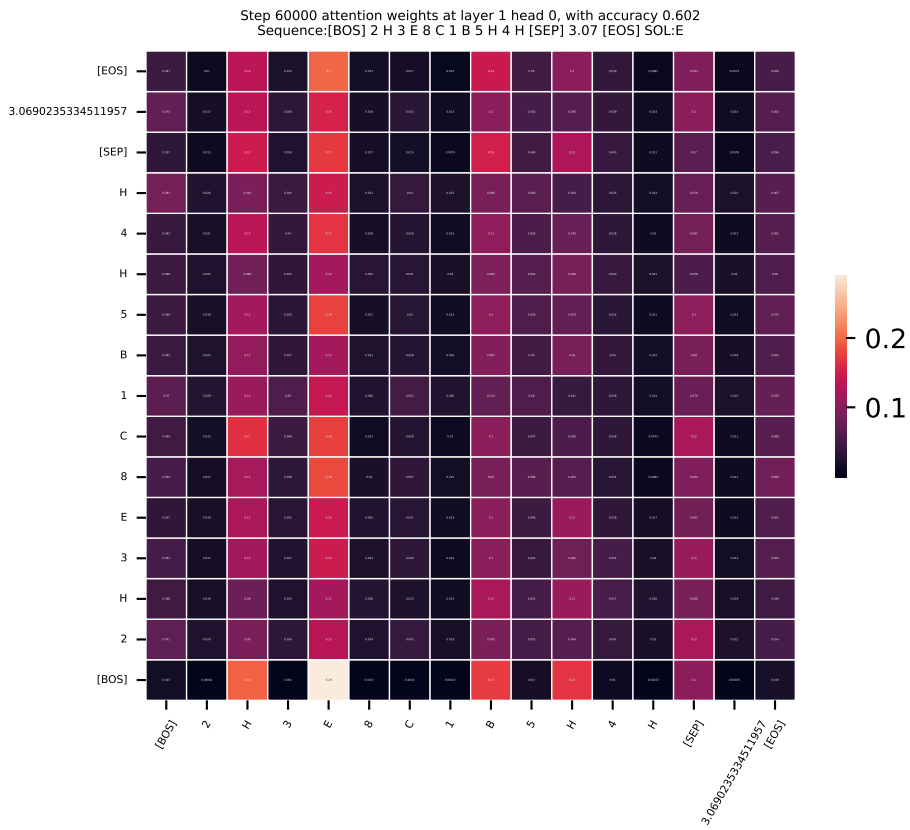

Step 60000 attention weights at layer 0 head 3, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

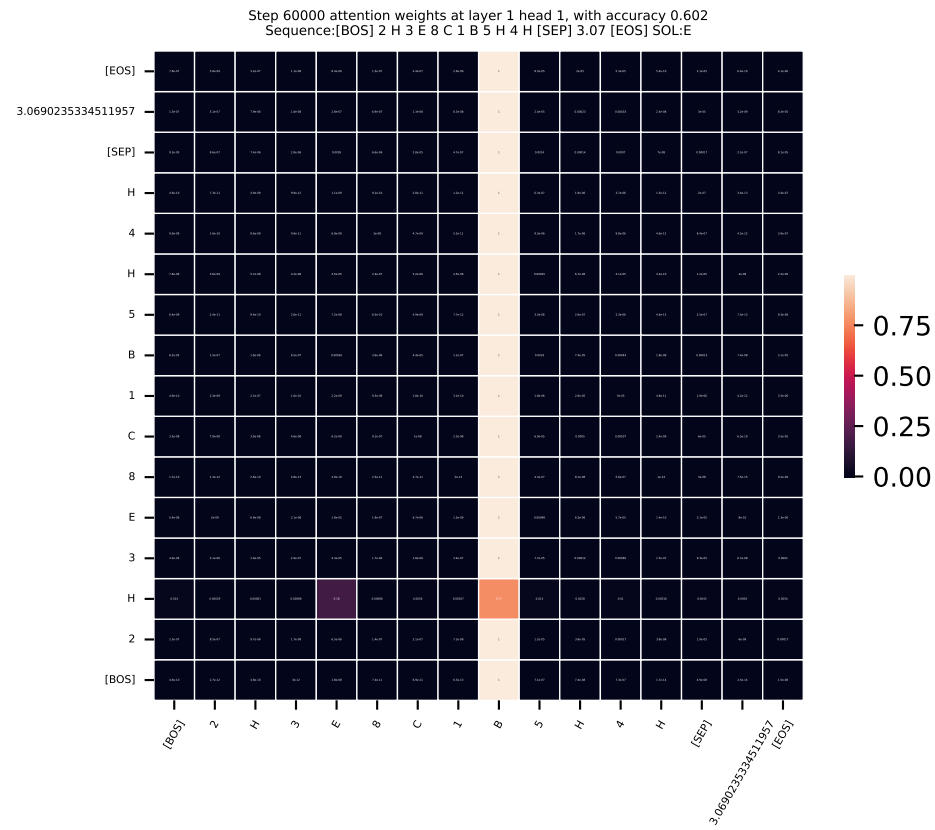

Step 60000 attention weights at layer 0 head 0, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

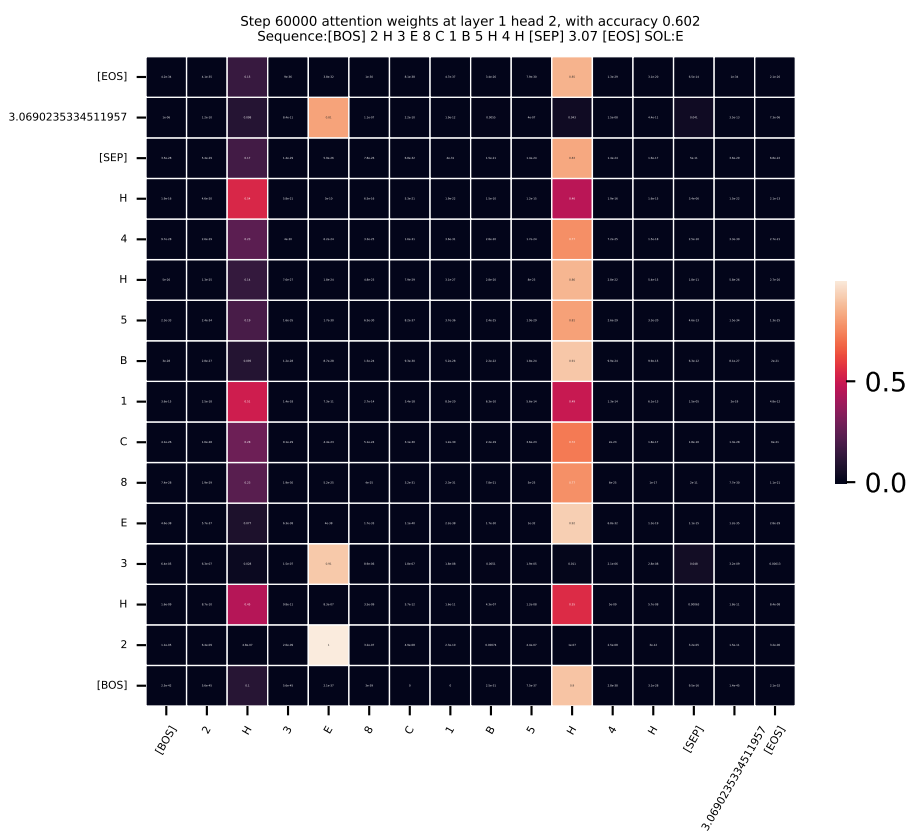

Step 60000 attention weights at layer 0 head 1, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

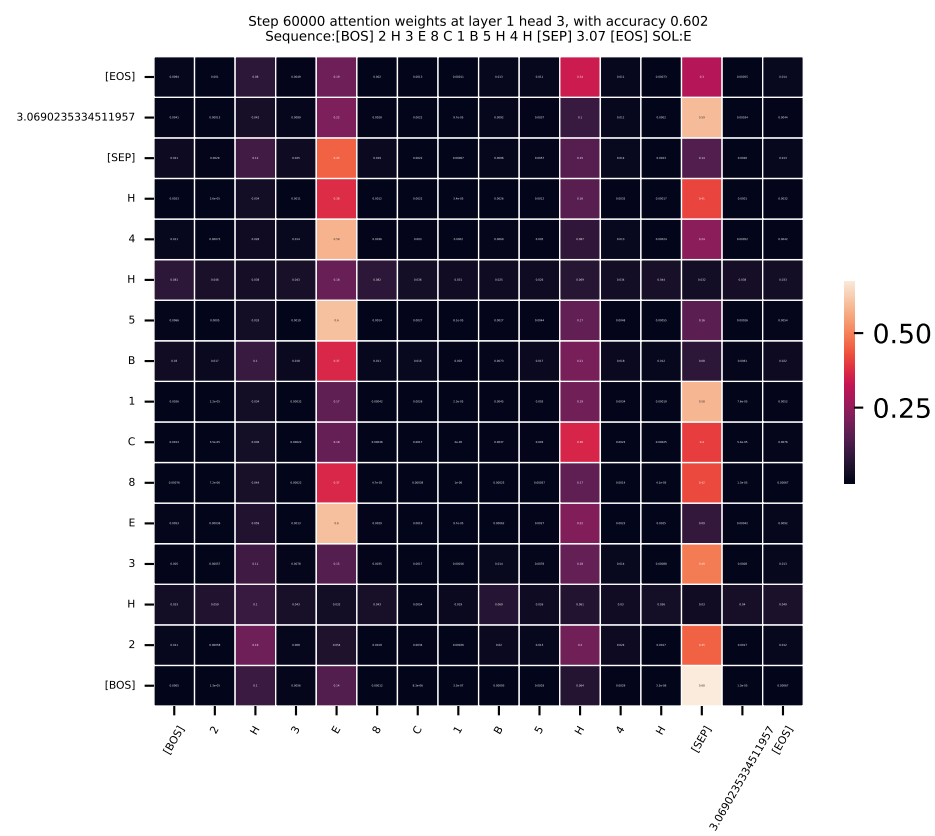

Step 60000 attention weights at layer 0 head 2, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

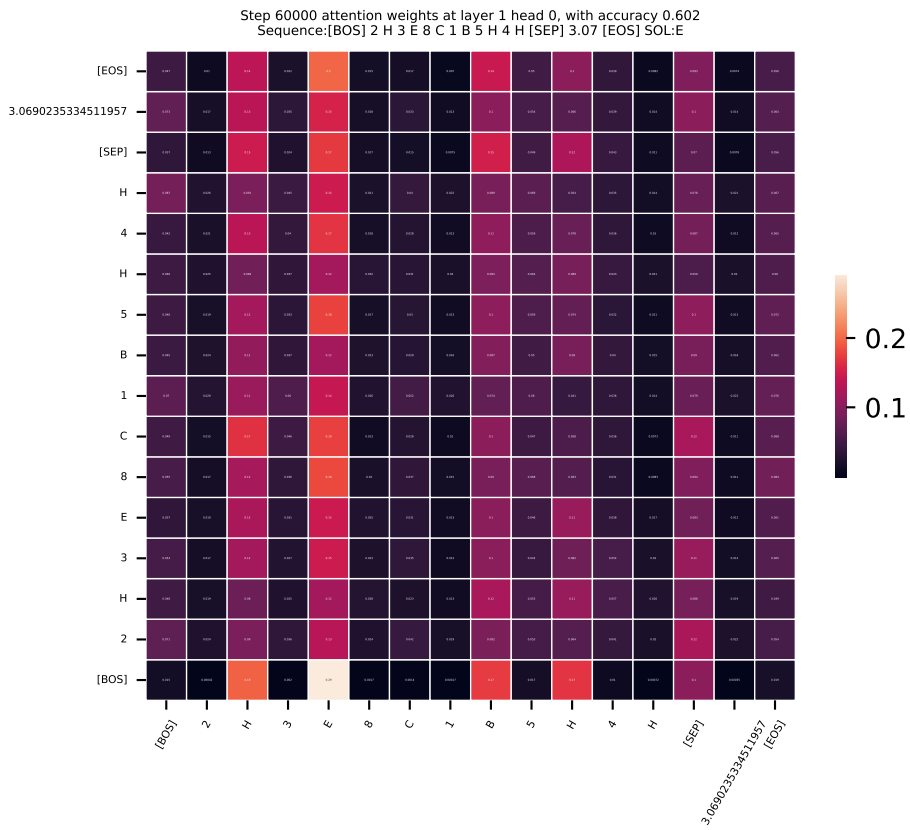

Step 60000 attention weights at layer 0 head 3, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

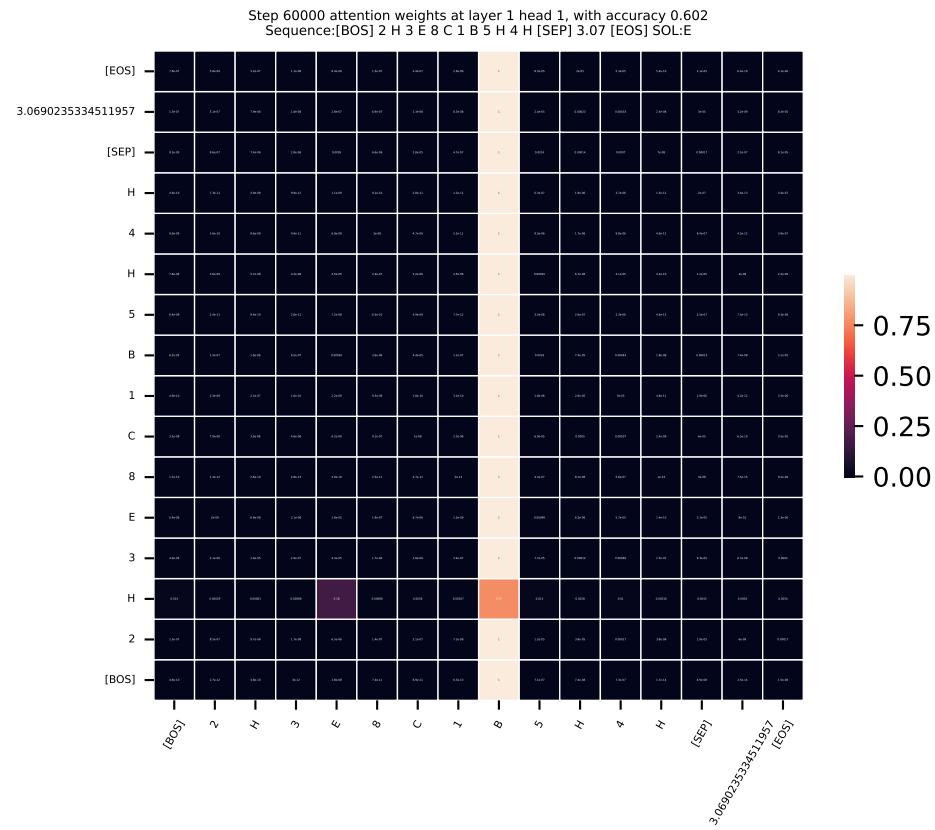

Step 60000 attention weights at layer 1 head 0, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.2
0.1

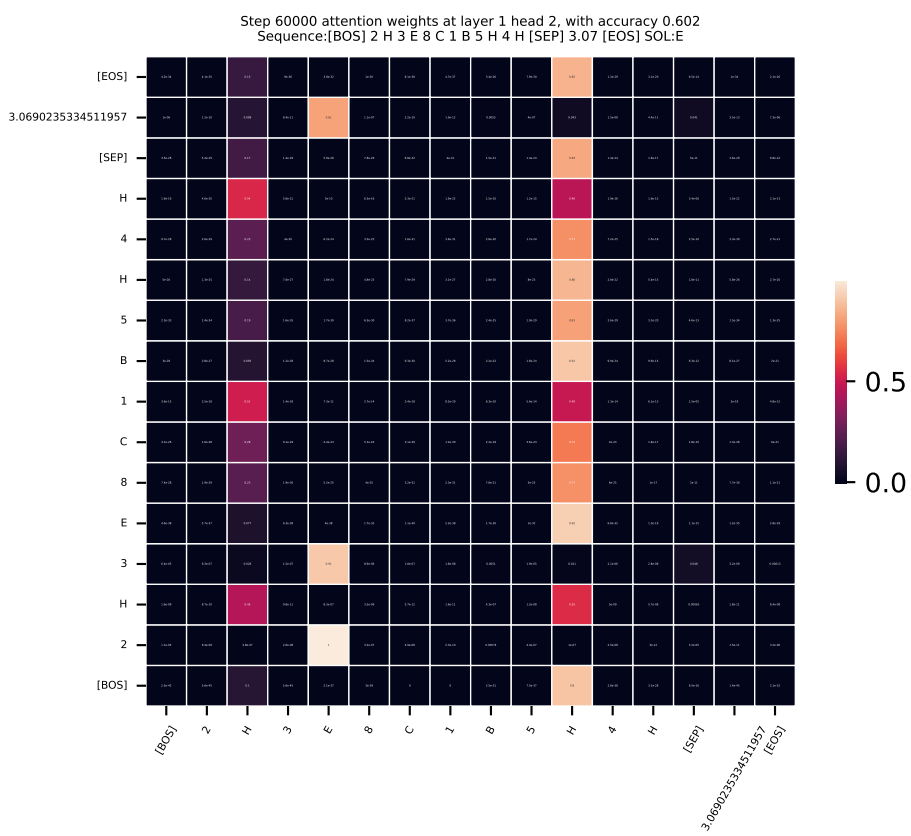

Step 60000 attention weights at layer 1 head 1, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

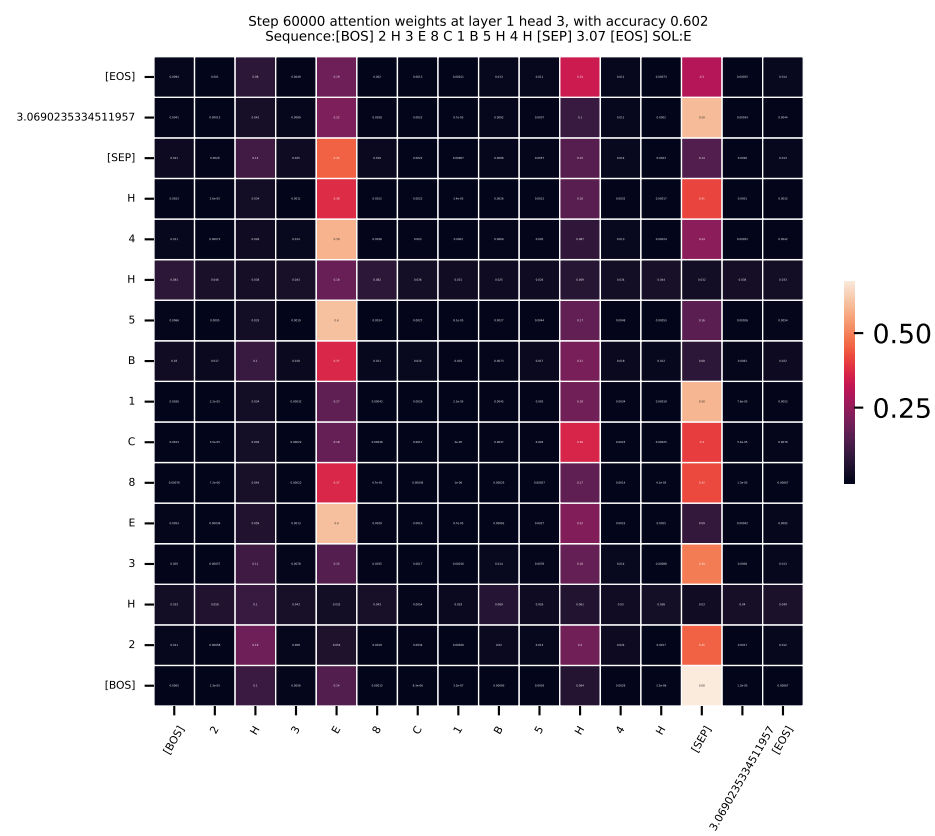

Step 60000 attention weights at layer 1 head 2, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.5
0.0

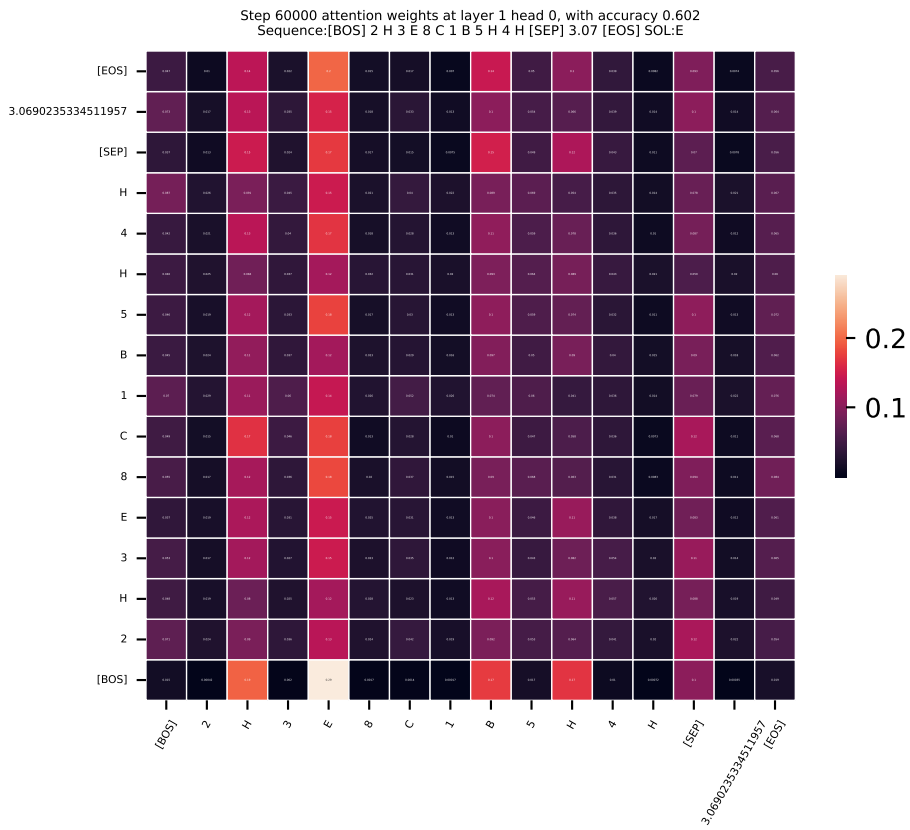

Step 60000 attention weights at layer 1 head 3, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.50
0.25

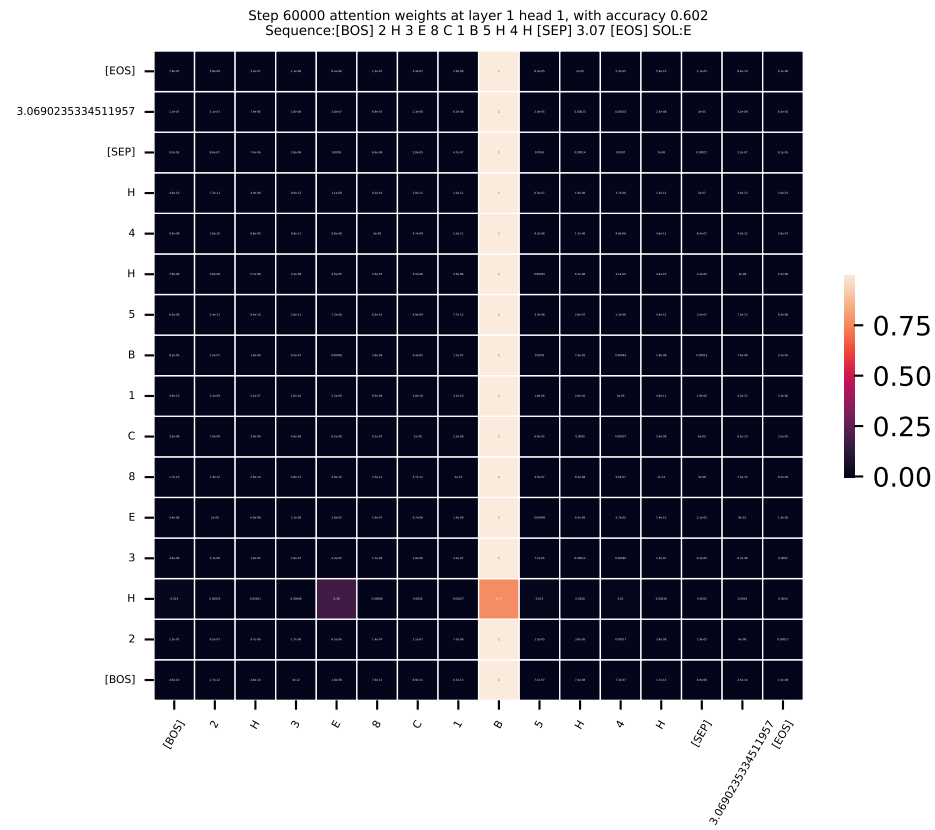

Step 60000 attention weights at layer 1 head 0, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.2
0.1

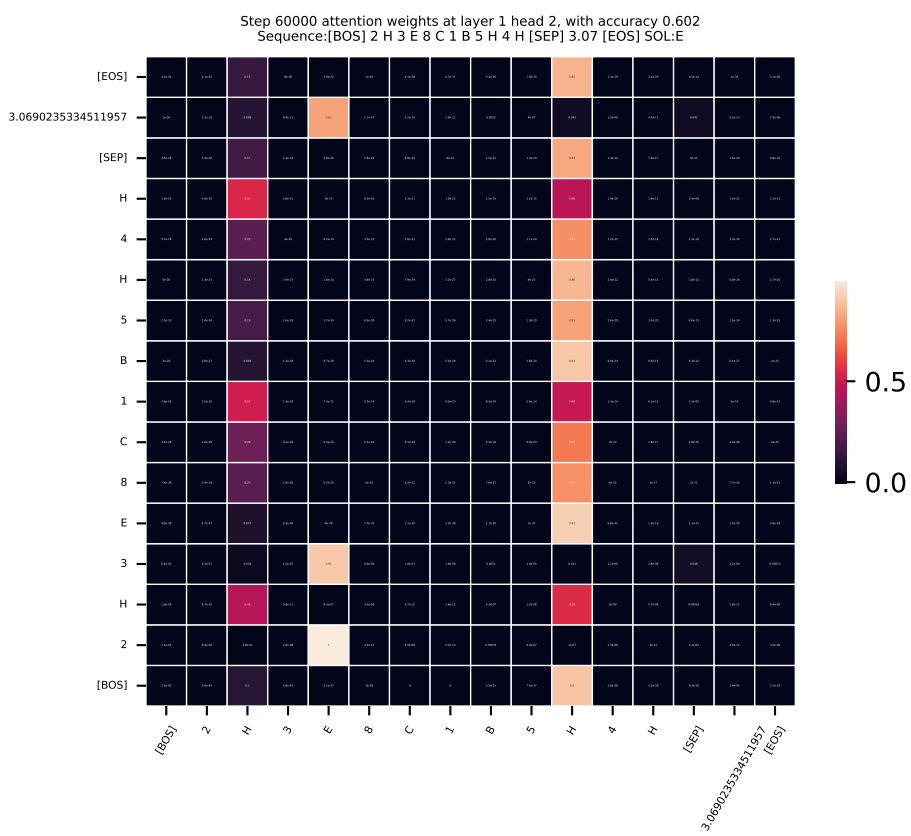

Step 60000 attention weights at layer 1 head 1, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

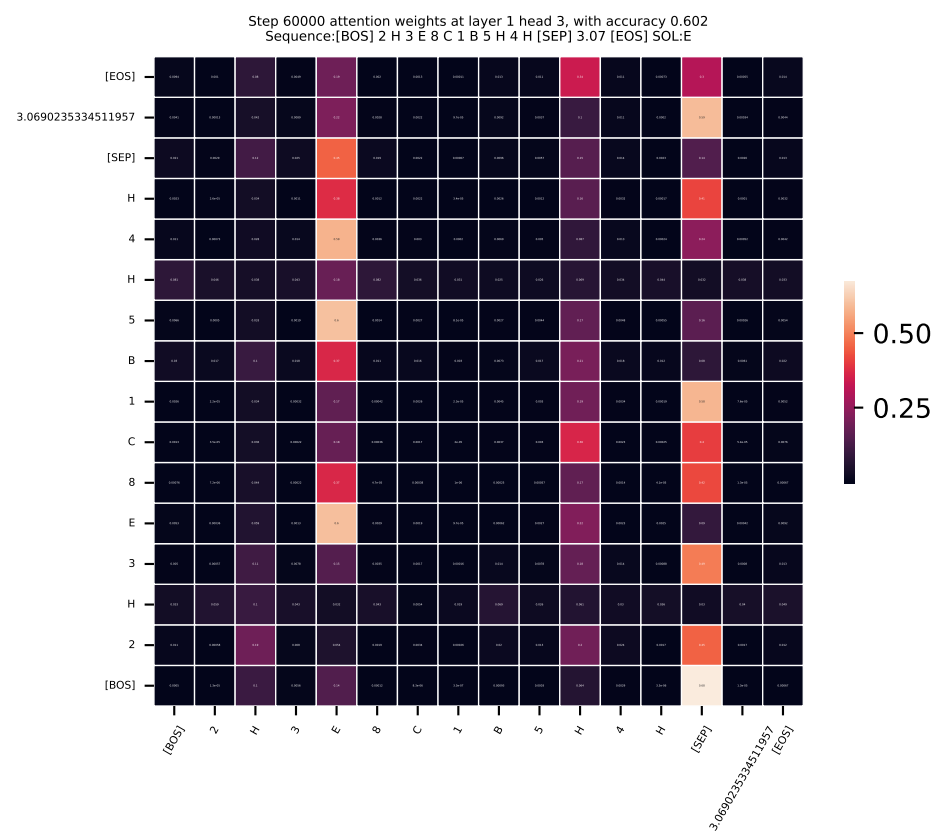

Step 60000 attention weights at layer 1 head 2, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.5
0.0

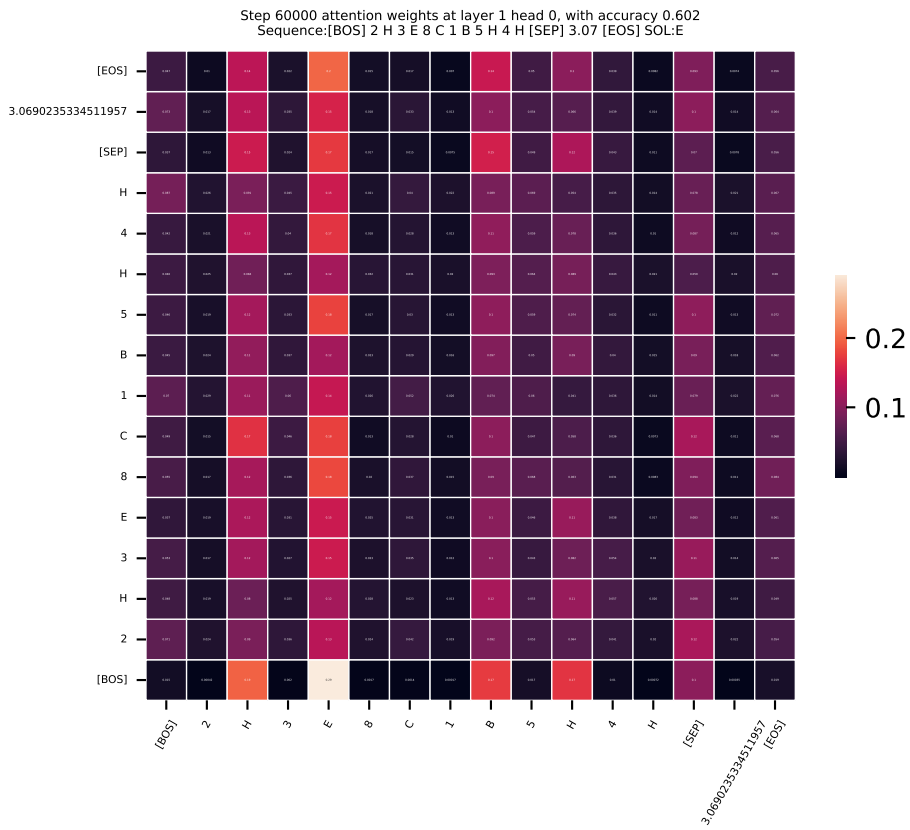

Step 60000 attention weights at layer 1 head 3, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.50
0.25

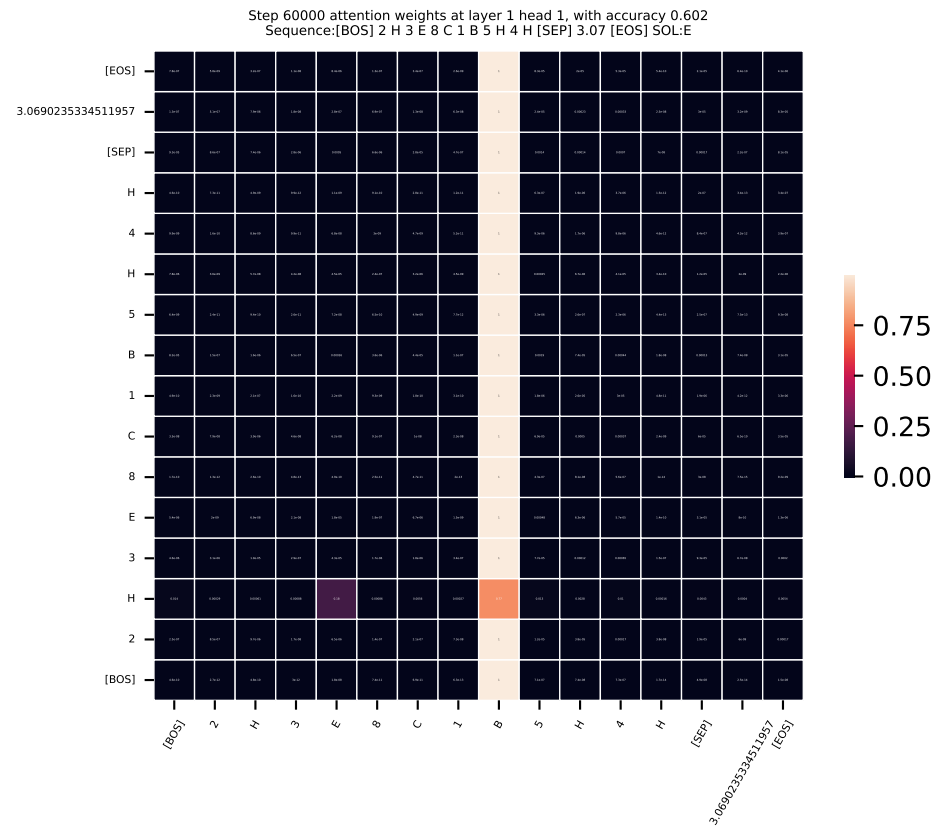

Step 60000 attention weights at layer 1 head 0, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.2
0.1

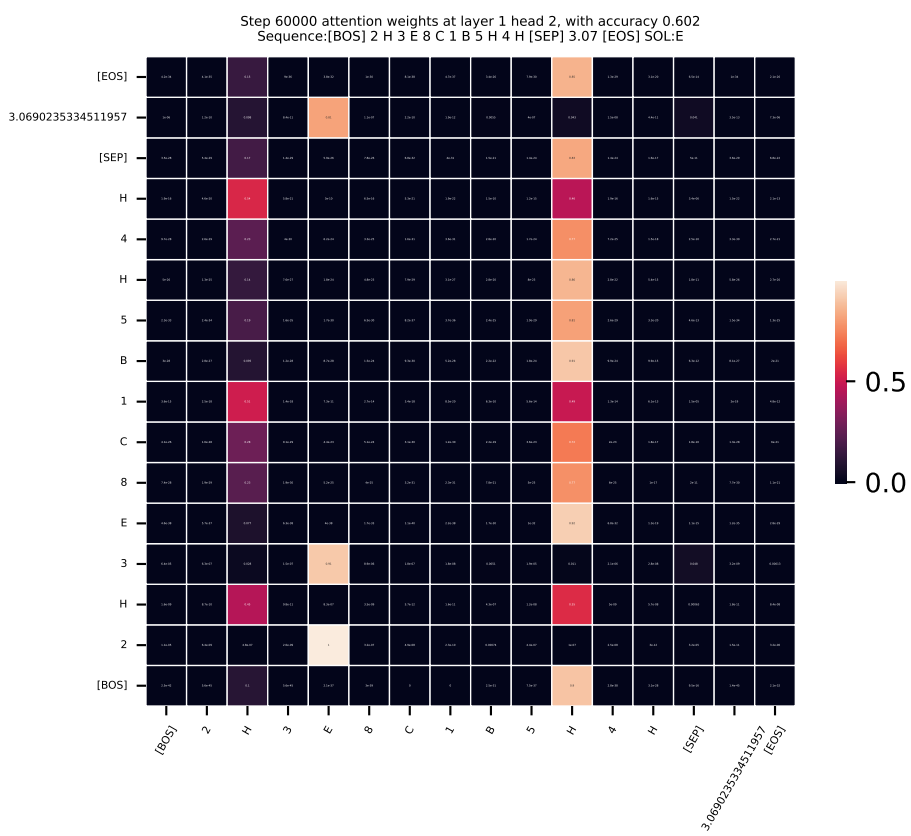

Step 60000 attention weights at layer 1 head 1, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

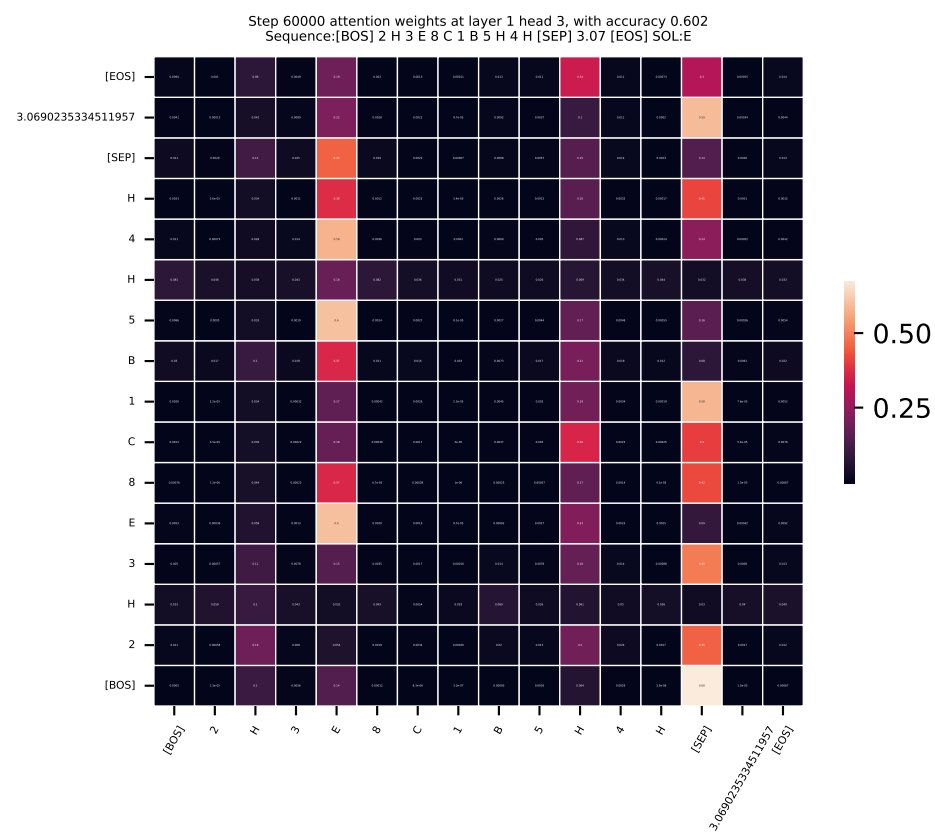

Step 60000 attention weights at layer 1 head 2, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.5
0.0

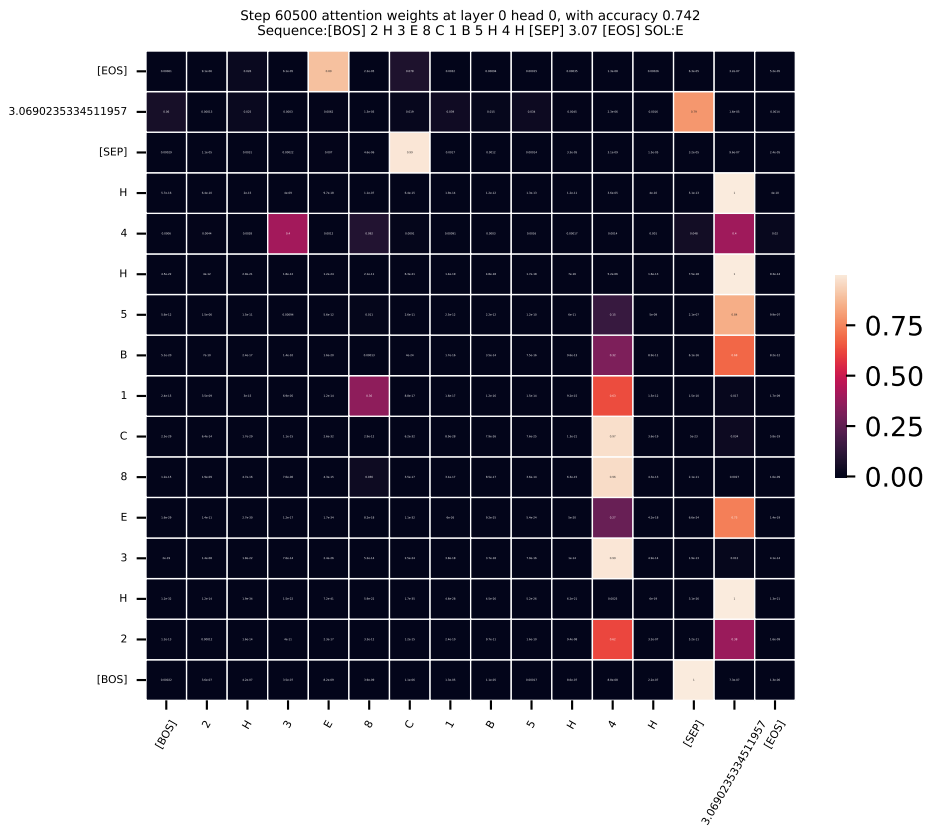

Step 60000 attention weights at layer 1 head 3, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.50
0.25

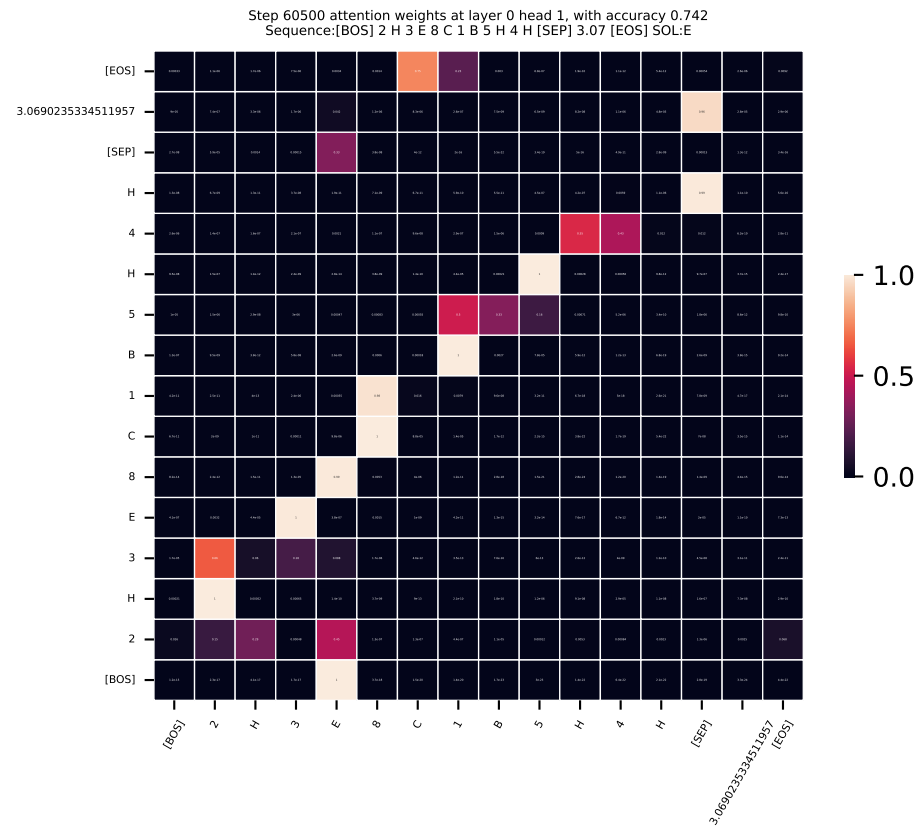

Step 60000 attention weights at layer 1 head 0, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.2
0.1

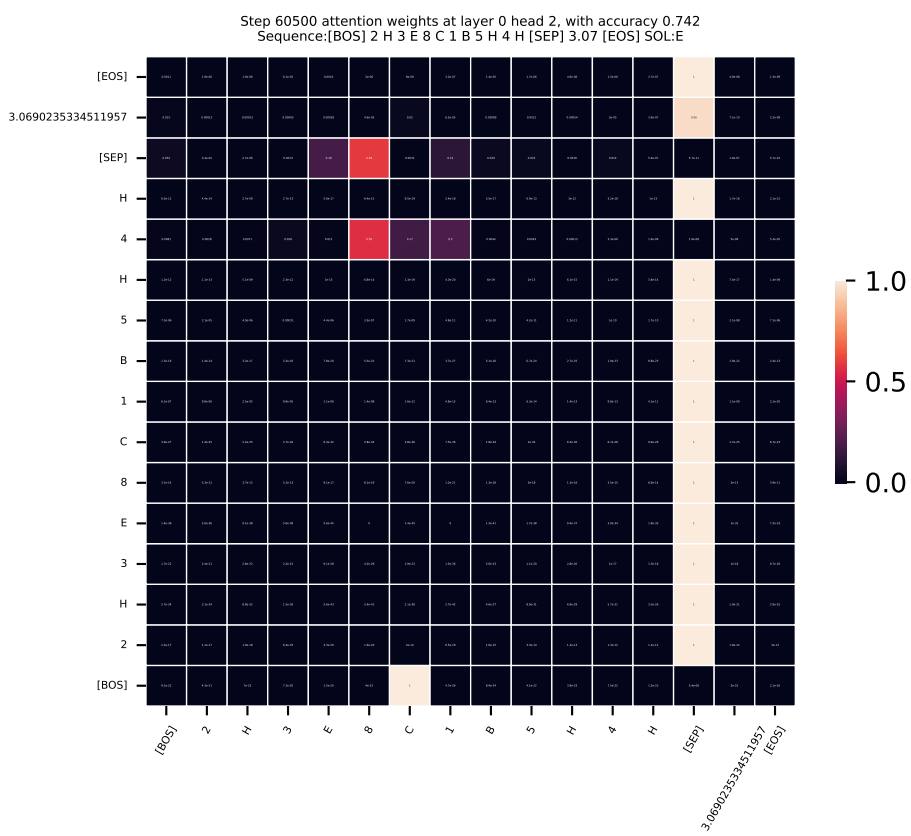

Step 60000 attention weights at layer 1 head 1, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

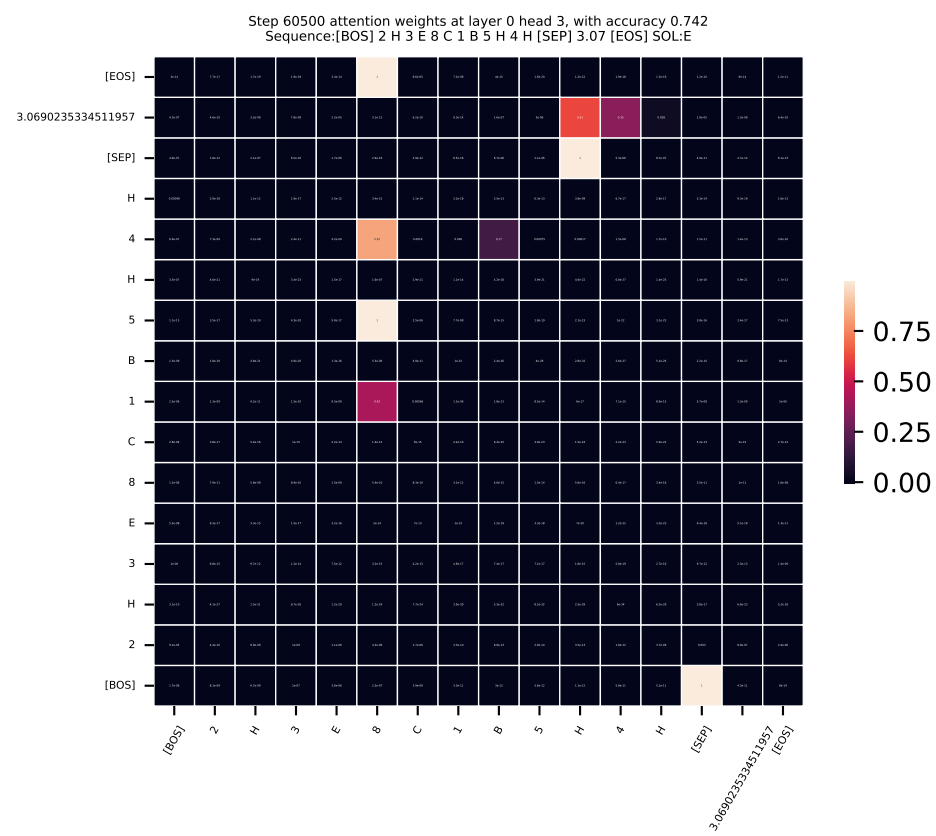

Step 60000 attention weights at layer 1 head 2, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.5
0.0

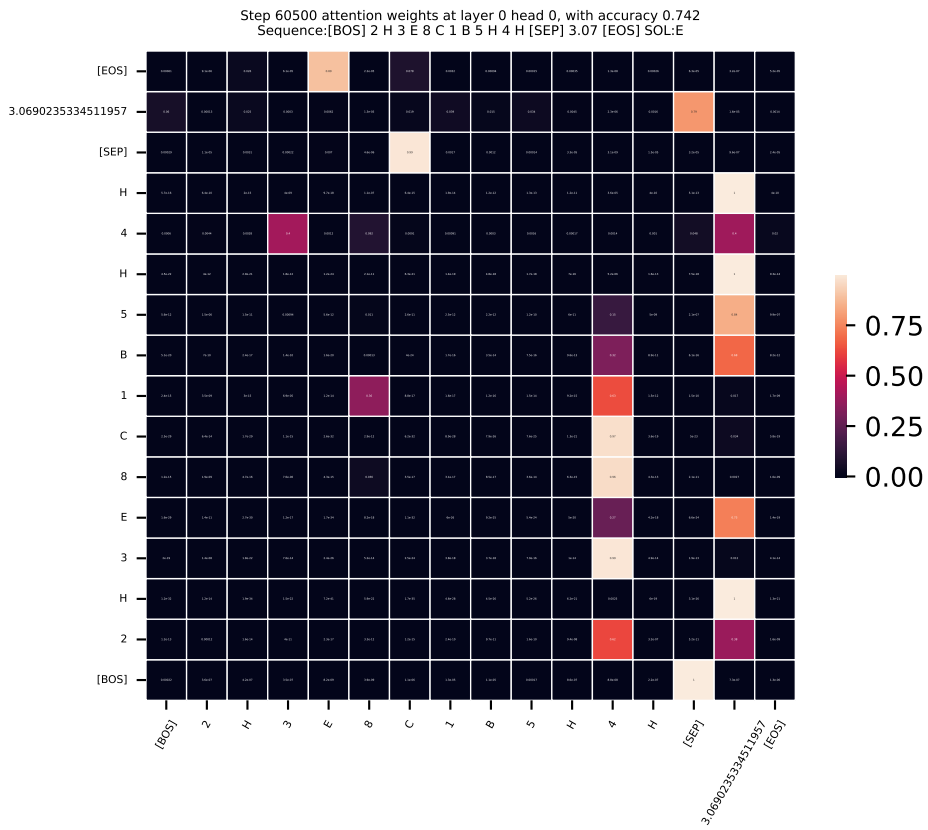

Step 60000 attention weights at layer 1 head 3, with accuracy 0.602
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.50
0.25

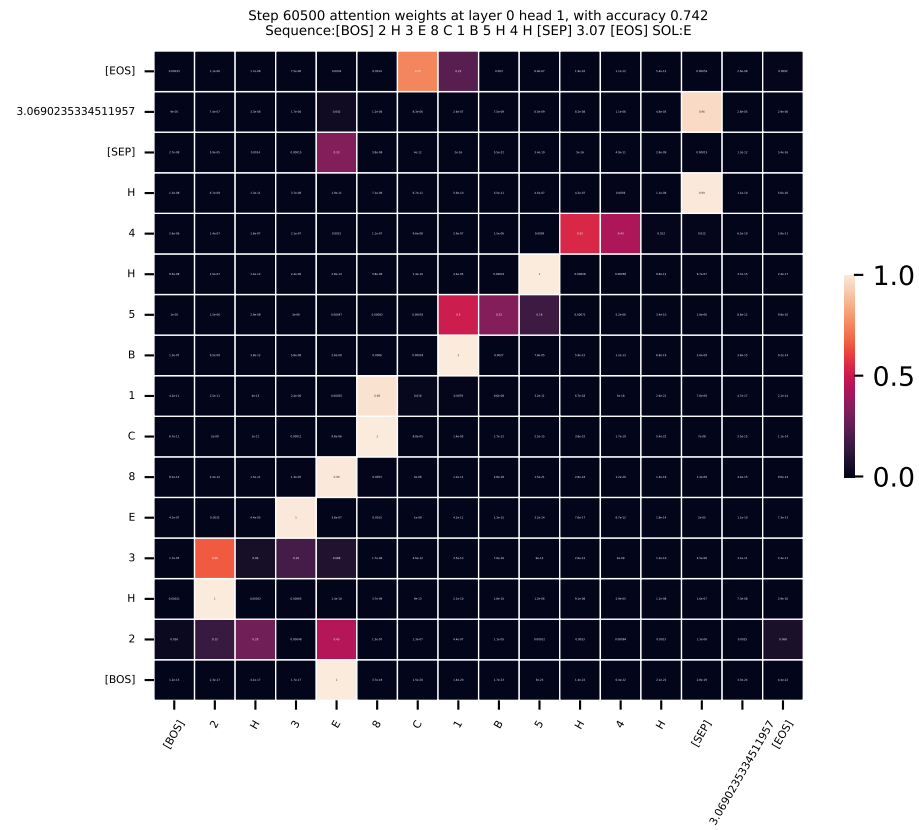

Step 60500 attention weights at layer 0 head 0, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

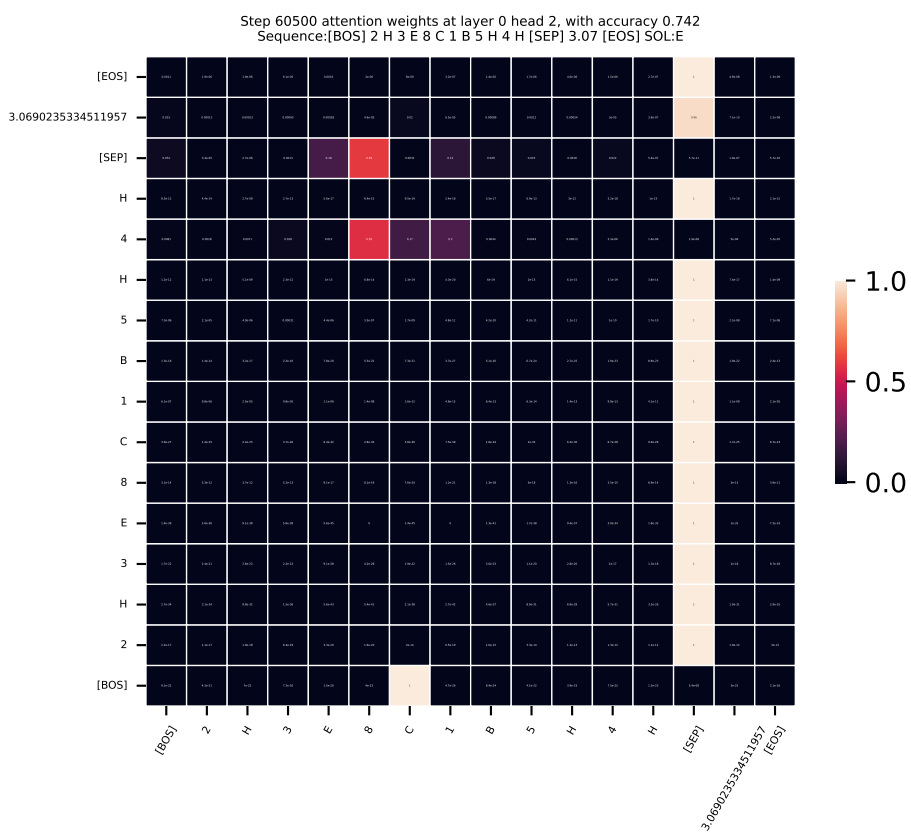

Step 60500 attention weights at layer 0 head 1, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

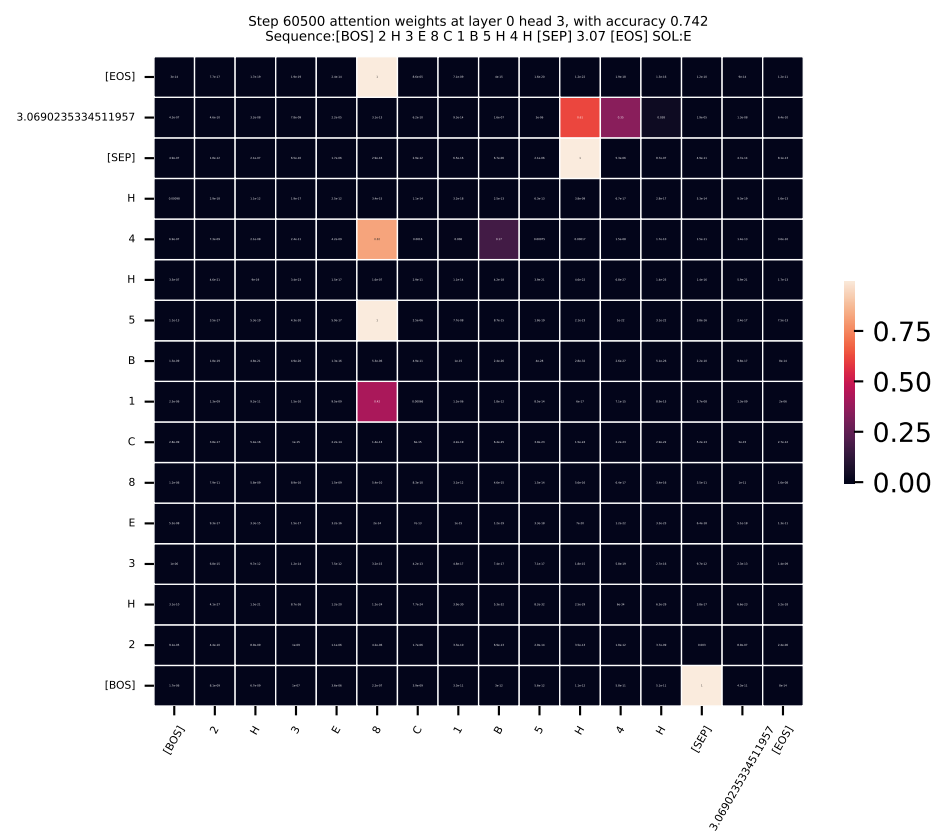

Step 60500 attention weights at layer 0 head 2, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

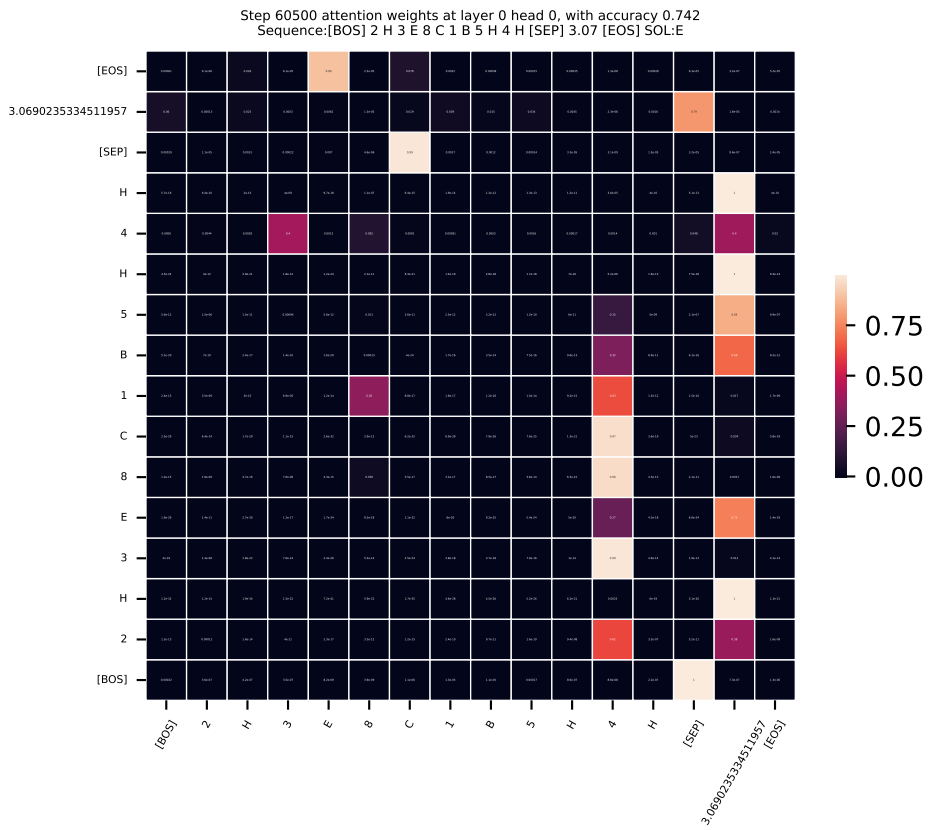

Step 60500 attention weights at layer 0 head 3, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

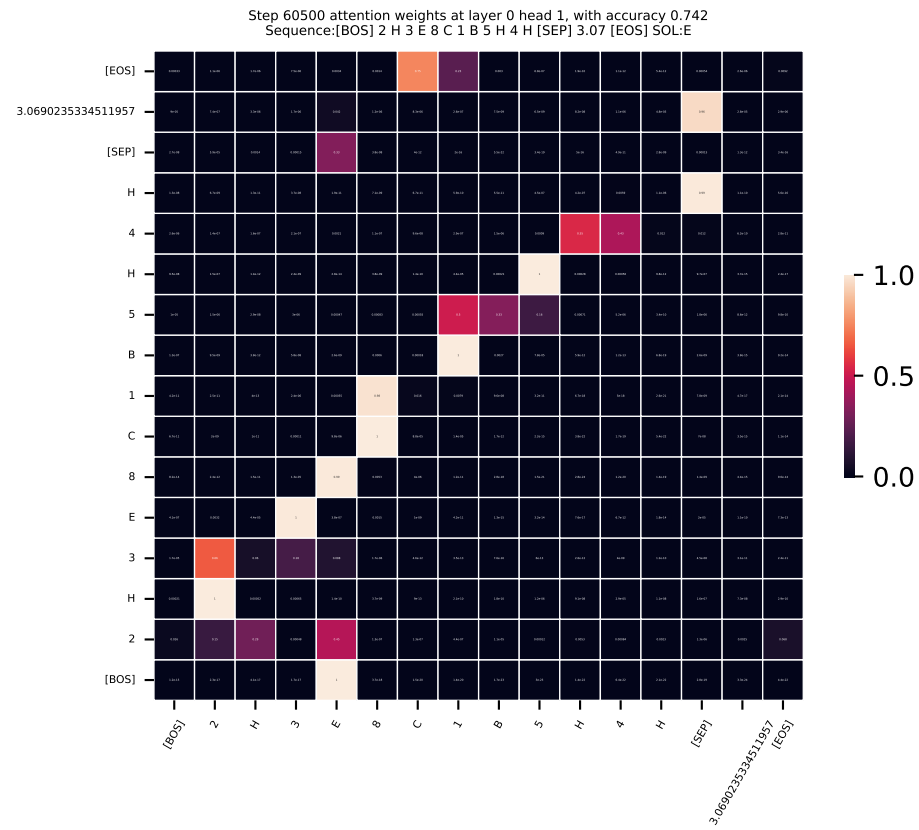

Step 60500 attention weights at layer 0 head 0, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

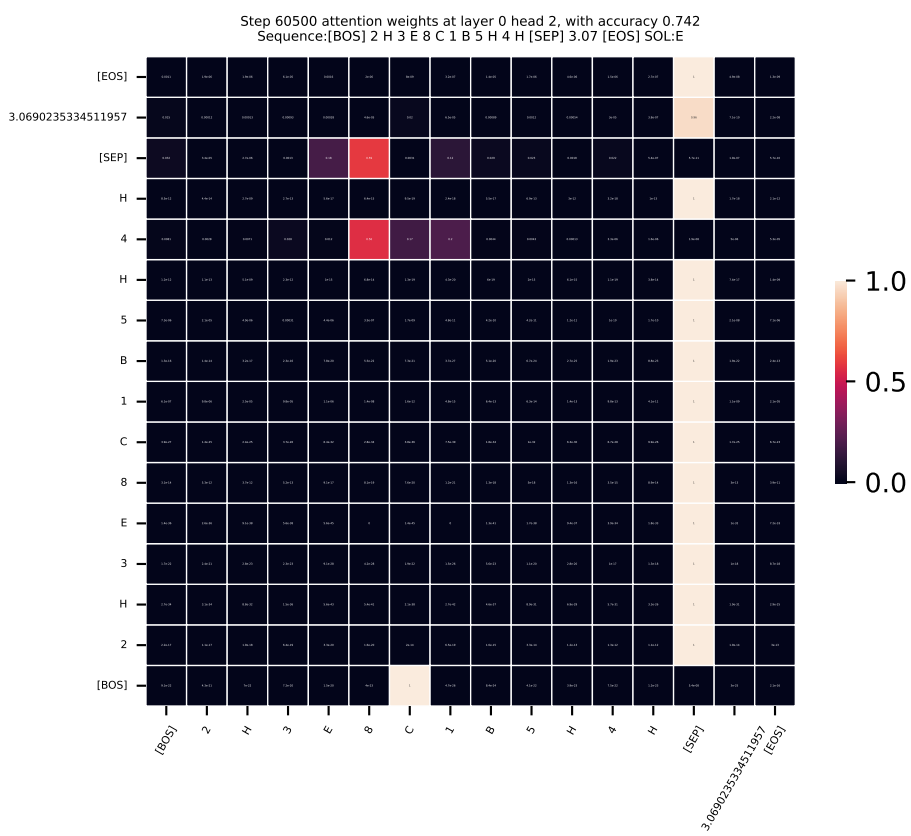

Step 60500 attention weights at layer 0 head 1, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

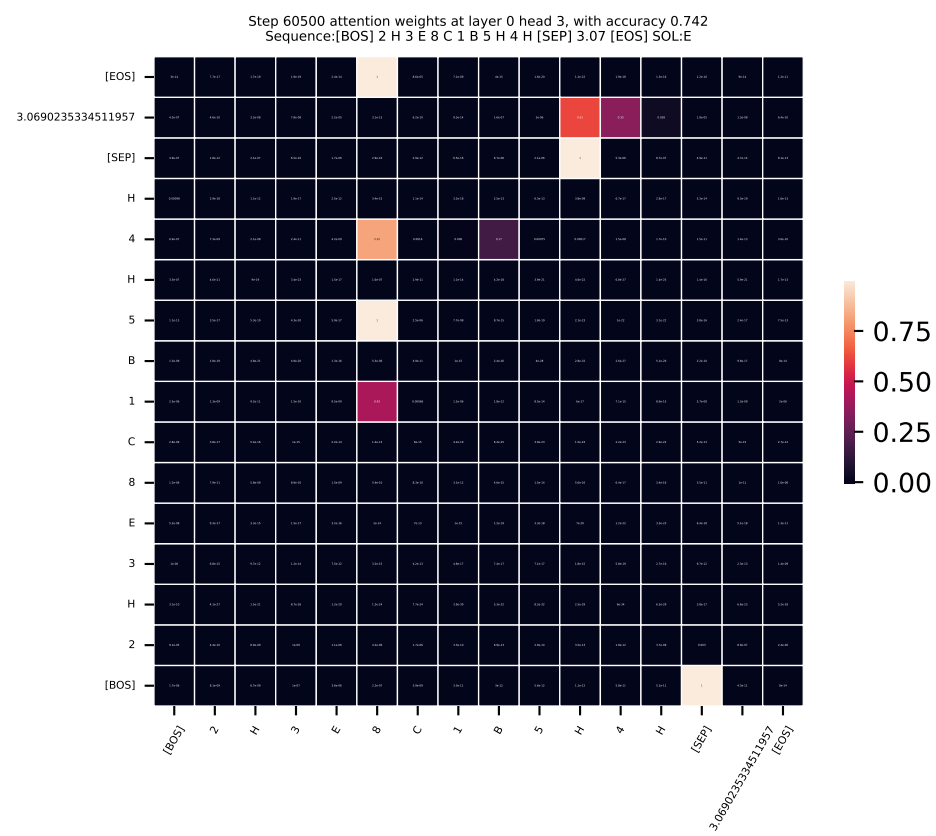

Step 60500 attention weights at layer 0 head 2, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

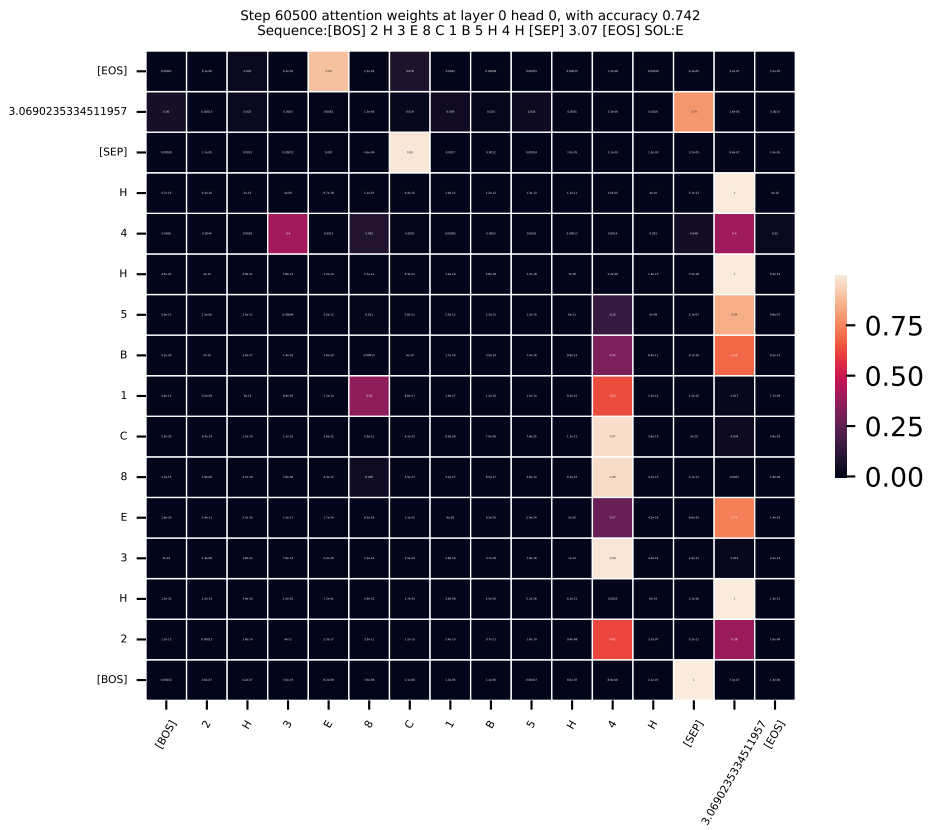

Step 60500 attention weights at layer 0 head 3, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

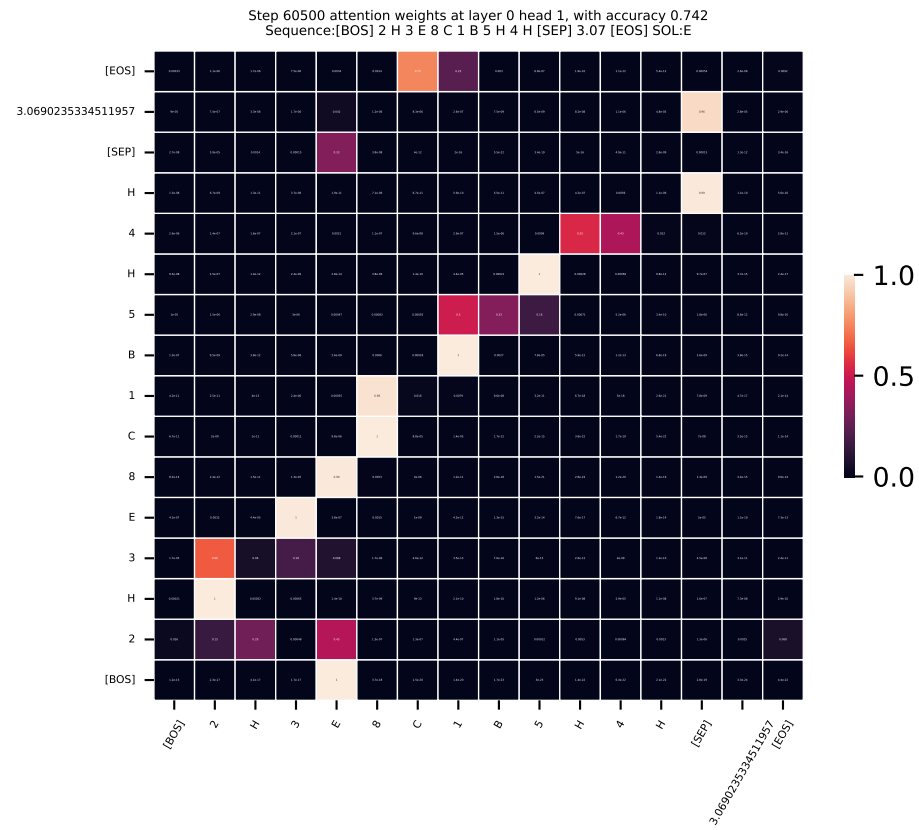

Step 60500 attention weights at layer 0 head 0, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

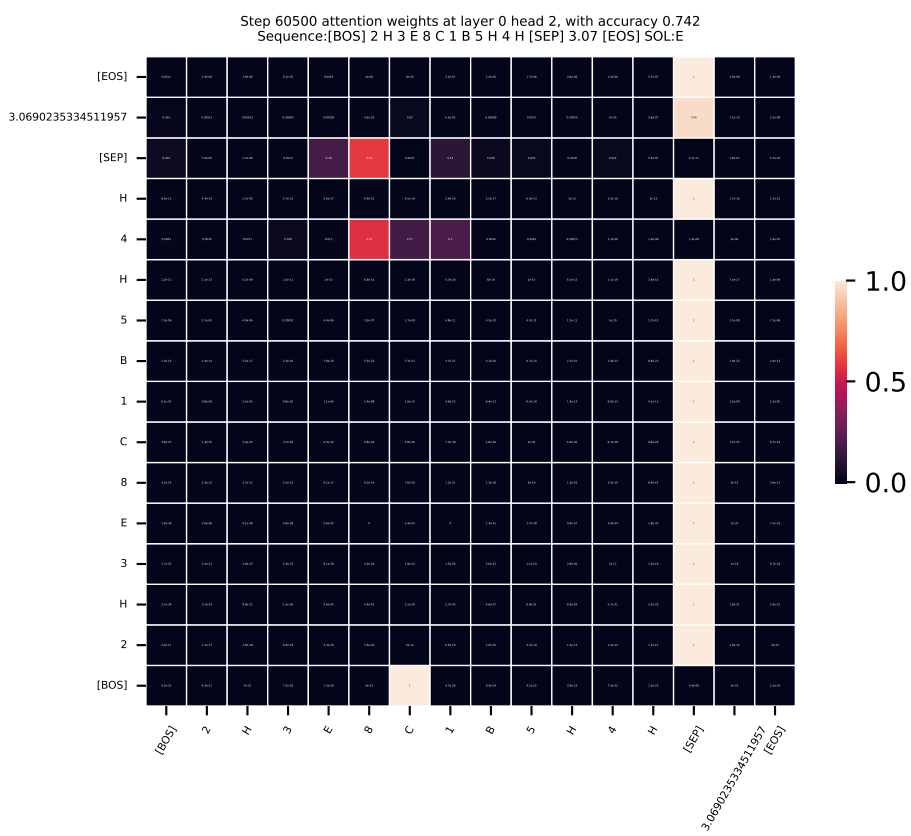

Step 60500 attention weights at layer 0 head 1, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

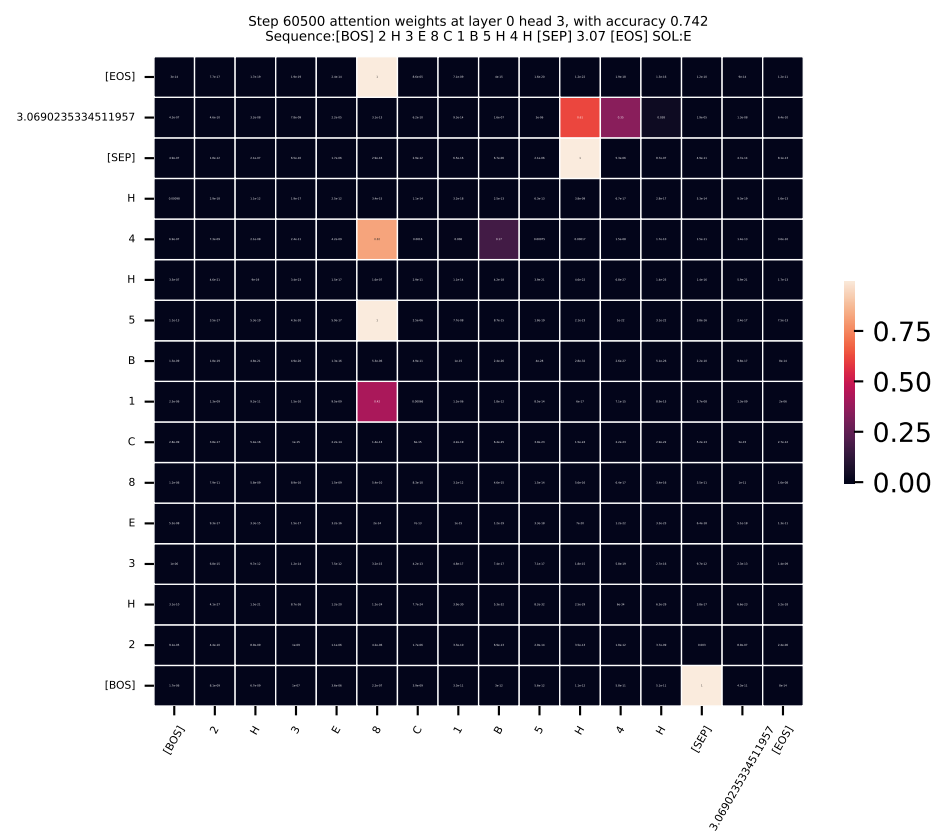

Step 60500 attention weights at layer 0 head 2, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

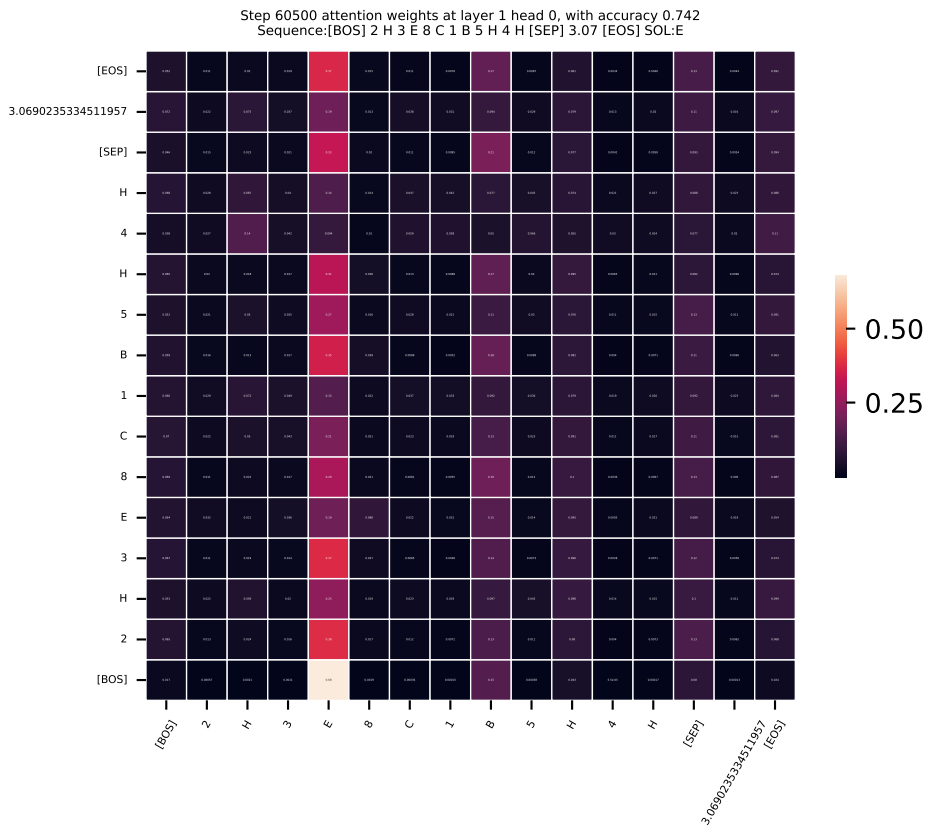

Step 60500 attention weights at layer 0 head 3, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

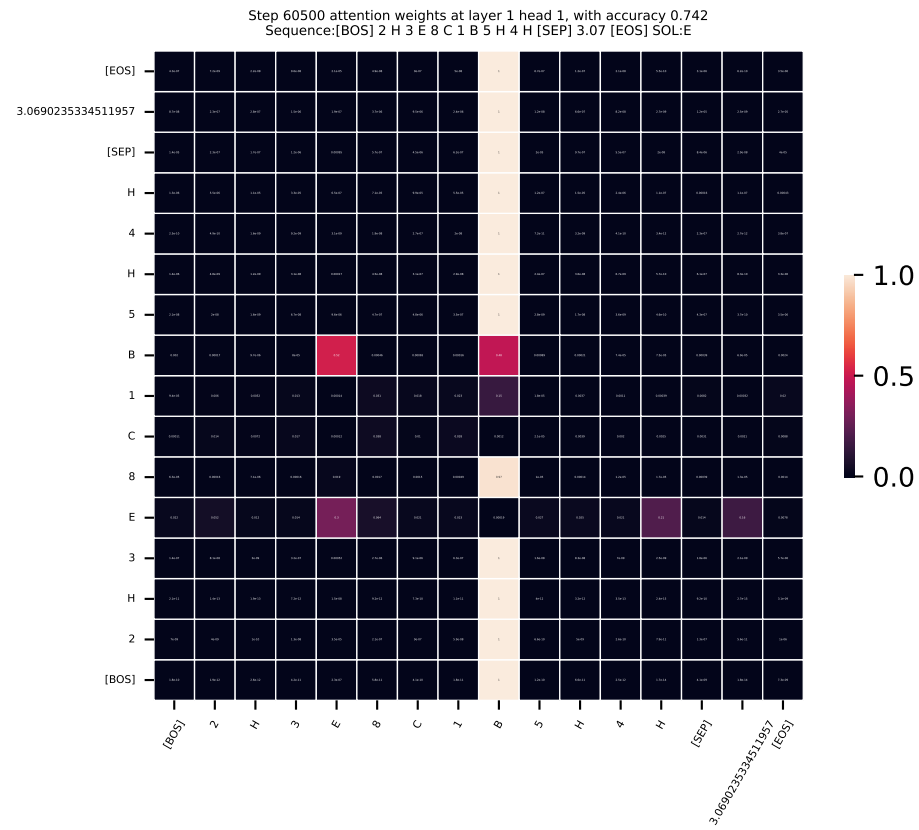

Step 60500 attention weights at layer 0 head 0, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

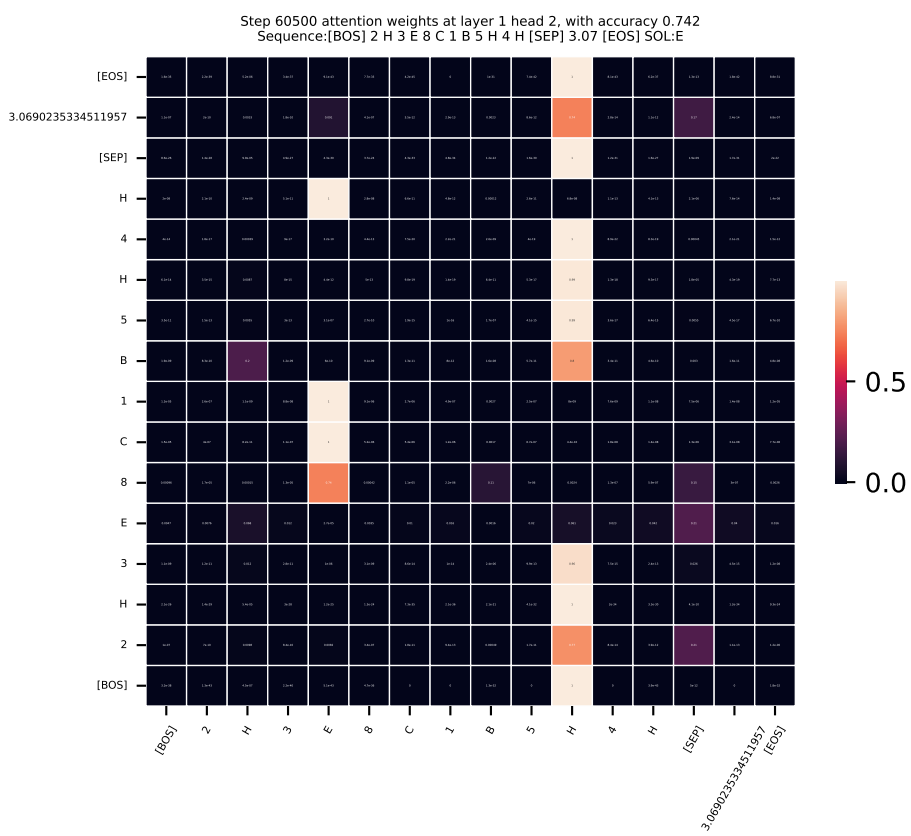

Step 60500 attention weights at layer 0 head 1, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

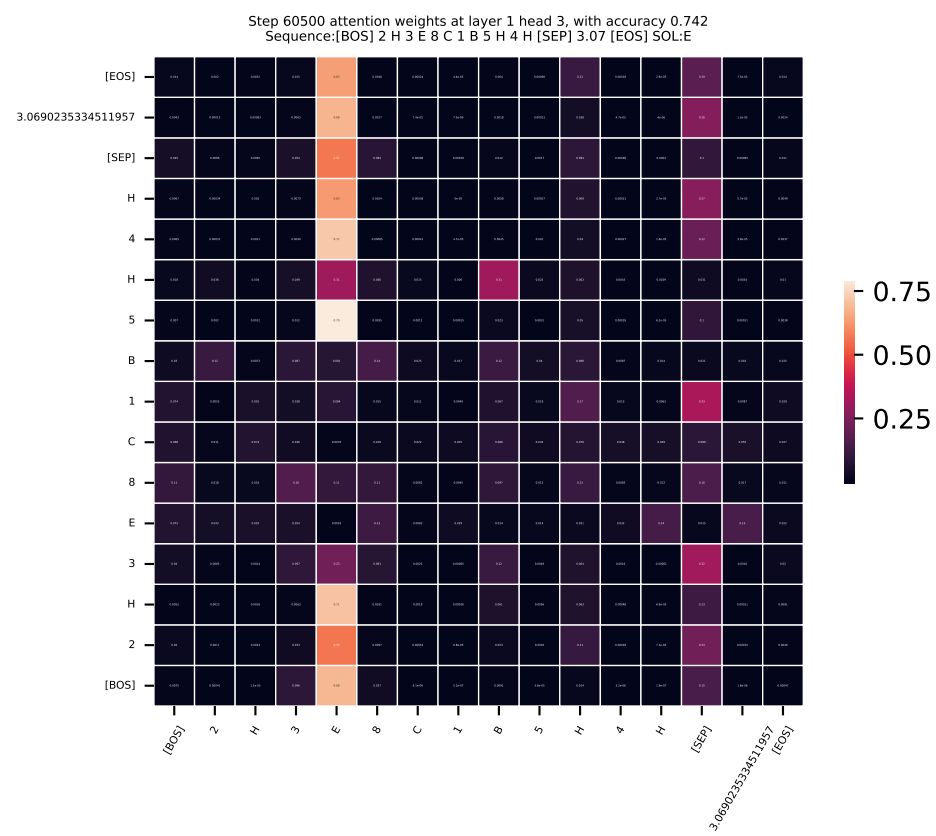

Step 60500 attention weights at layer 0 head 2, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

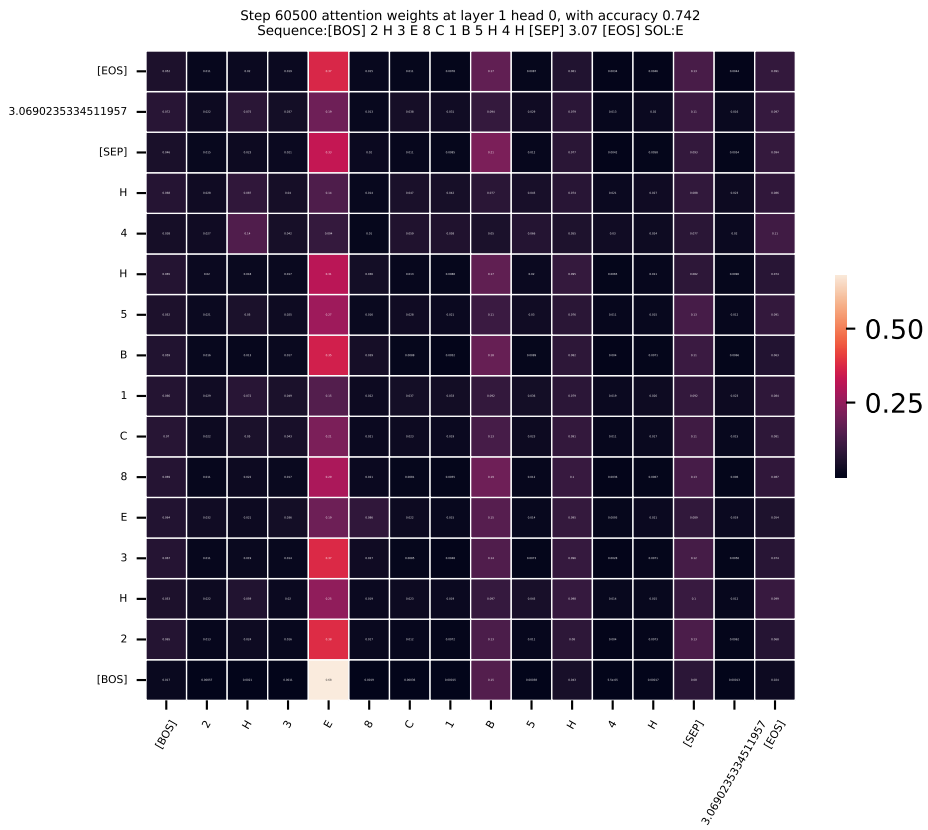

Step 60500 attention weights at layer 0 head 3, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

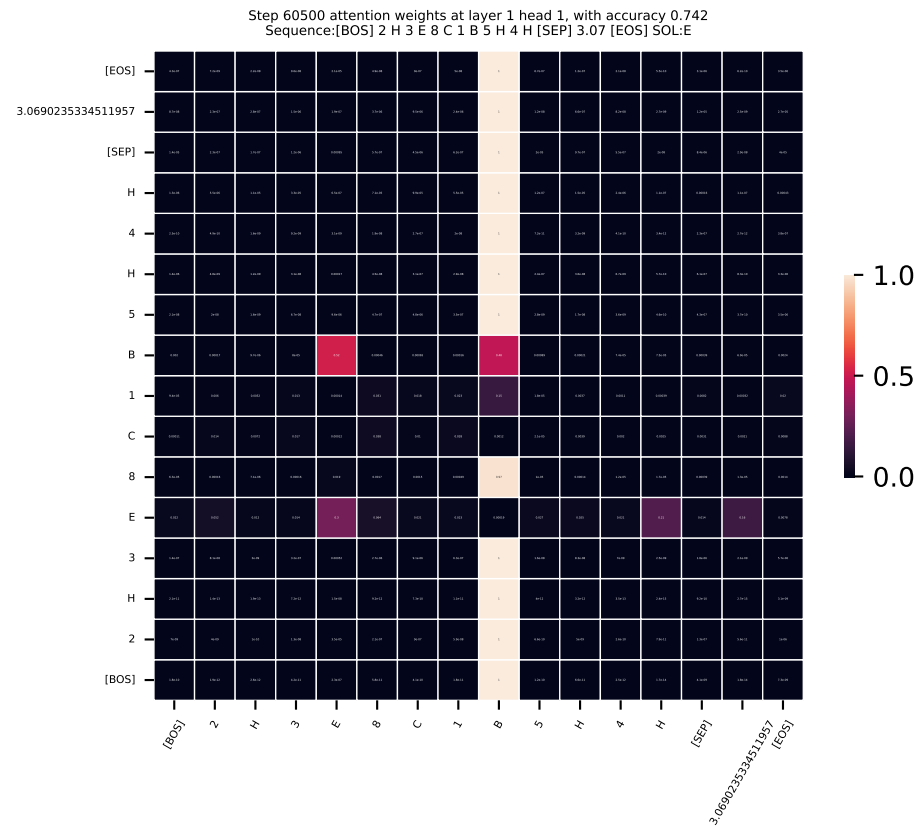

Step 60500 attention weights at layer 1 head 0, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.50
0.25

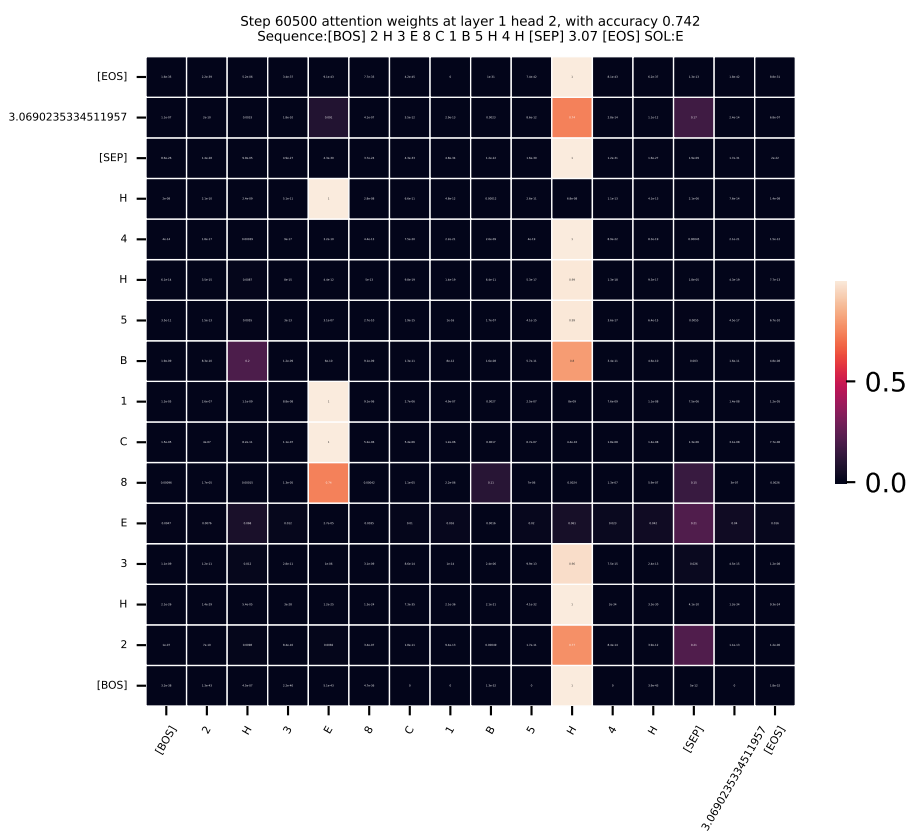

Step 60500 attention weights at layer 1 head 1, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

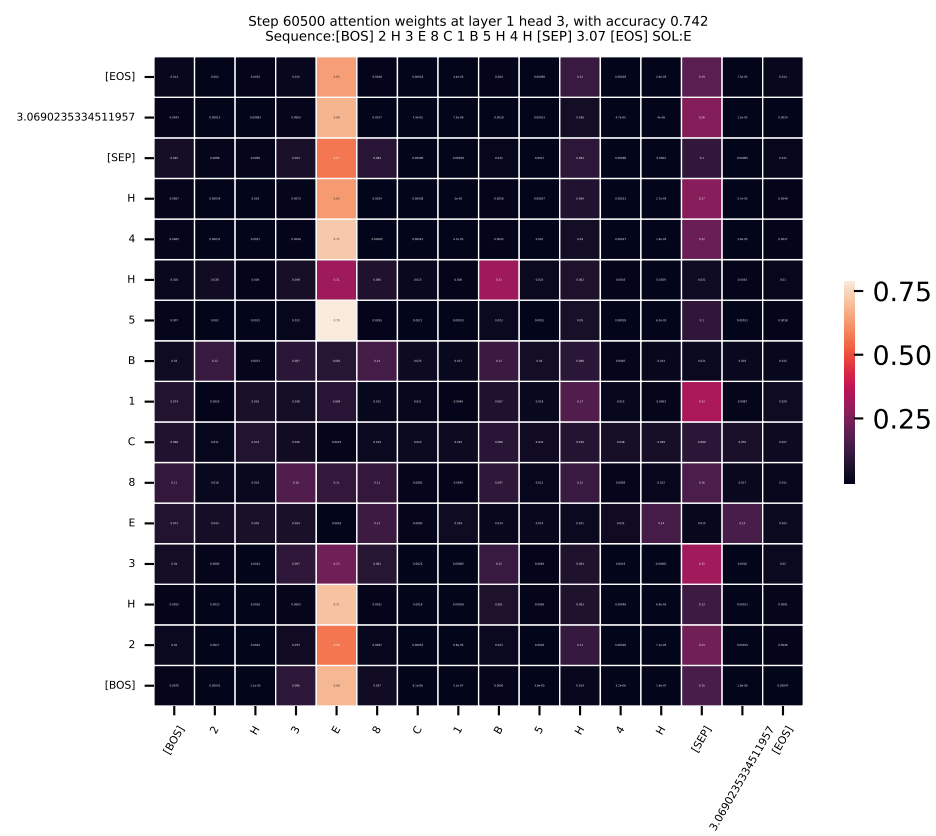

Step 60500 attention weights at layer 1 head 2, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.5
0.0

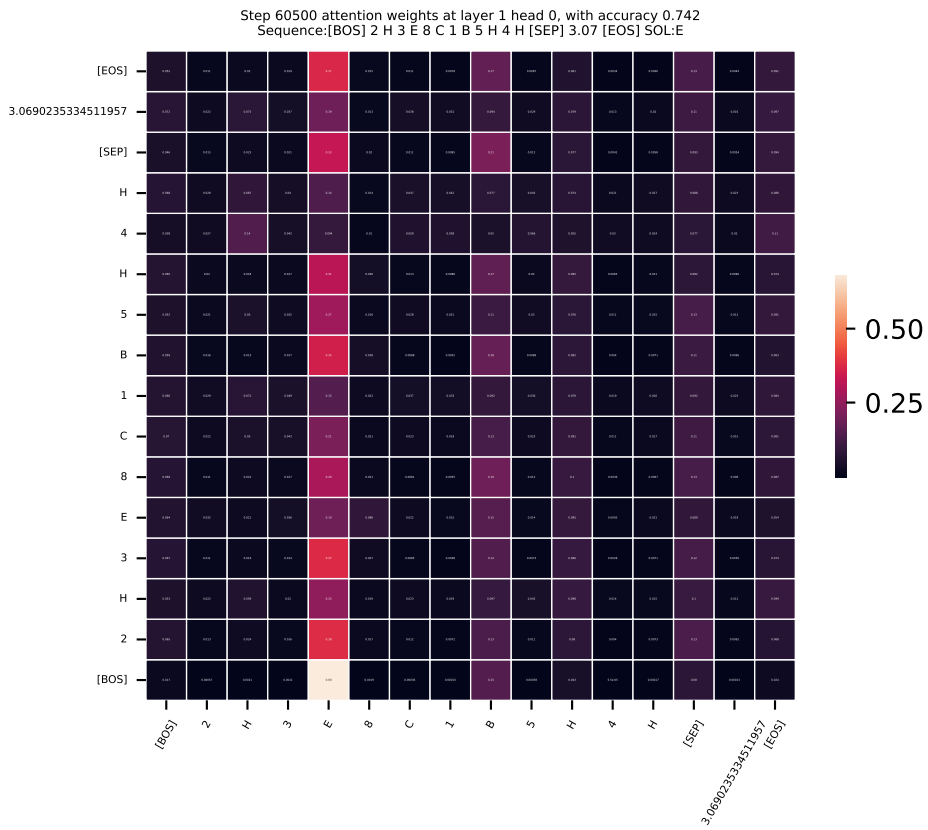

Step 60500 attention weights at layer 1 head 3, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

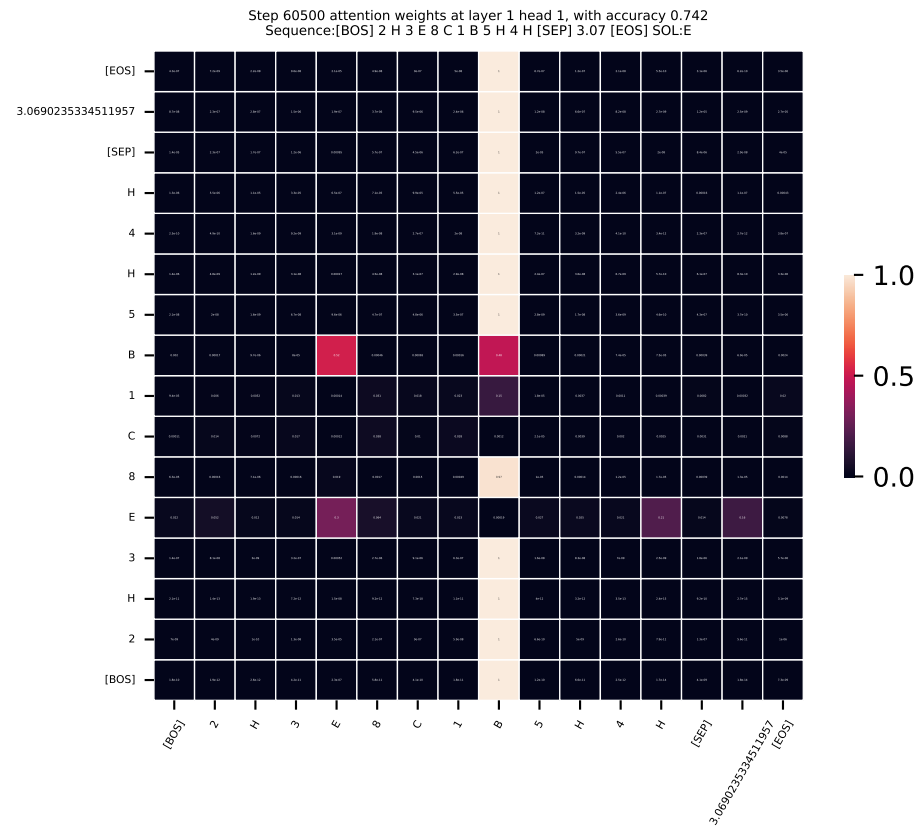

Step 60500 attention weights at layer 1 head 0, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.50
0.25

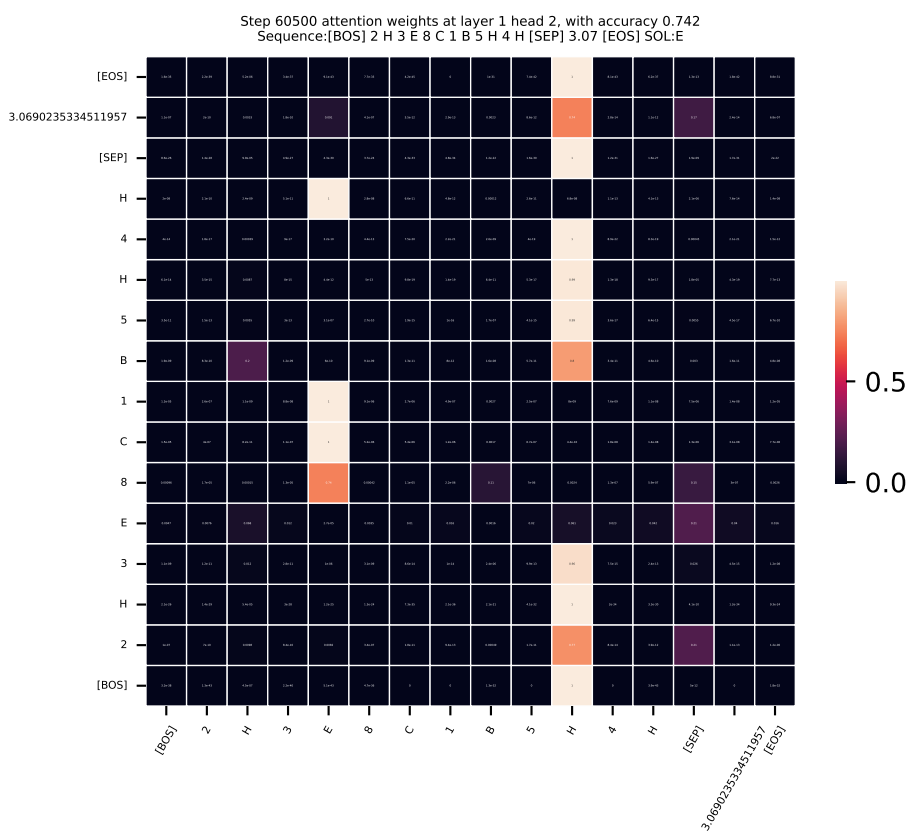

Step 60500 attention weights at layer 1 head 1, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

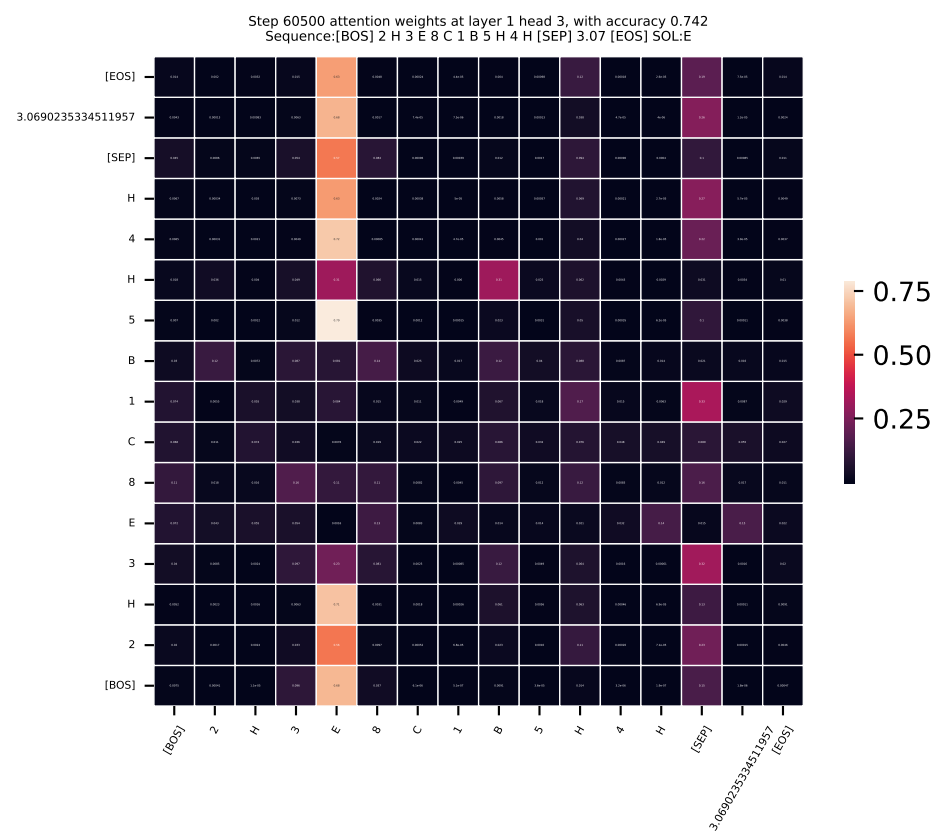

Step 60500 attention weights at layer 1 head 2, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.5
0.0

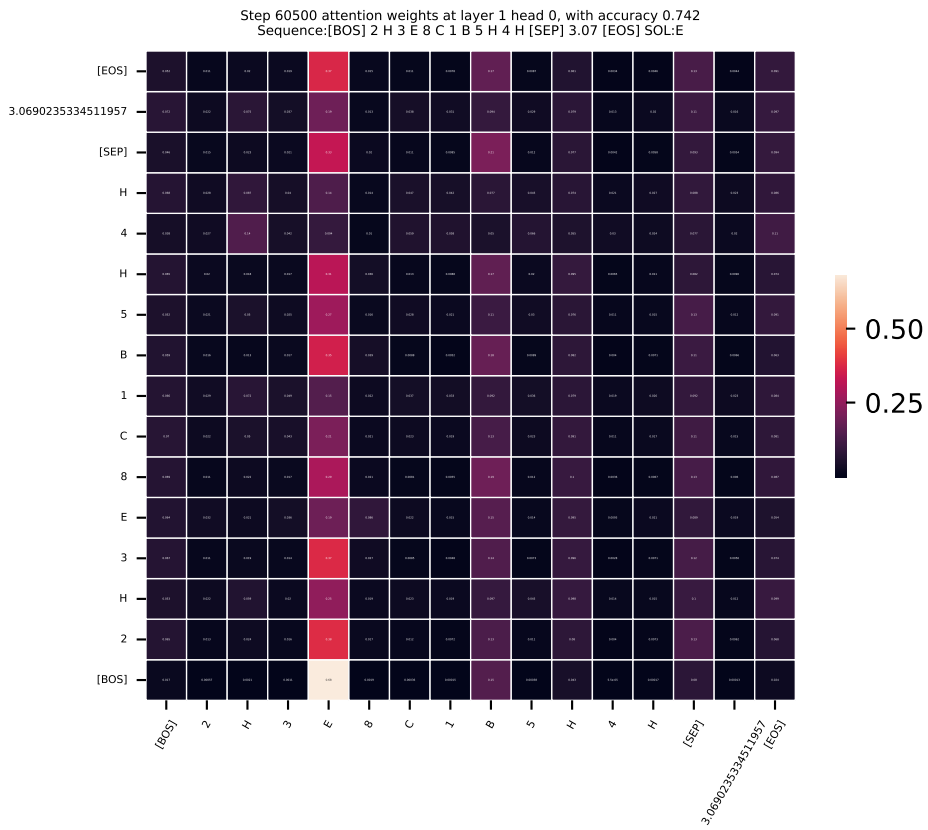

Step 60500 attention weights at layer 1 head 3, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

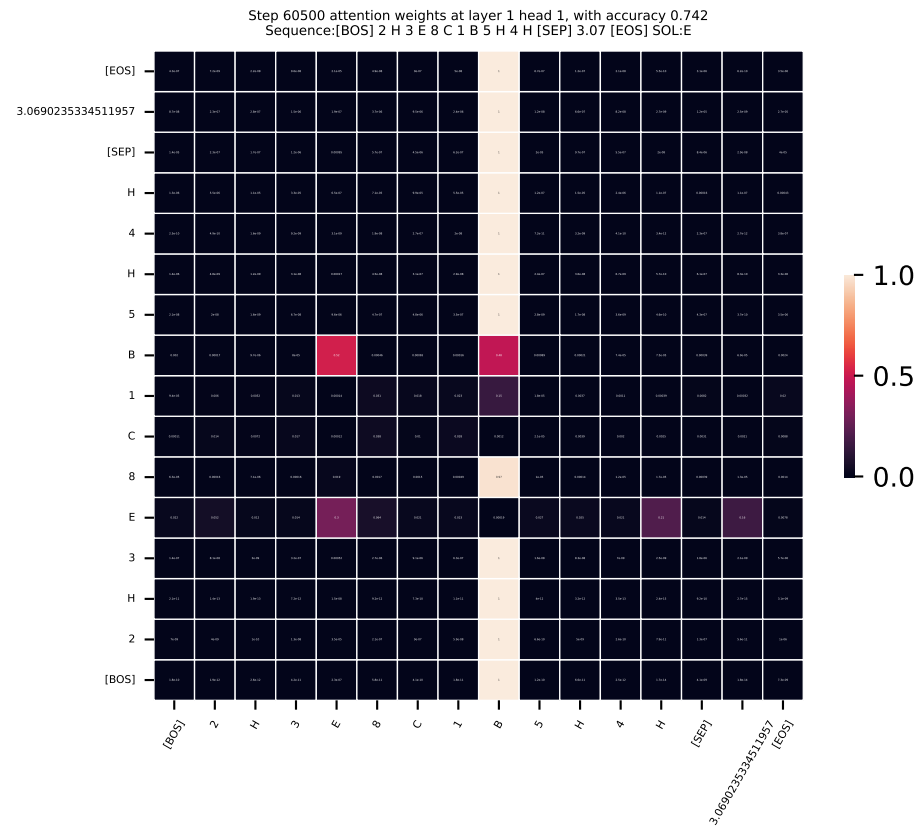

Step 60500 attention weights at layer 1 head 0, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.50
0.25

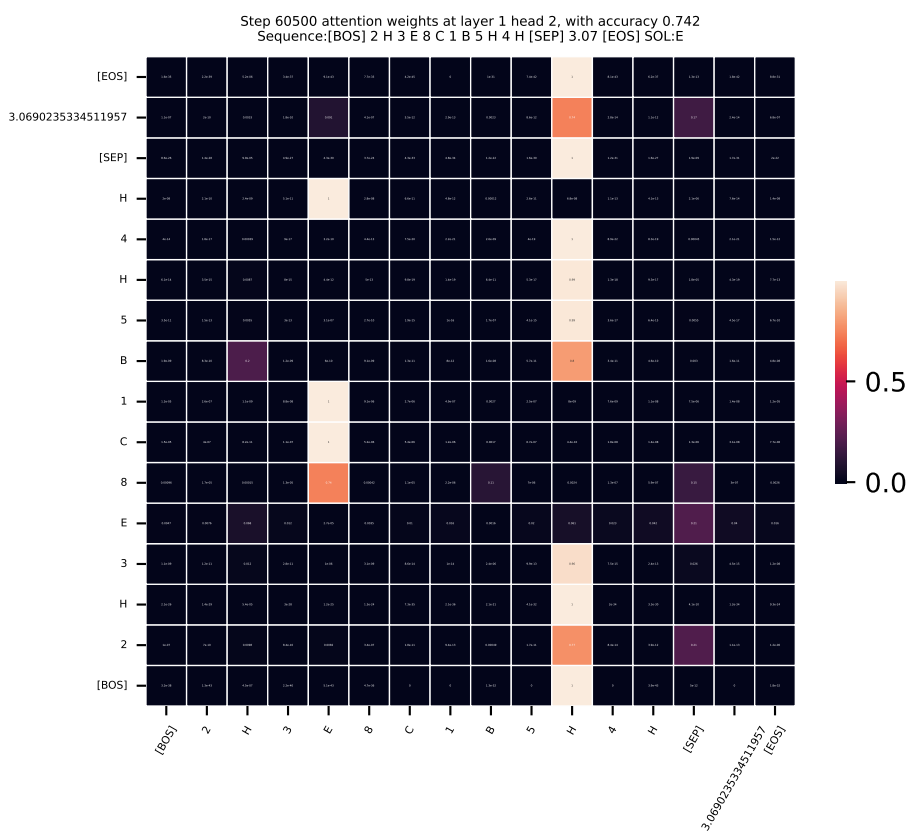

Step 60500 attention weights at layer 1 head 1, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

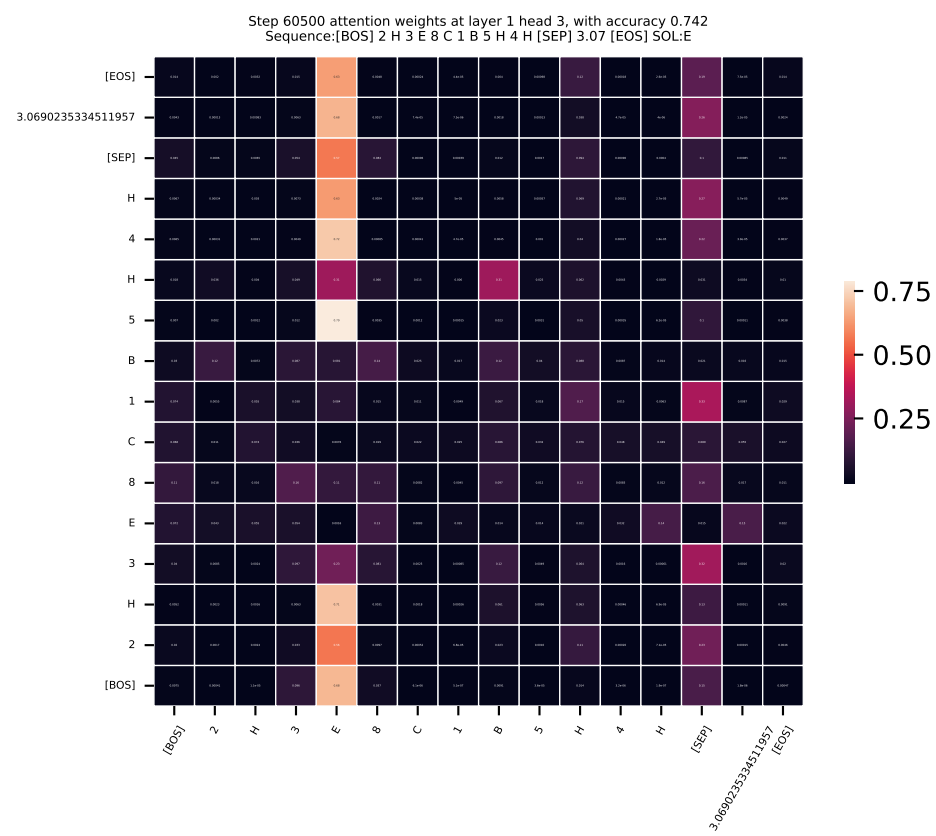

Step 60500 attention weights at layer 1 head 2, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.5
0.0

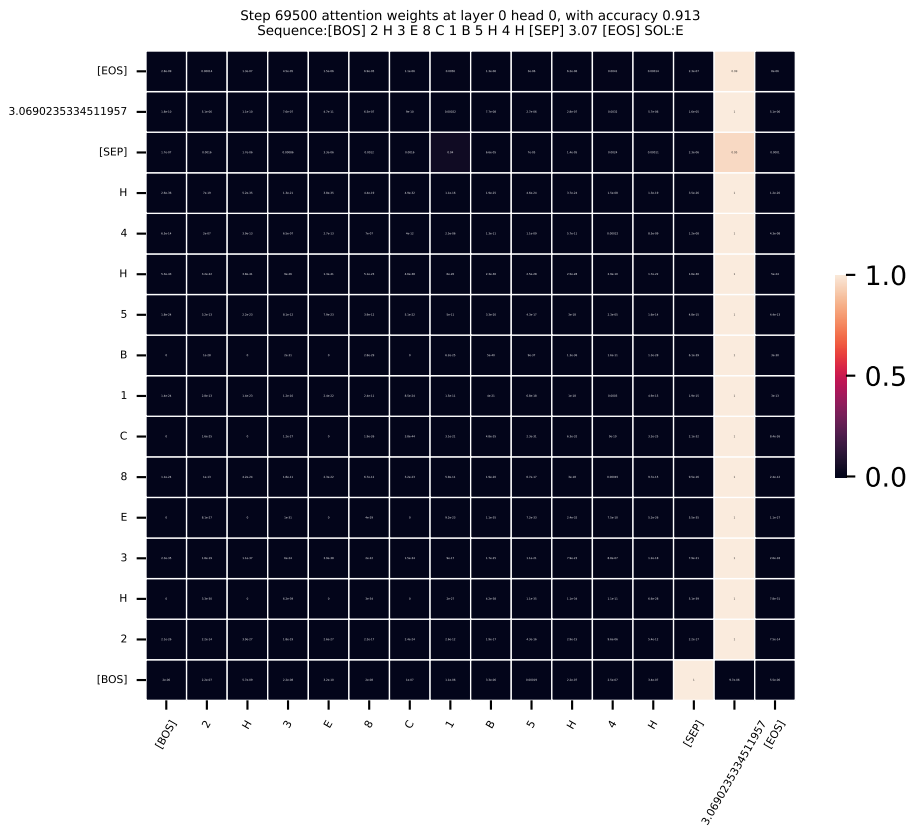

Step 60500 attention weights at layer 1 head 3, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

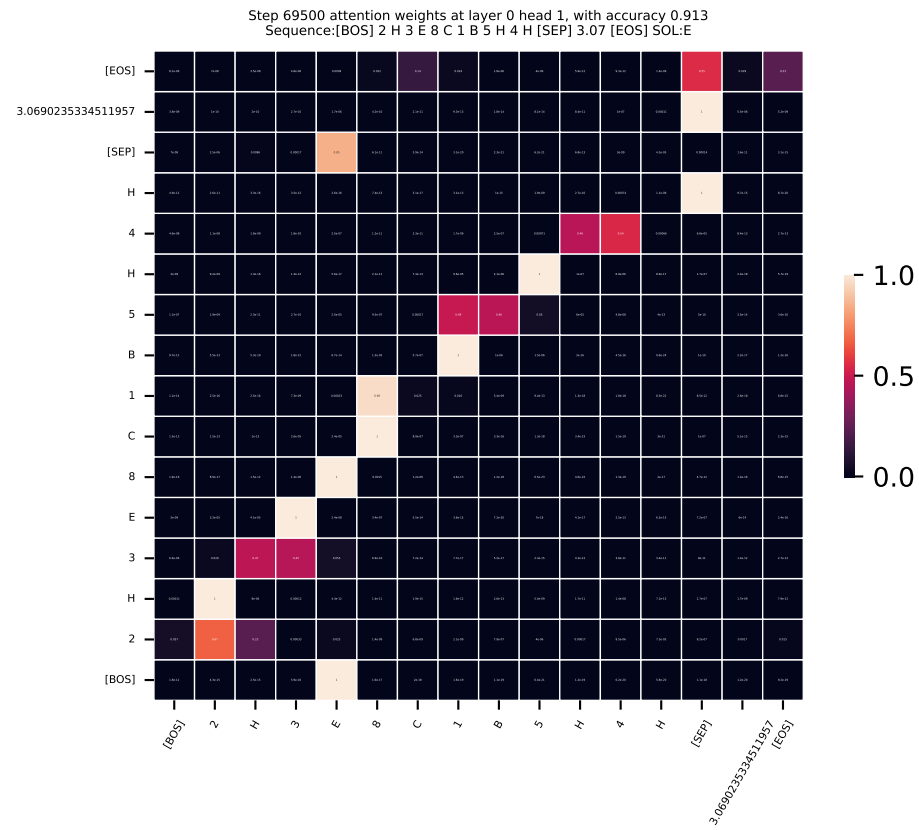

Step 60500 attention weights at layer 1 head 0, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.50
0.25

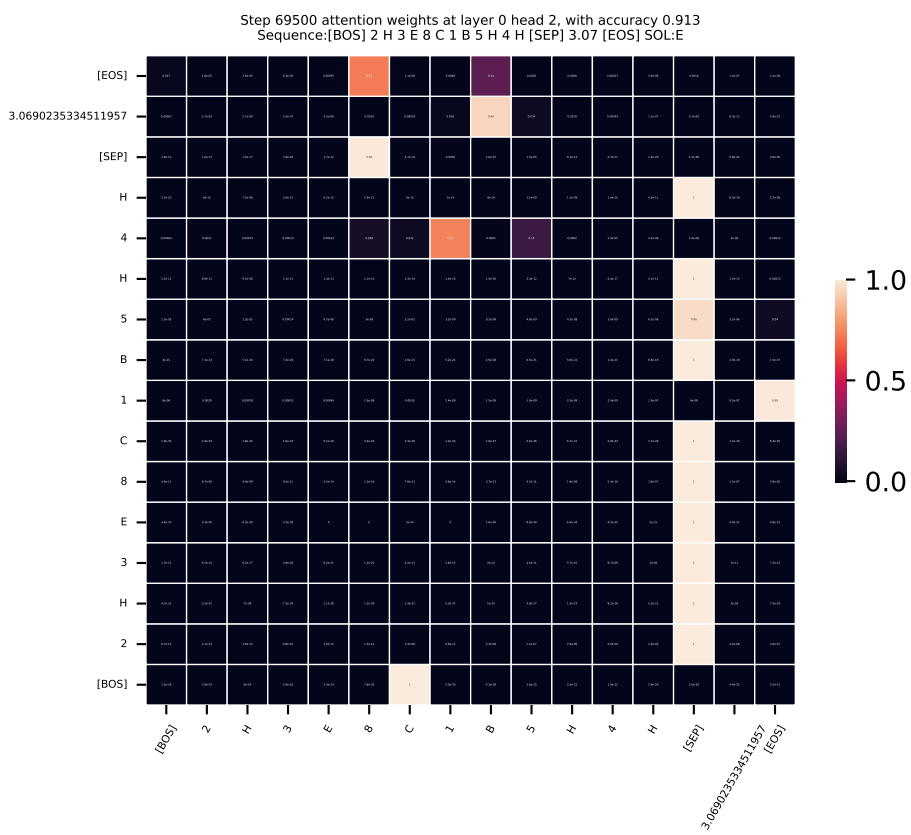

Step 60500 attention weights at layer 1 head 1, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

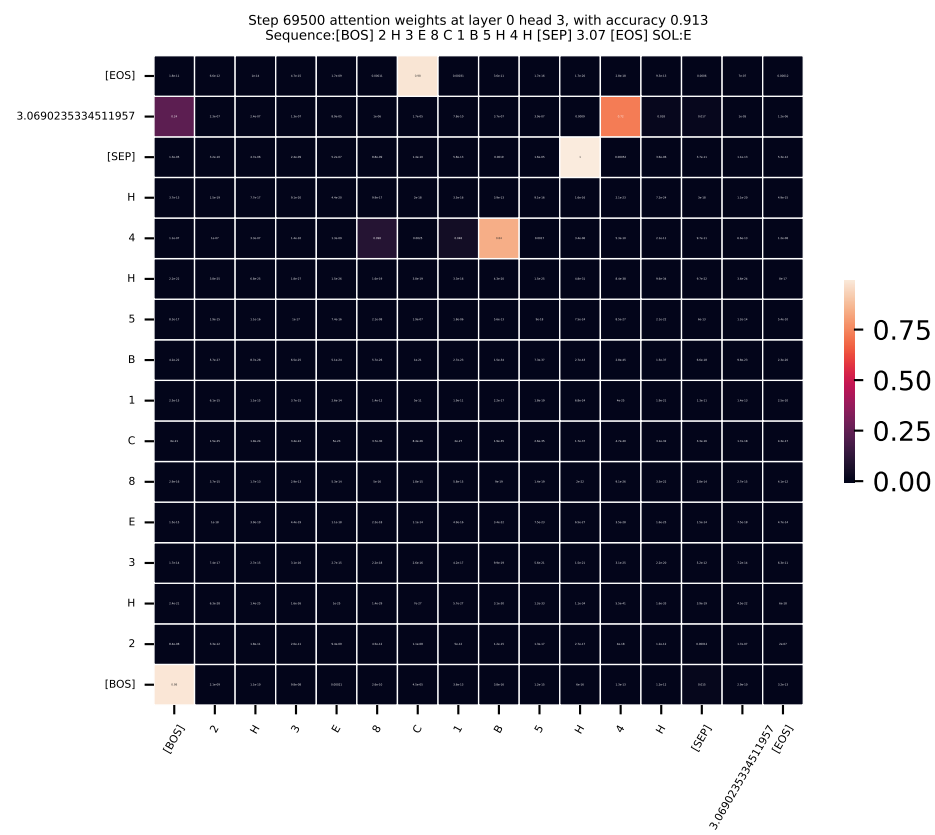

Step 60500 attention weights at layer 1 head 2, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.5
0.0

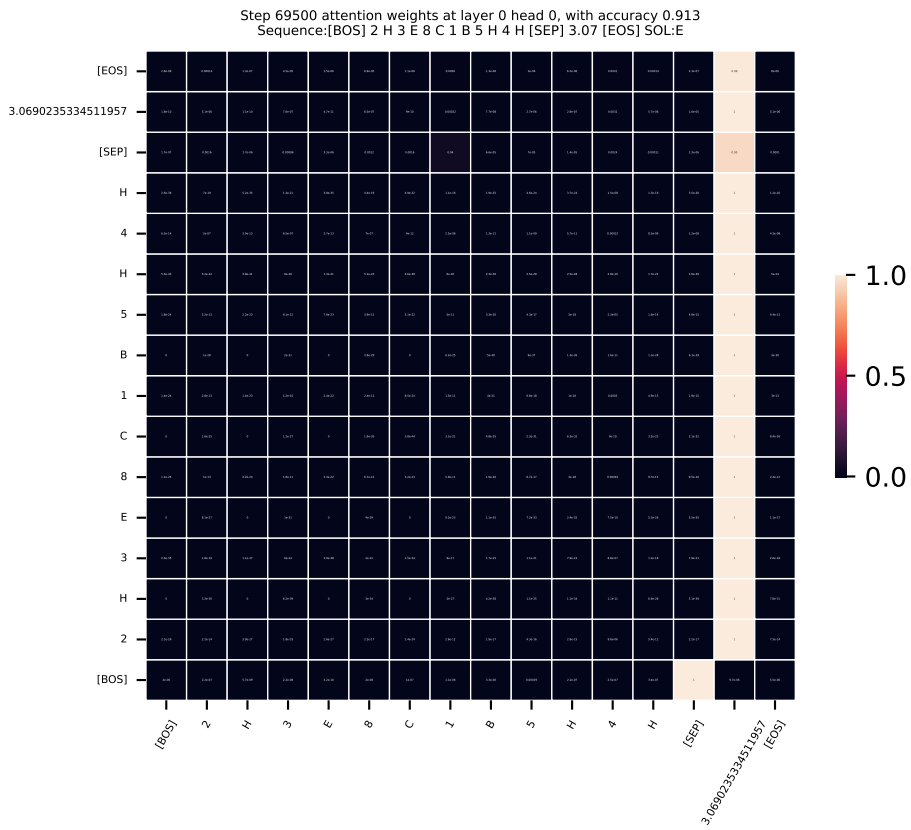

Step 60500 attention weights at layer 1 head 3, with accuracy 0.742
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

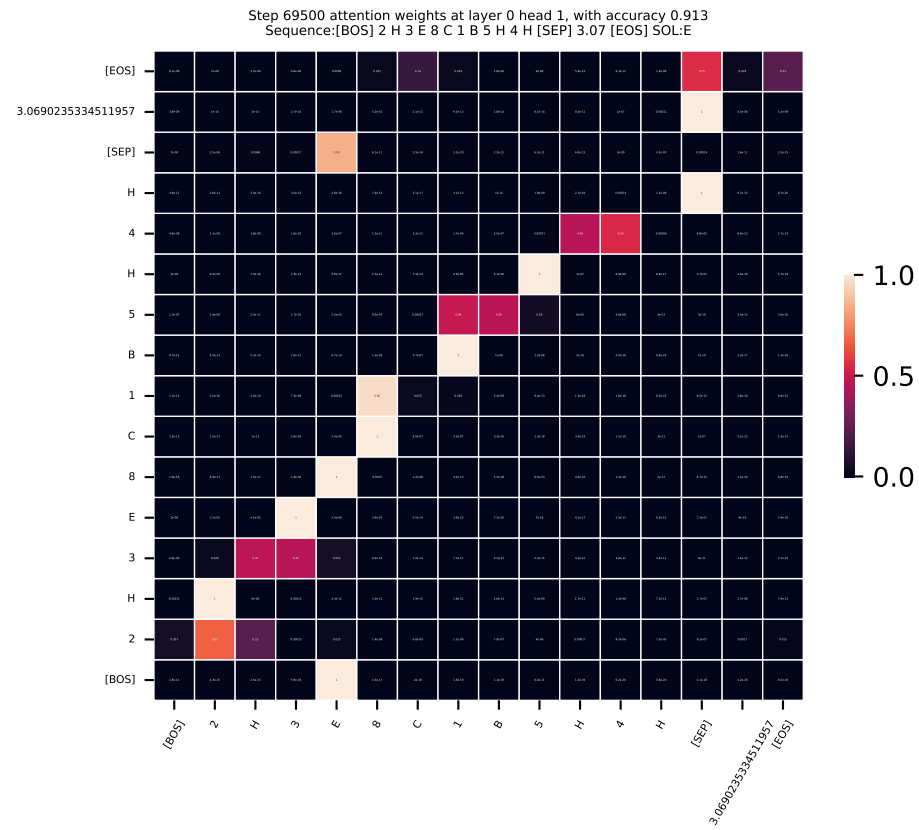

Step 69500 attention weights at layer 0 head 0, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

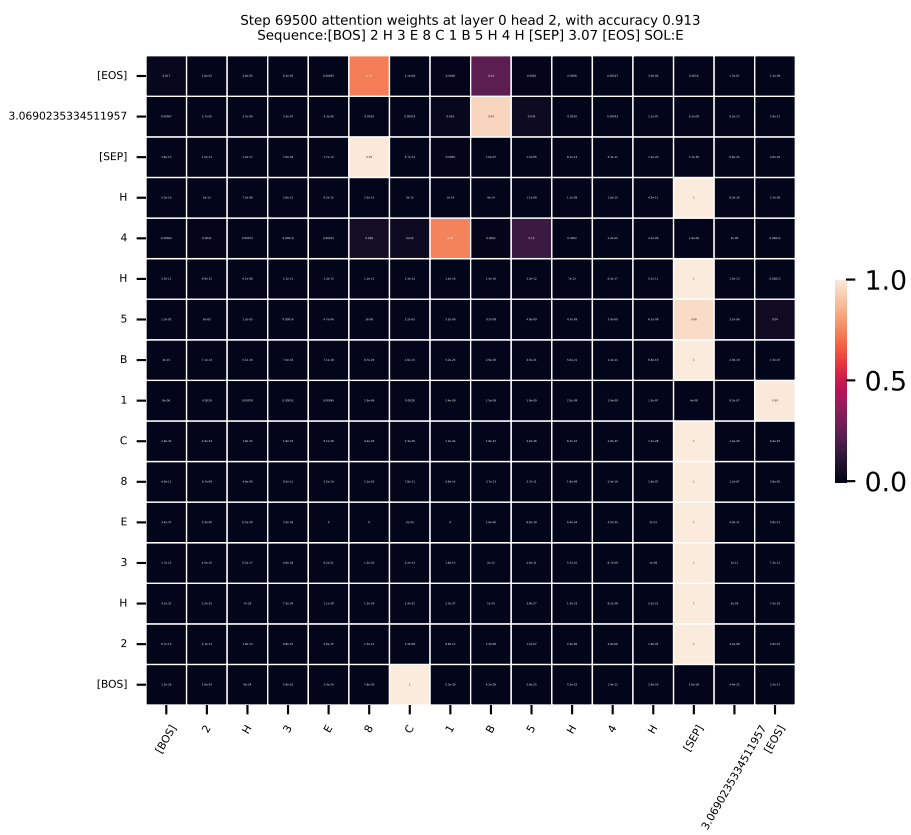

Step 69500 attention weights at layer 0 head 1, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

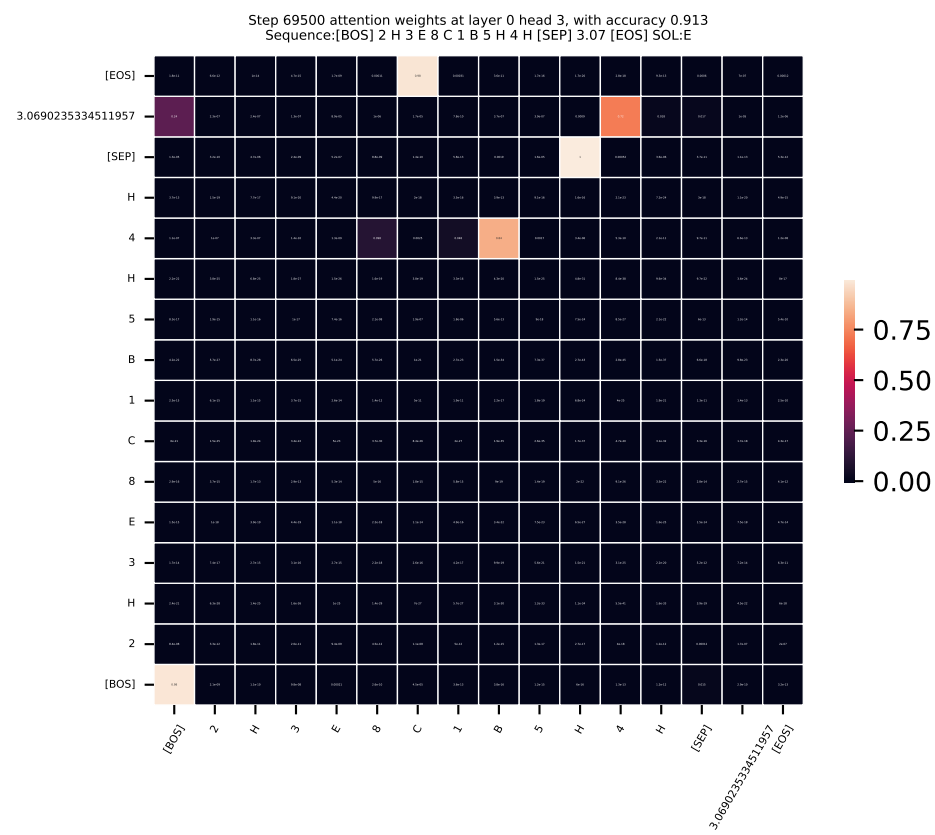

Step 69500 attention weights at layer 0 head 2, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

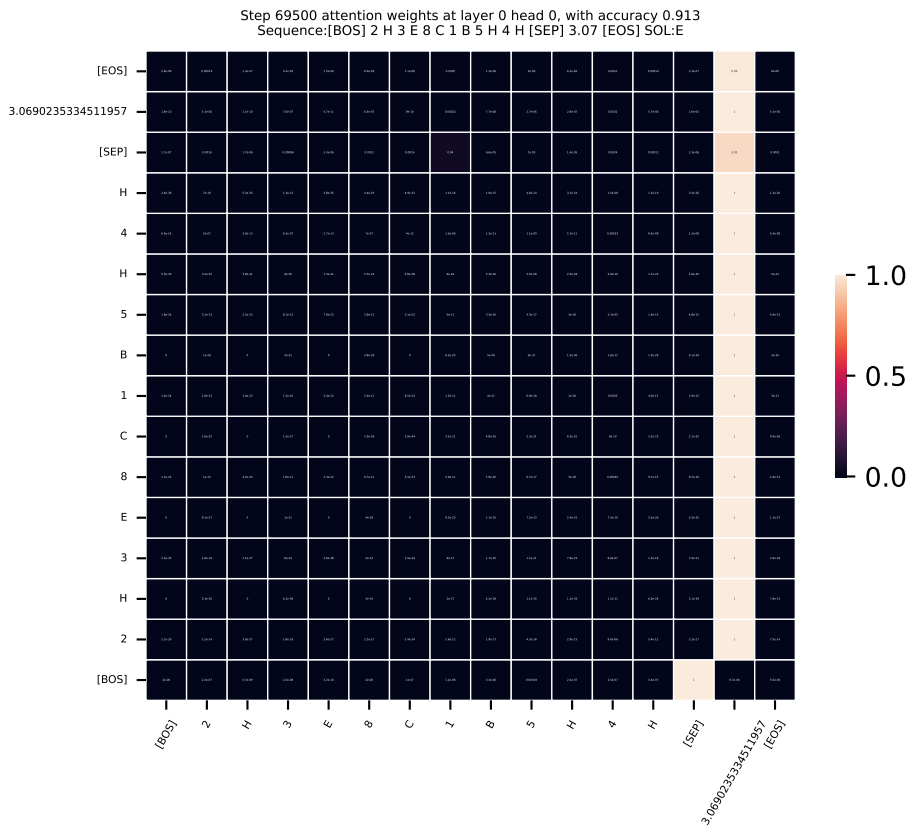

Step 69500 attention weights at layer 0 head 3, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.75
0.50
0.25
0.00

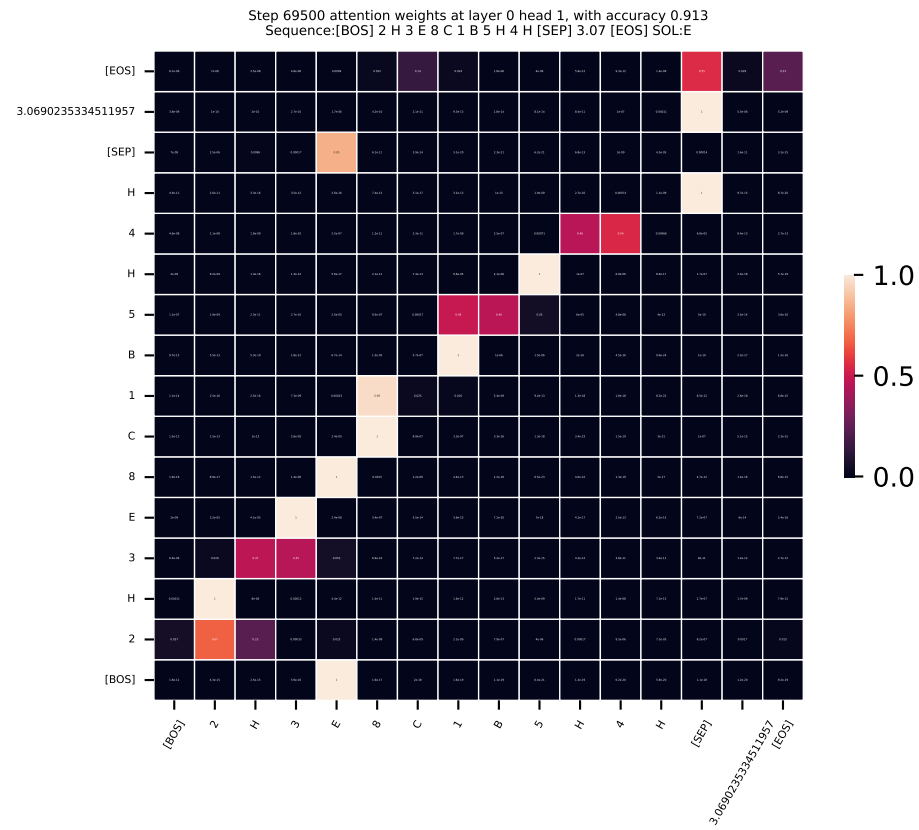

Step 69500 attention weights at layer 0 head 0, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

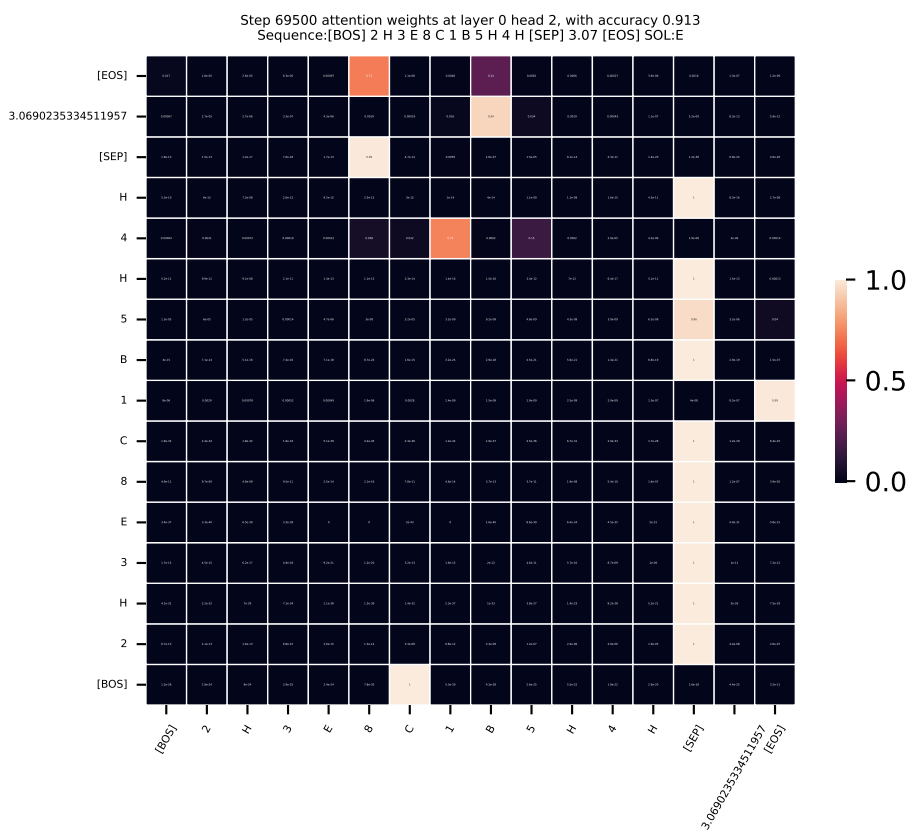

Step 69500 attention weights at layer 0 head 1, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

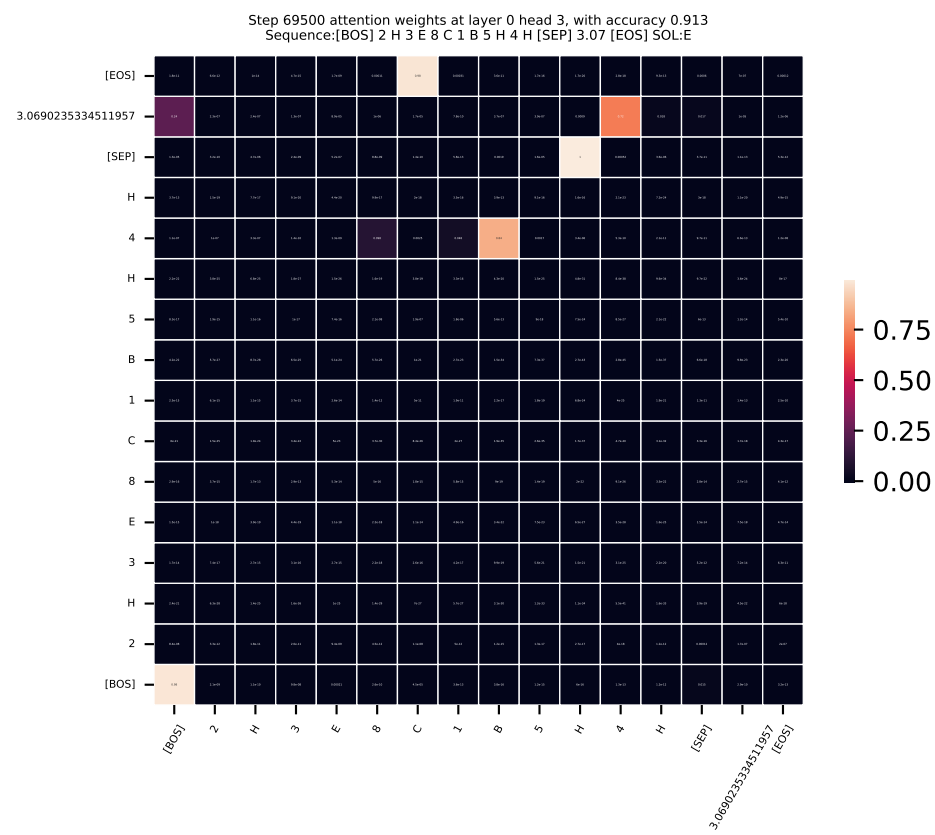

Step 69500 attention weights at layer 0 head 2, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

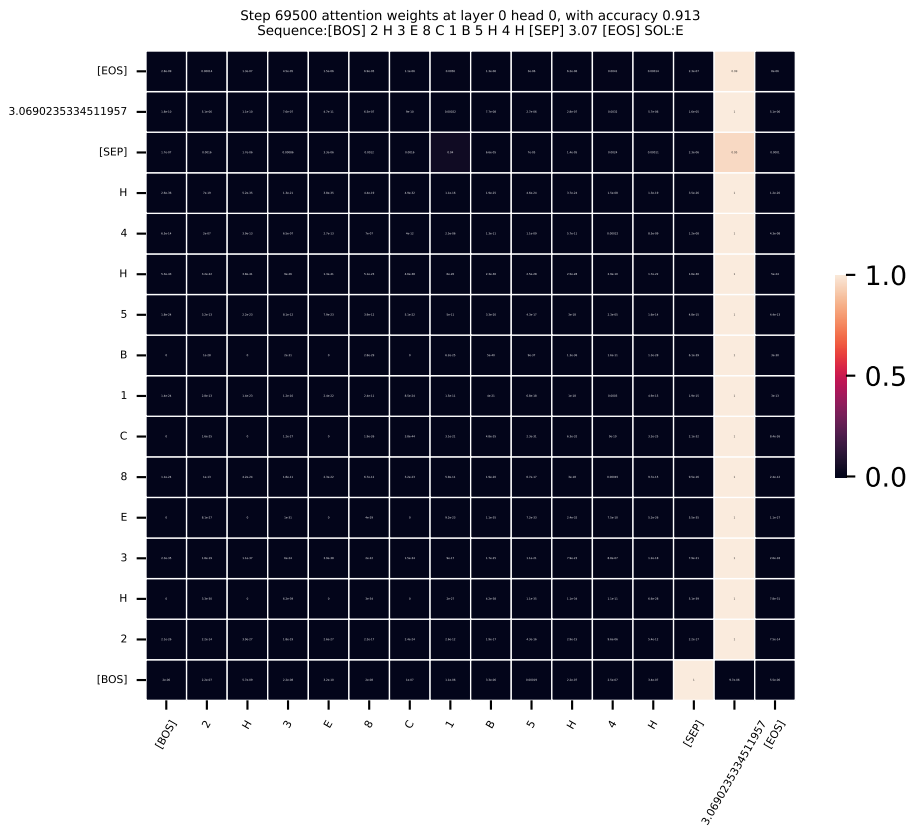

Step 69500 attention weights at layer 0 head 3, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.75
0.50
0.25
0.00

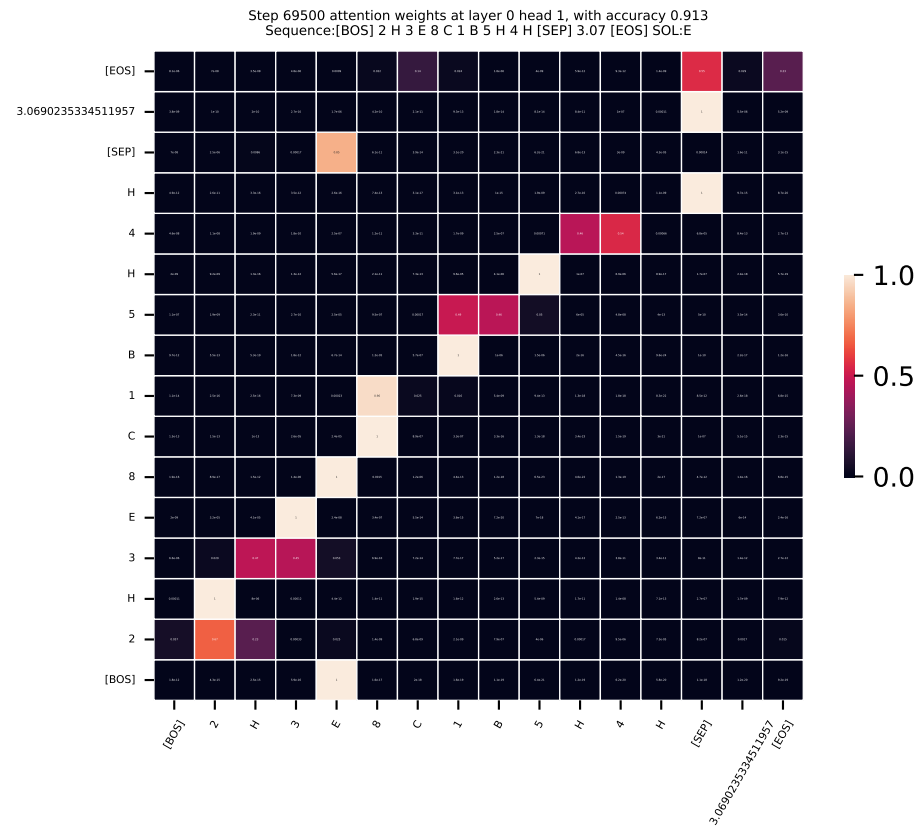

Step 69500 attention weights at layer 0 head 0, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

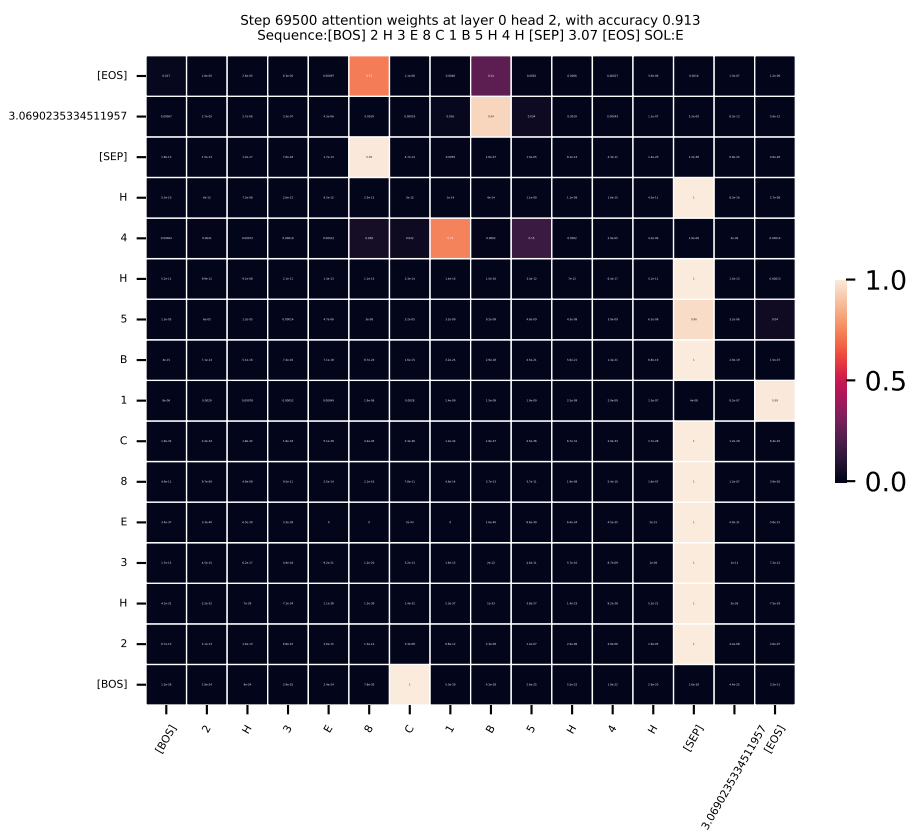

Step 69500 attention weights at layer 0 head 1, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

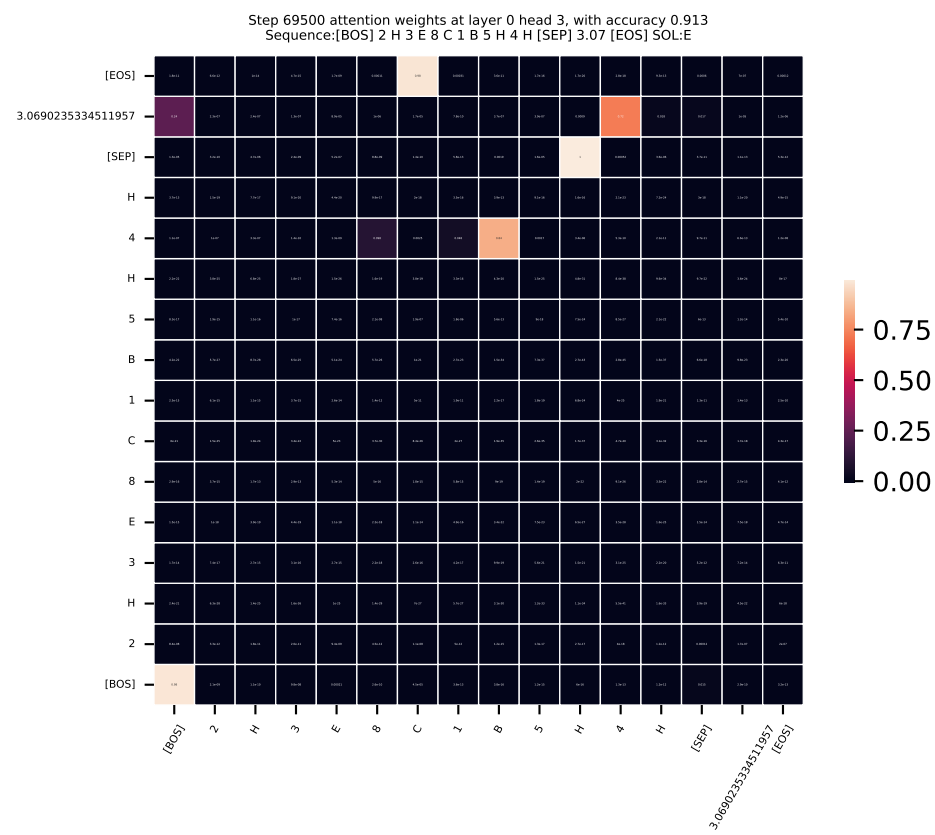

Step 69500 attention weights at layer 0 head 2, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

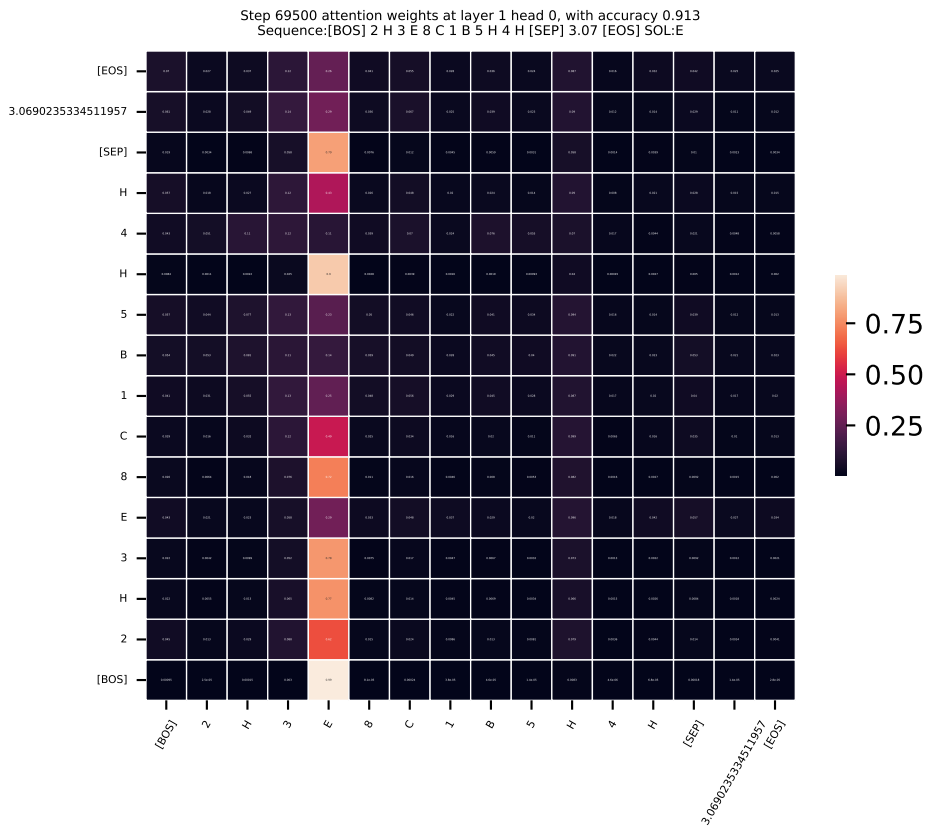

Step 69500 attention weights at layer 0 head 3, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.75
0.50
0.25
0.00

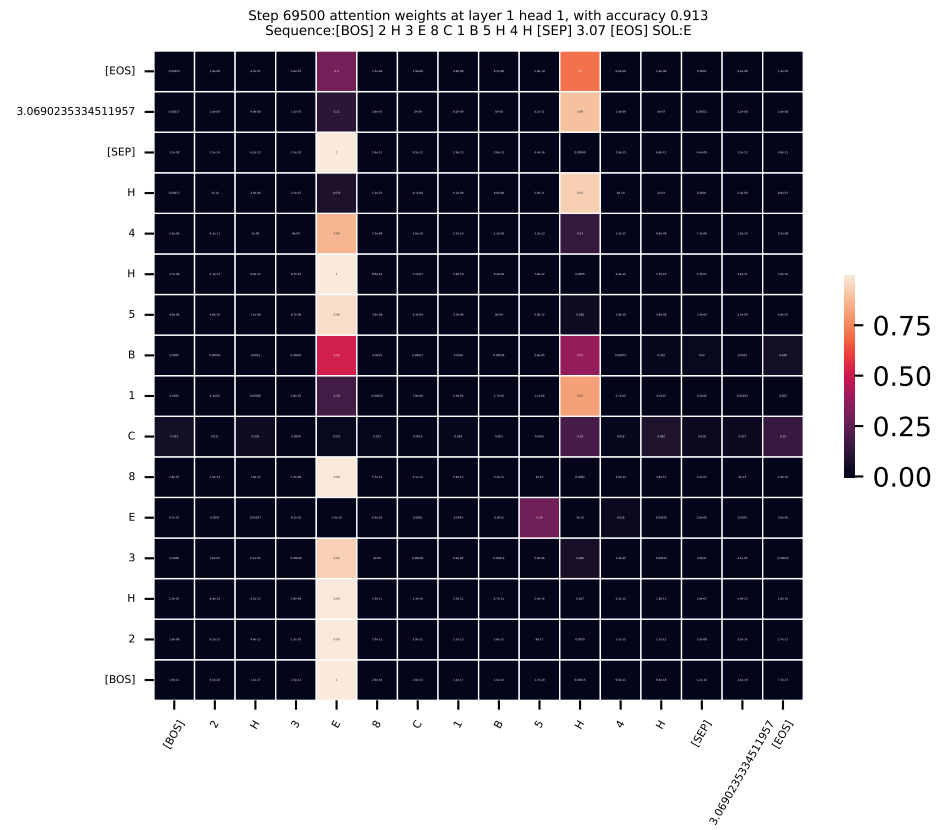

Step 69500 attention weights at layer 0 head 0, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

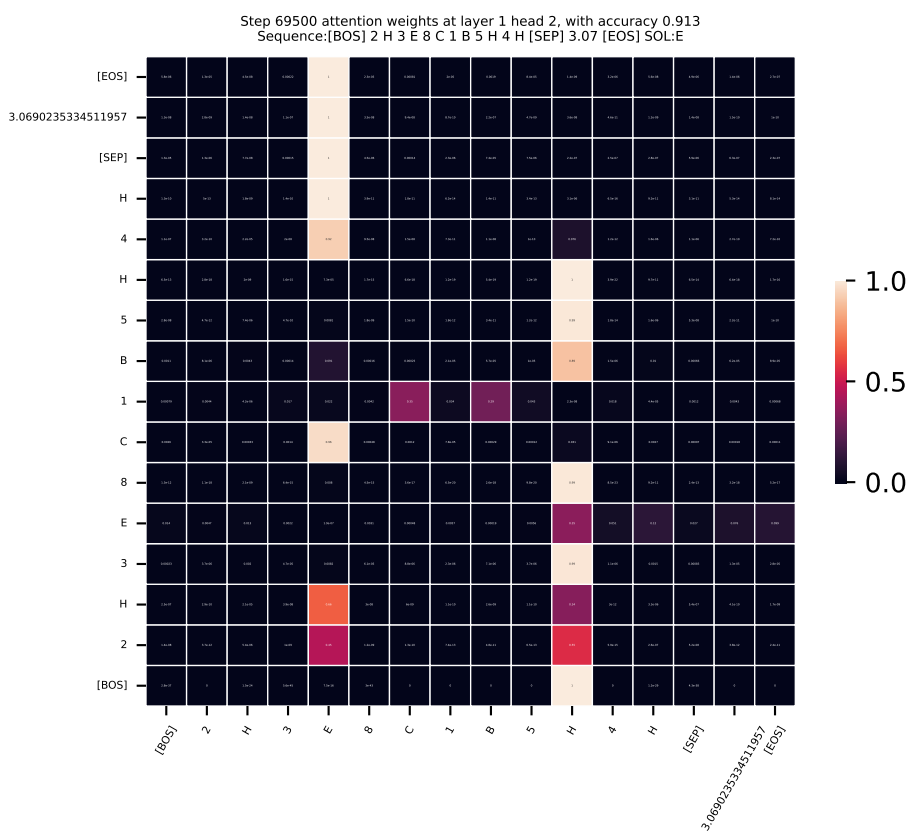

Step 69500 attention weights at layer 0 head 1, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

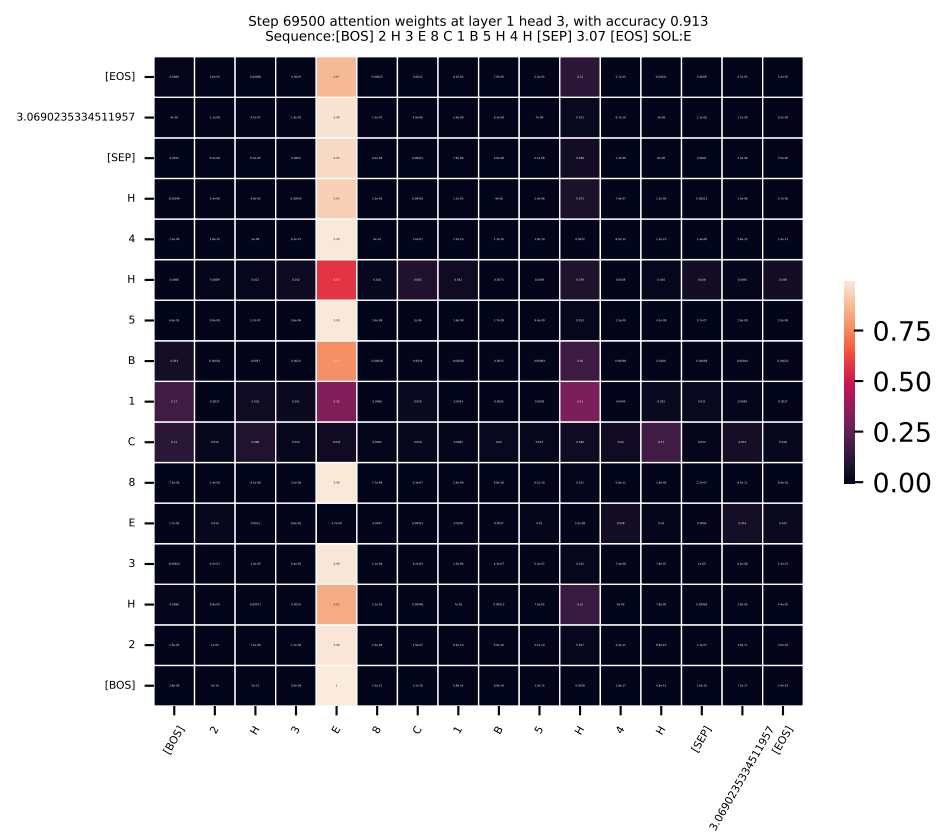

Step 69500 attention weights at layer 0 head 2, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

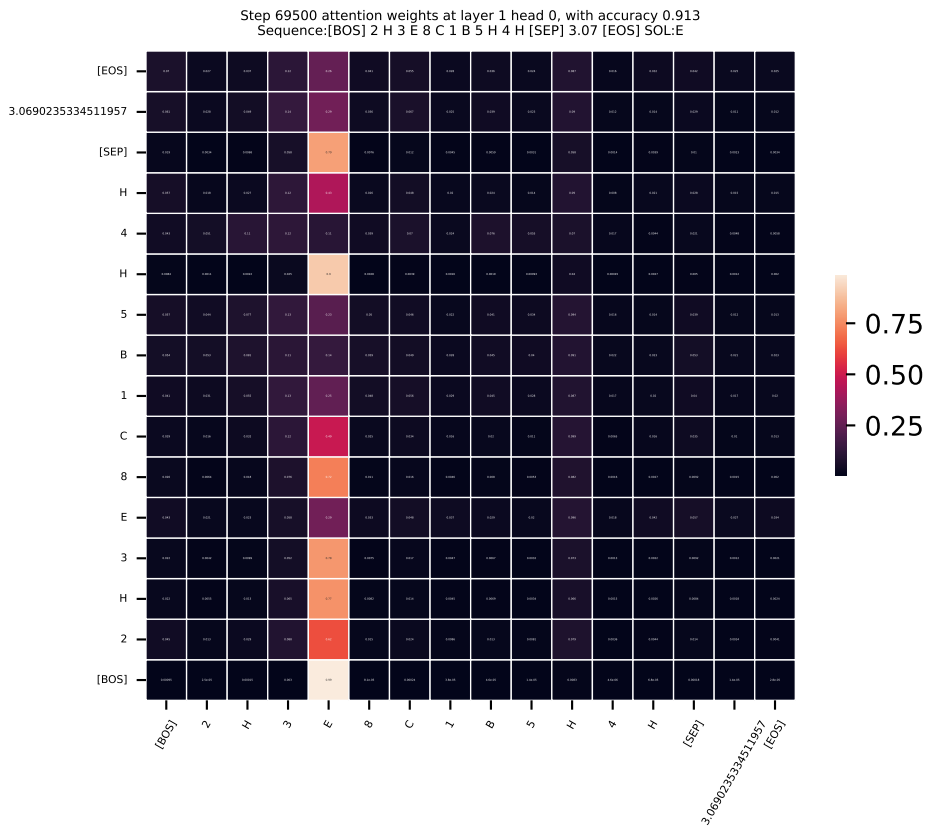

Step 69500 attention weights at layer 0 head 3, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.75
0.50
0.25
0.00

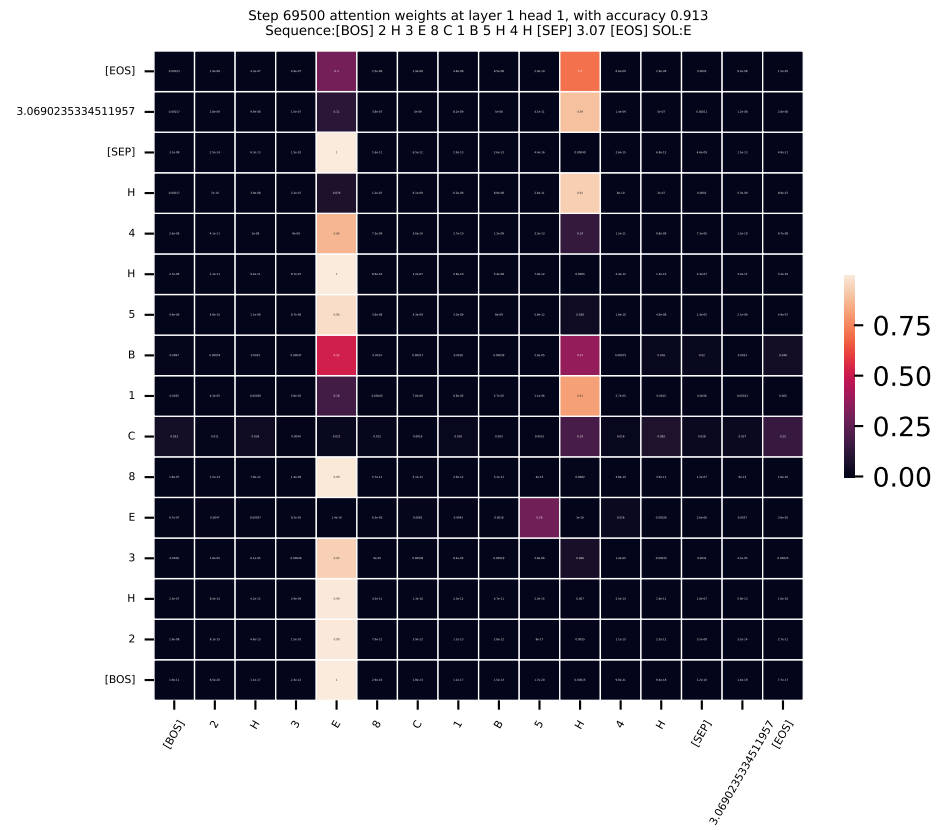

Step 69500 attention weights at layer 1 head 0, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

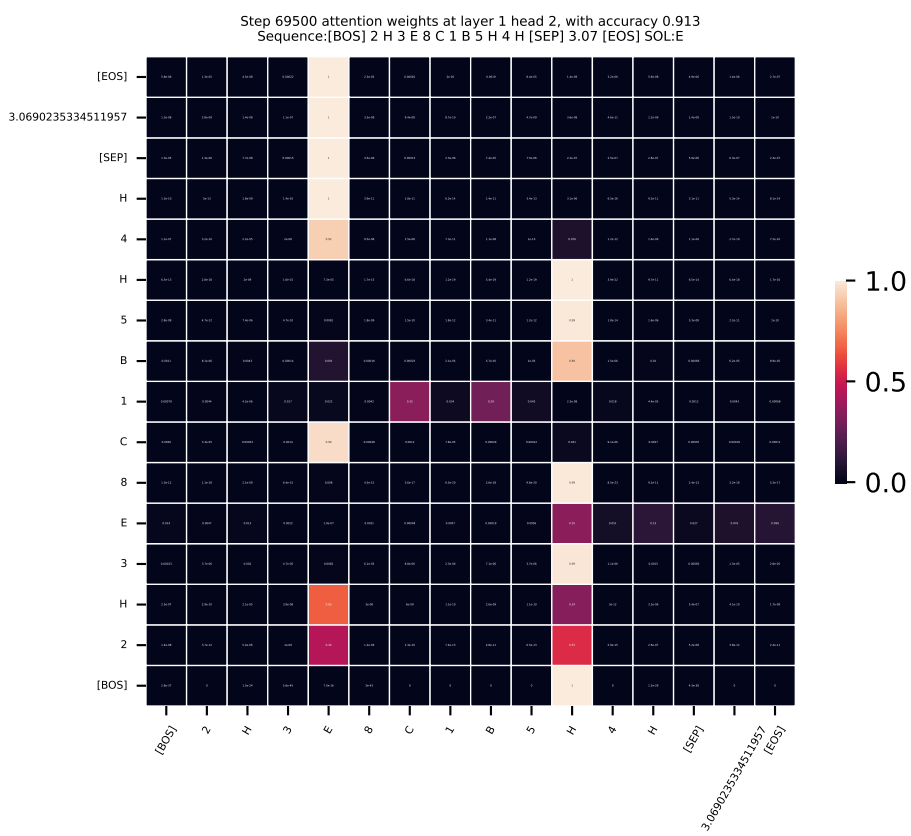

Step 69500 attention weights at layer 1 head 1, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

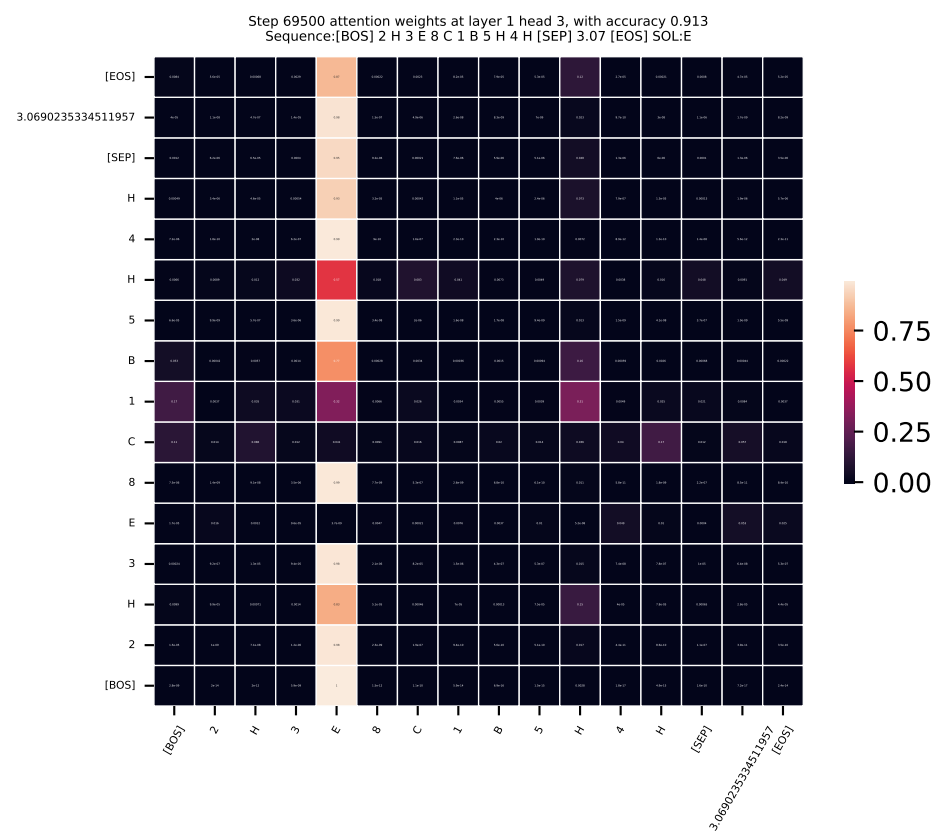

Step 69500 attention weights at layer 1 head 2, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

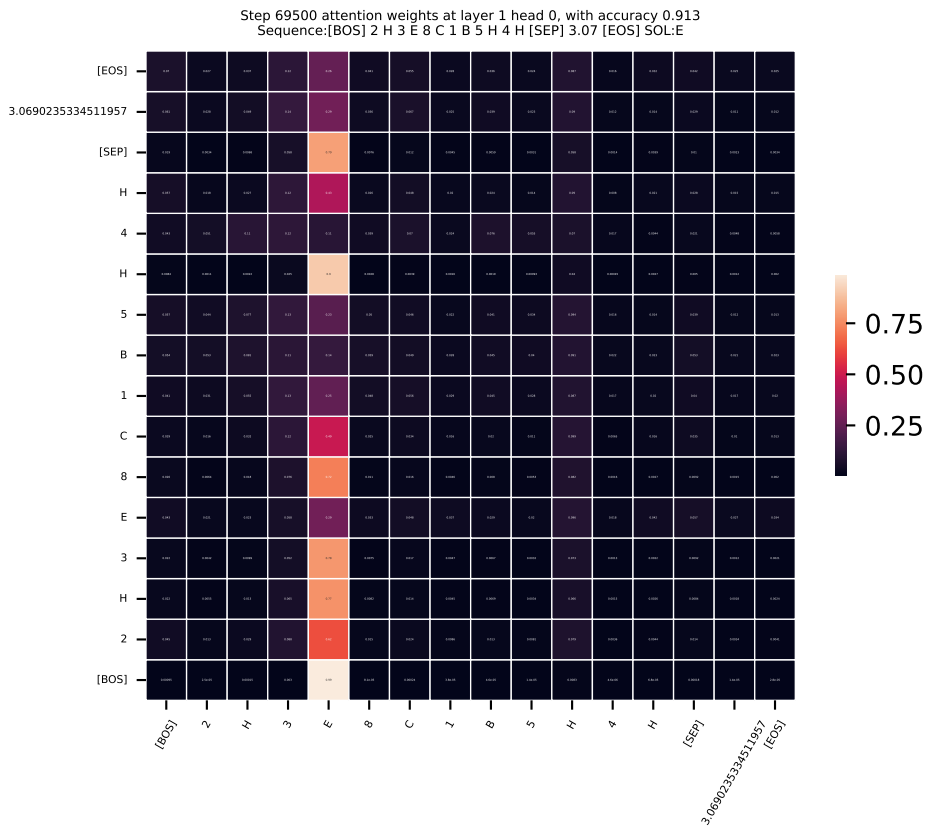

Step 69500 attention weights at layer 1 head 3, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

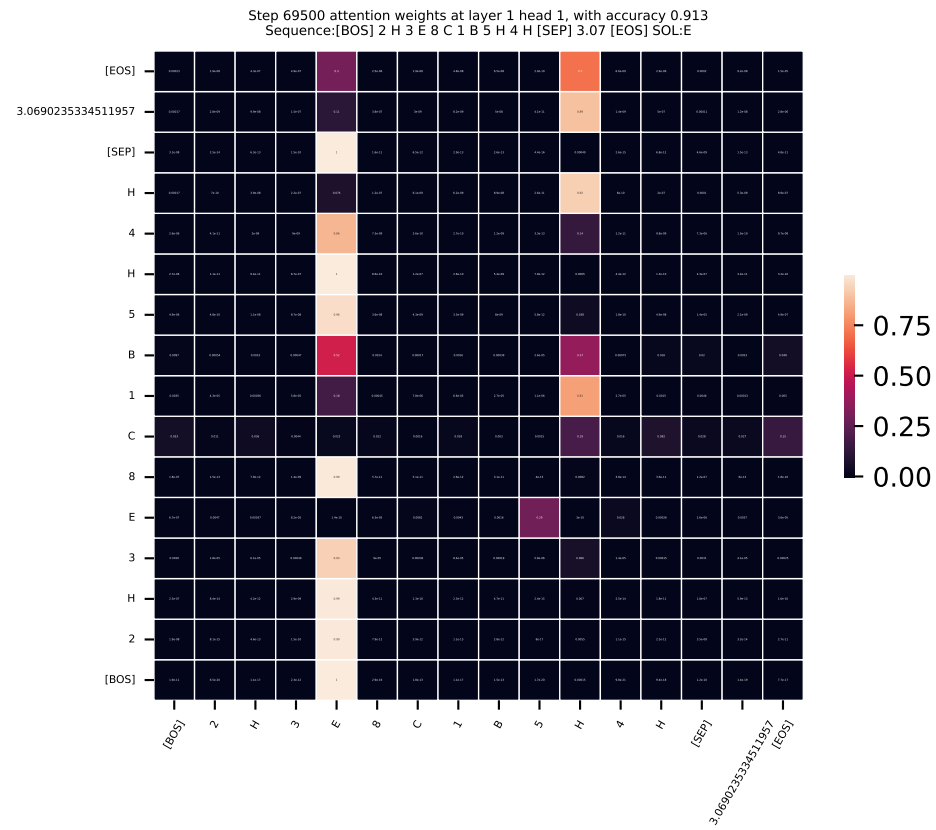

Step 69500 attention weights at layer 1 head 0, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

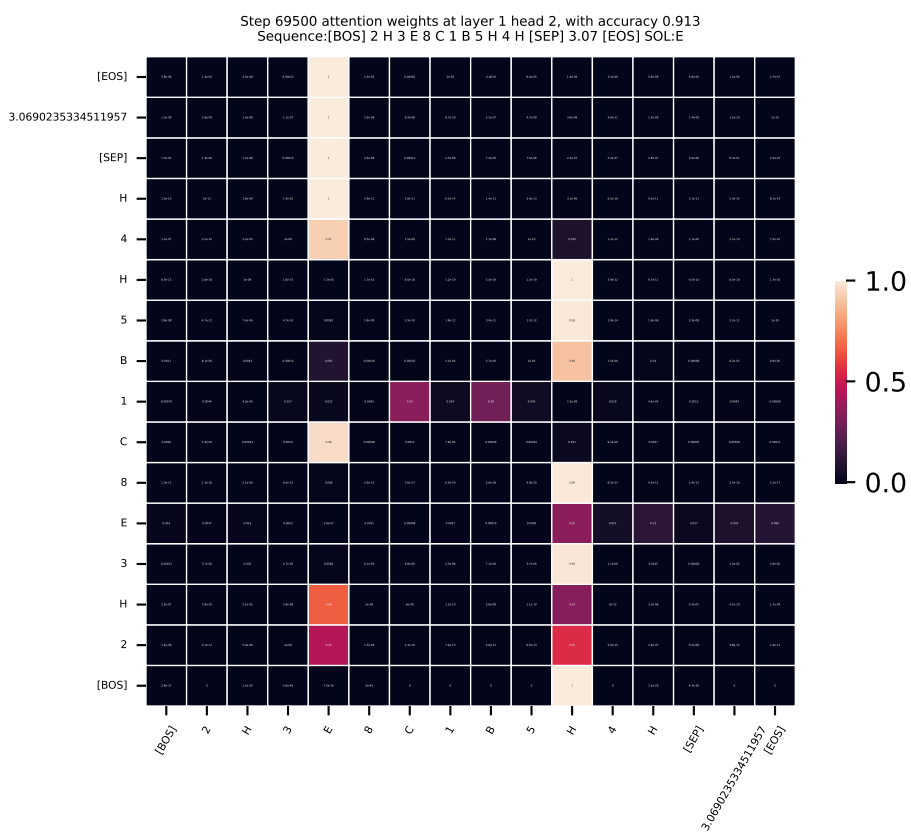

Step 69500 attention weights at layer 1 head 1, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

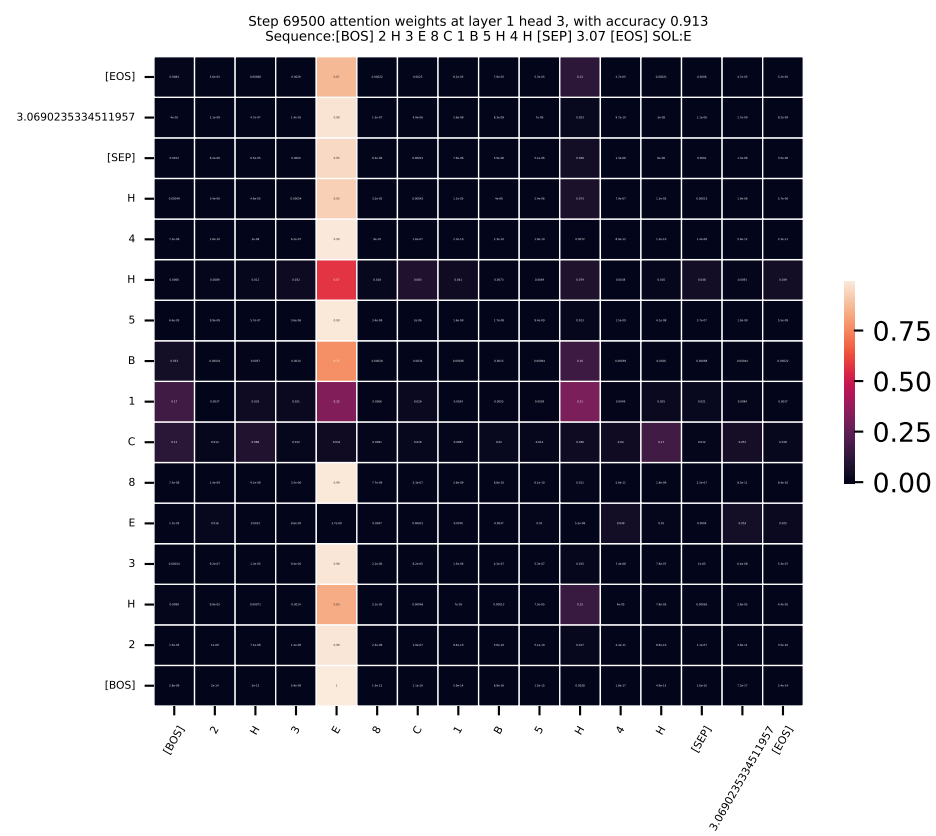

Step 69500 attention weights at layer 1 head 2, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

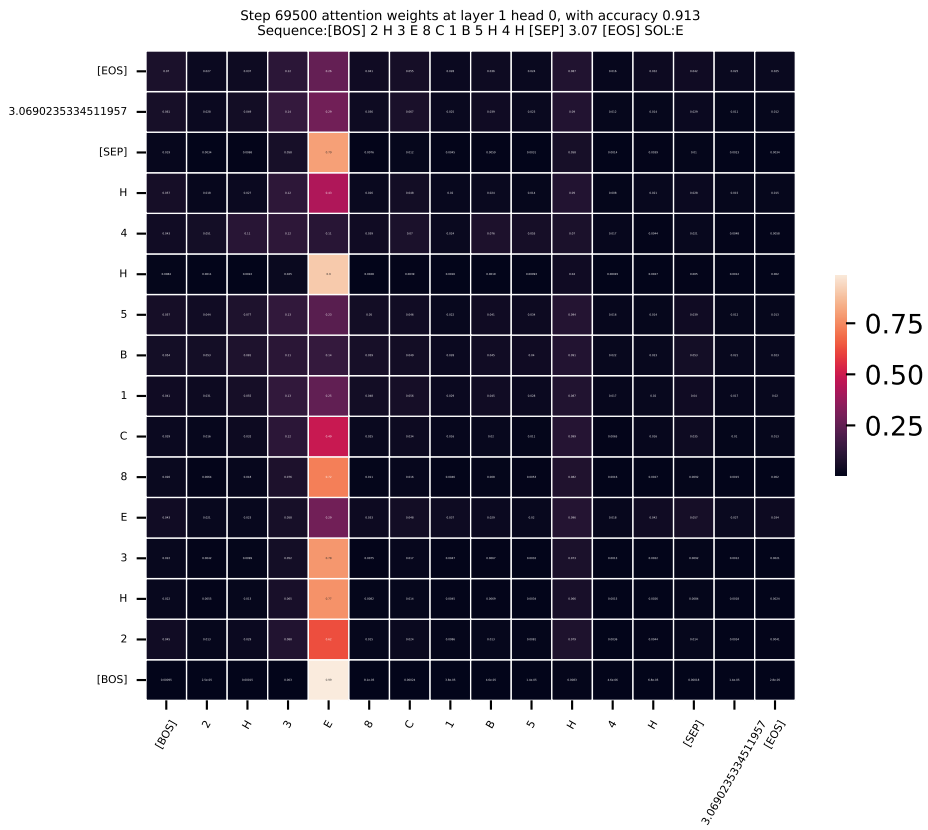

Step 69500 attention weights at layer 1 head 3, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

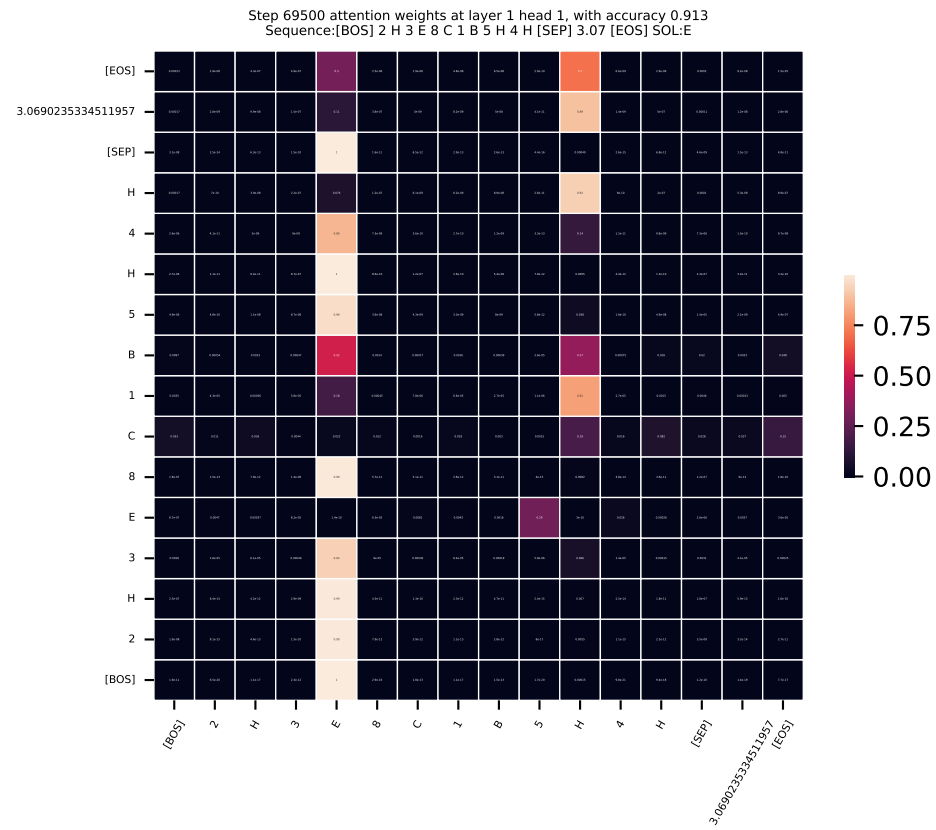

Step 69500 attention weights at layer 1 head 0, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

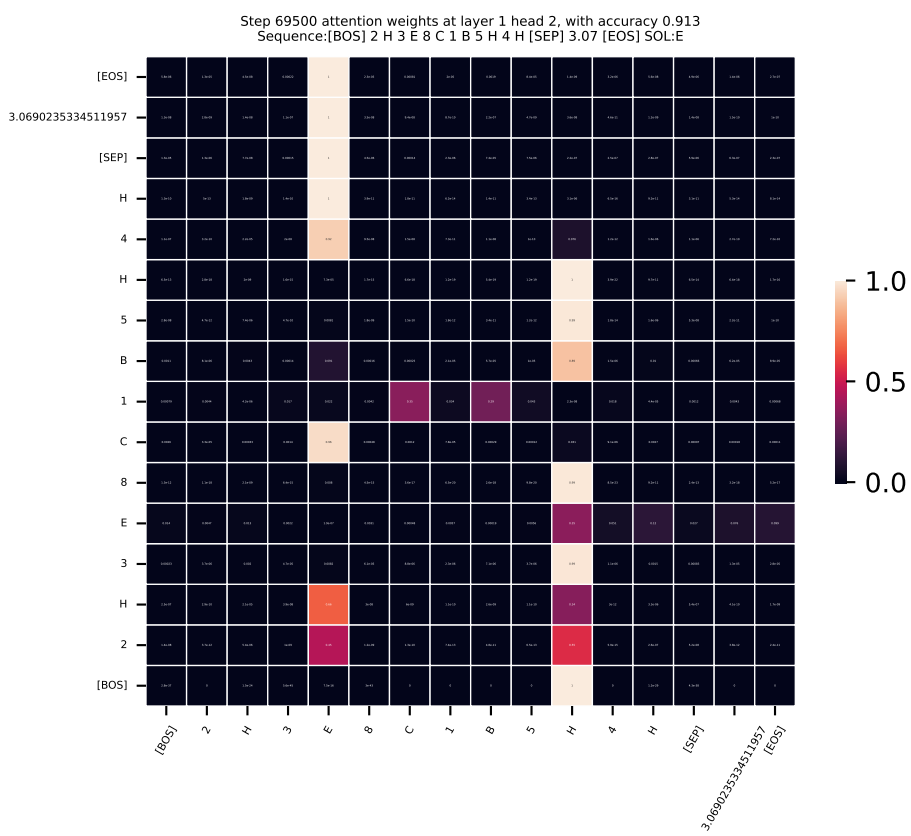

Step 69500 attention weights at layer 1 head 1, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

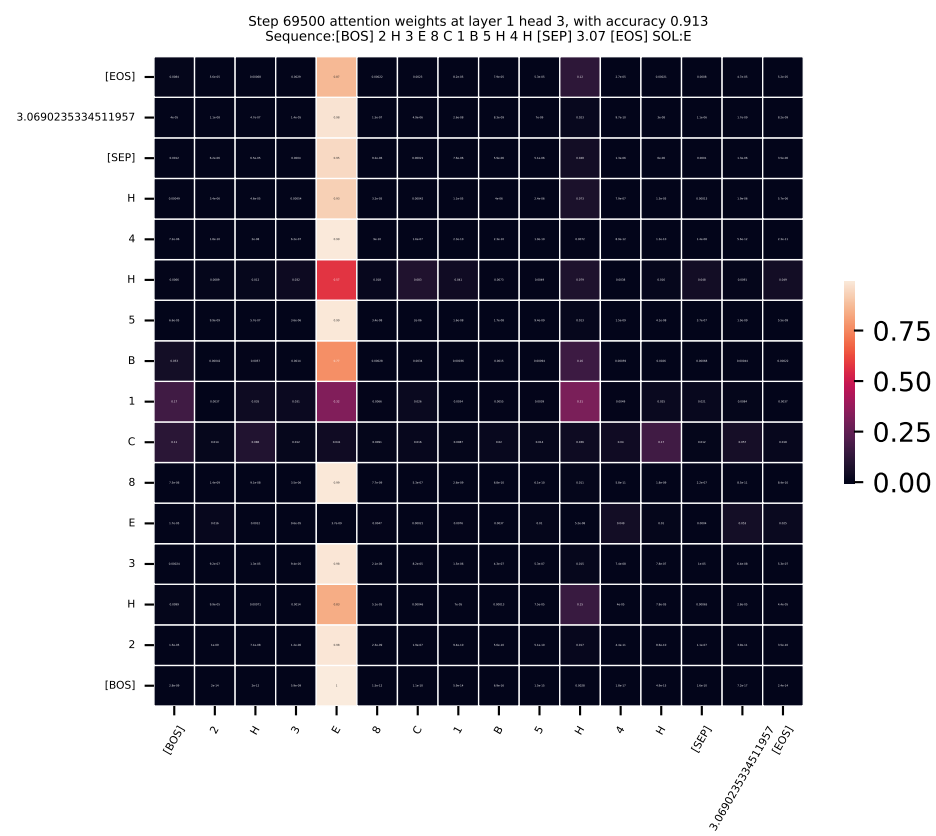

Step 69500 attention weights at layer 1 head 2, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

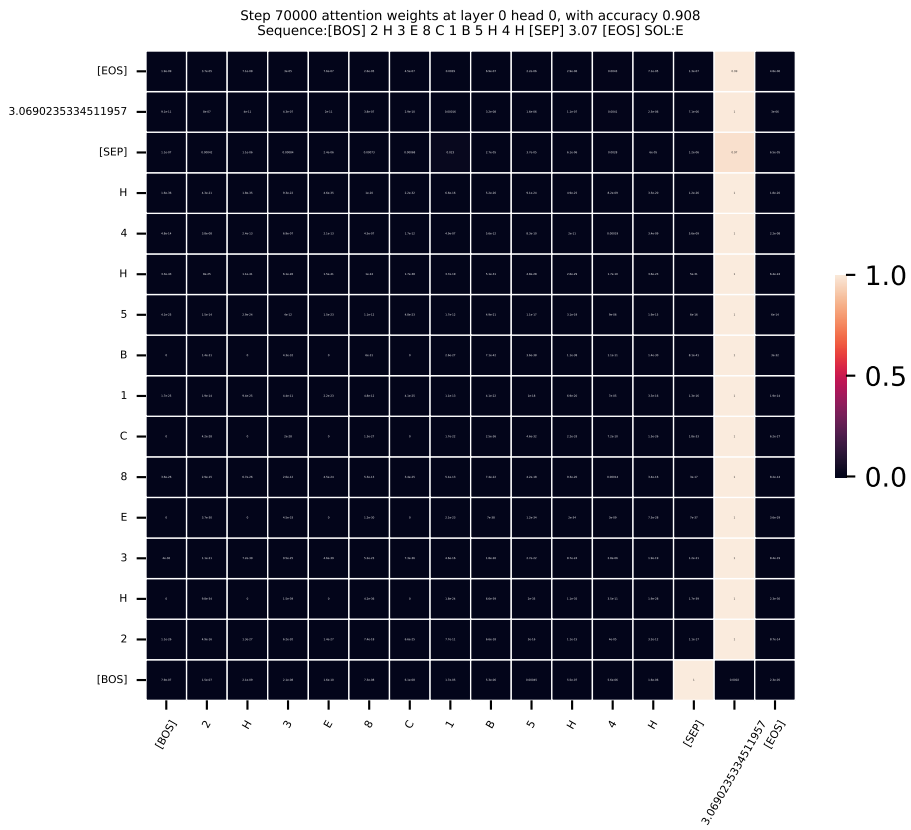

Step 69500 attention weights at layer 1 head 3, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

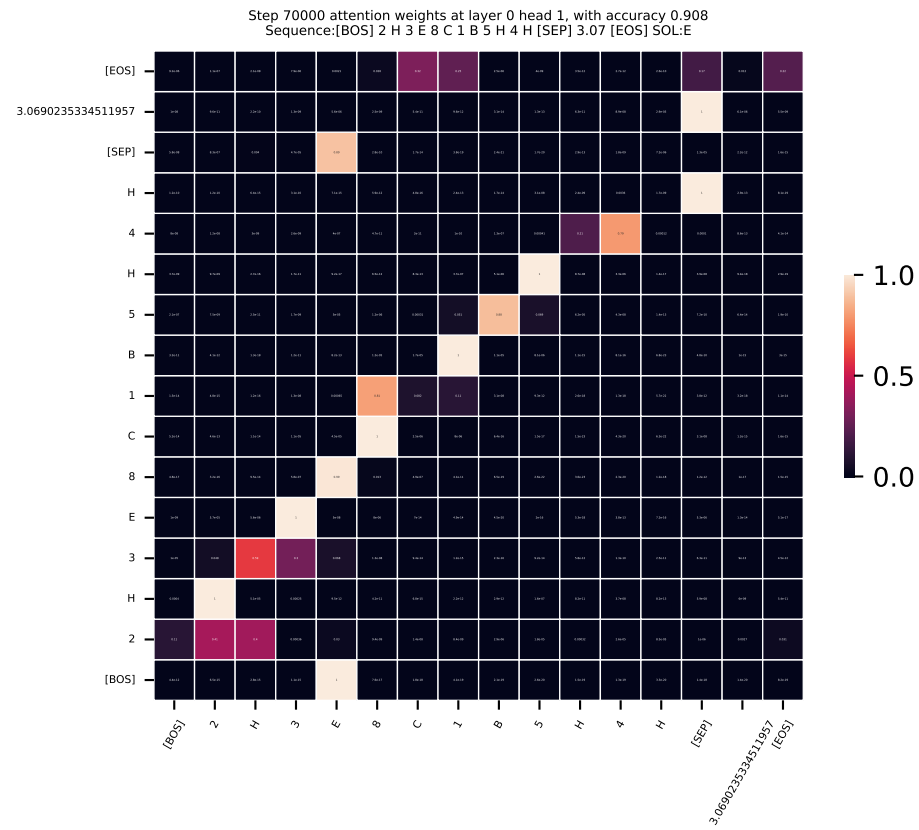

Step 69500 attention weights at layer 1 head 0, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

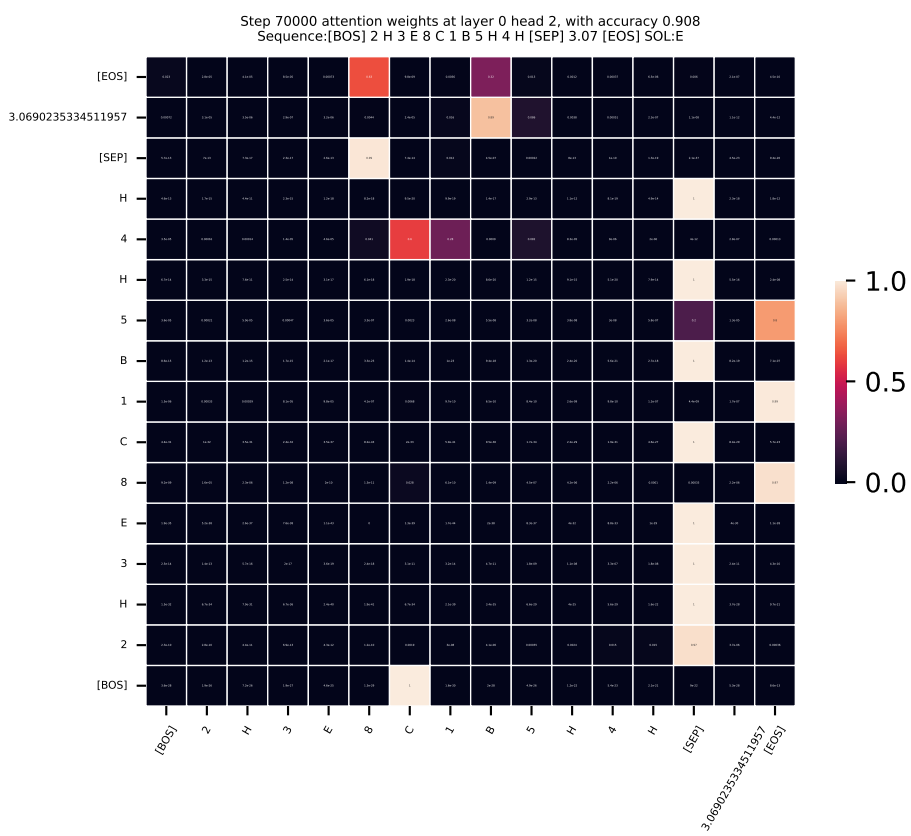

Step 69500 attention weights at layer 1 head 1, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

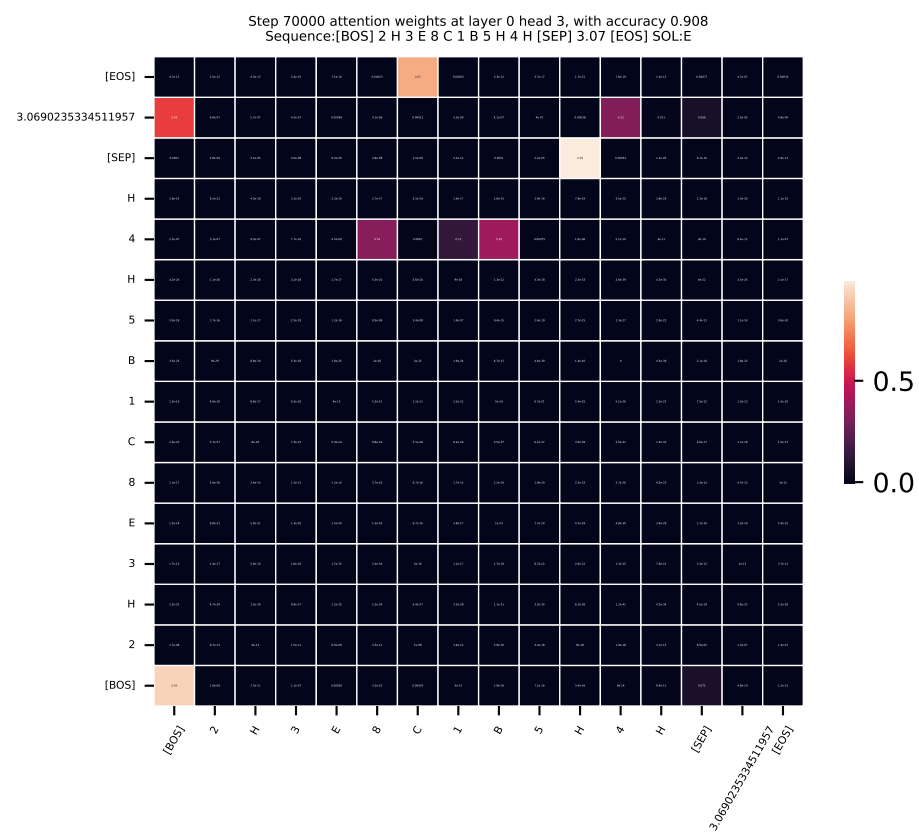

Step 69500 attention weights at layer 1 head 2, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

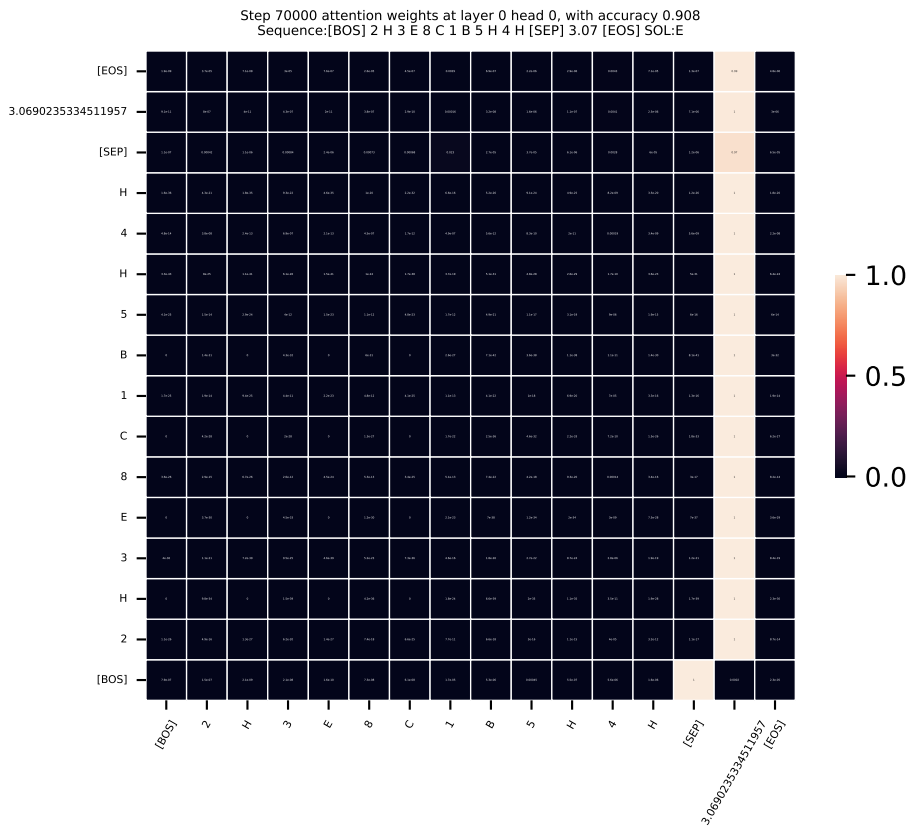

Step 69500 attention weights at layer 1 head 3, with accuracy 0.913
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

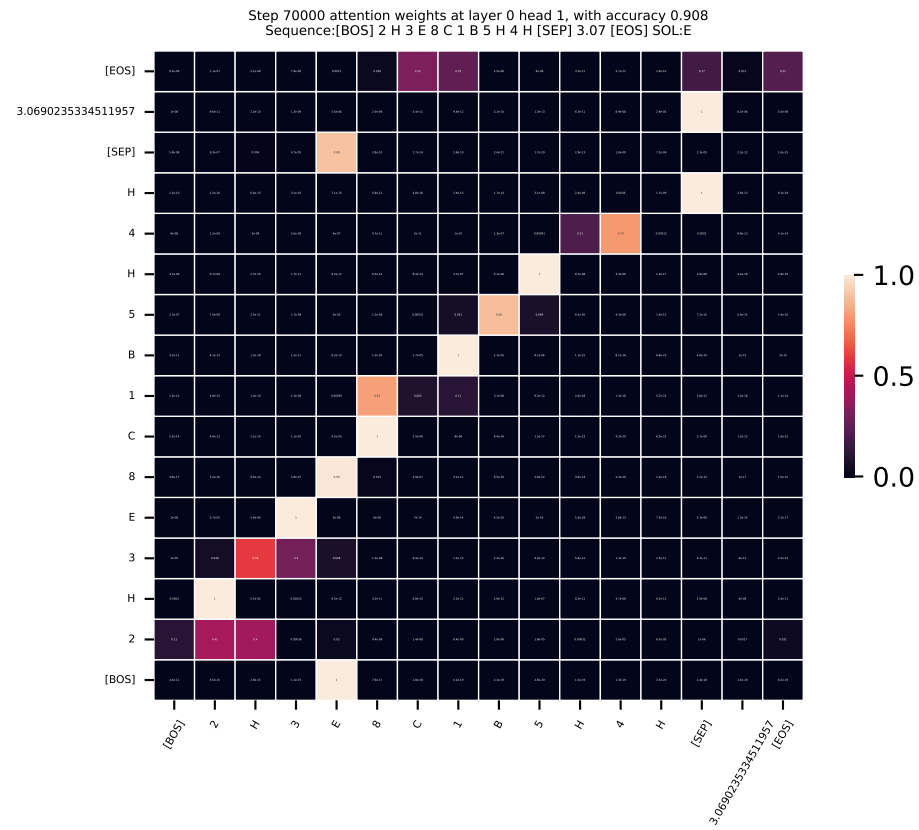

Step 70000 attention weights at layer 0 head 0, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

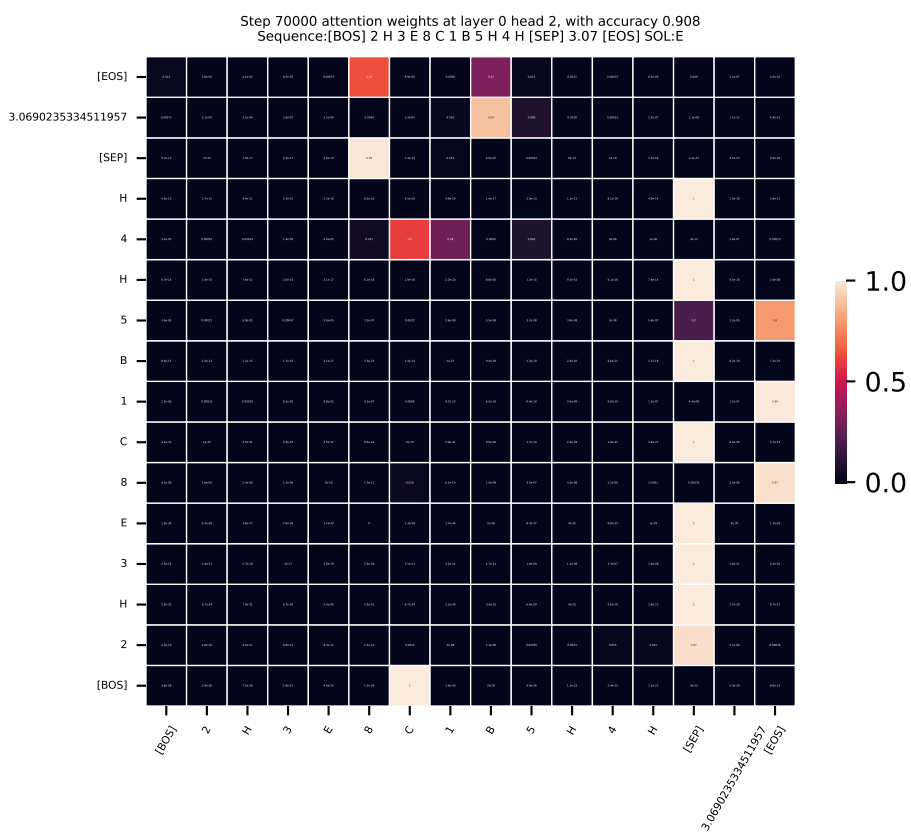

Step 70000 attention weights at layer 0 head 1, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

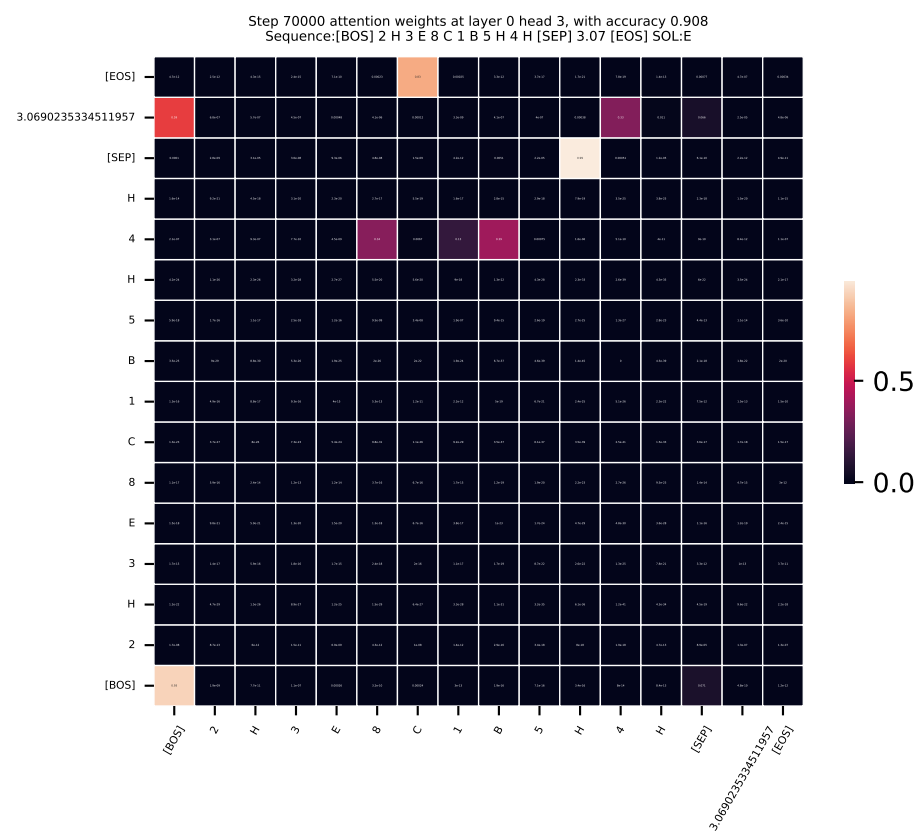

Step 70000 attention weights at layer 0 head 2, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

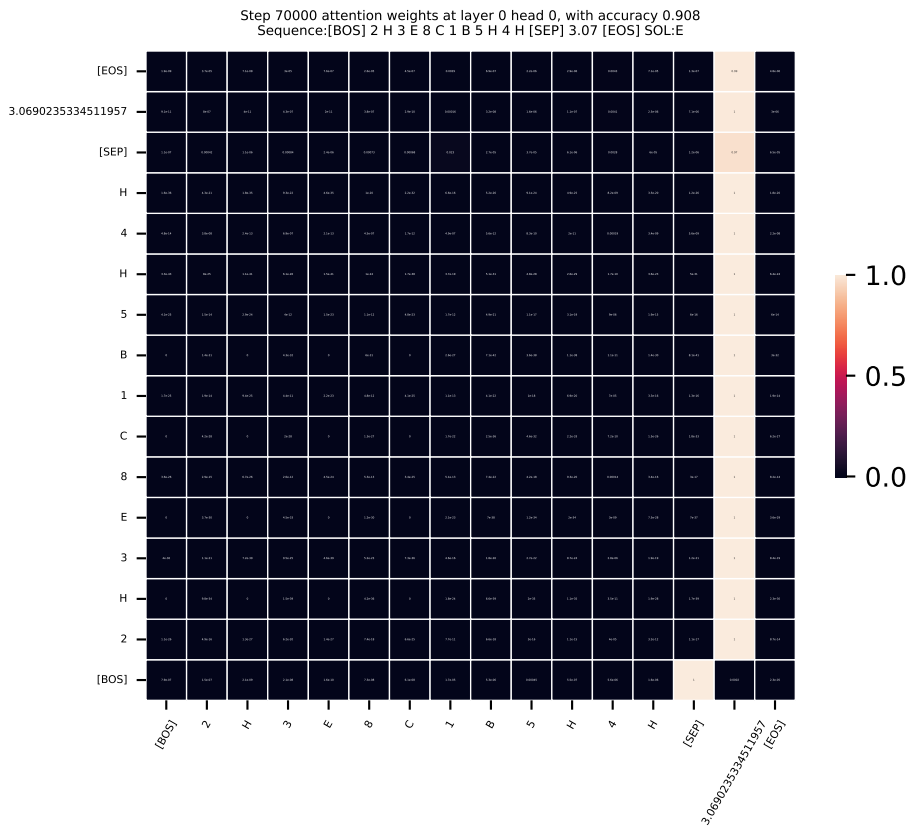

Step 70000 attention weights at layer 0 head 3, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.5
0.0

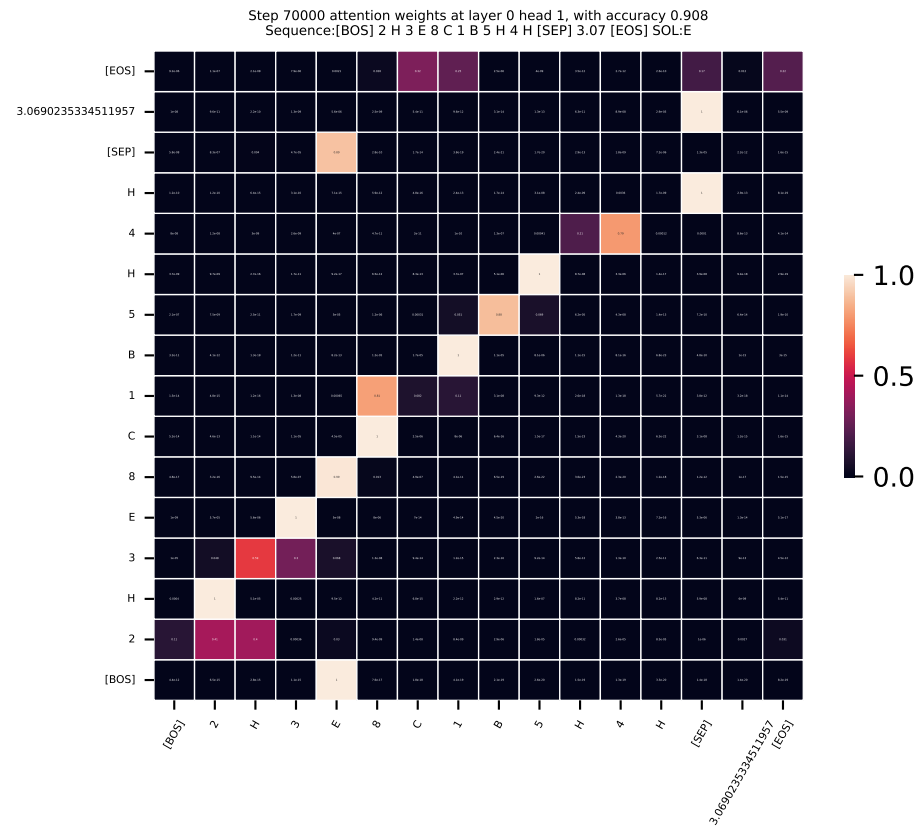

Step 70000 attention weights at layer 0 head 0, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

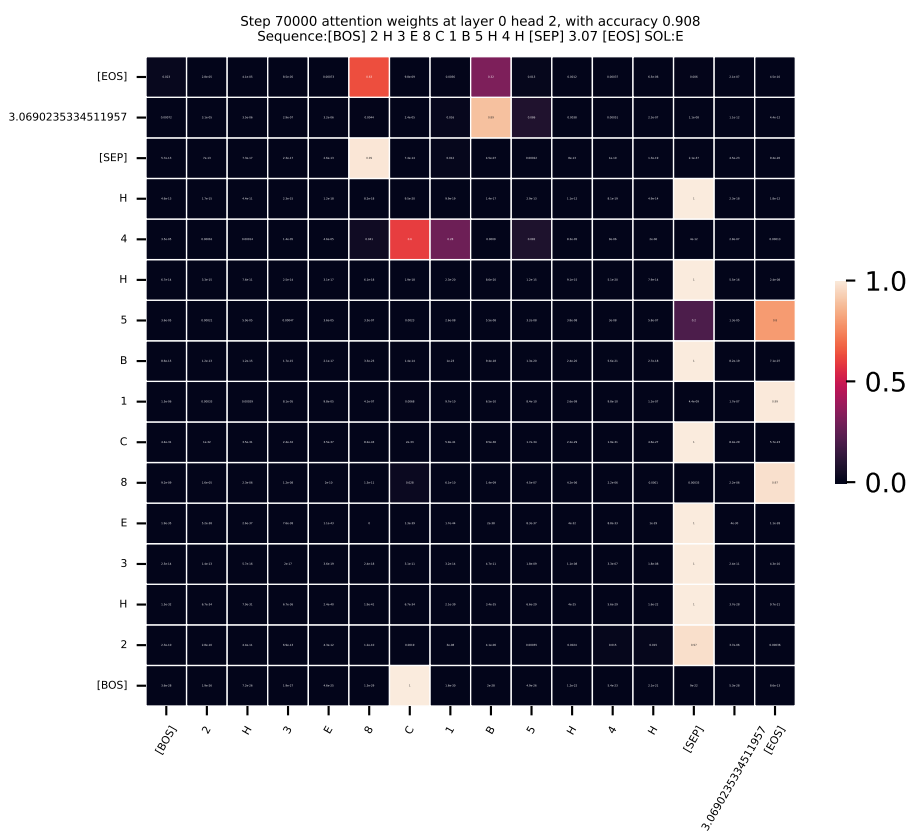

Step 70000 attention weights at layer 0 head 1, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

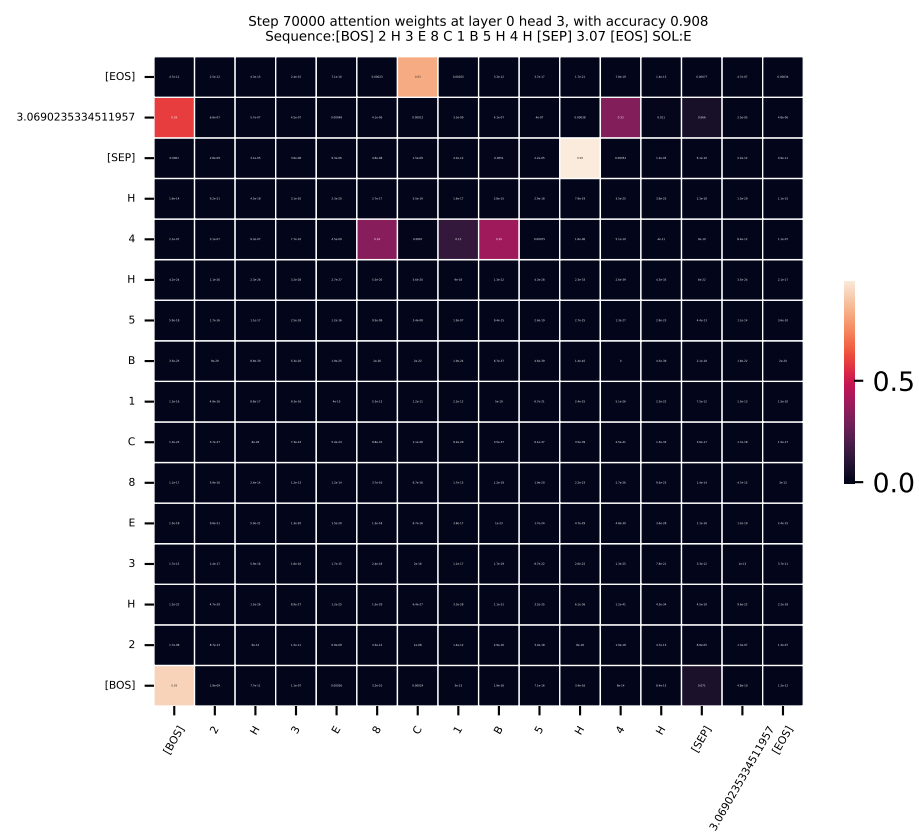

Step 70000 attention weights at layer 0 head 2, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

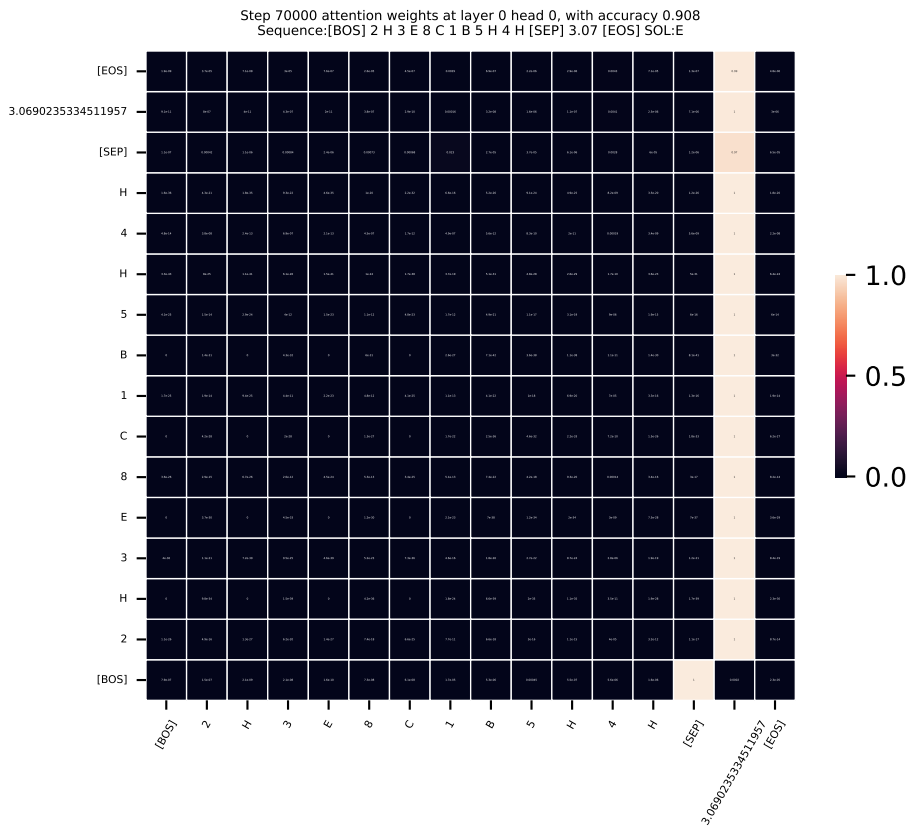

Step 70000 attention weights at layer 0 head 3, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.5
0.0

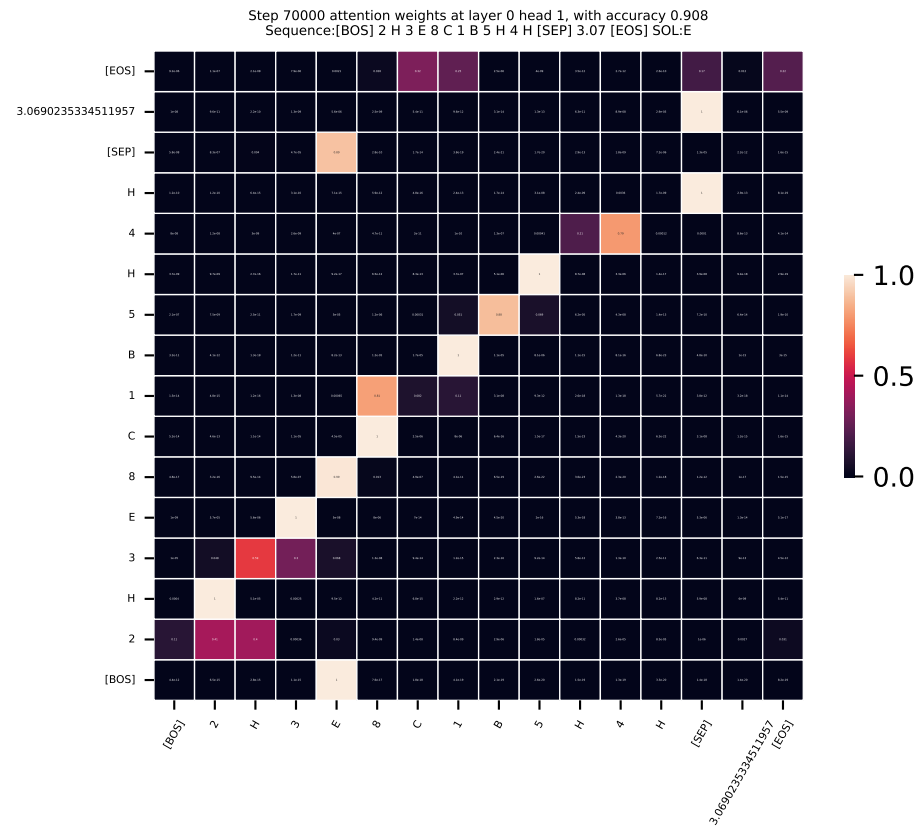

Step 70000 attention weights at layer 0 head 0, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

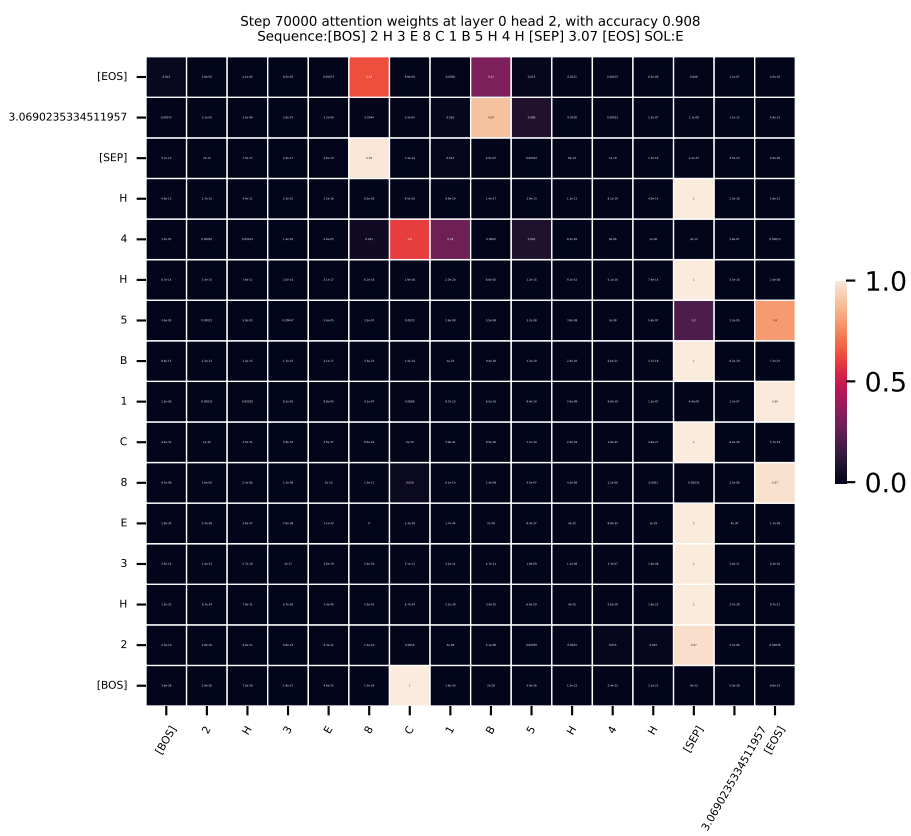

Step 70000 attention weights at layer 0 head 1, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

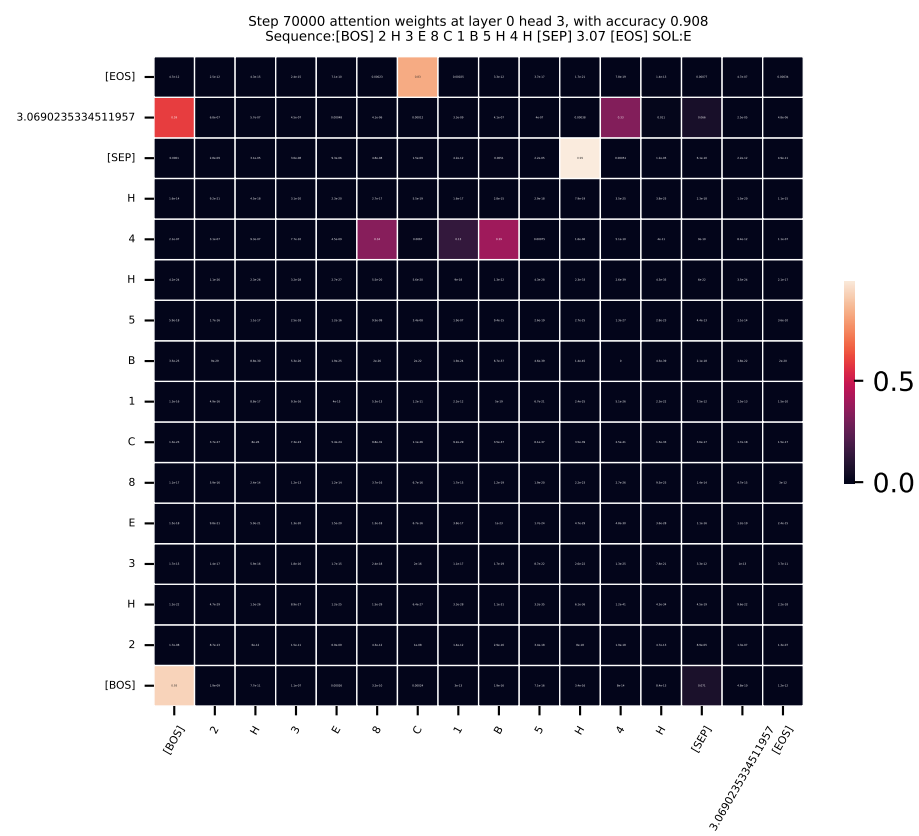

Step 70000 attention weights at layer 0 head 2, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

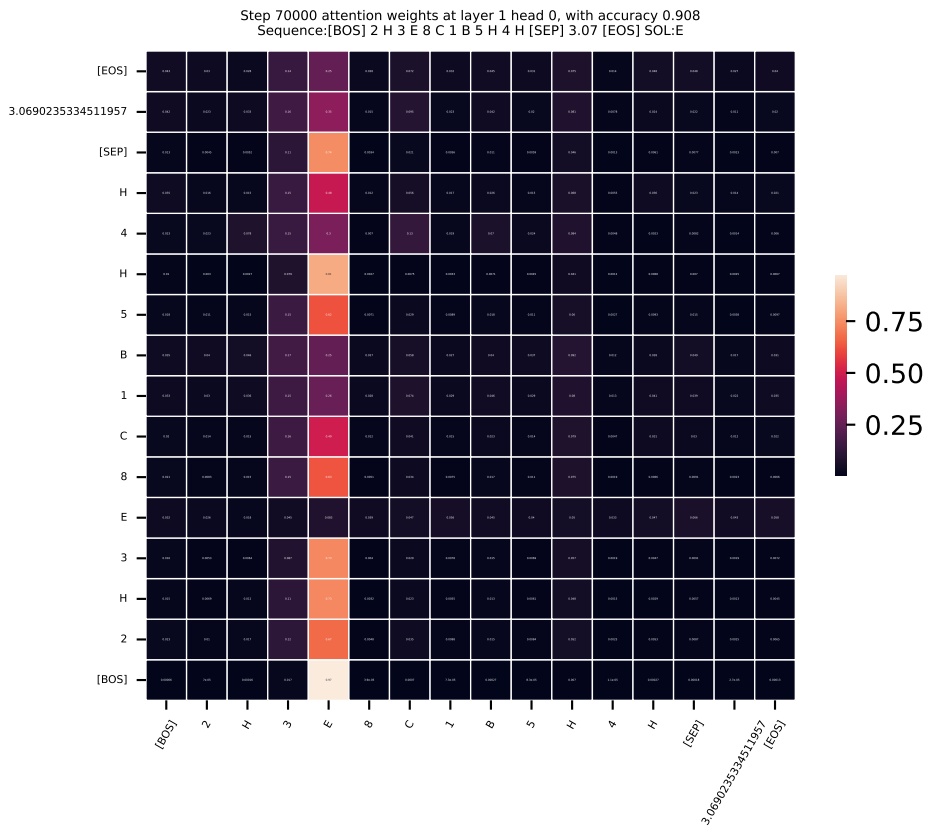

Step 70000 attention weights at layer 0 head 3, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.5
0.0

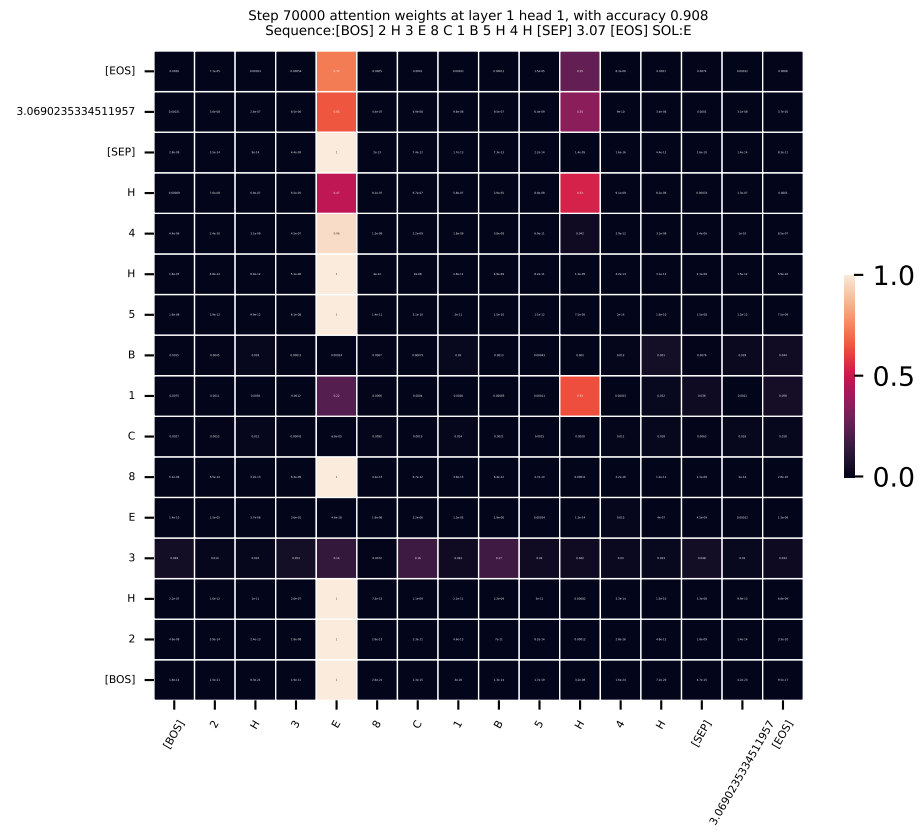

Step 70000 attention weights at layer 0 head 0, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

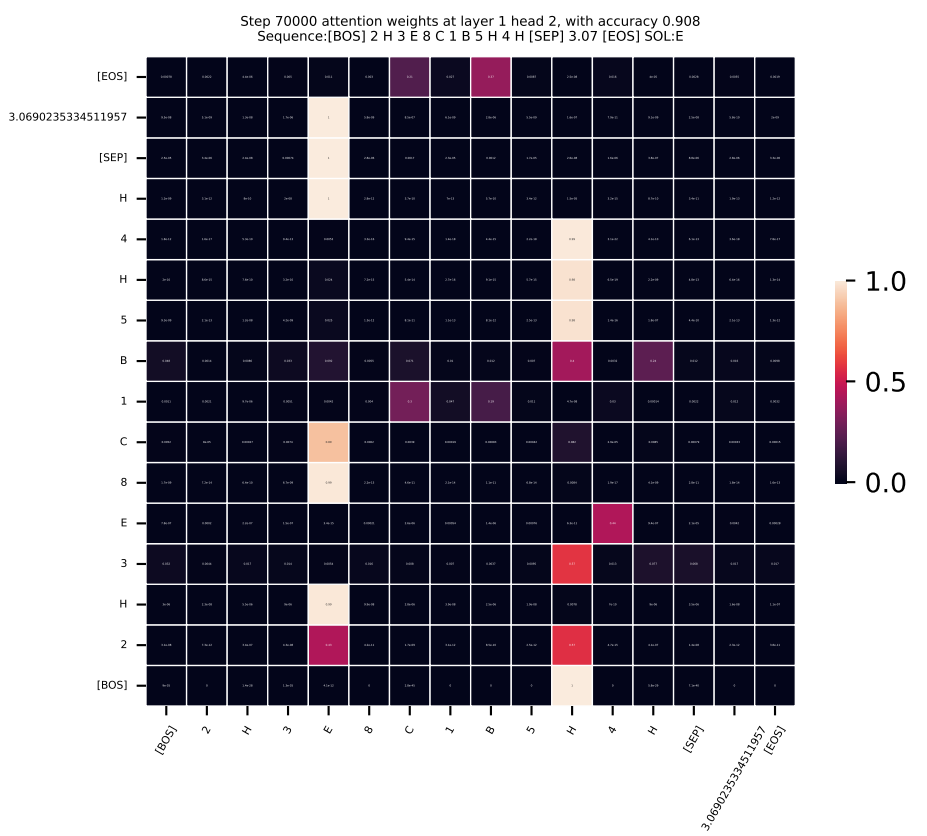

Step 70000 attention weights at layer 0 head 1, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

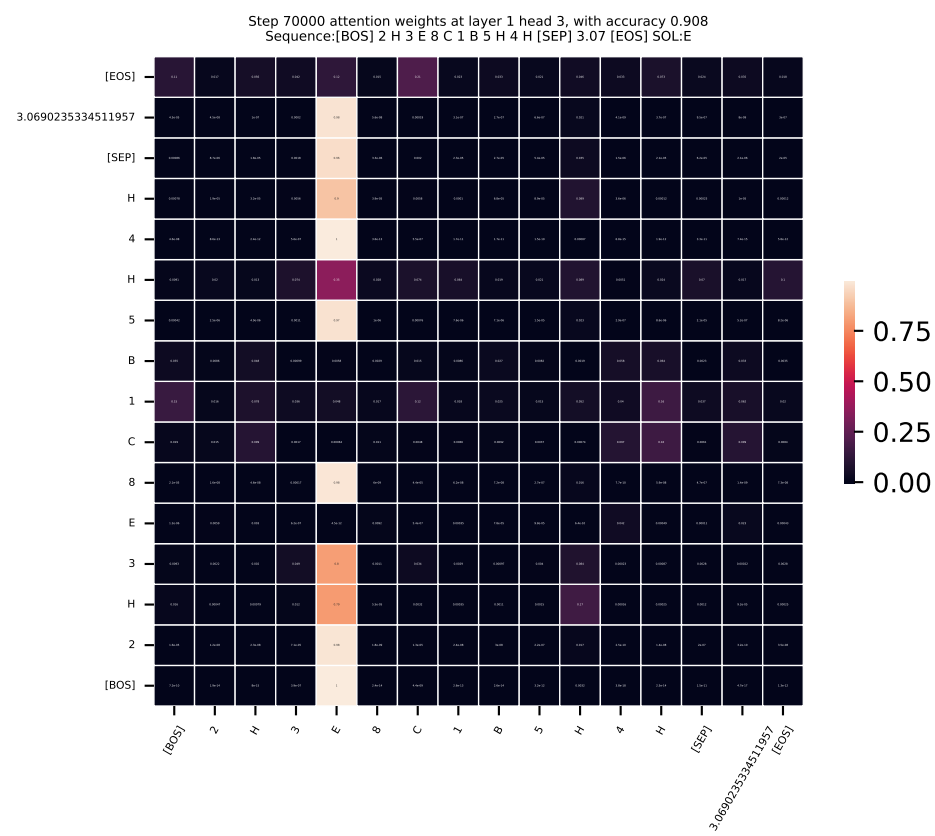

Step 70000 attention weights at layer 0 head 2, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

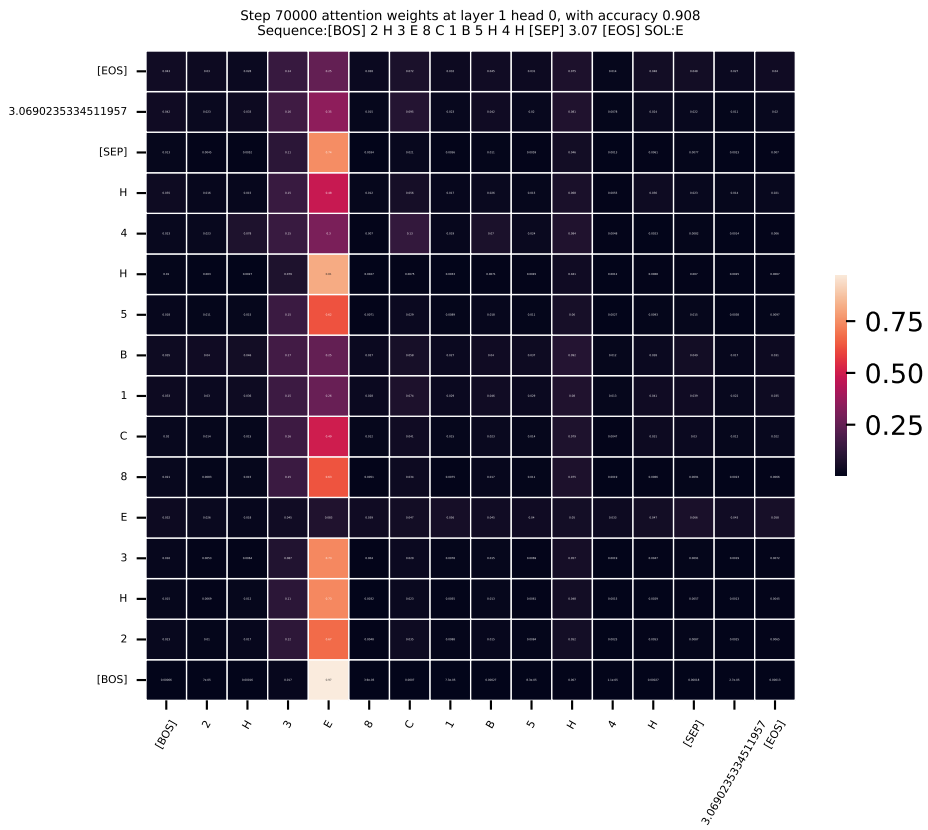

Step 70000 attention weights at layer 0 head 3, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.5
0.0

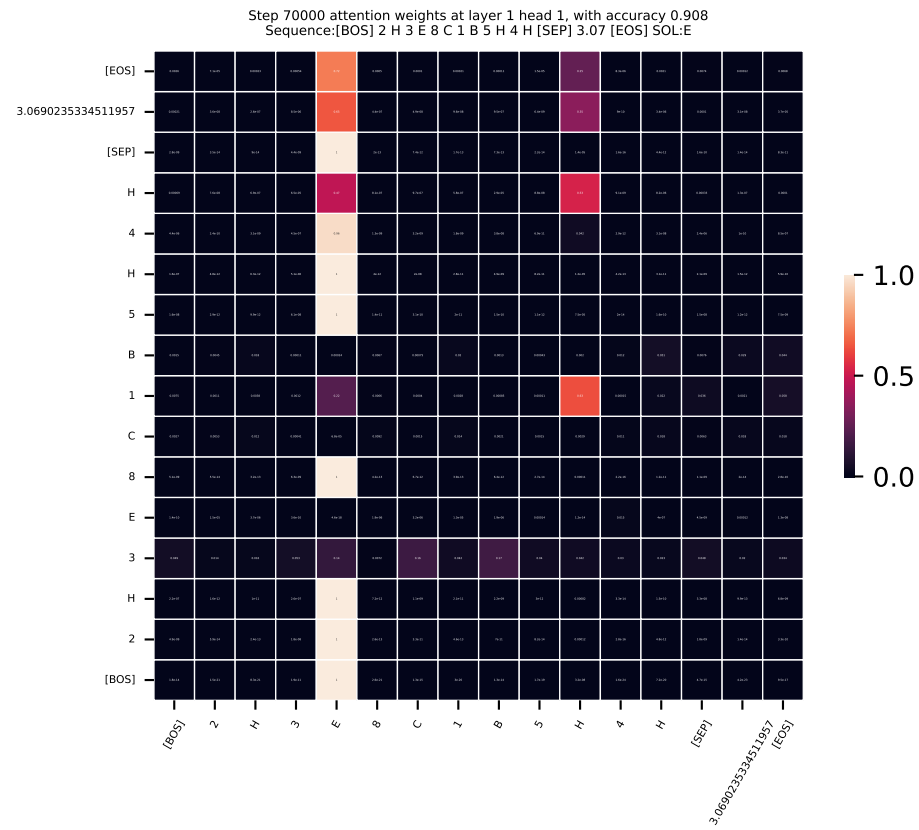

Step 70000 attention weights at layer 1 head 0, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

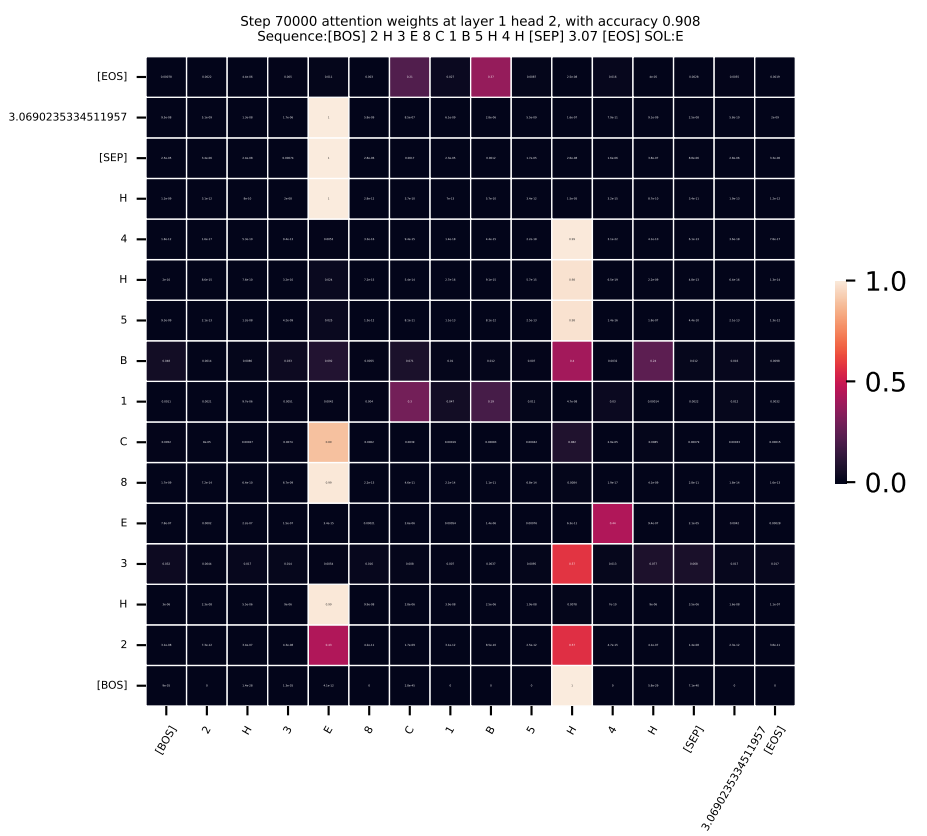

Step 70000 attention weights at layer 1 head 1, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

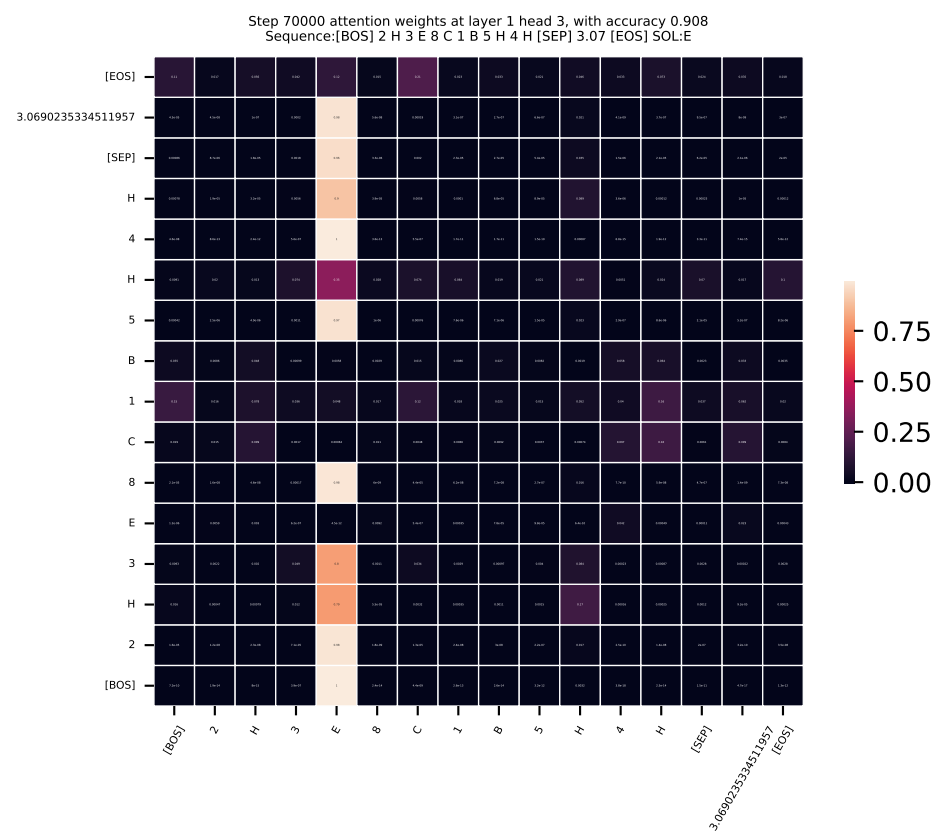

Step 70000 attention weights at layer 1 head 2, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

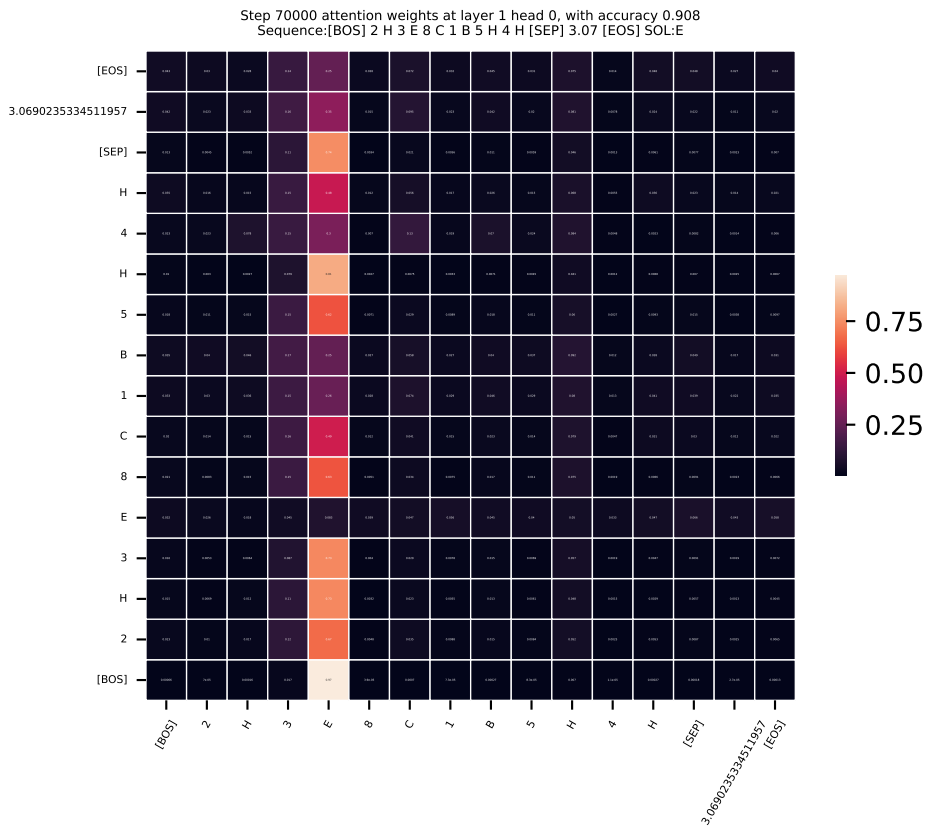

Step 70000 attention weights at layer 1 head 3, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

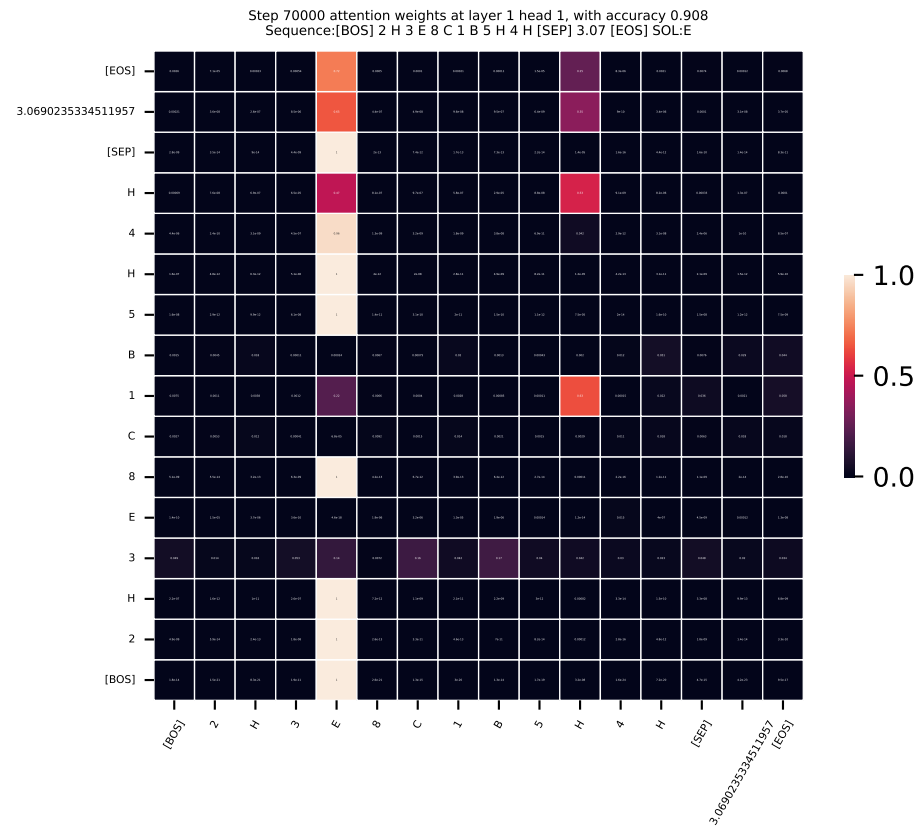

Step 70000 attention weights at layer 1 head 0, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

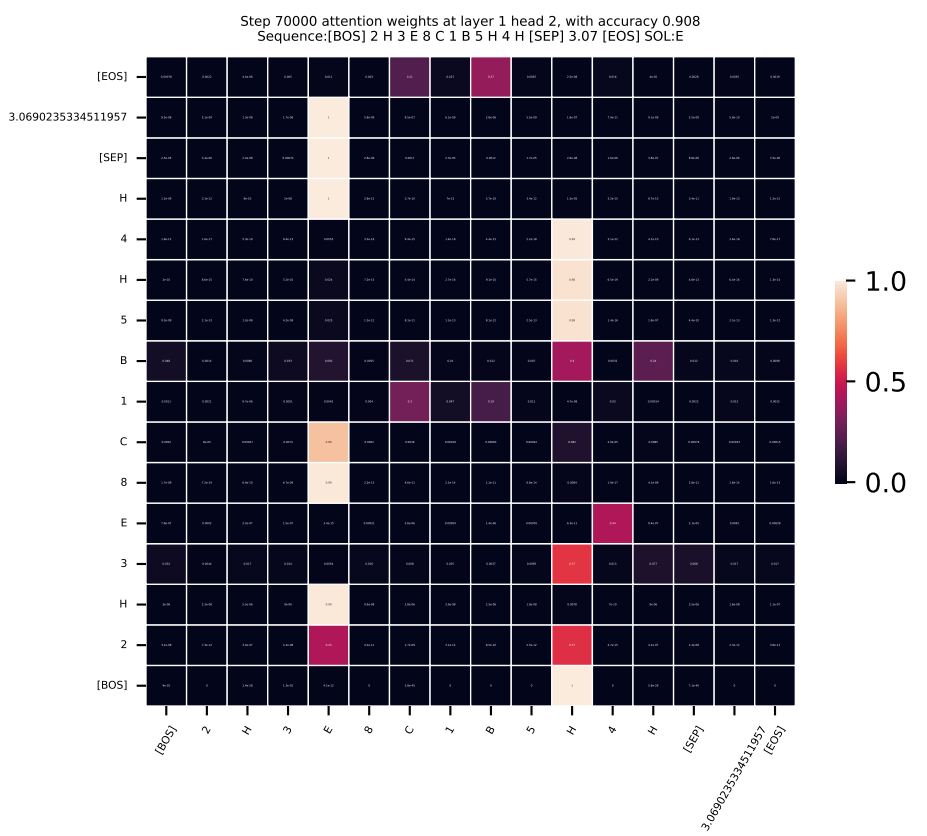

Step 70000 attention weights at layer 1 head 1, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

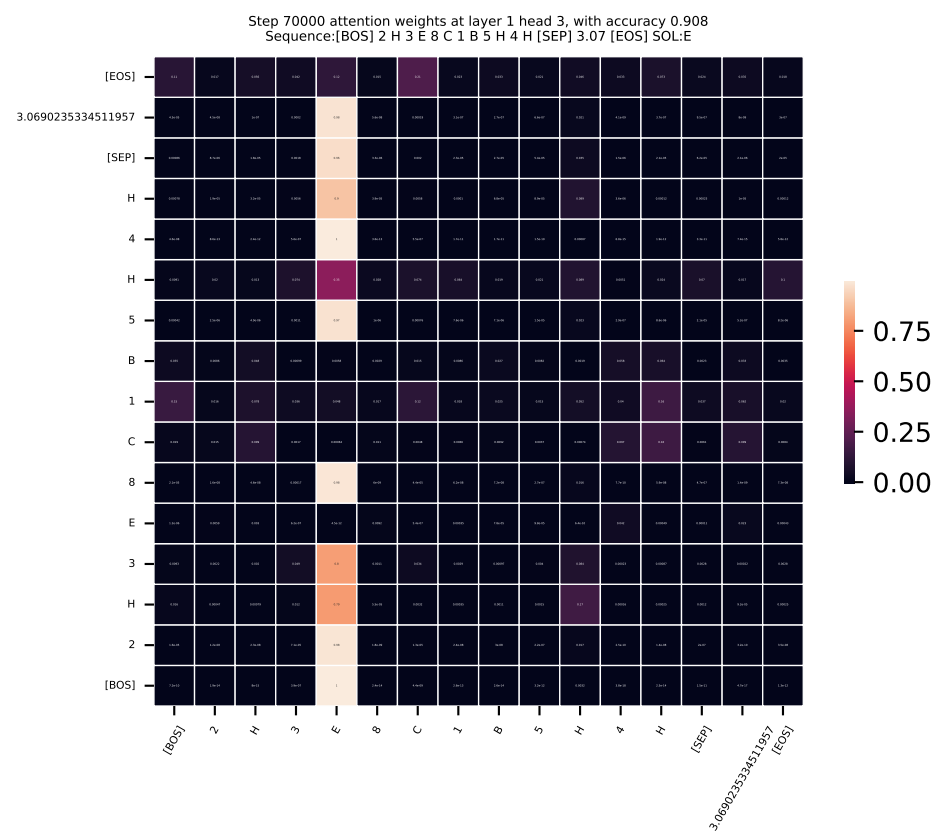

Step 70000 attention weights at layer 1 head 2, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

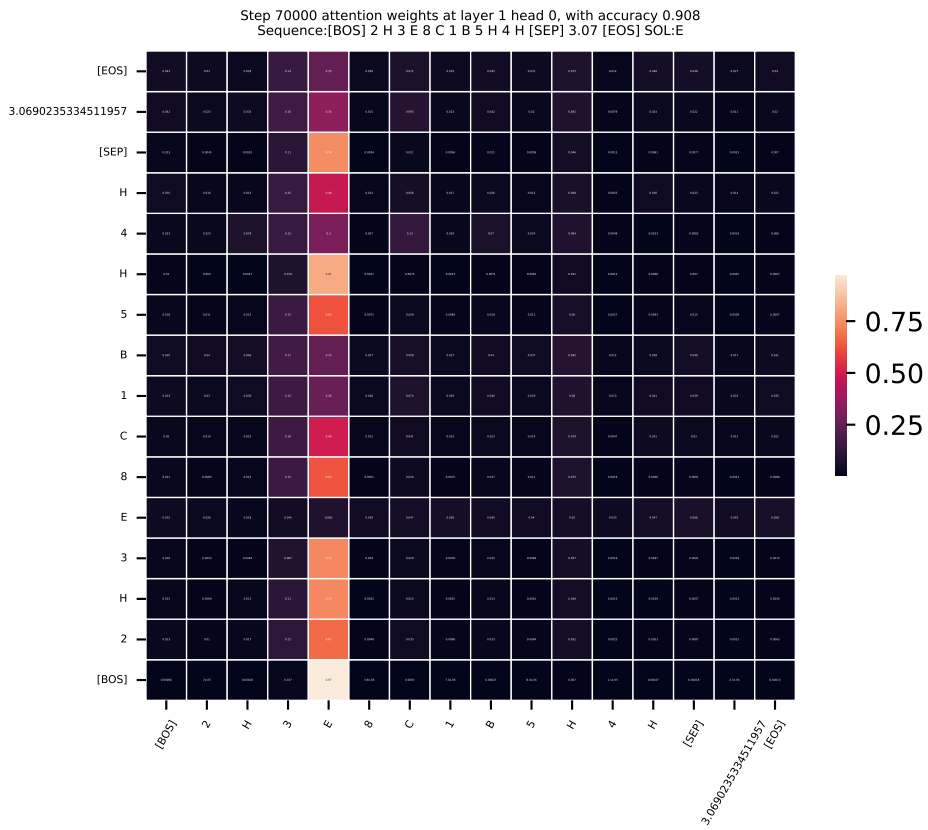

Step 70000 attention weights at layer 1 head 3, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

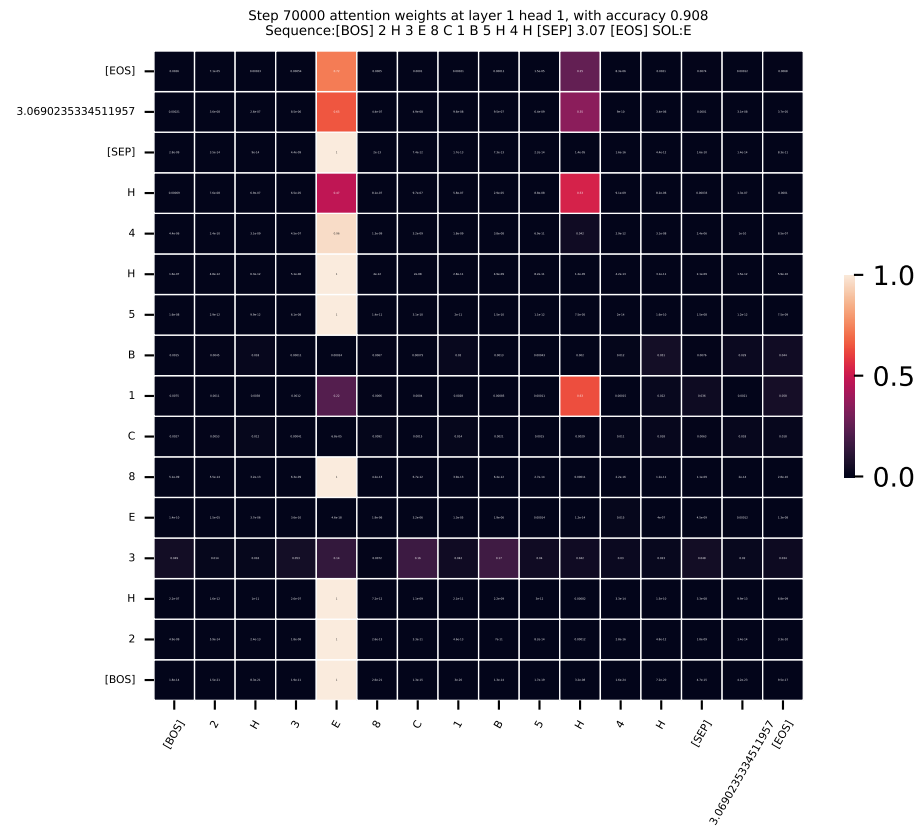

Step 70000 attention weights at layer 1 head 0, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

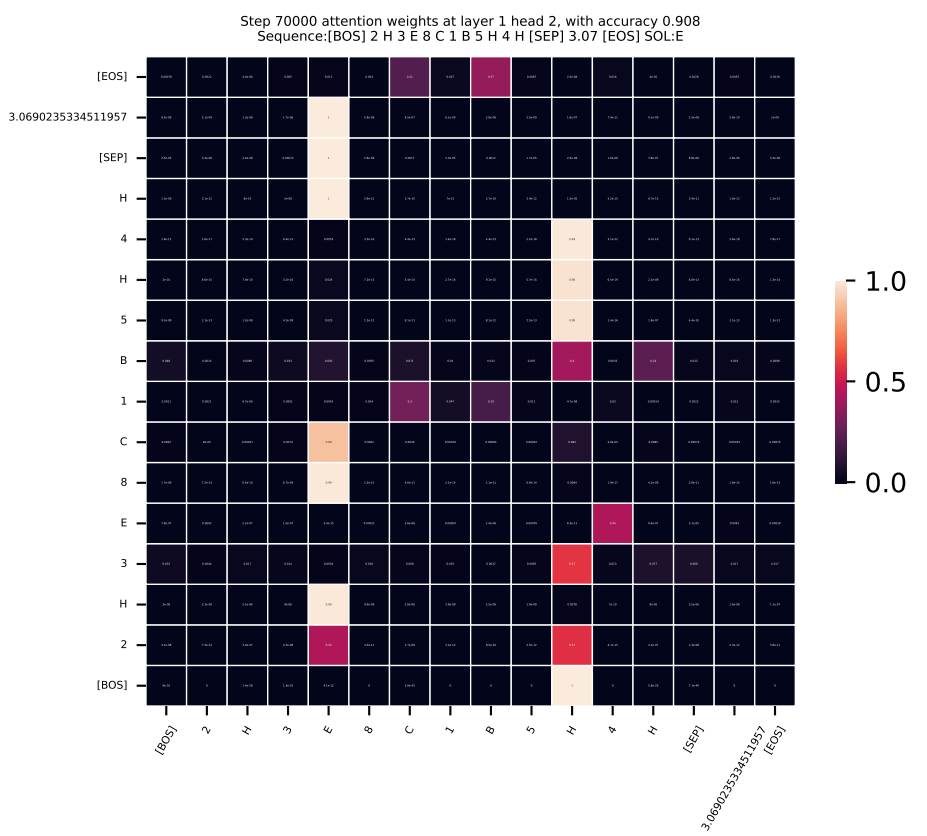

Step 70000 attention weights at layer 1 head 1, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

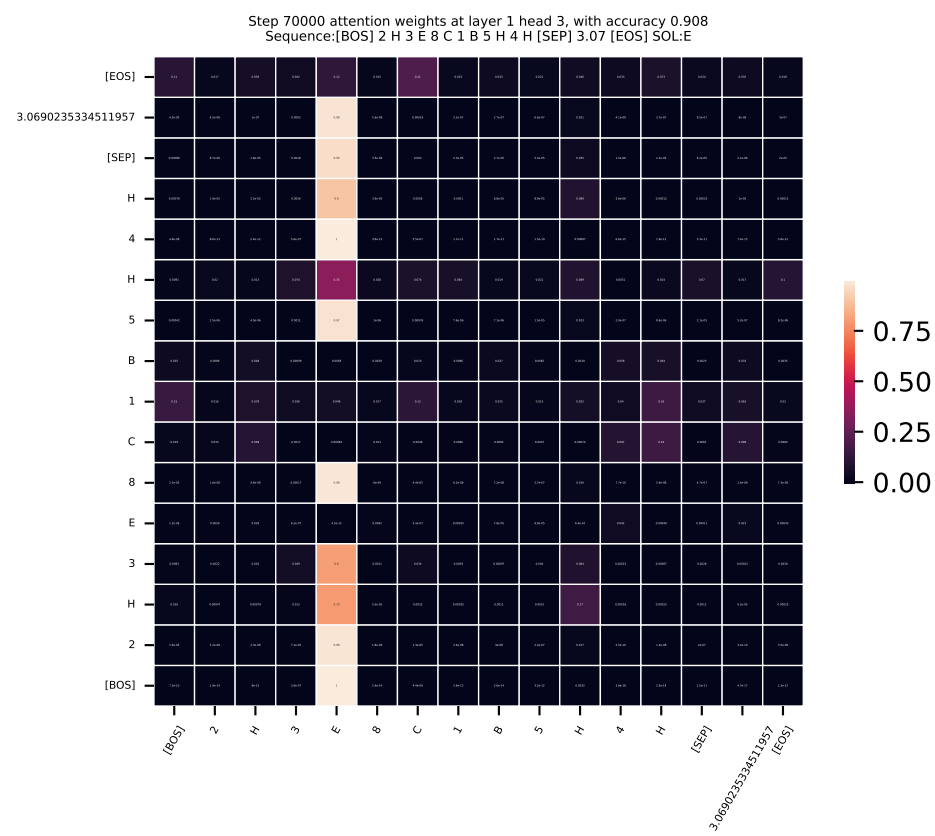

Step 70000 attention weights at layer 1 head 2, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

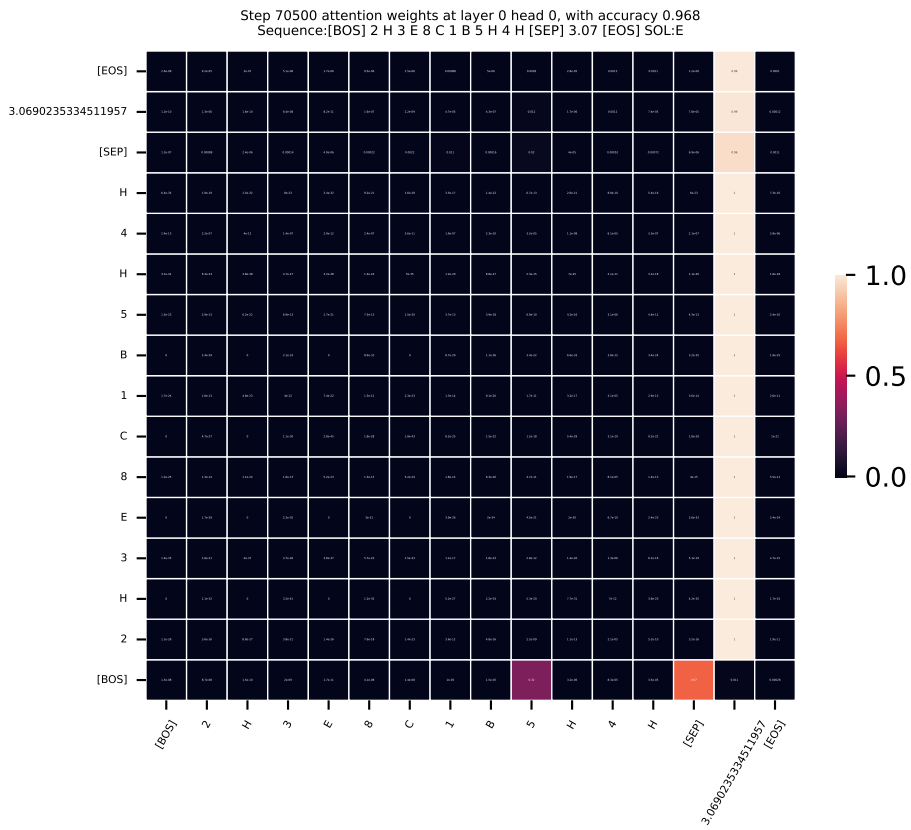

Step 70000 attention weights at layer 1 head 3, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

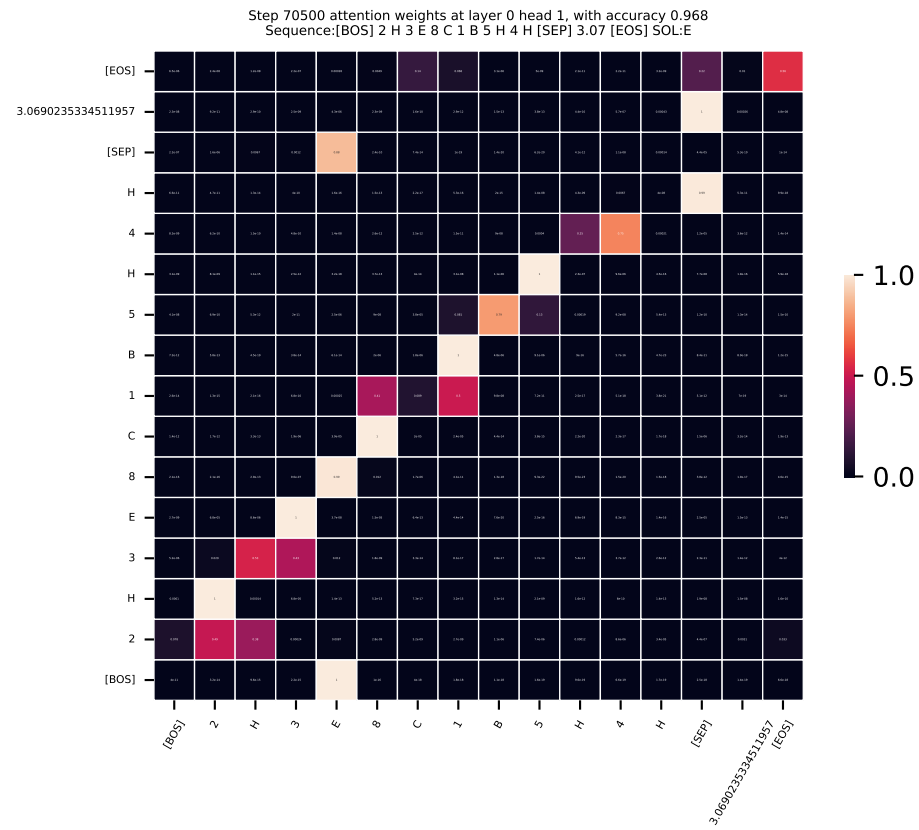

Step 70000 attention weights at layer 1 head 0, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

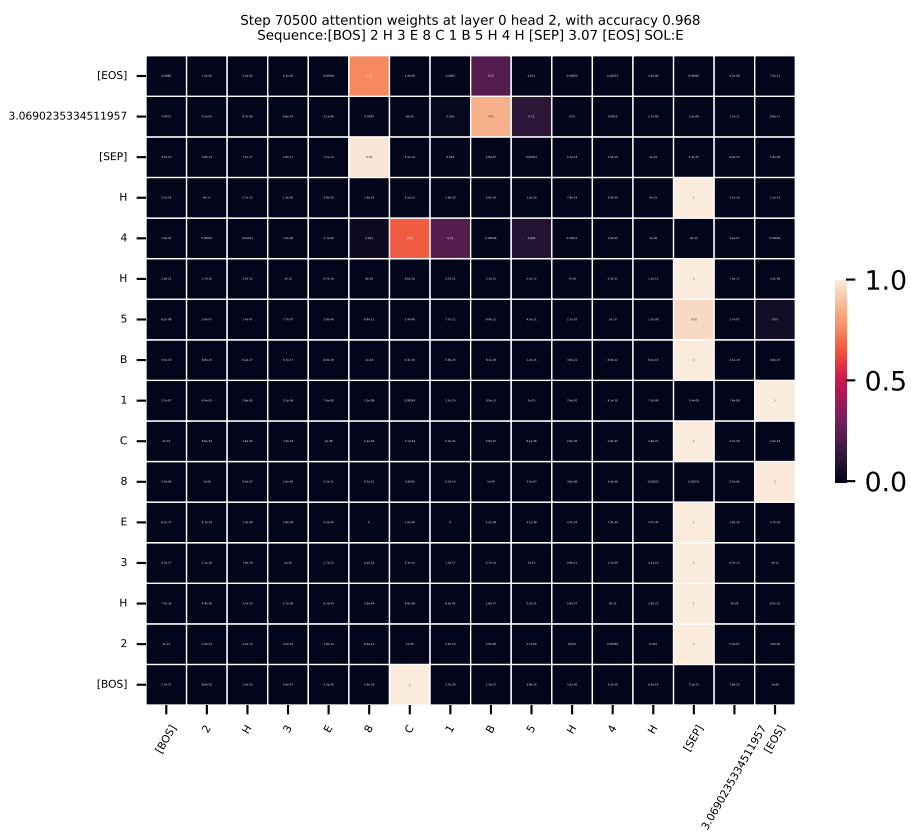

Step 70000 attention weights at layer 1 head 1, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

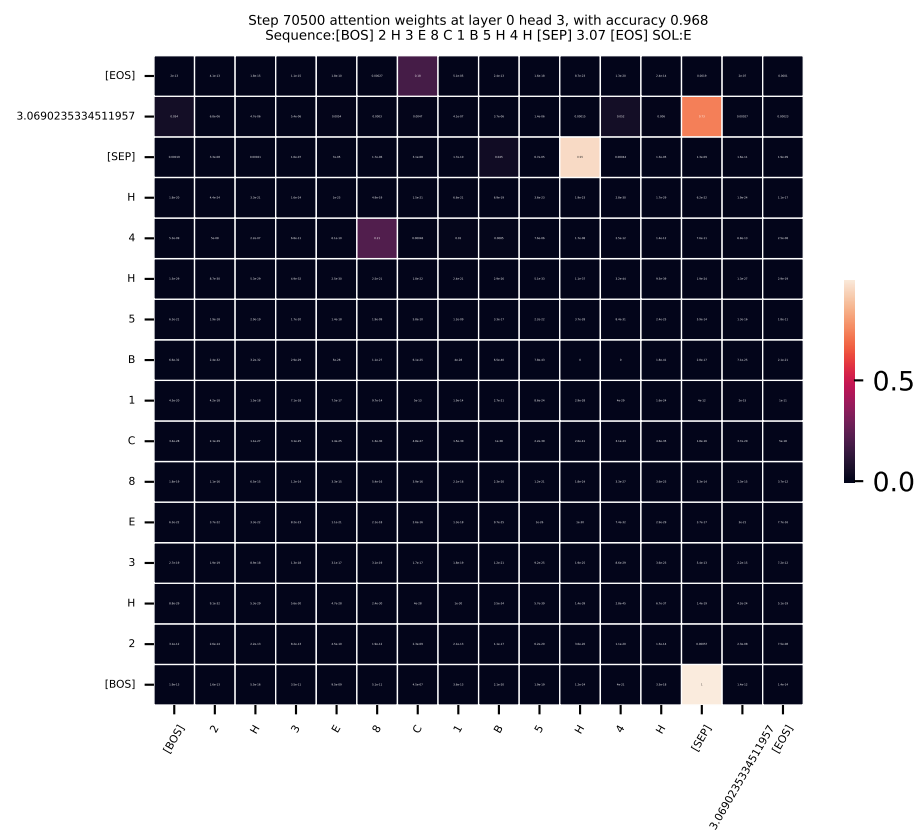

Step 70000 attention weights at layer 1 head 2, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

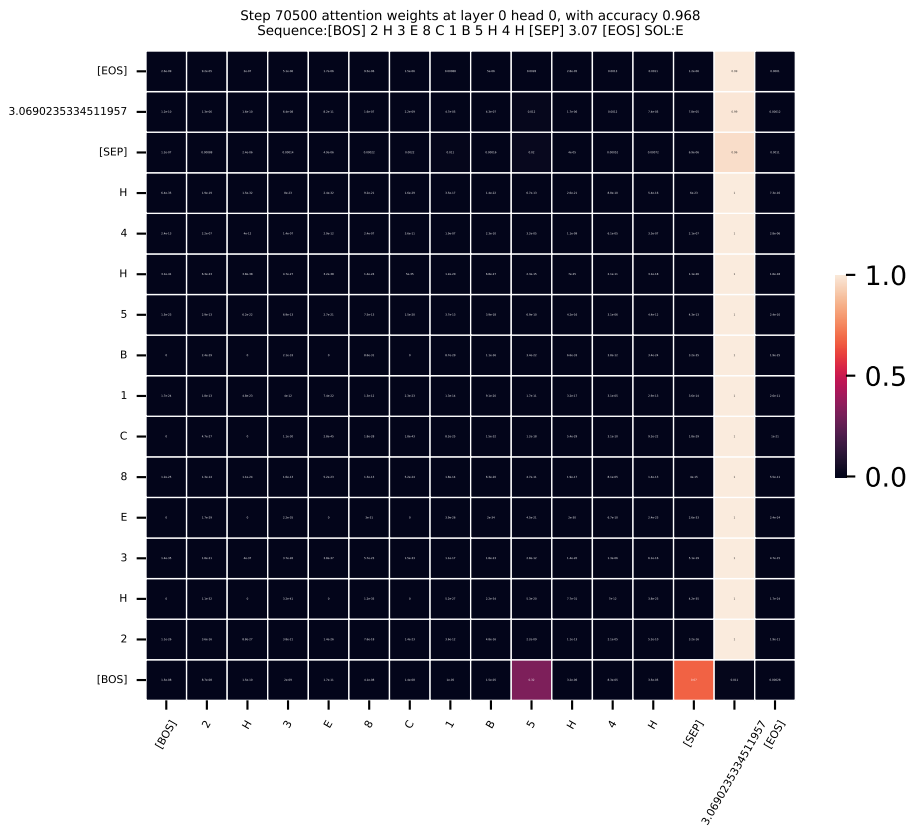

Step 70000 attention weights at layer 1 head 3, with accuracy 0.908
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

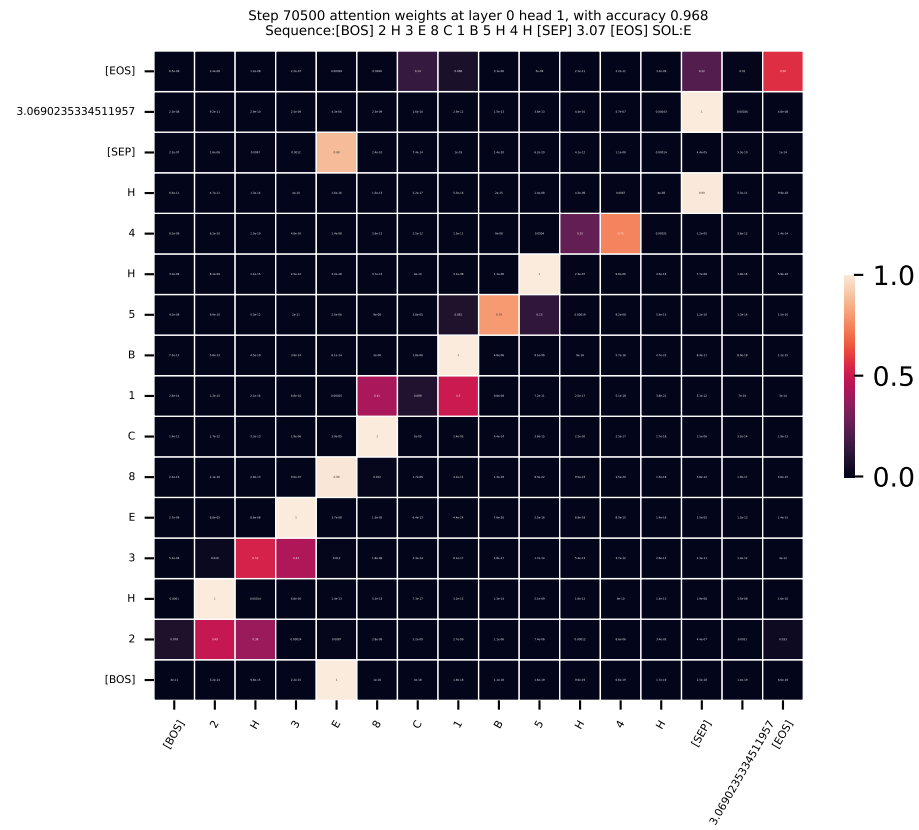

Step 70500 attention weights at layer 0 head 0, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

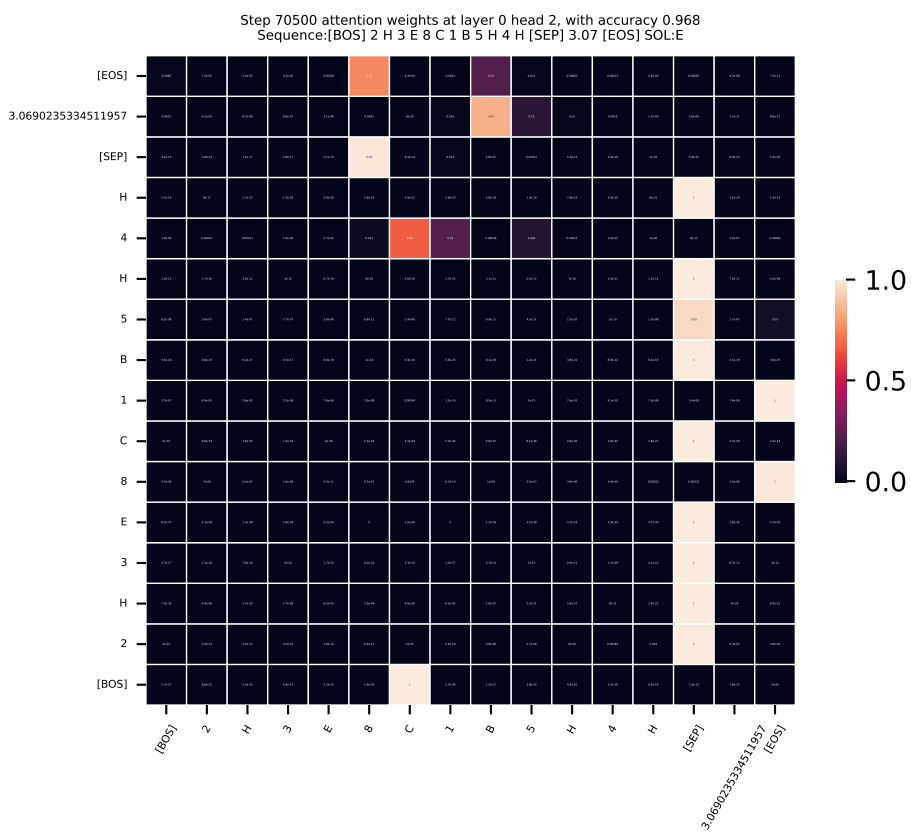

Step 70500 attention weights at layer 0 head 1, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

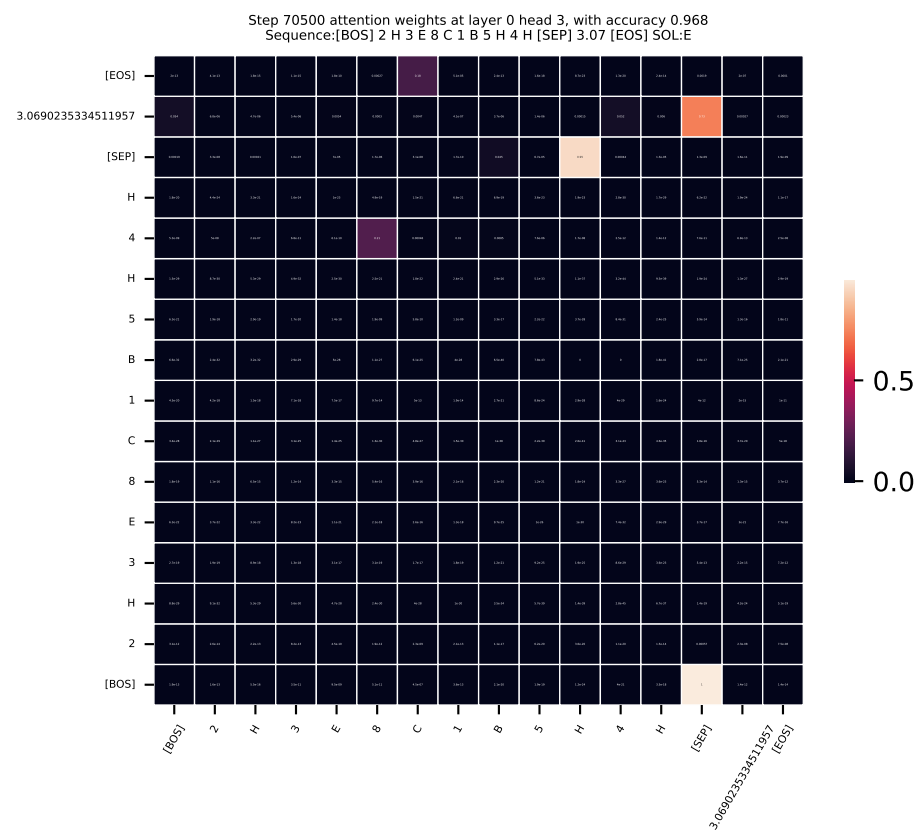

Step 70500 attention weights at layer 0 head 2, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

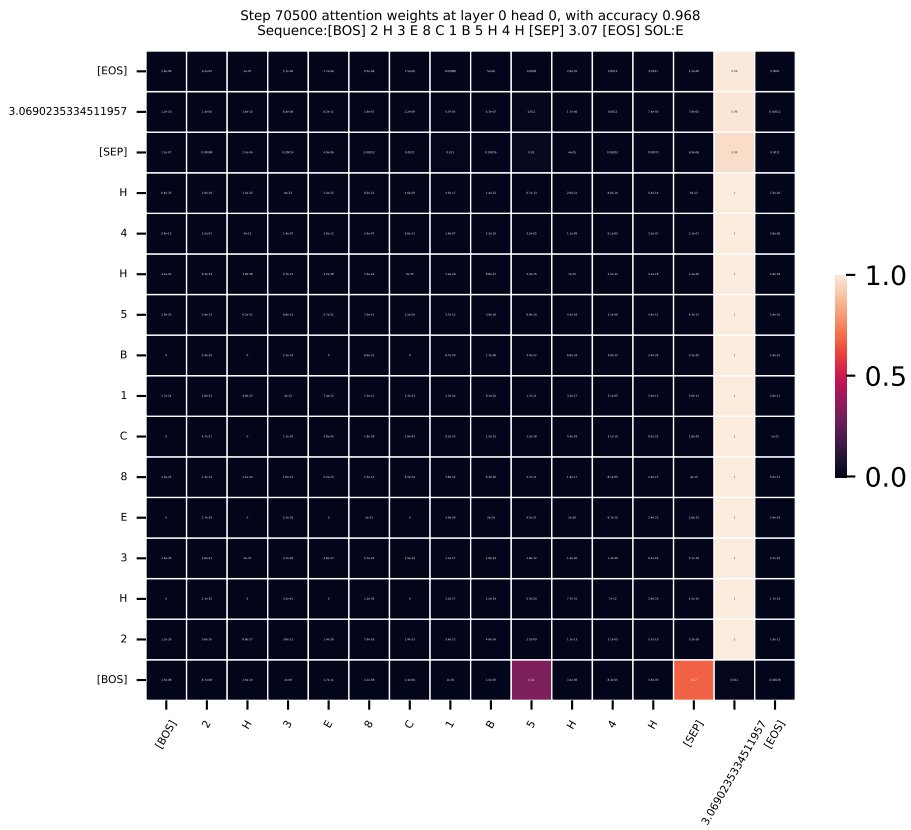

Step 70500 attention weights at layer 0 head 3, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.5
0.0

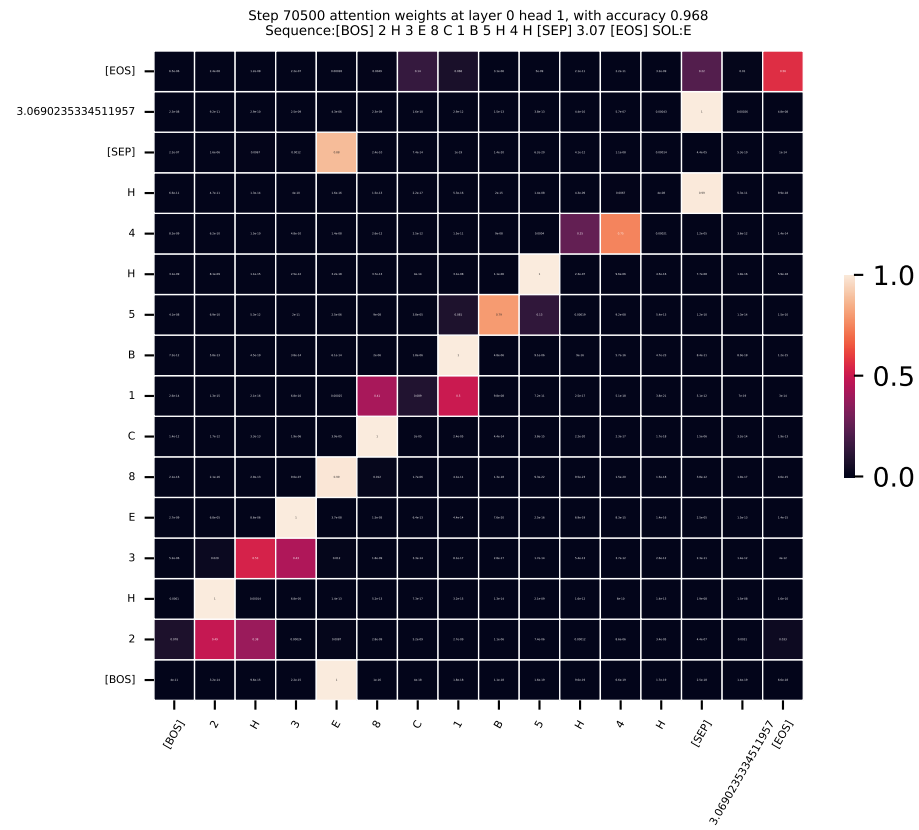

Step 70500 attention weights at layer 0 head 0, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

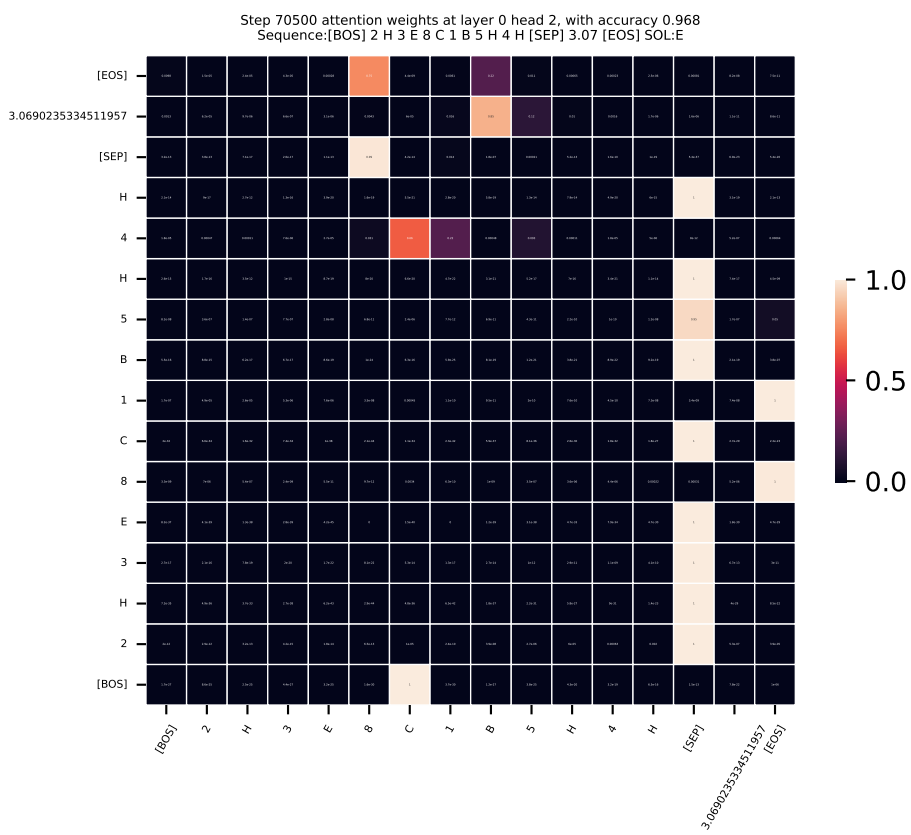

Step 70500 attention weights at layer 0 head 1, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

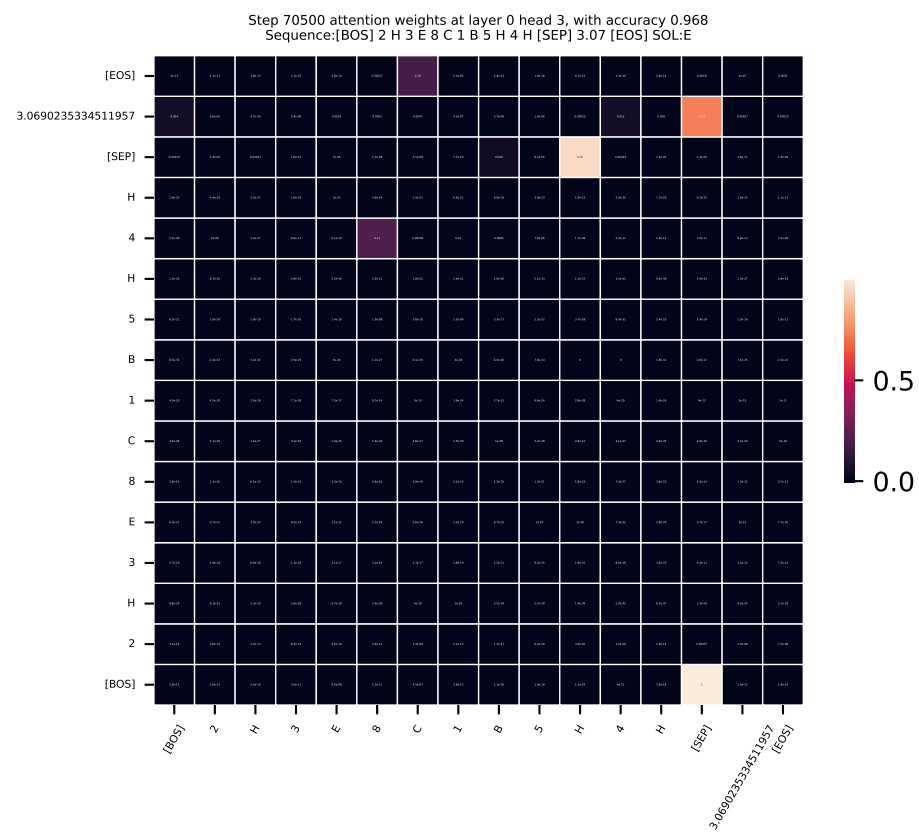

Step 70500 attention weights at layer 0 head 2, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

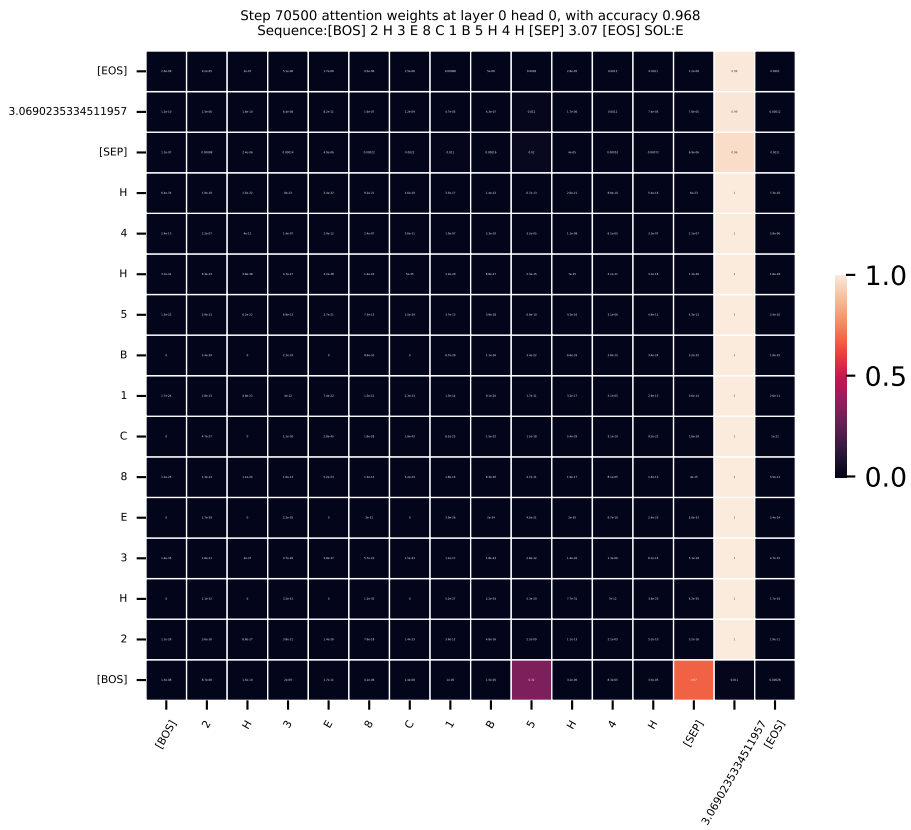

Step 70500 attention weights at layer 0 head 3, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.5
0.0

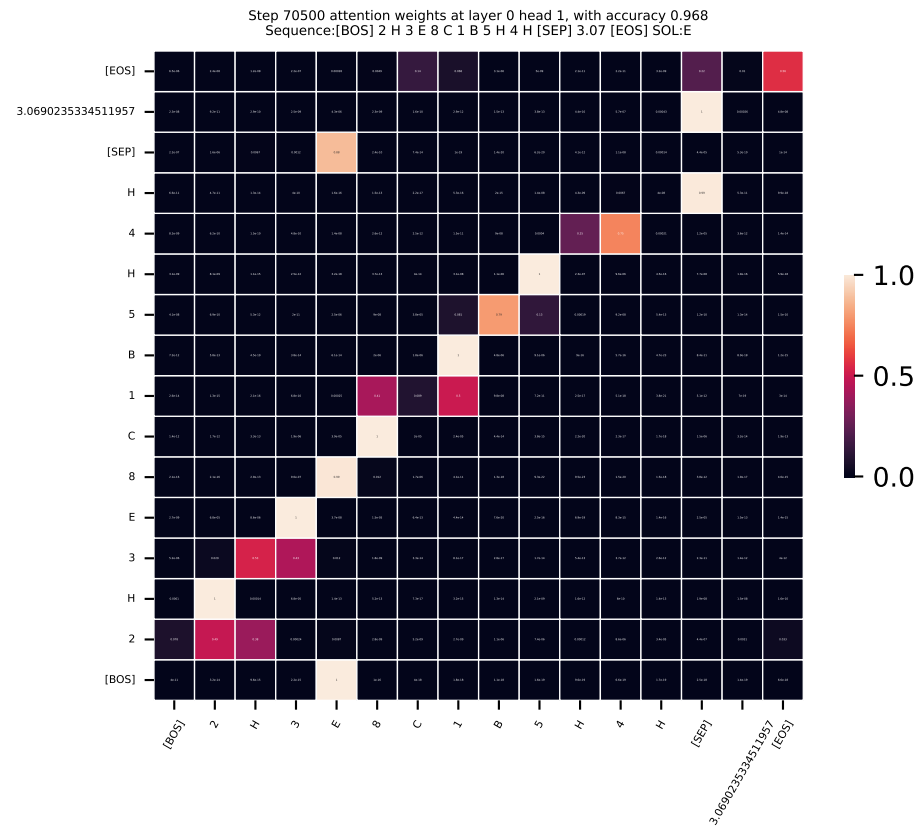

Step 70500 attention weights at layer 0 head 0, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

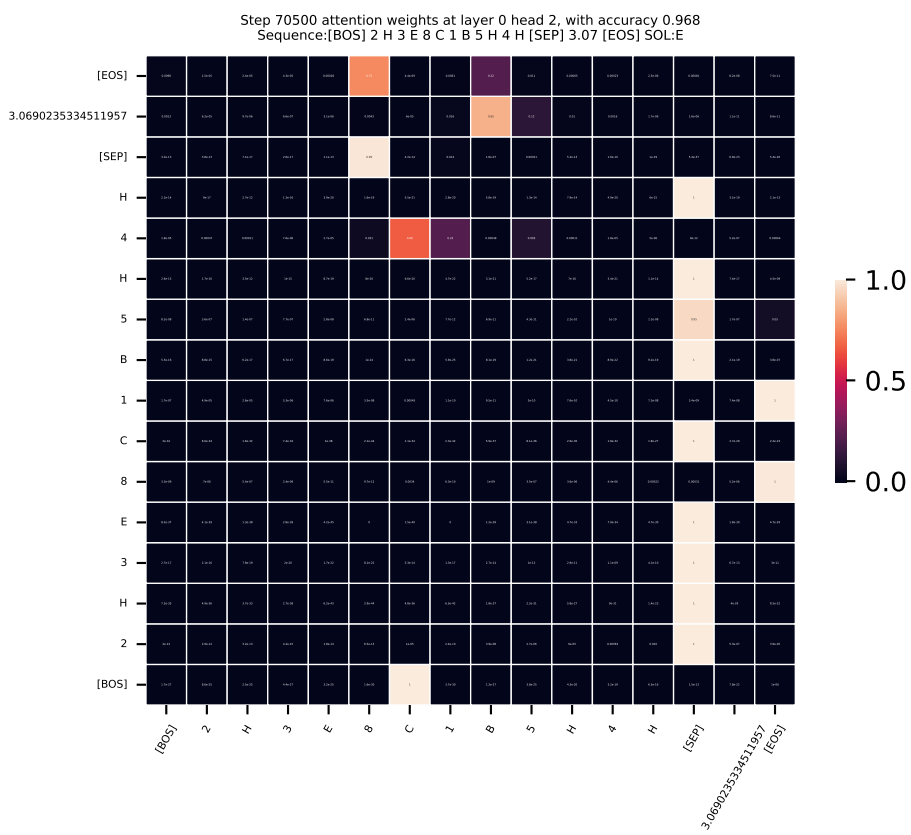

Step 70500 attention weights at layer 0 head 1, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

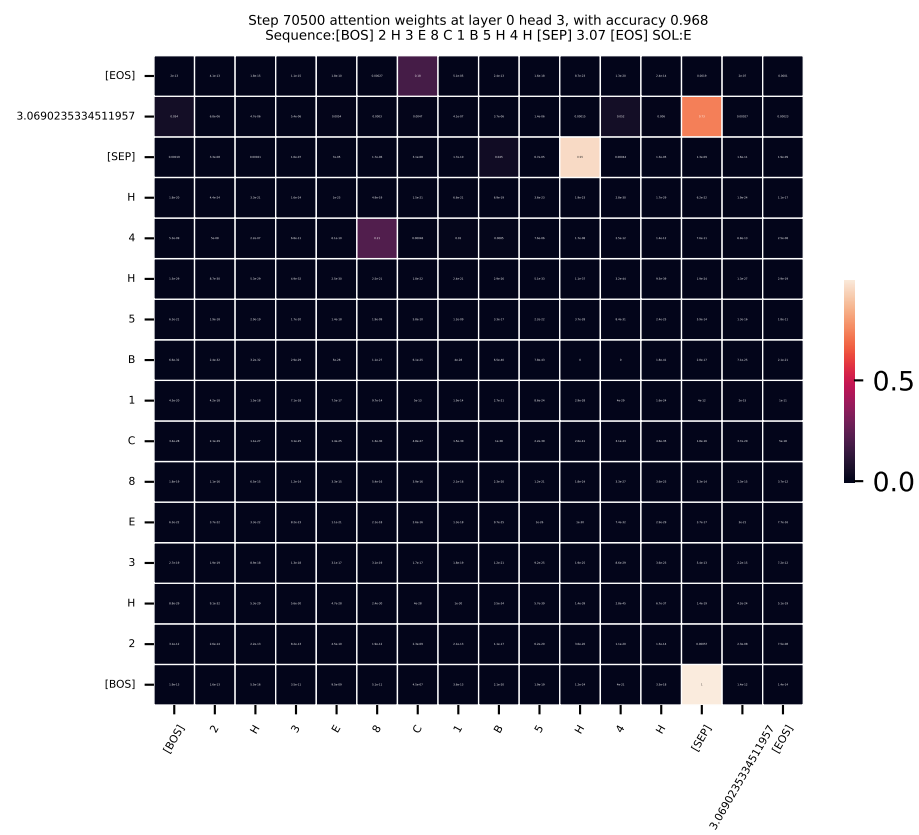

Step 70500 attention weights at layer 0 head 2, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

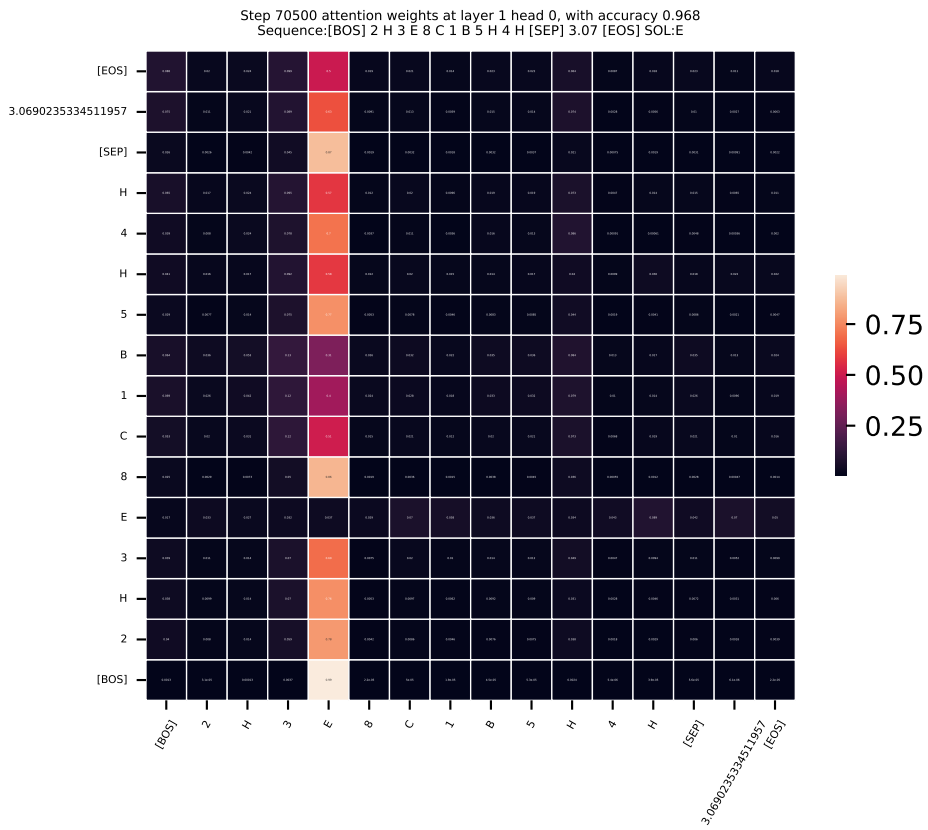

Step 70500 attention weights at layer 0 head 3, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.5
0.0

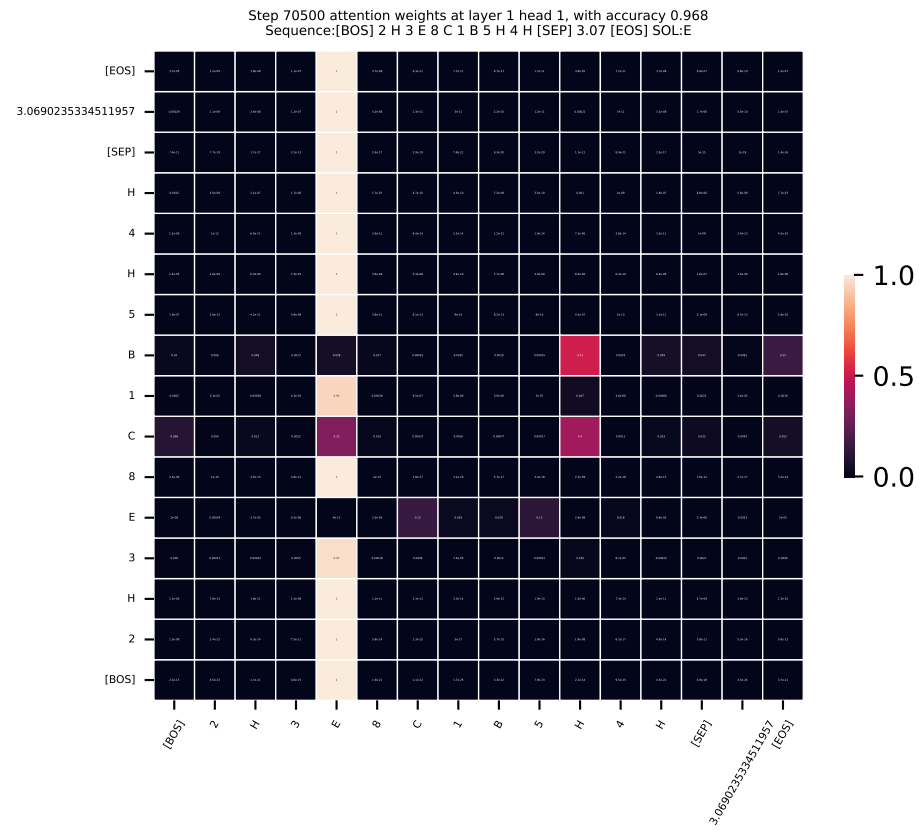

Step 70500 attention weights at layer 0 head 0, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

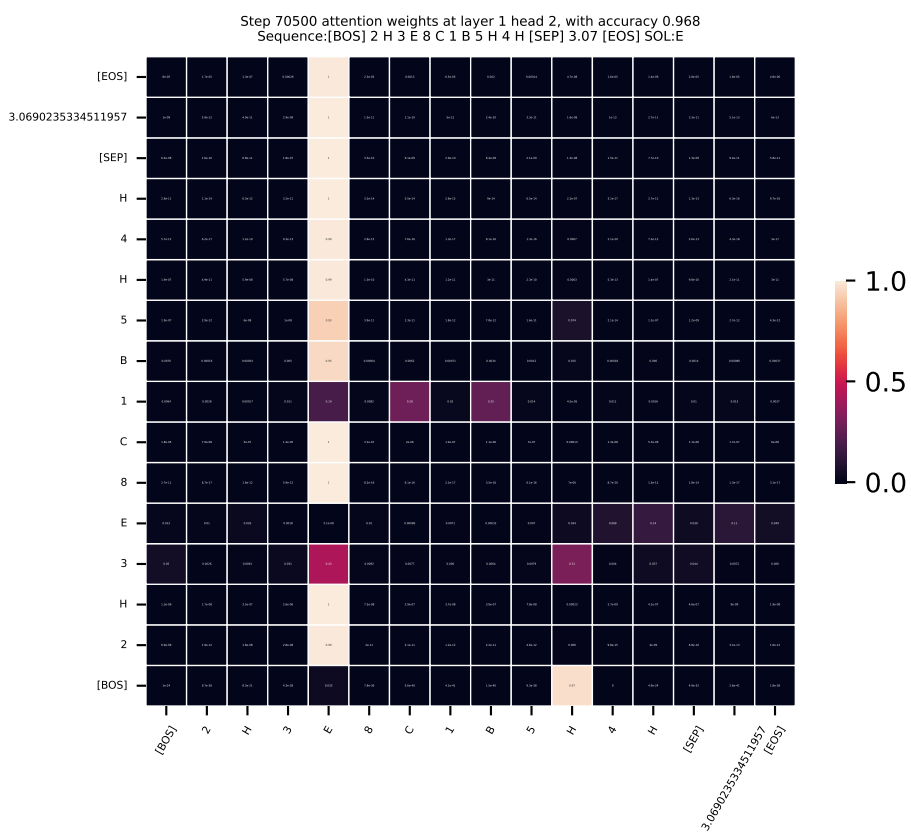

Step 70500 attention weights at layer 0 head 1, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

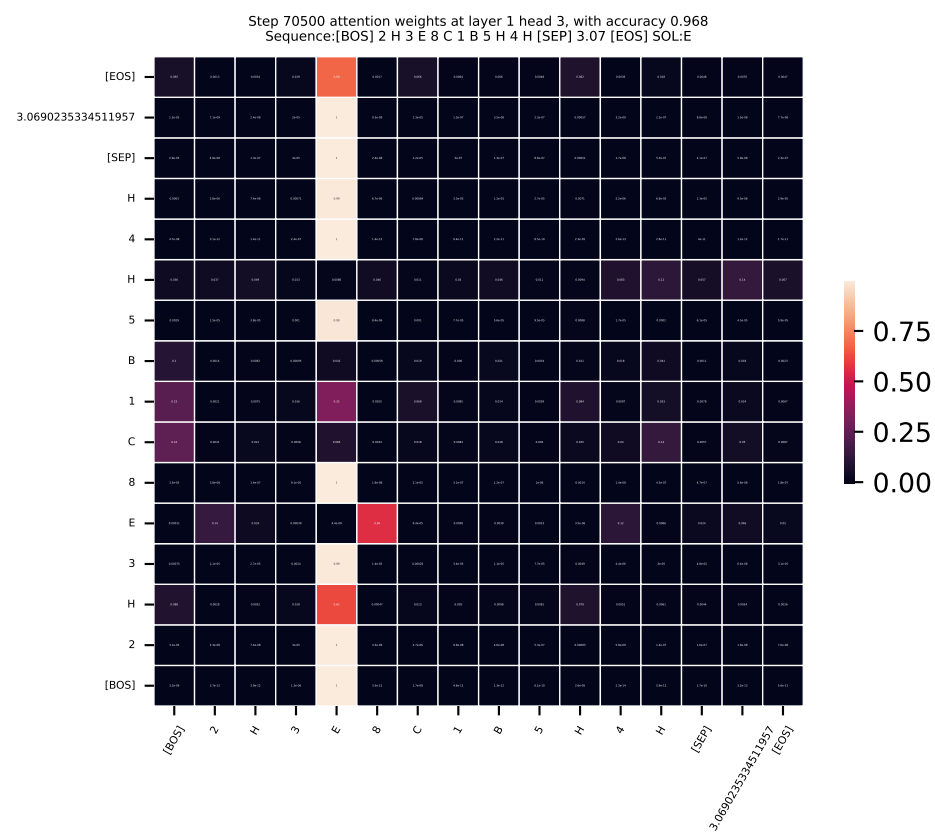

Step 70500 attention weights at layer 0 head 2, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

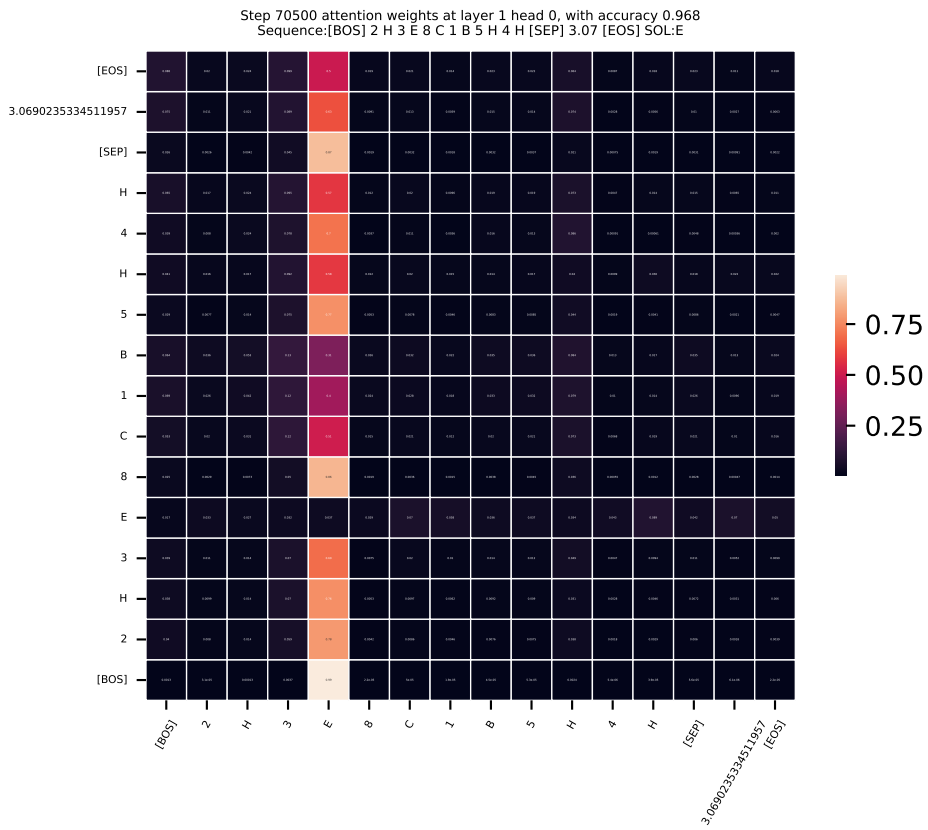

Step 70500 attention weights at layer 0 head 3, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.5
0.0

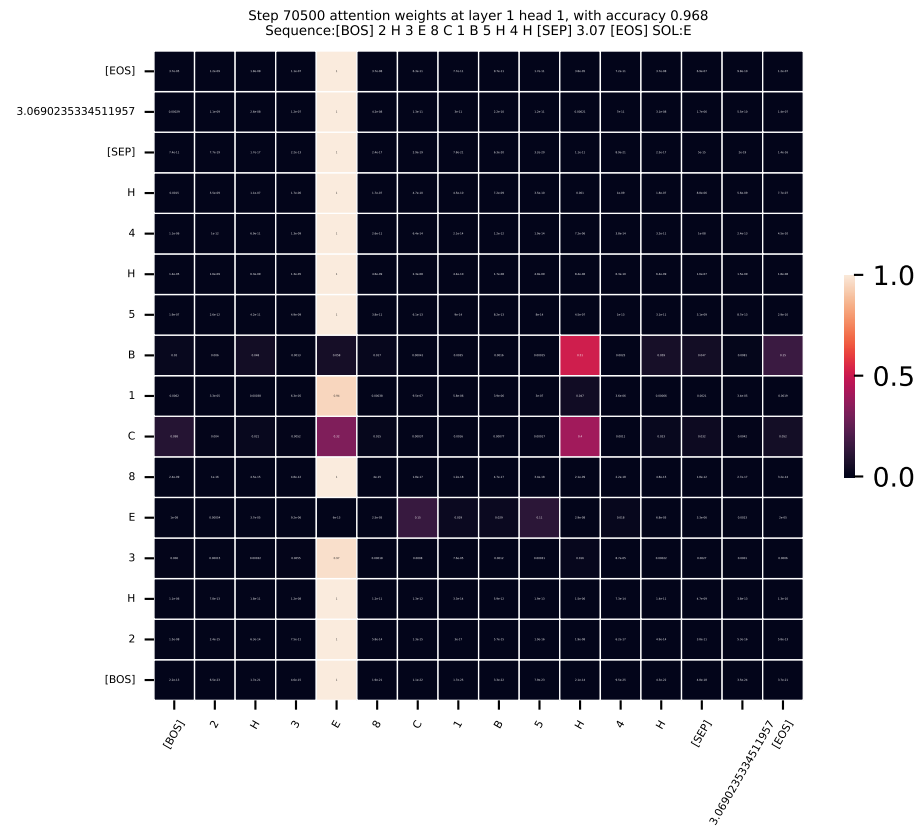

Step 70500 attention weights at layer 1 head 0, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

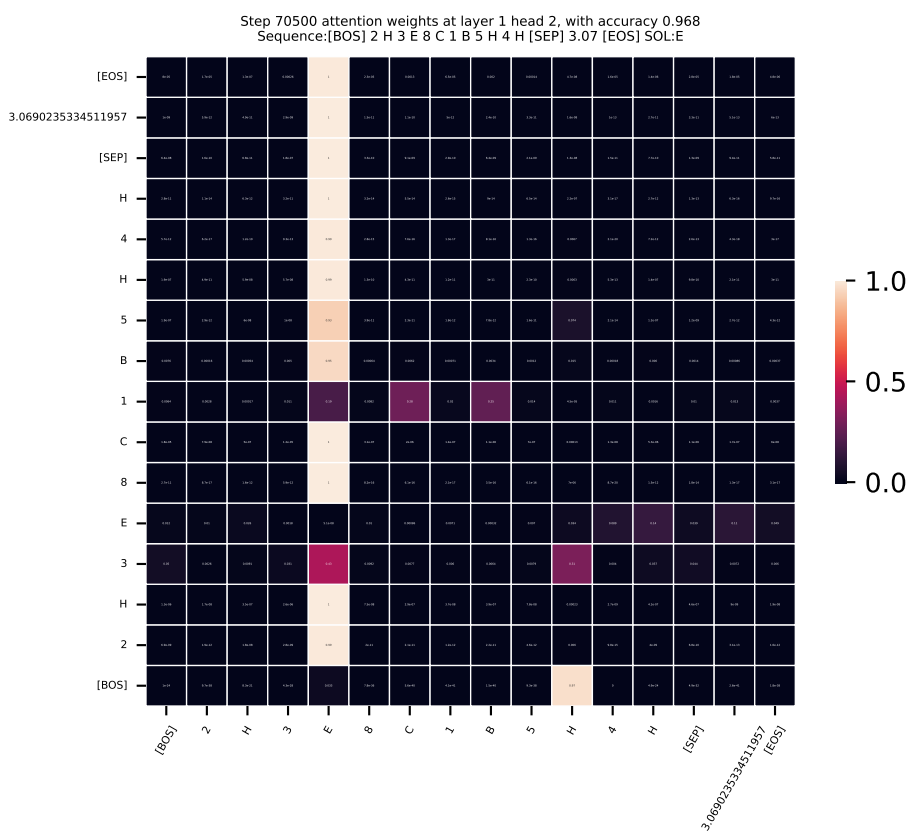

Step 70500 attention weights at layer 1 head 1, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

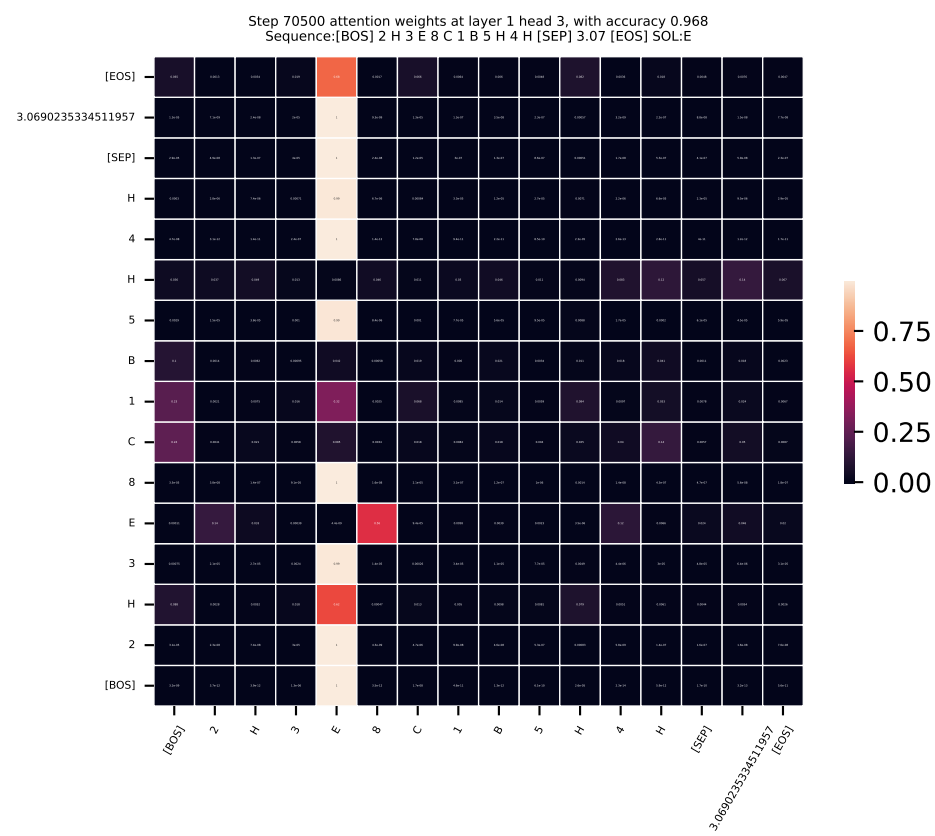

Step 70500 attention weights at layer 1 head 2, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

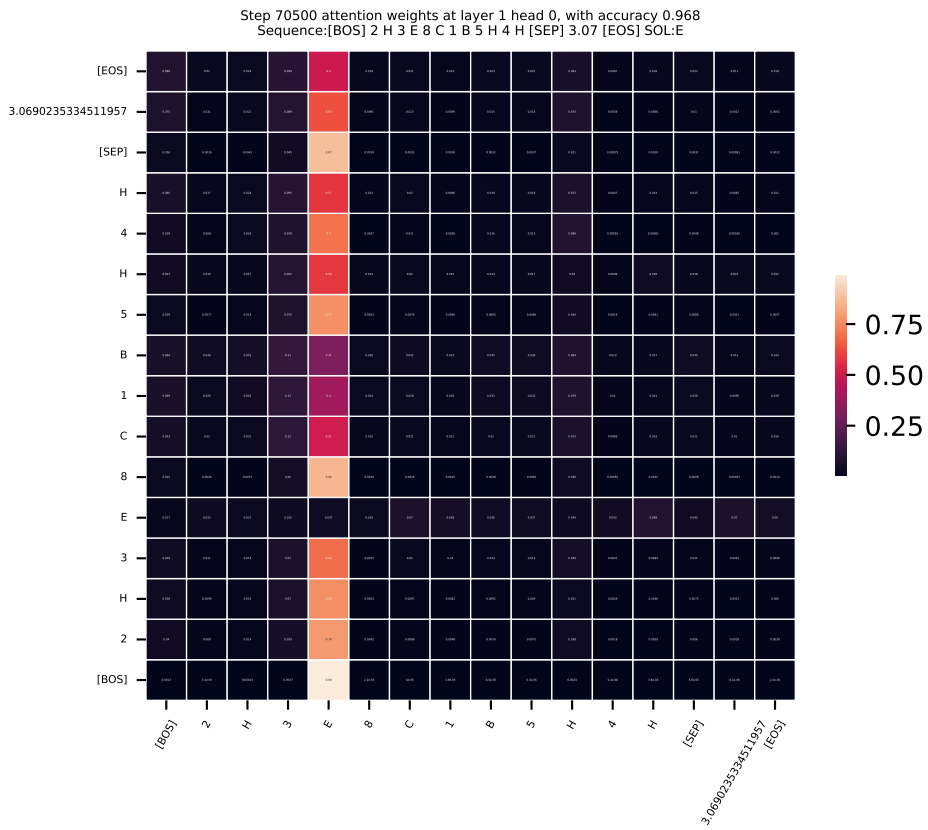

Step 70500 attention weights at layer 1 head 3, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

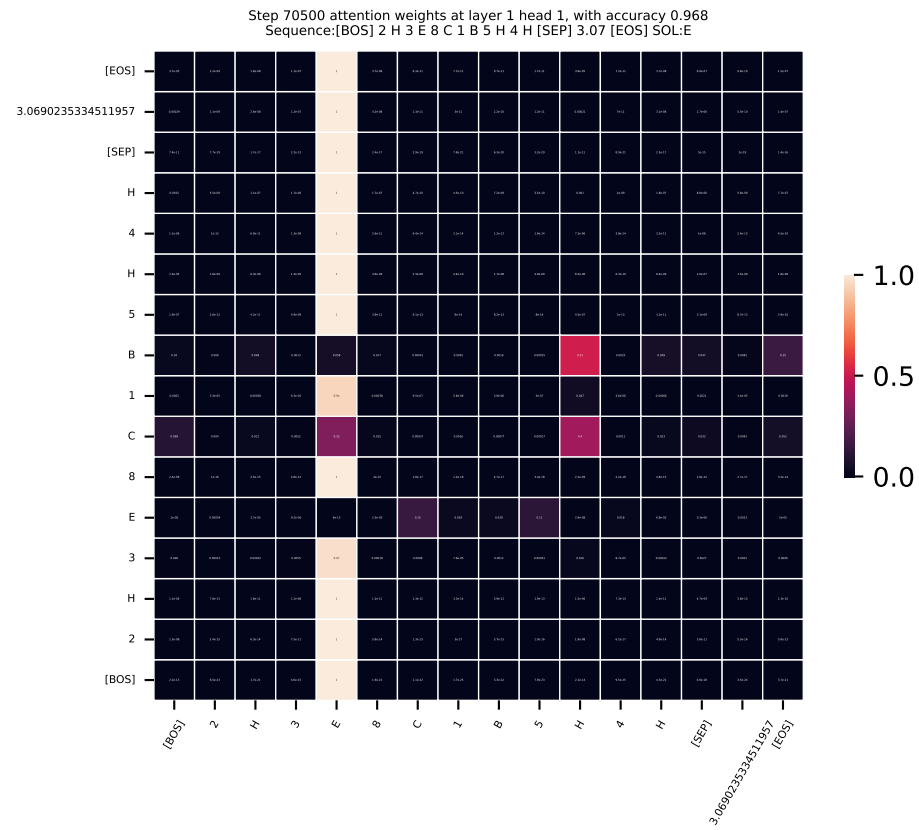

Step 70500 attention weights at layer 1 head 0, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

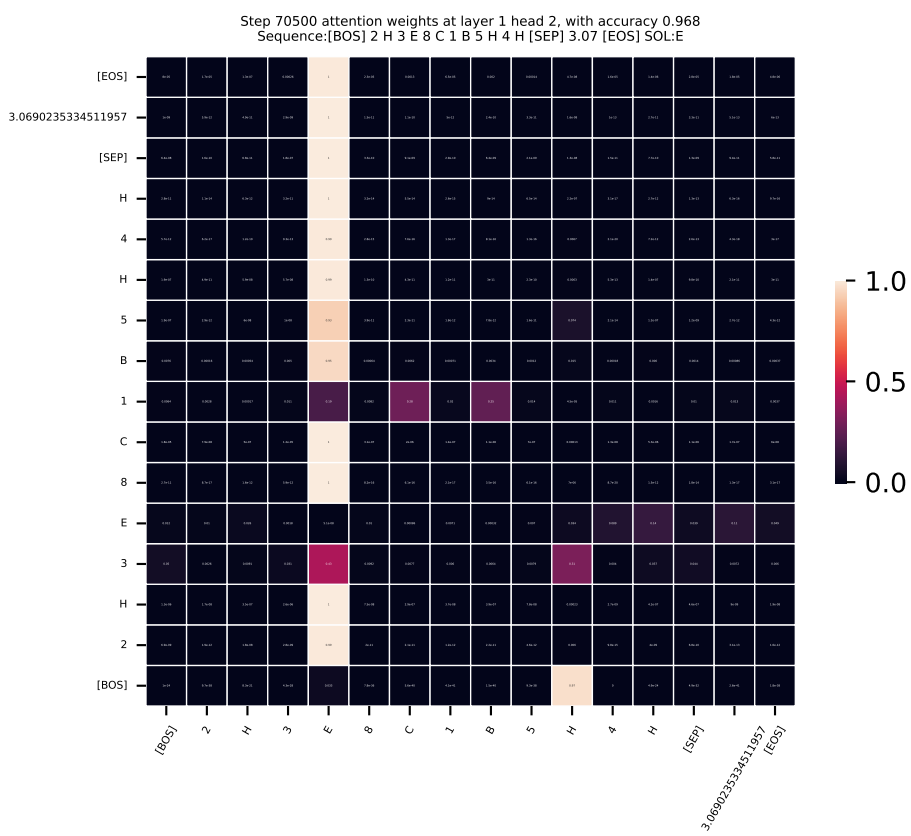

Step 70500 attention weights at layer 1 head 1, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

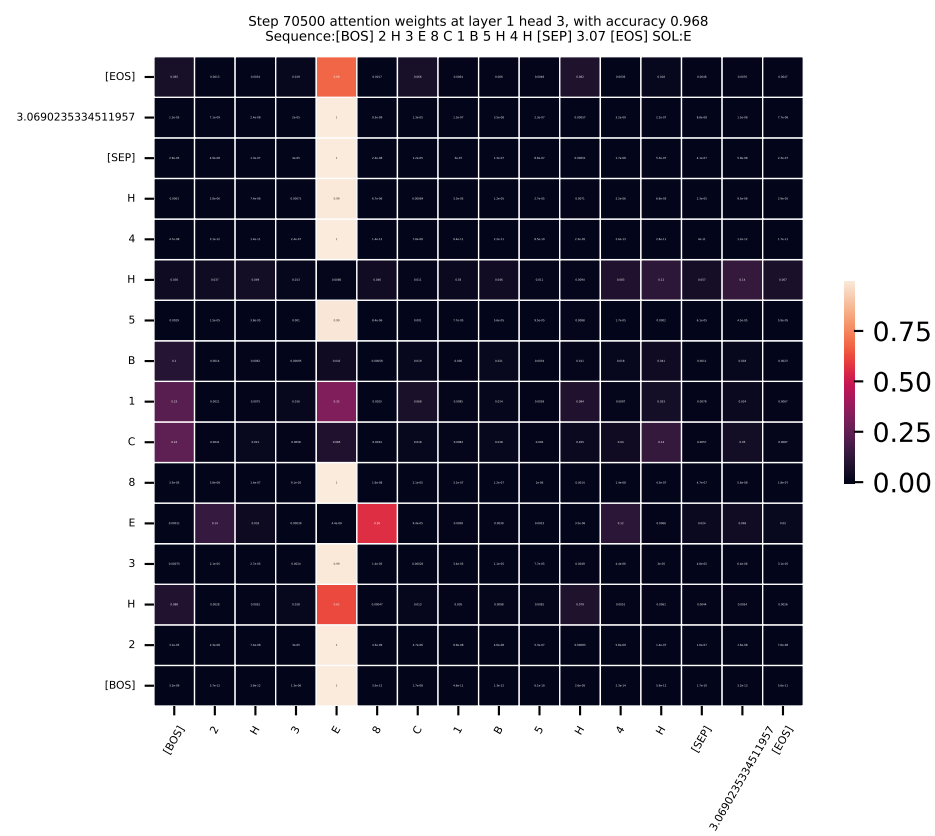

Step 70500 attention weights at layer 1 head 2, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

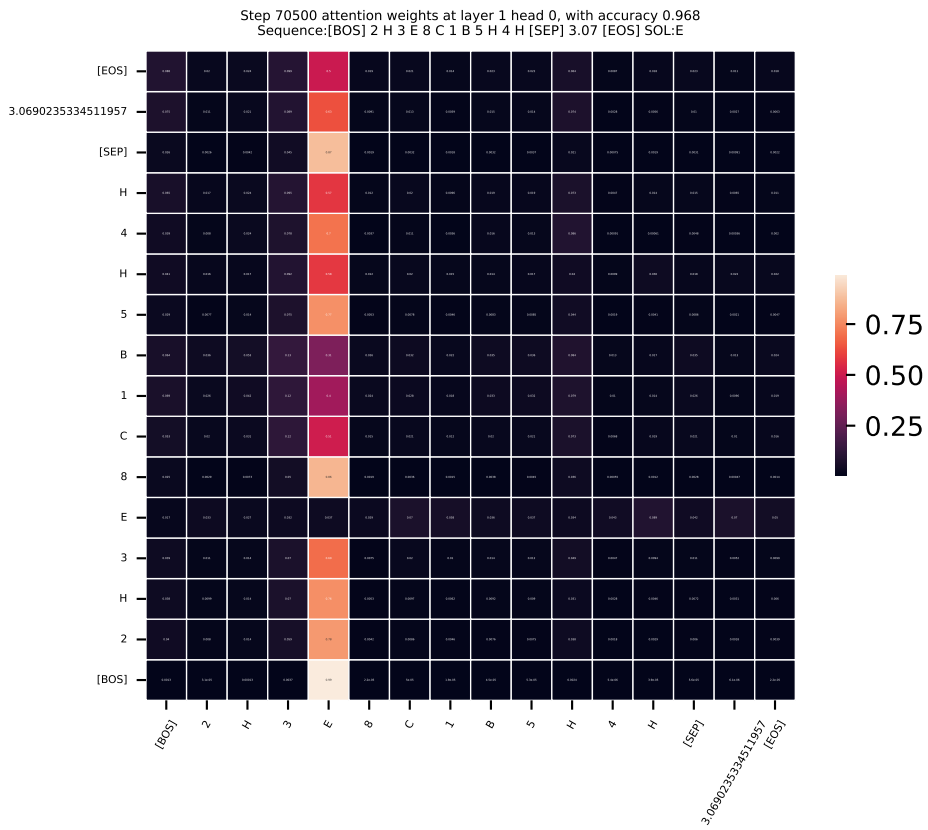

Step 70500 attention weights at layer 1 head 3, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

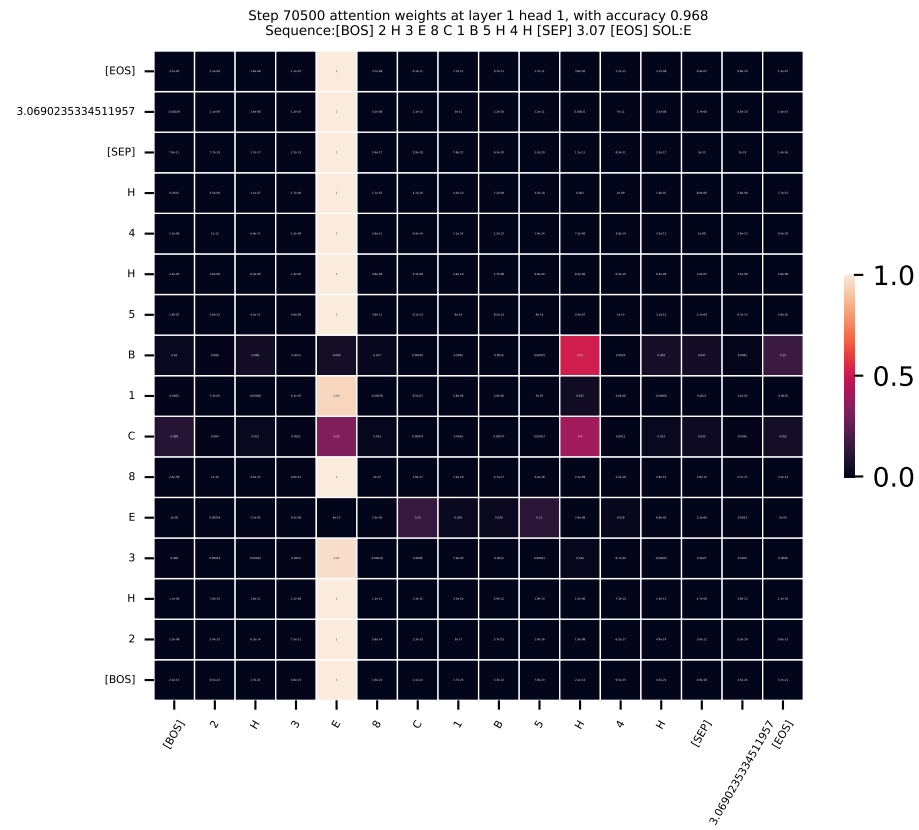

Step 70500 attention weights at layer 1 head 0, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

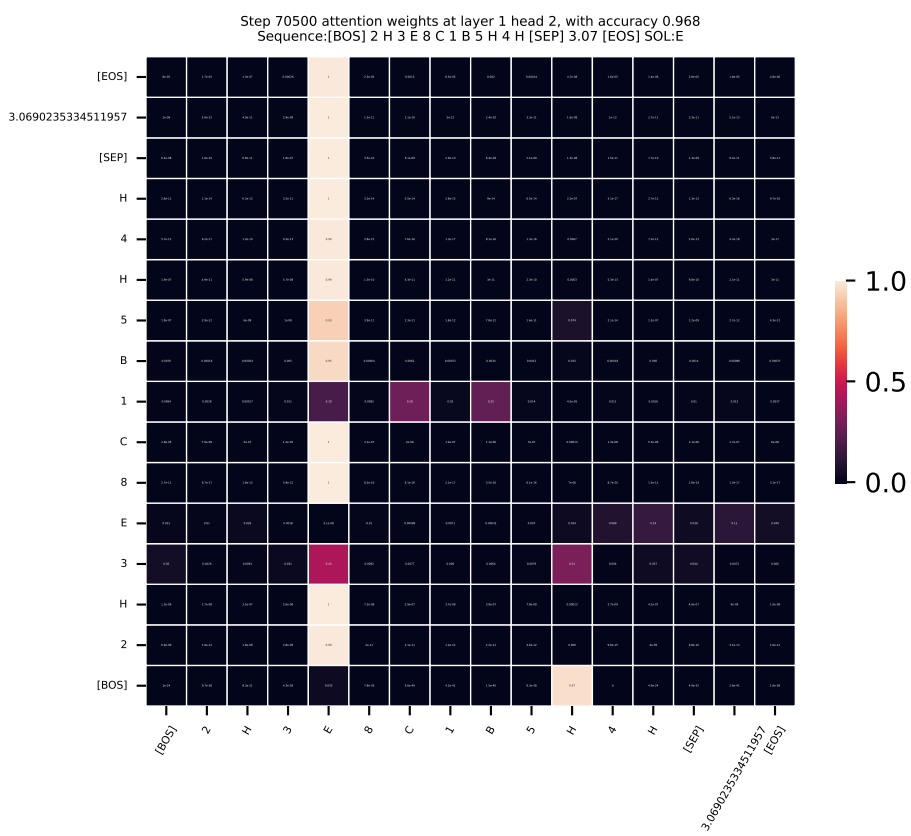

Step 70500 attention weights at layer 1 head 1, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

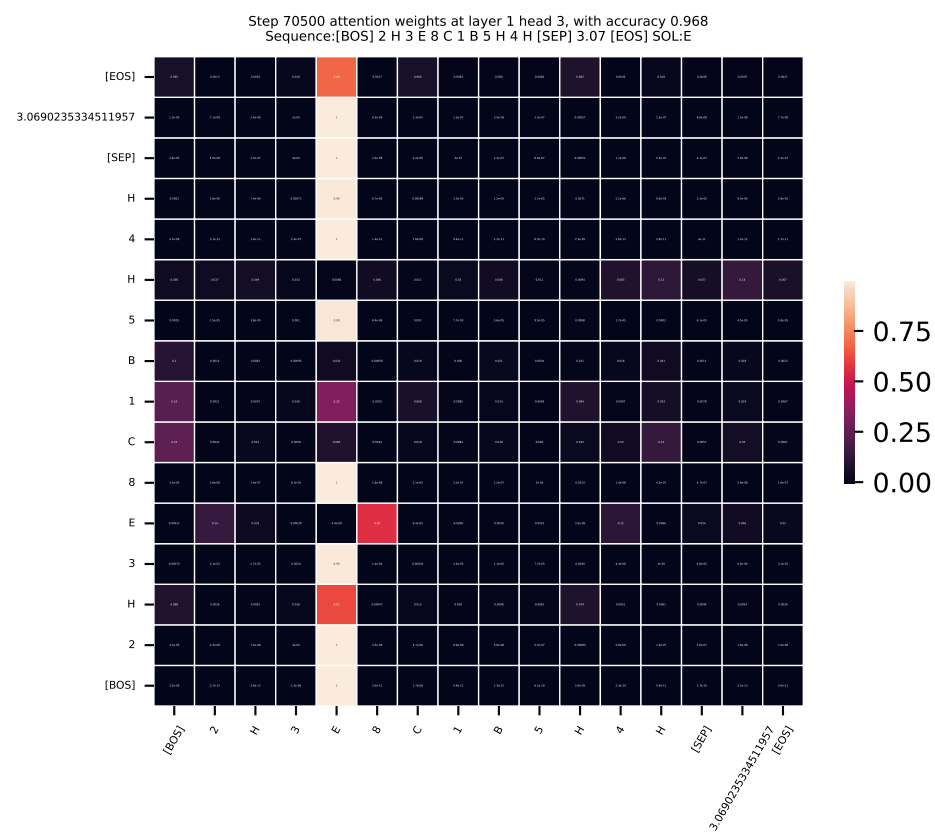

Step 70500 attention weights at layer 1 head 2, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

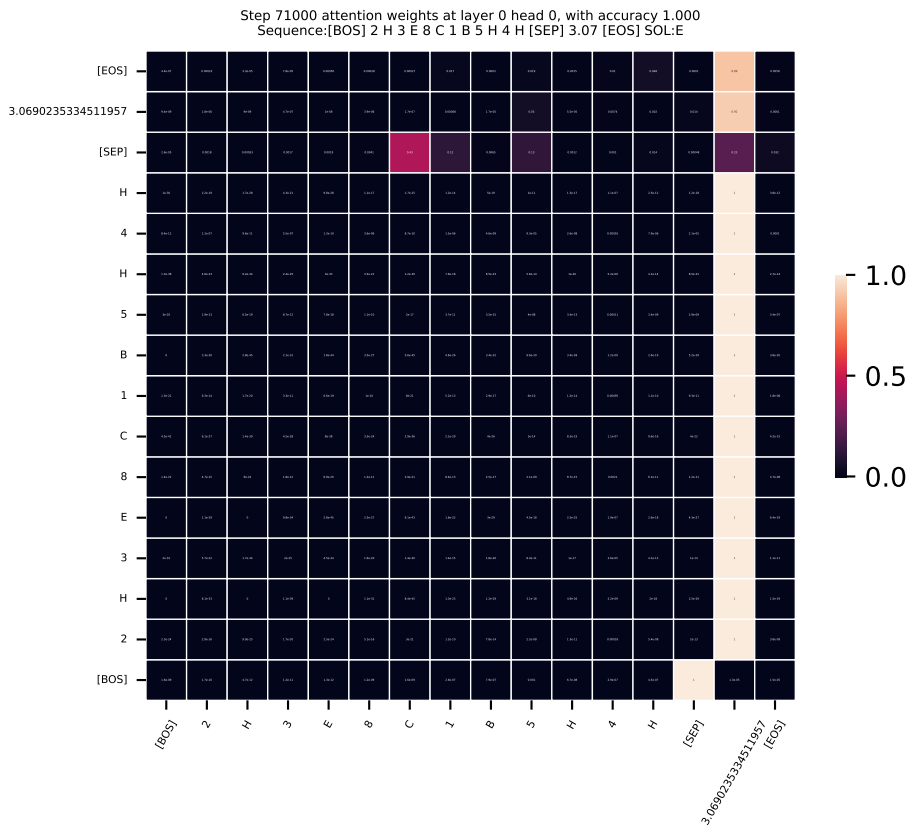

Step 70500 attention weights at layer 1 head 3, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

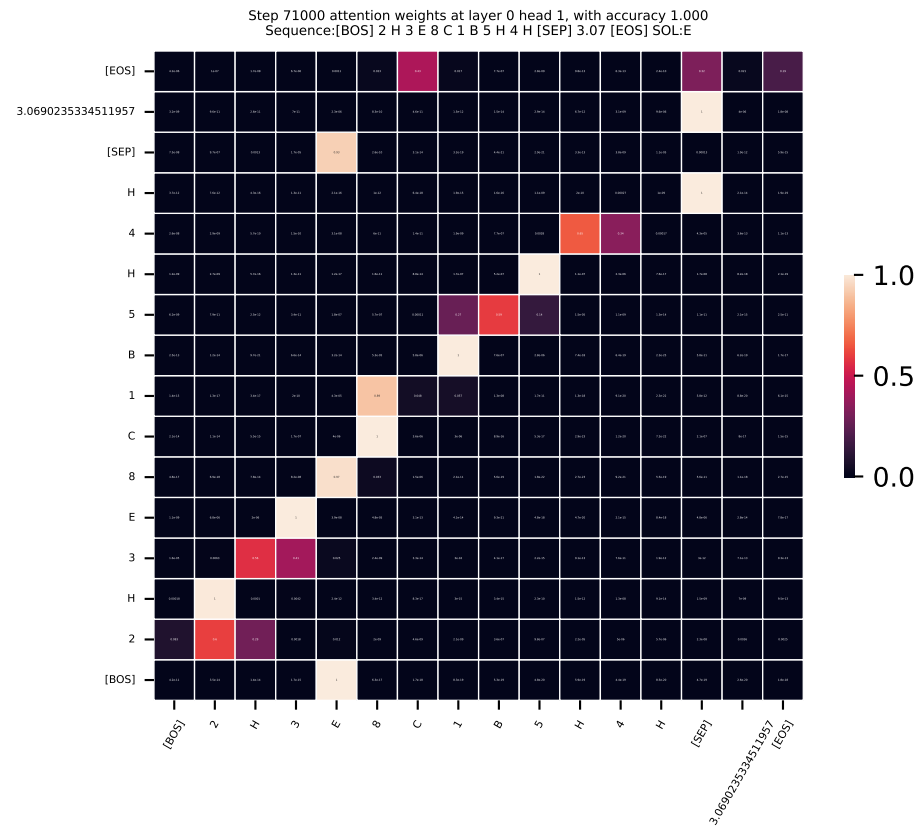

Step 70500 attention weights at layer 1 head 0, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

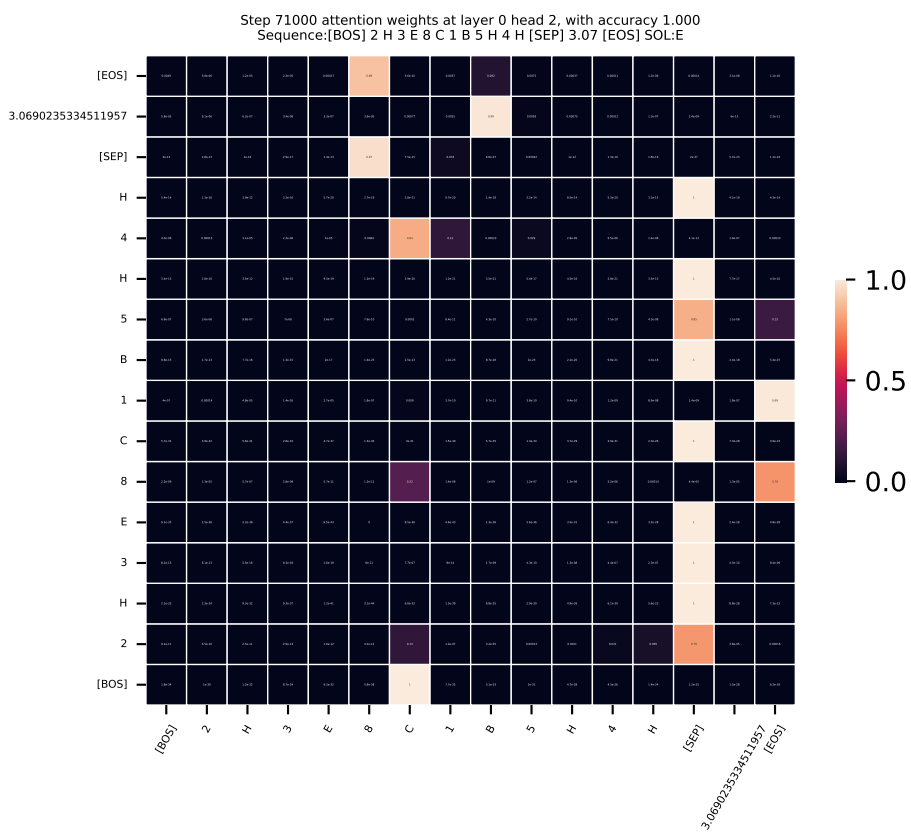

Step 70500 attention weights at layer 1 head 1, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

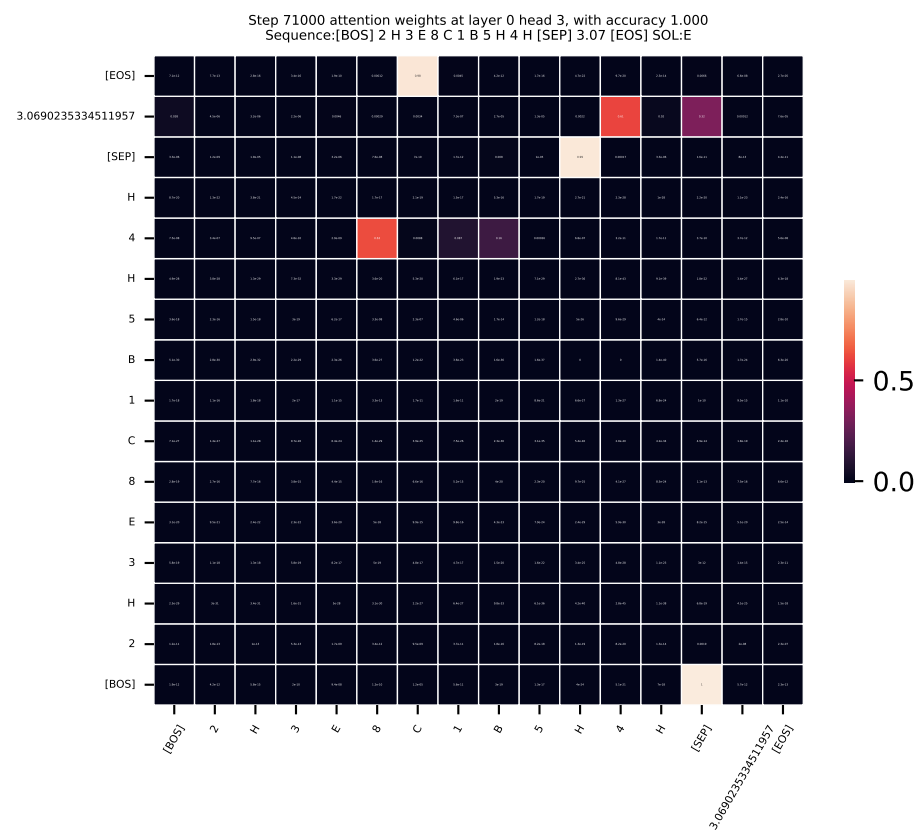

Step 70500 attention weights at layer 1 head 2, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

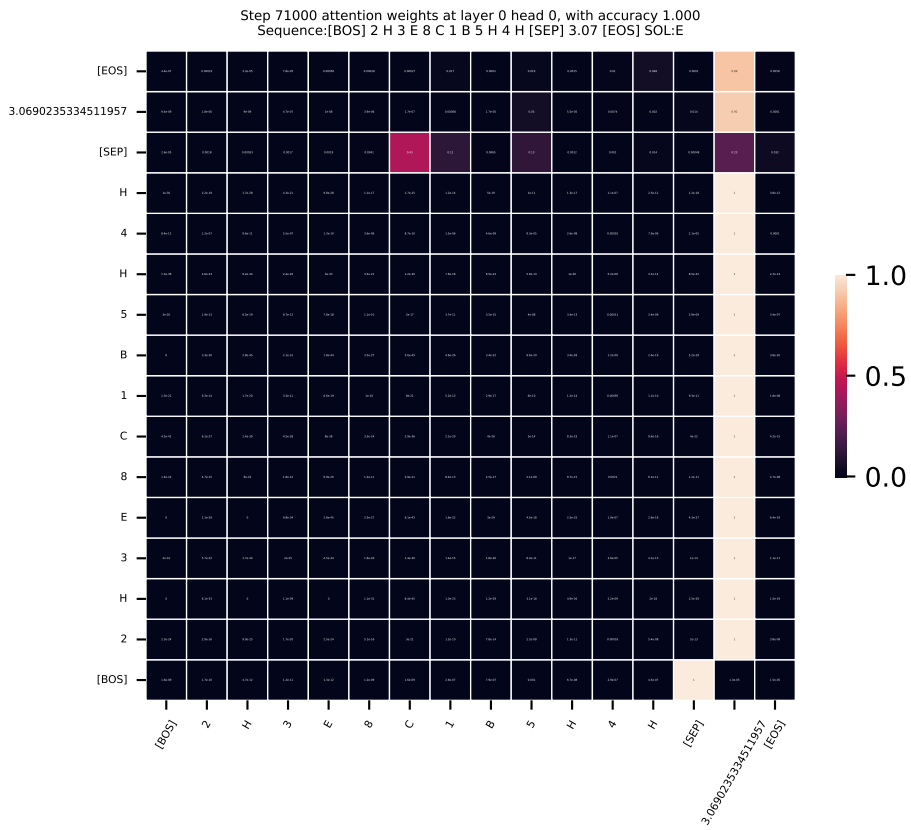

Step 70500 attention weights at layer 1 head 3, with accuracy 0.968
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

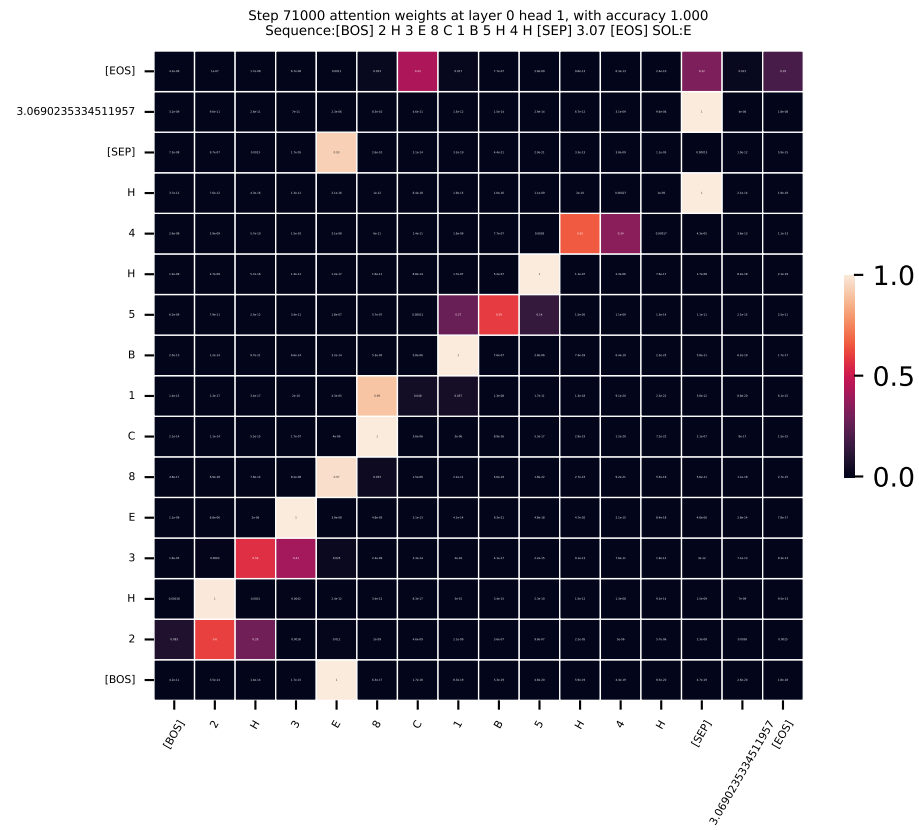

Step 71000 attention weights at layer 0 head 0, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

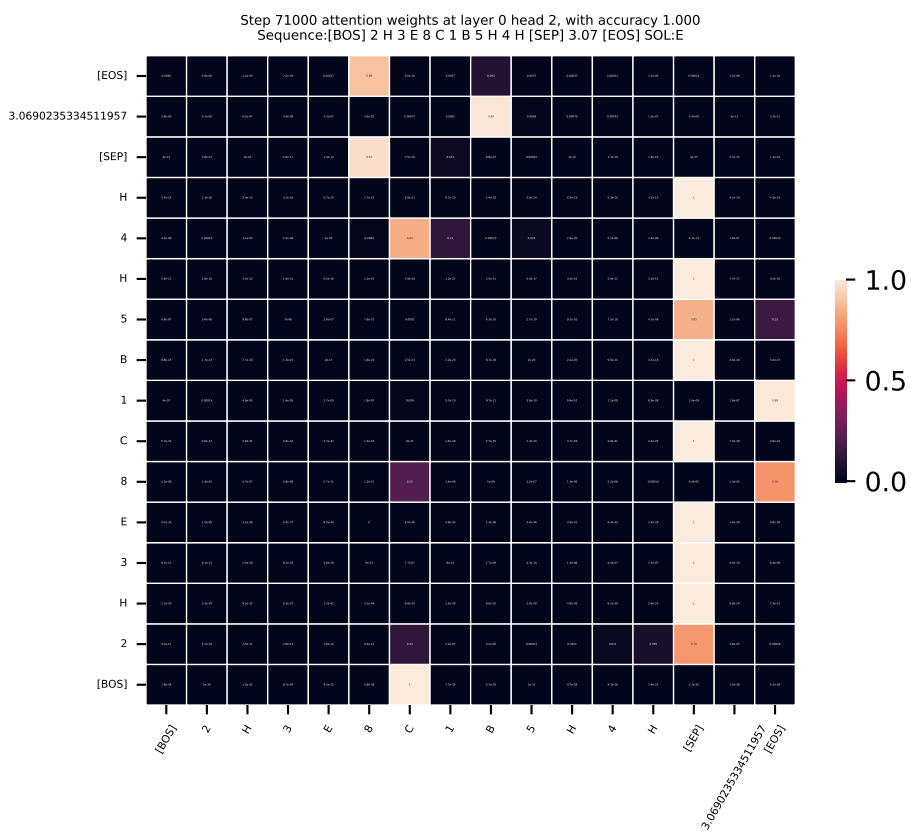

Step 71000 attention weights at layer 0 head 1, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

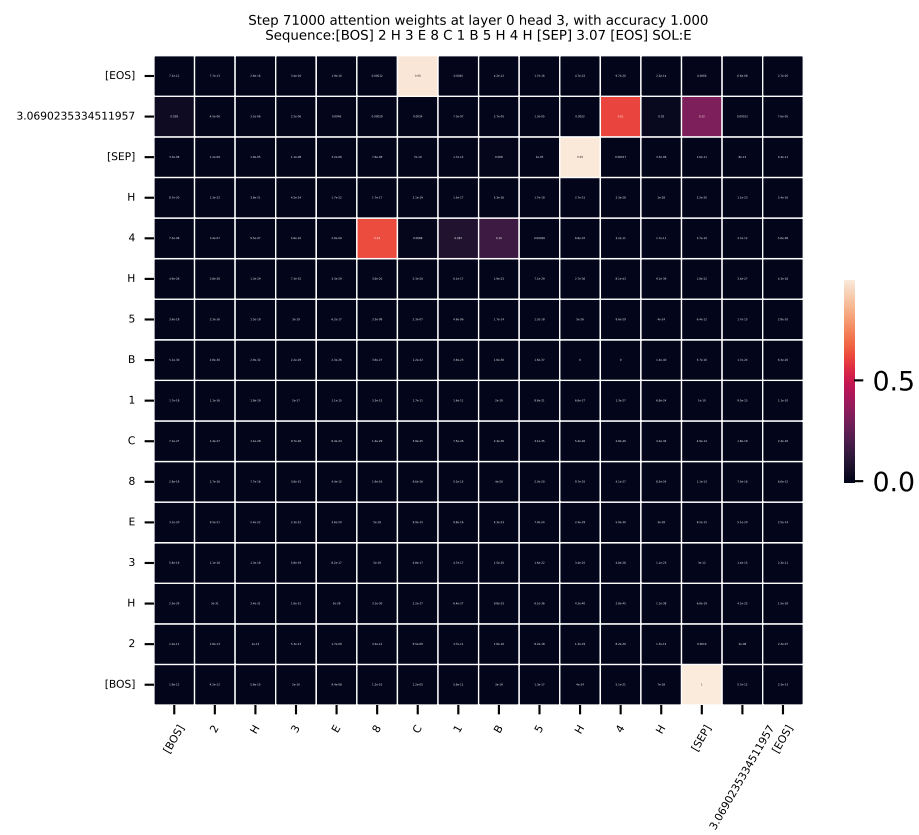

Step 71000 attention weights at layer 0 head 2, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

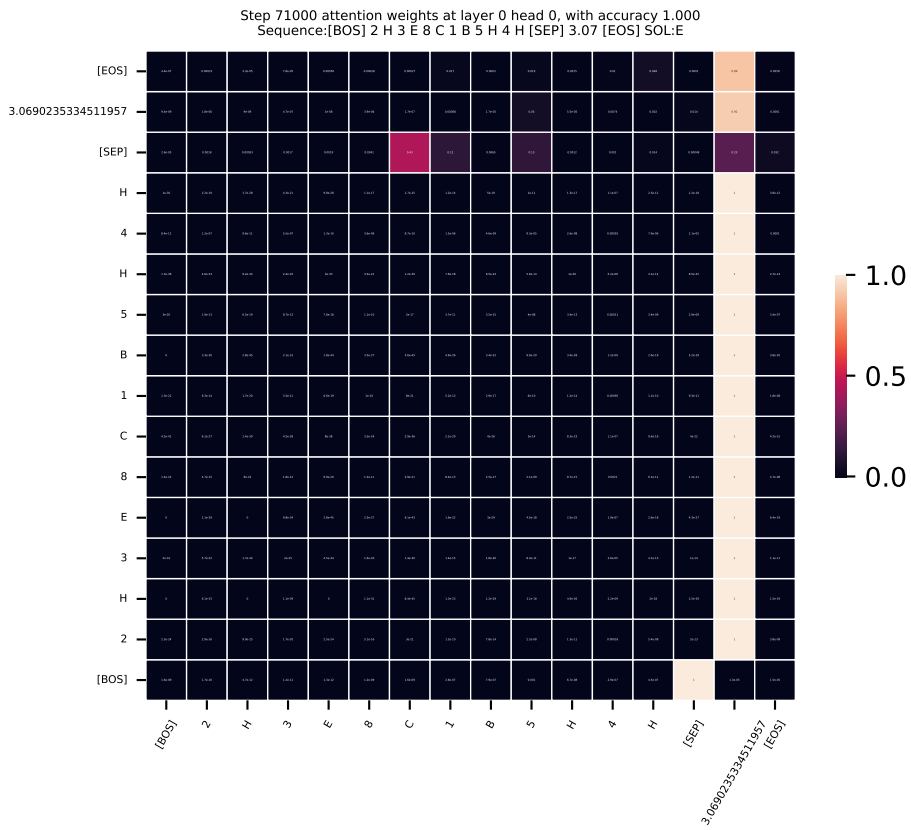

Step 71000 attention weights at layer 0 head 3, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.5
0.0

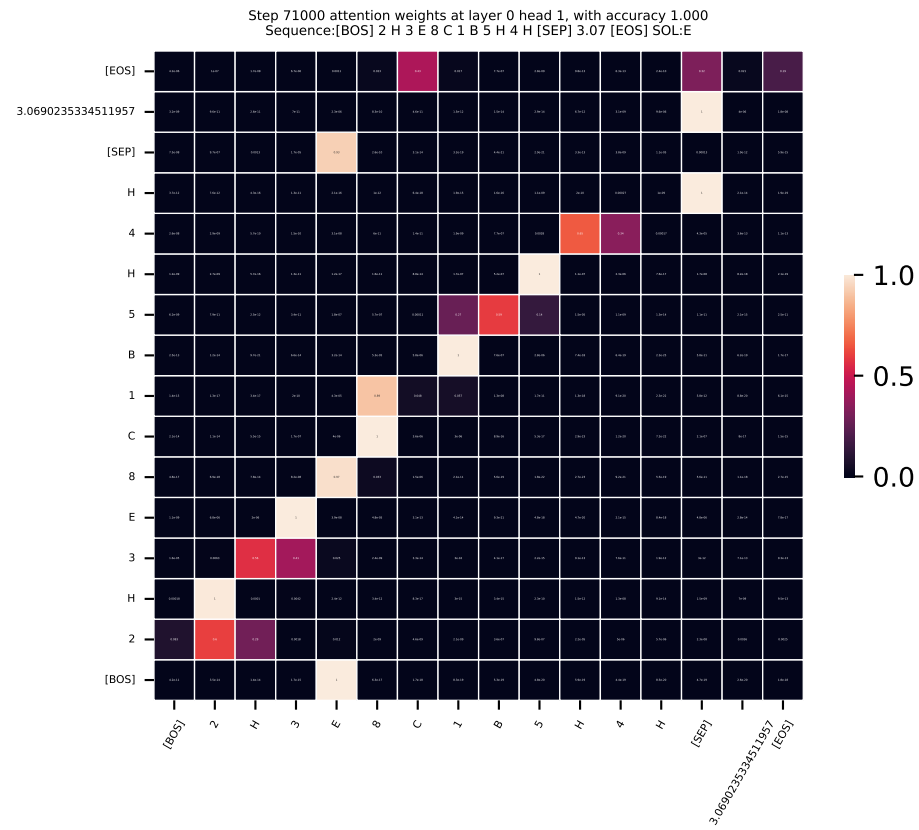

Step 71000 attention weights at layer 0 head 0, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

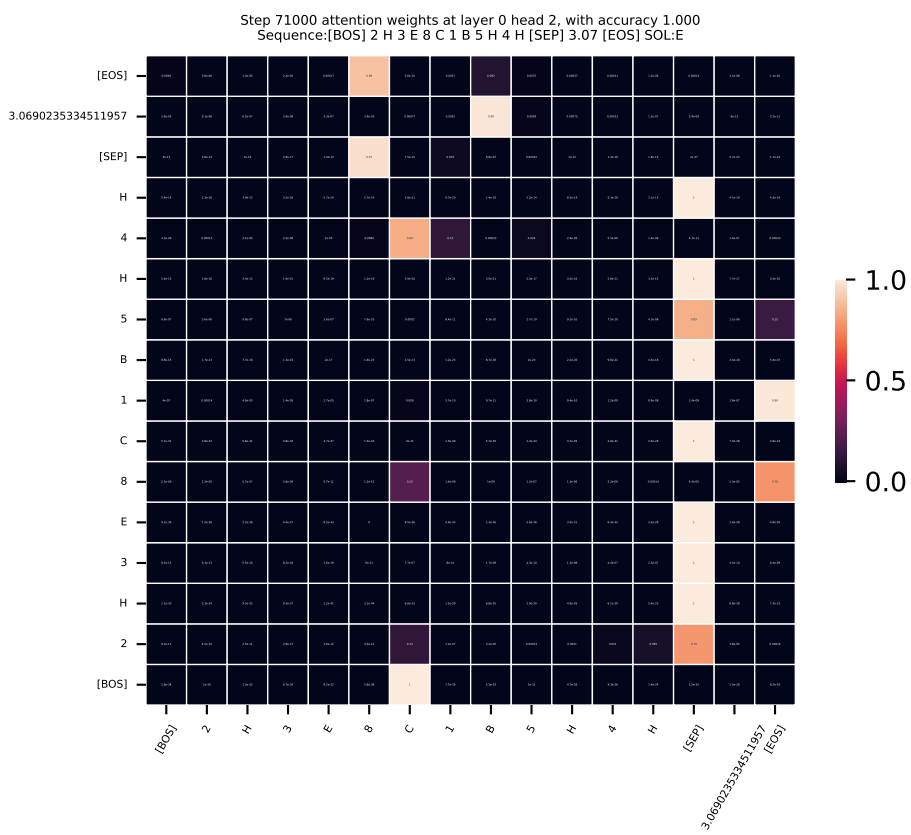

Step 71000 attention weights at layer 0 head 1, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

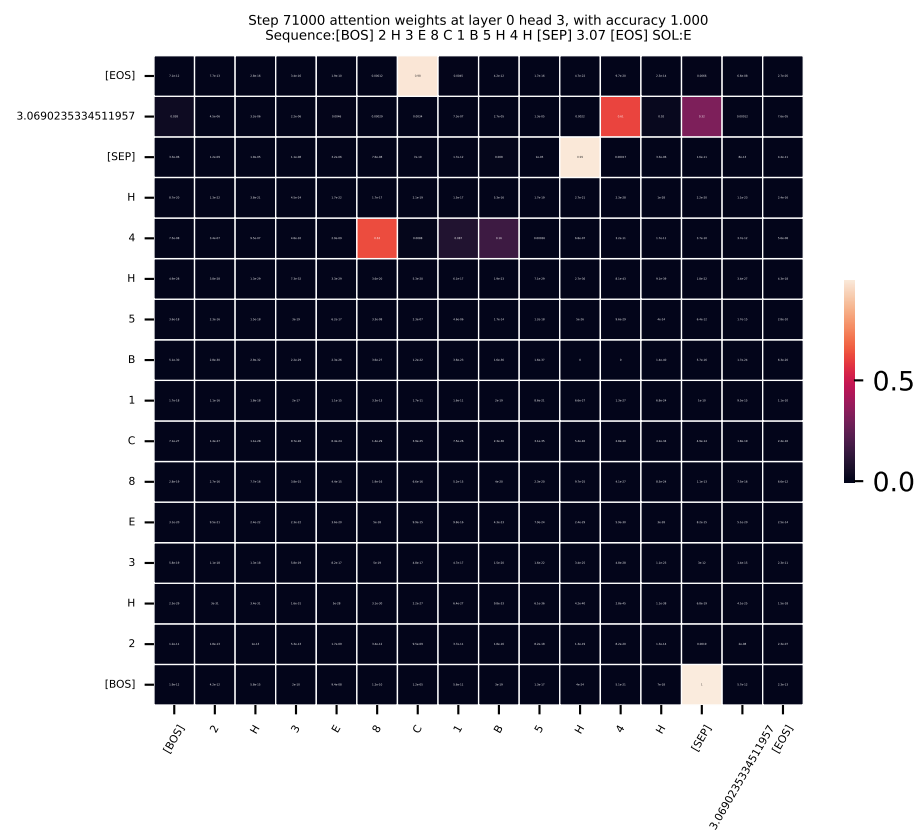

Step 71000 attention weights at layer 0 head 2, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

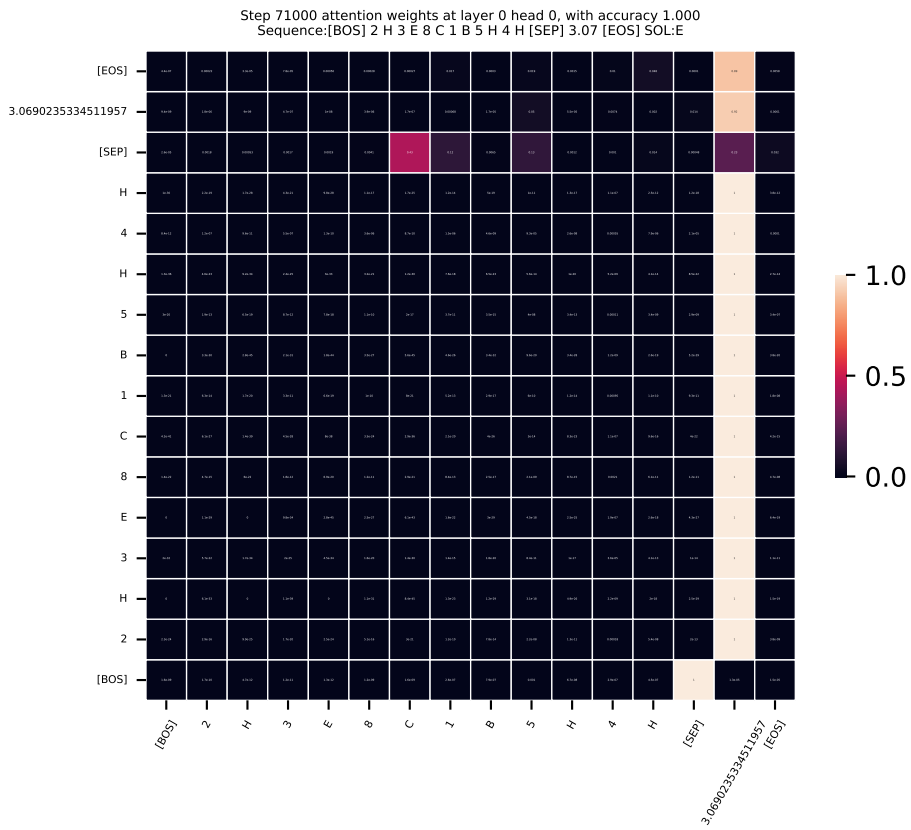

Step 71000 attention weights at layer 0 head 3, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.5
0.0

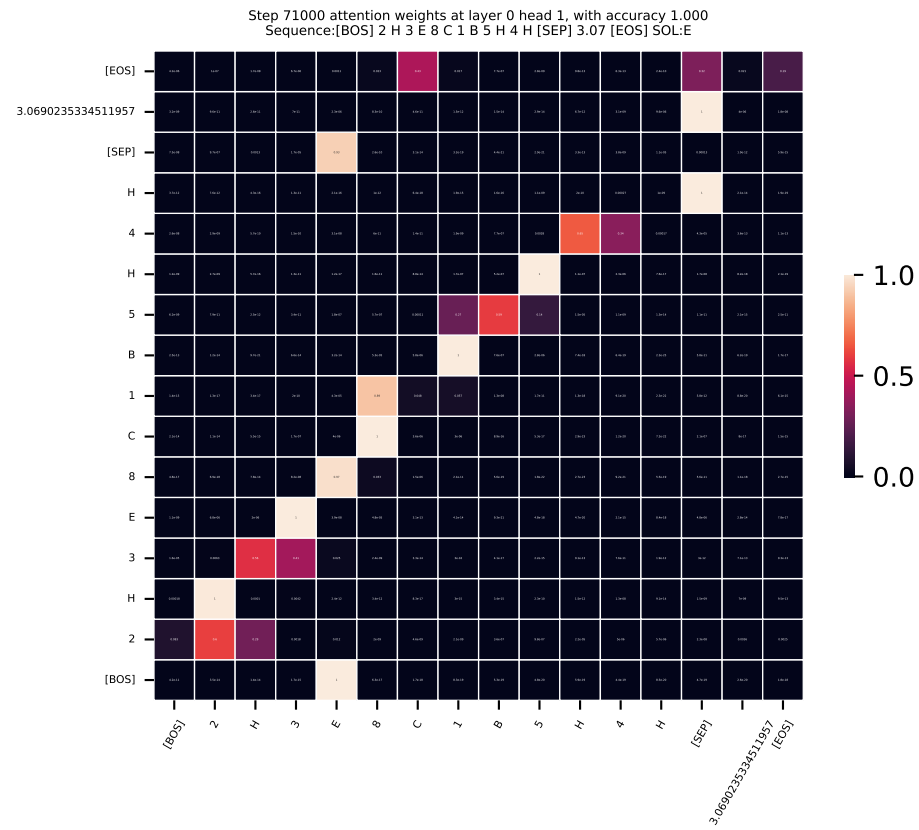

Step 71000 attention weights at layer 0 head 0, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

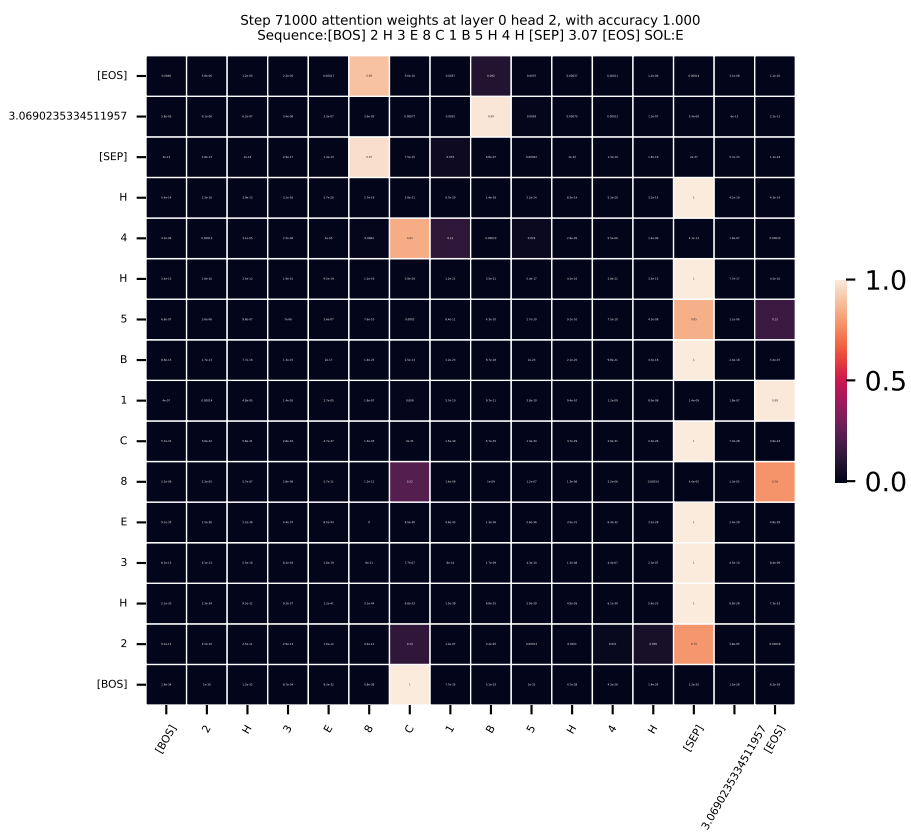

Step 71000 attention weights at layer 0 head 1, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

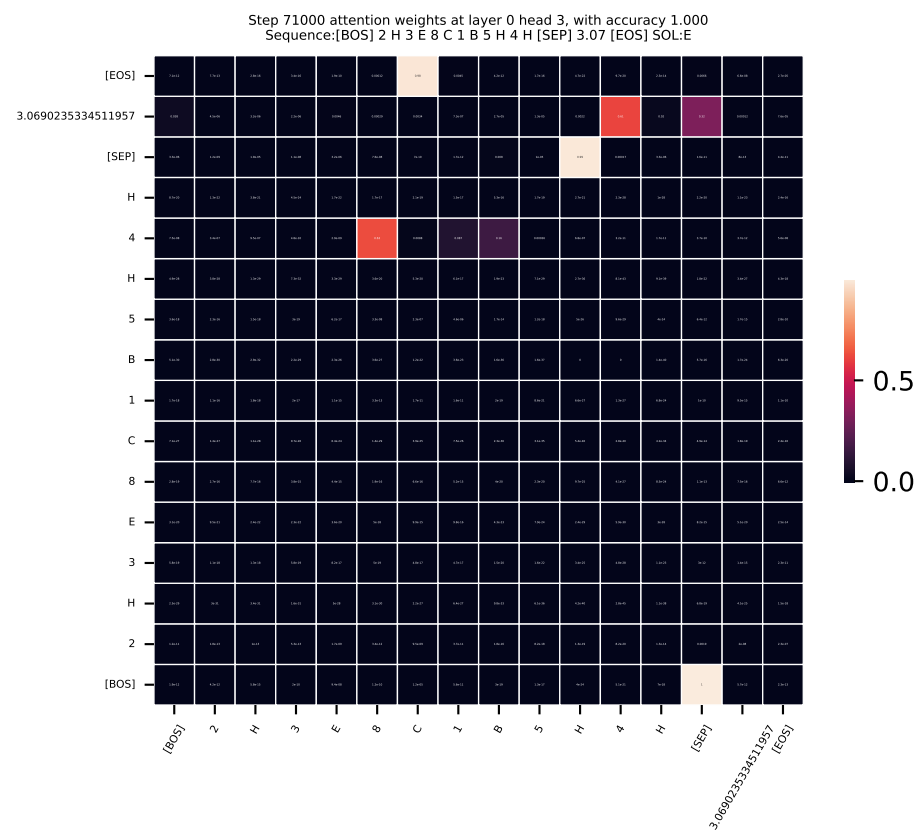

Step 71000 attention weights at layer 0 head 2, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

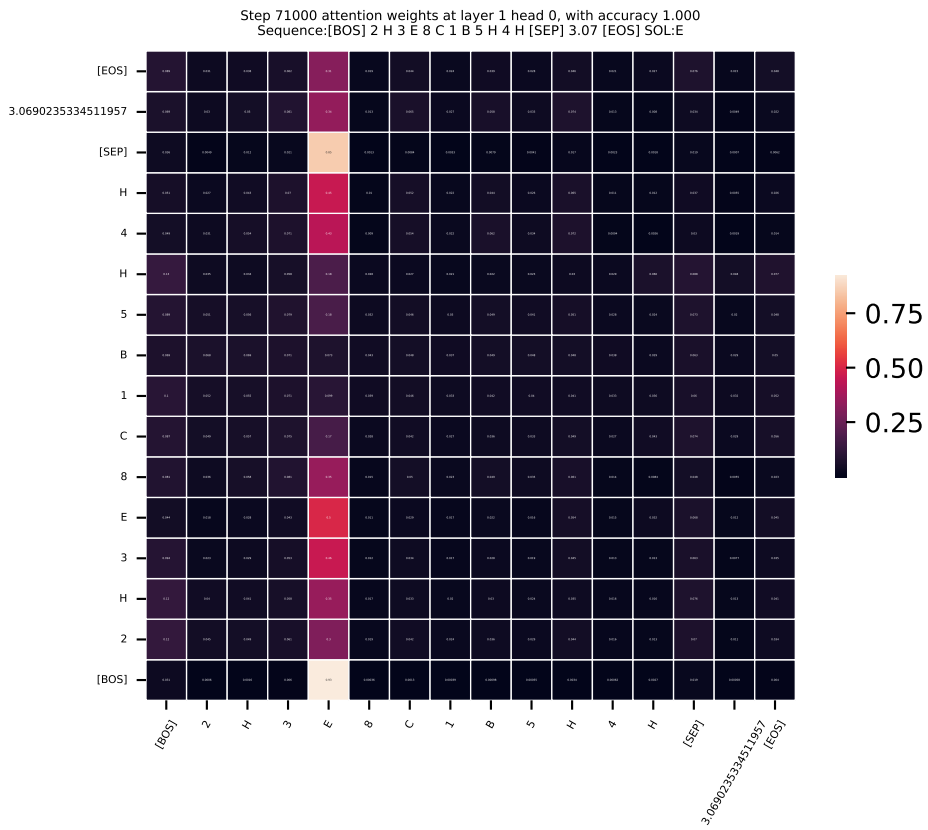

Step 71000 attention weights at layer 0 head 3, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.5
0.0

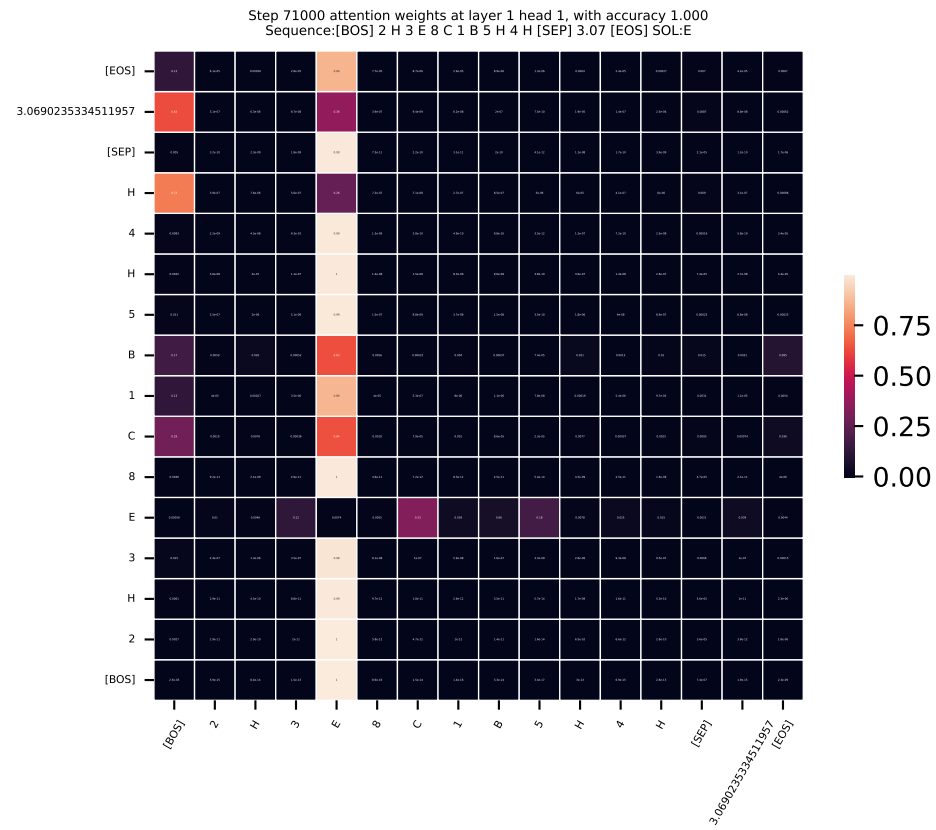

Step 71000 attention weights at layer 0 head 0, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

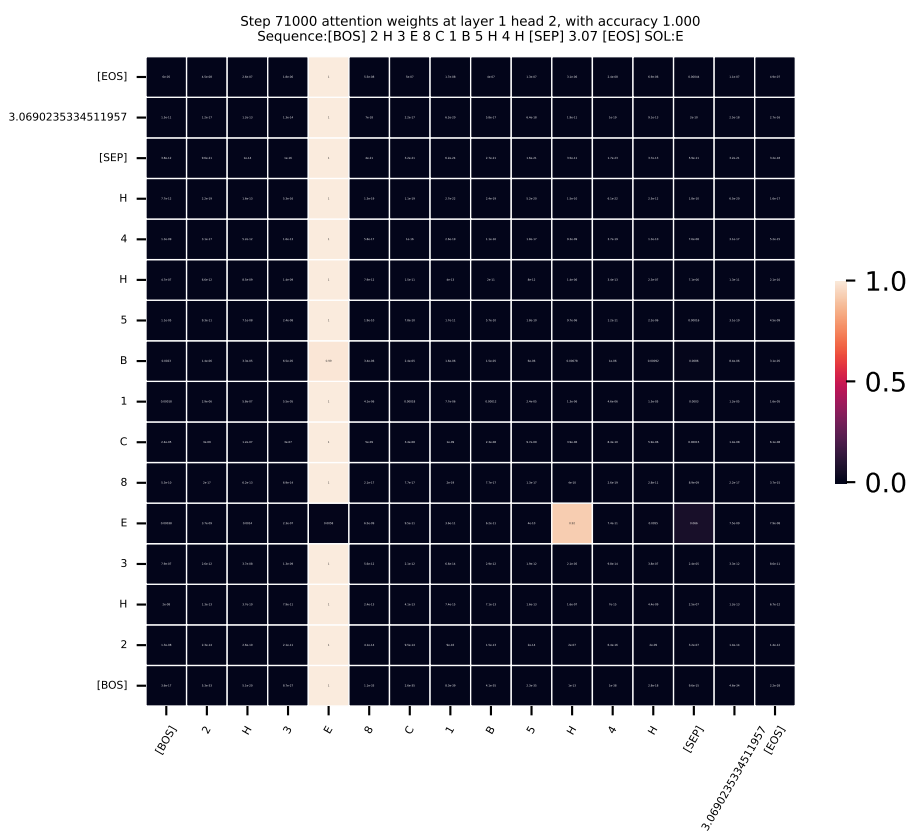

Step 71000 attention weights at layer 0 head 1, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

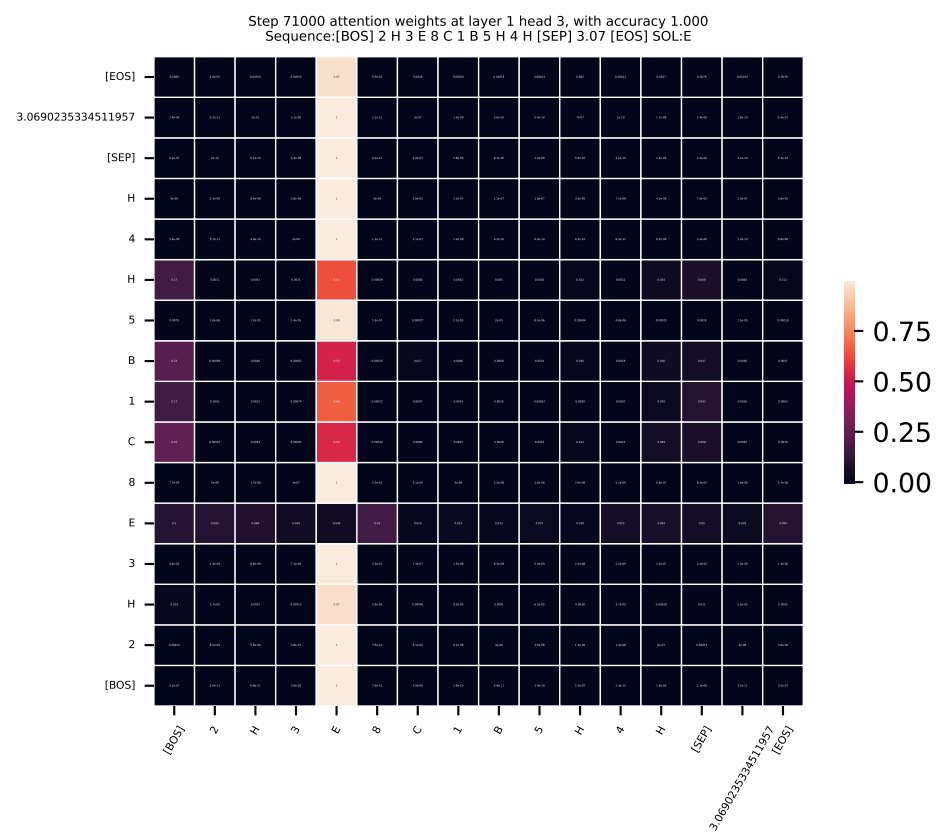

Step 71000 attention weights at layer 0 head 2, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
1.0
0.5
0.0

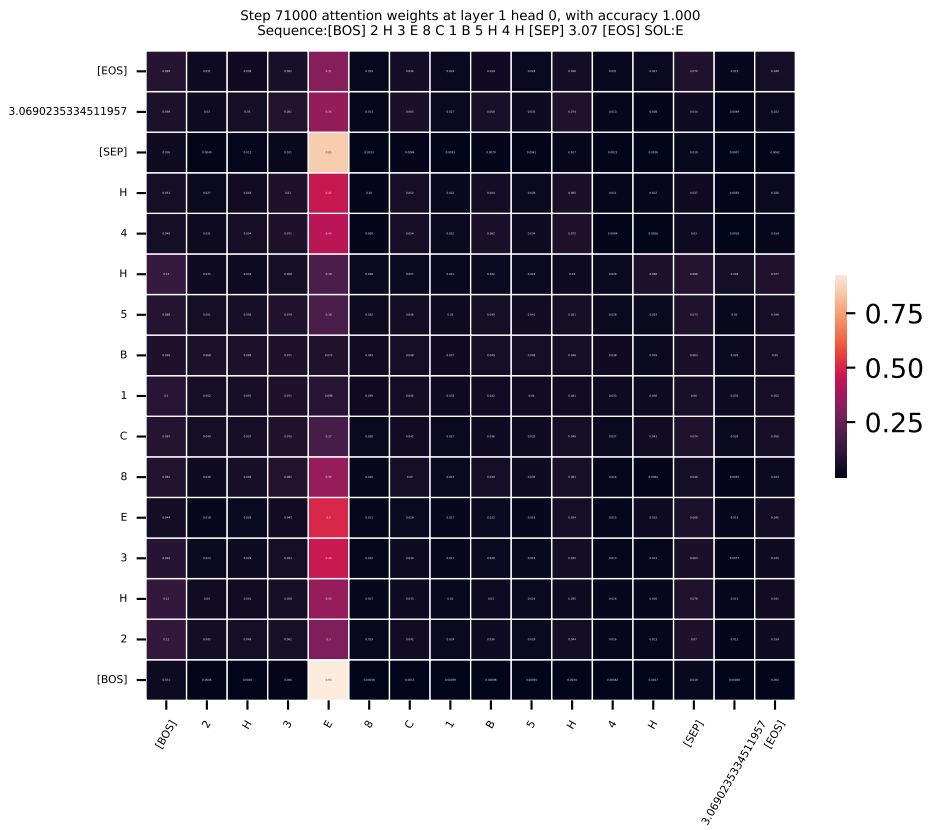

Step 71000 attention weights at layer 0 head 3, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.0690235334511957 [EOS]
0.5
0.0

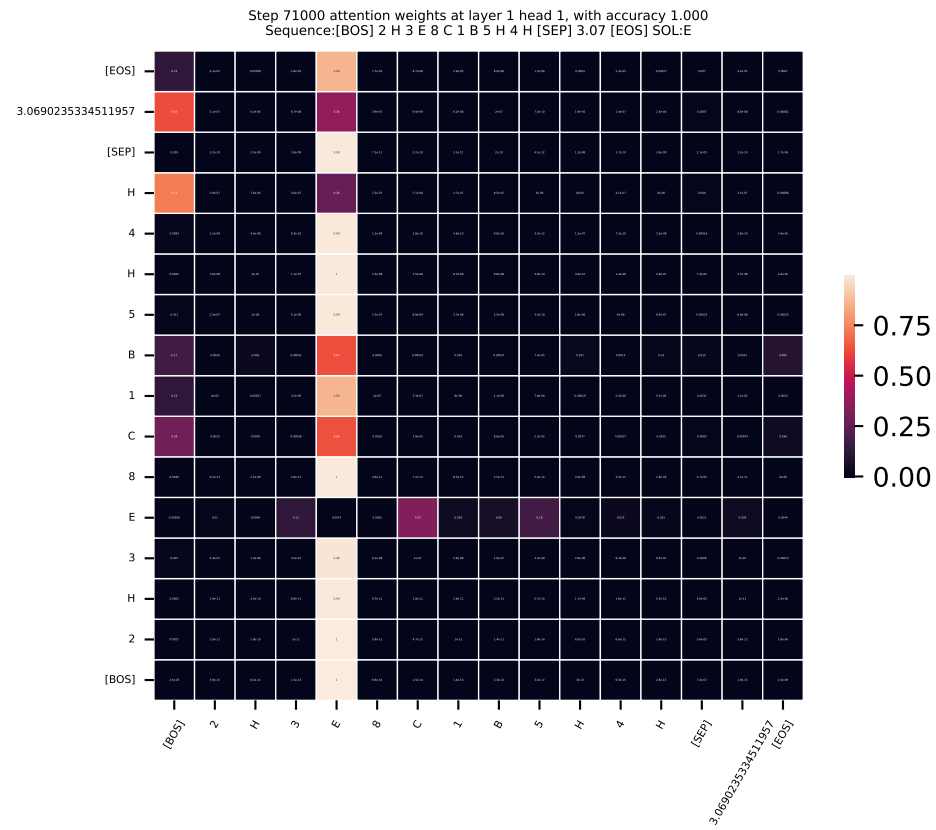

Step 71000 attention weights at layer 1 head 0, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

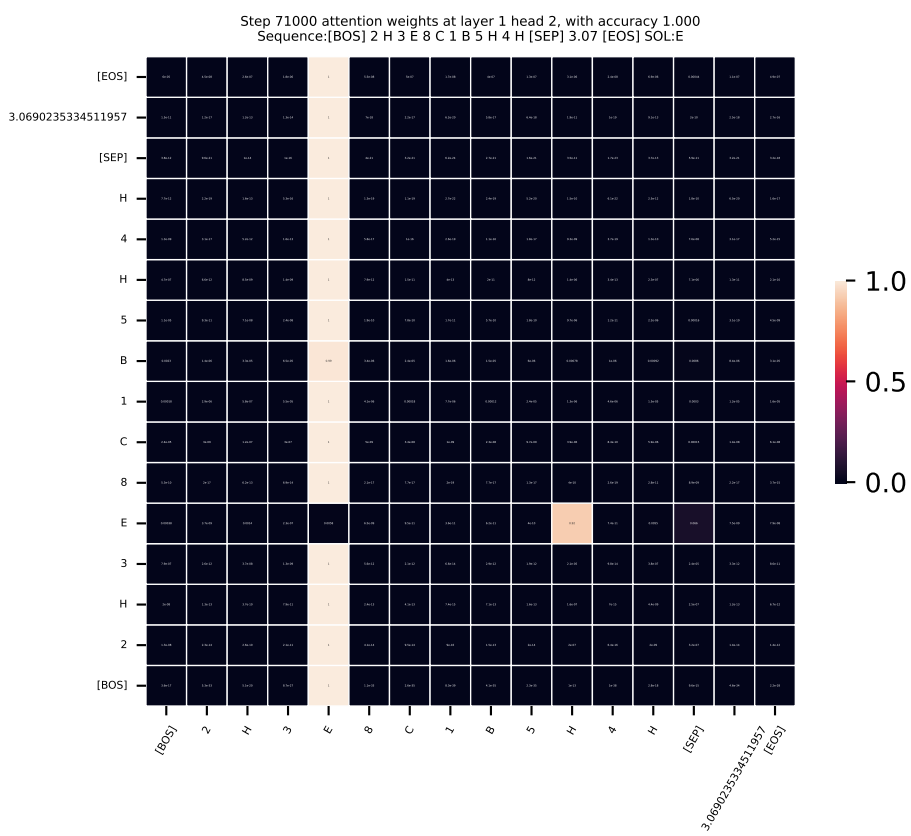

Step 71000 attention weights at layer 1 head 1, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

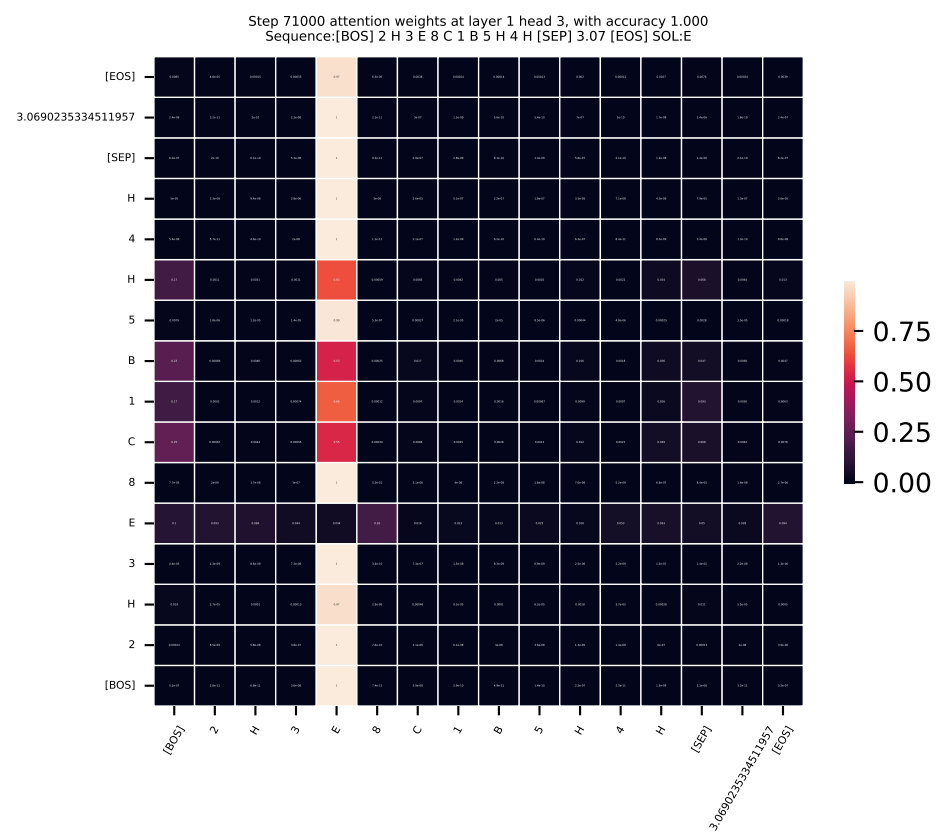

Step 71000 attention weights at layer 1 head 2, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

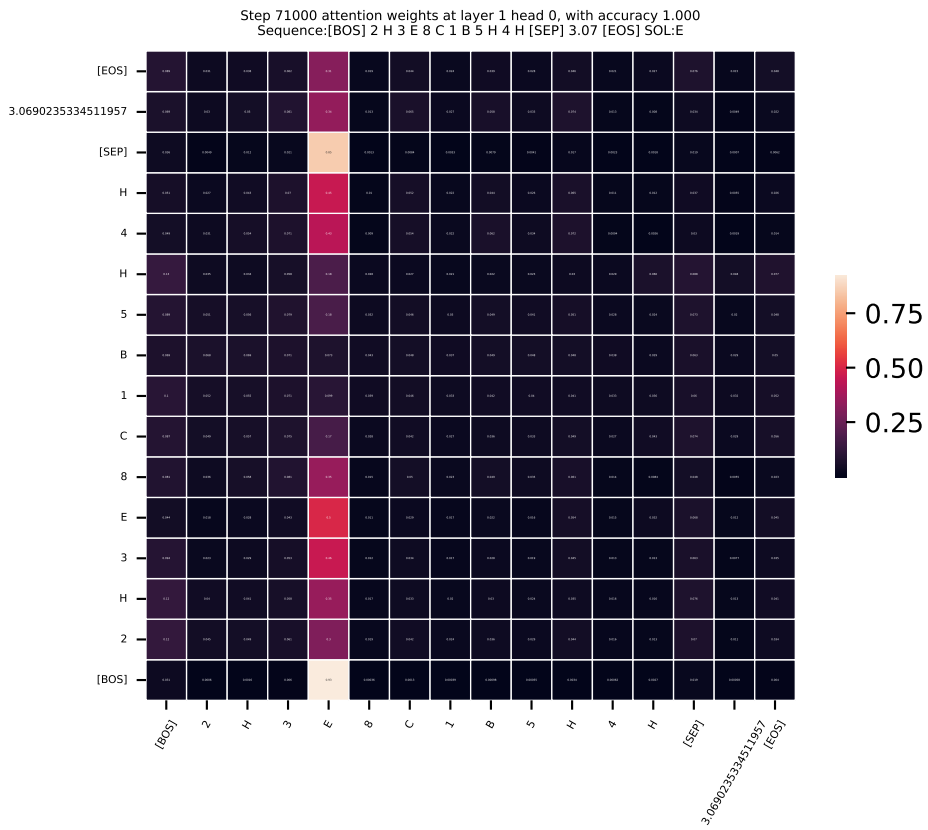

Step 71000 attention weights at layer 1 head 3, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

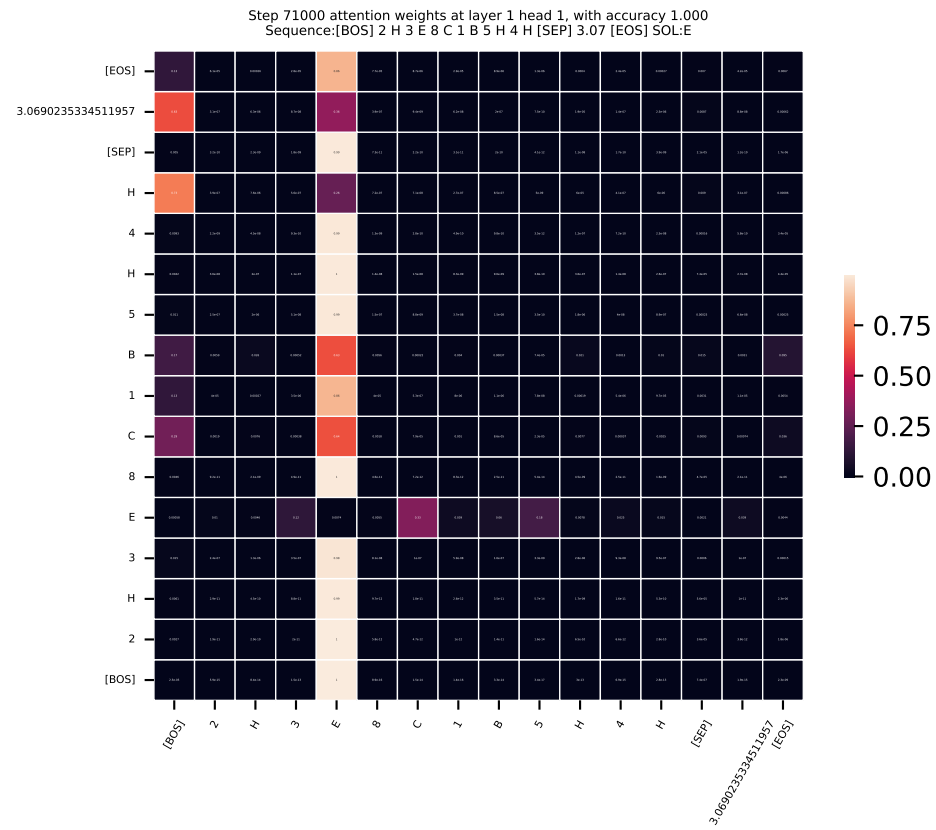

Step 71000 attention weights at layer 1 head 0, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

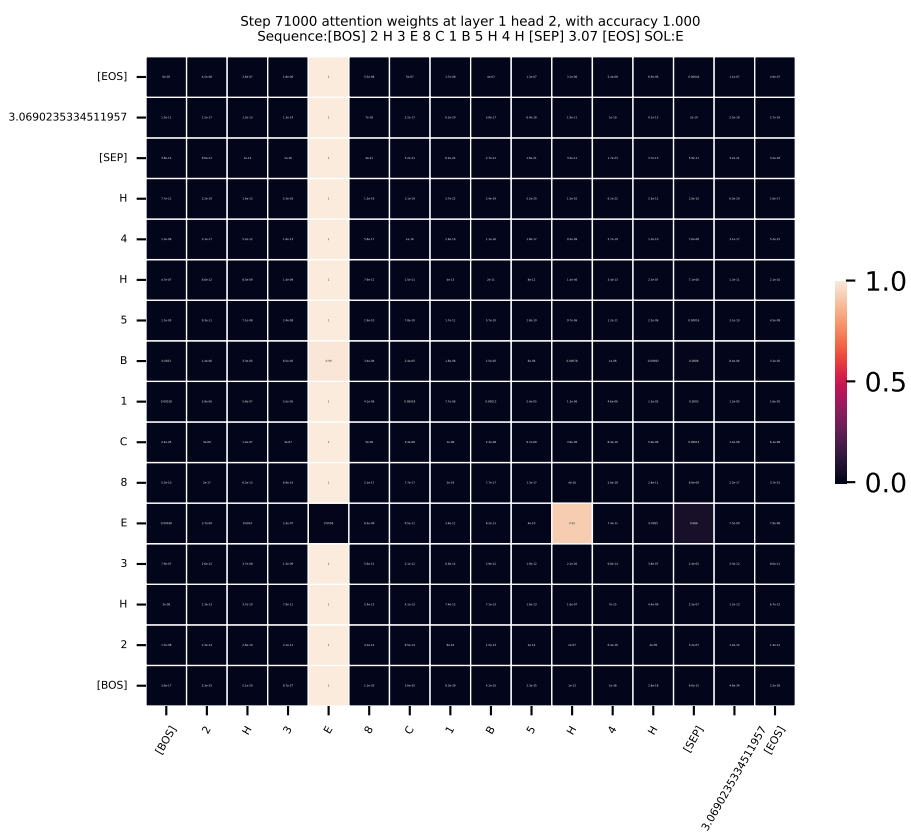

Step 71000 attention weights at layer 1 head 1, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

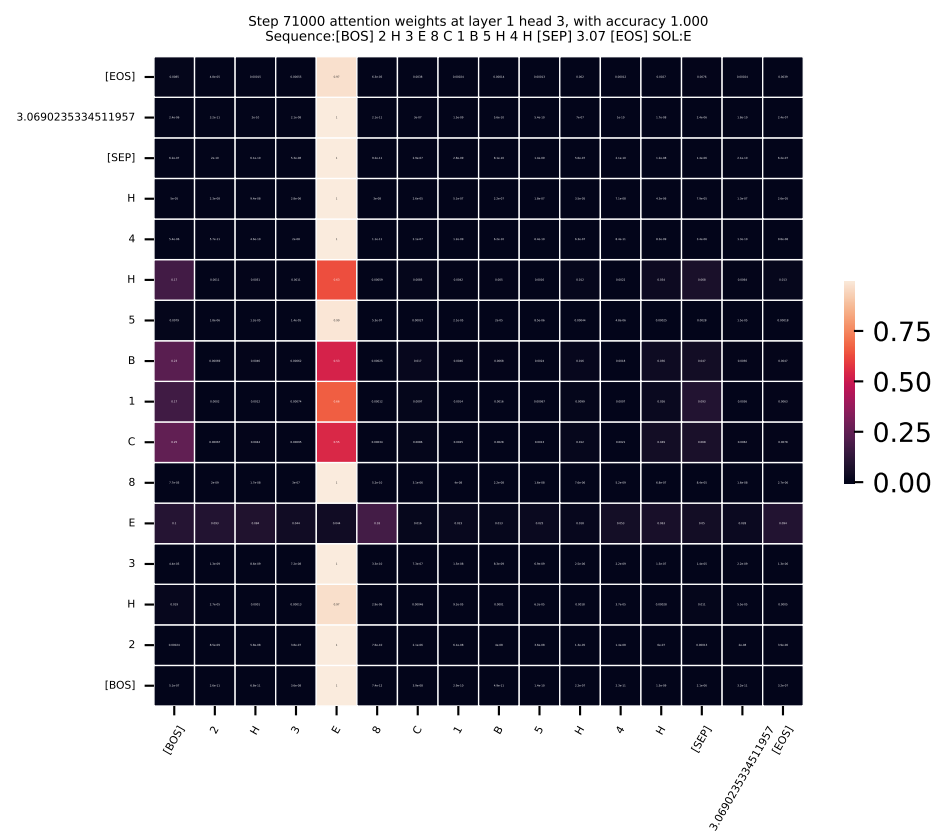

Step 71000 attention weights at layer 1 head 2, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0

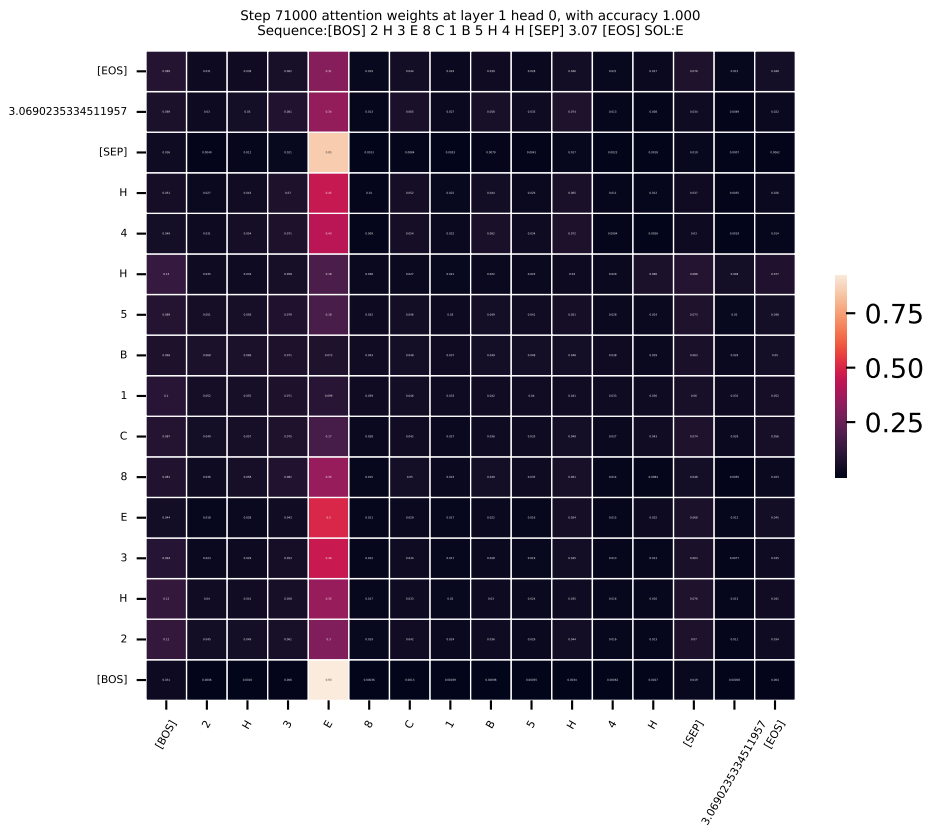

Step 71000 attention weights at layer 1 head 3, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

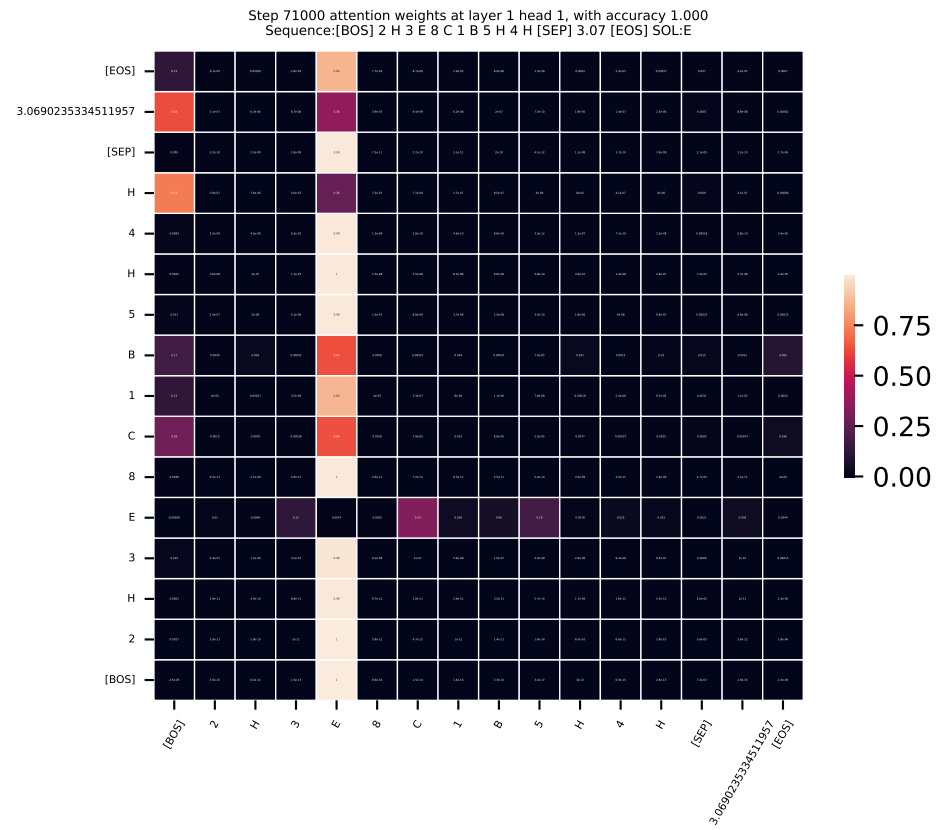

Step 71000 attention weights at layer 1 head 0, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25

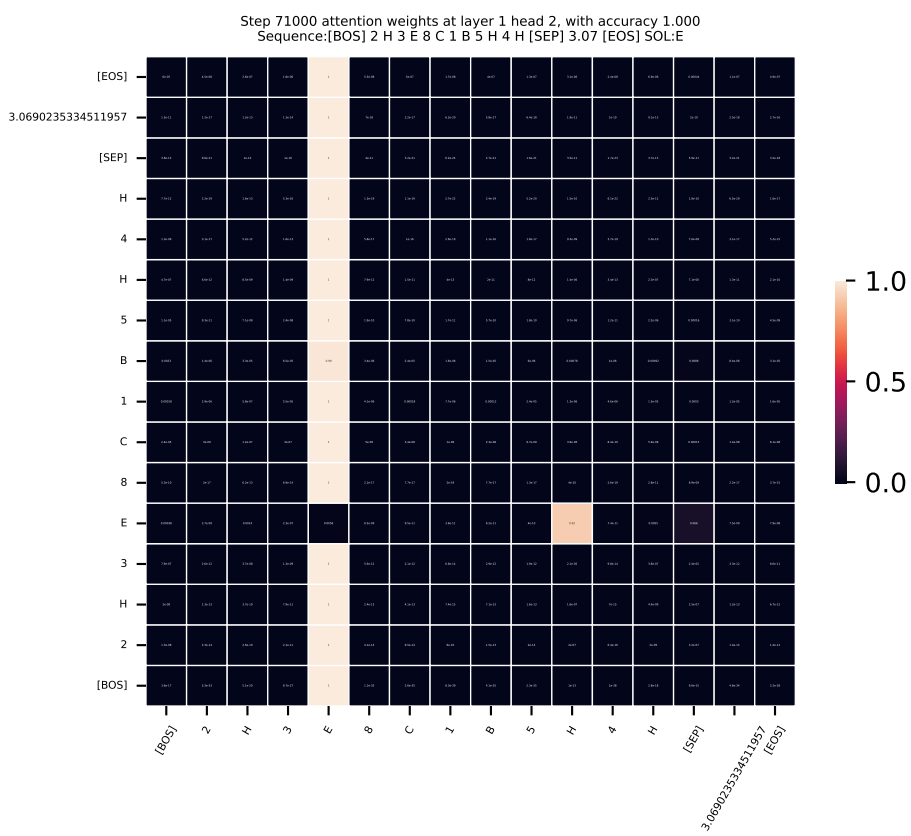

Step 71000 attention weights at layer 1 head 1, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00

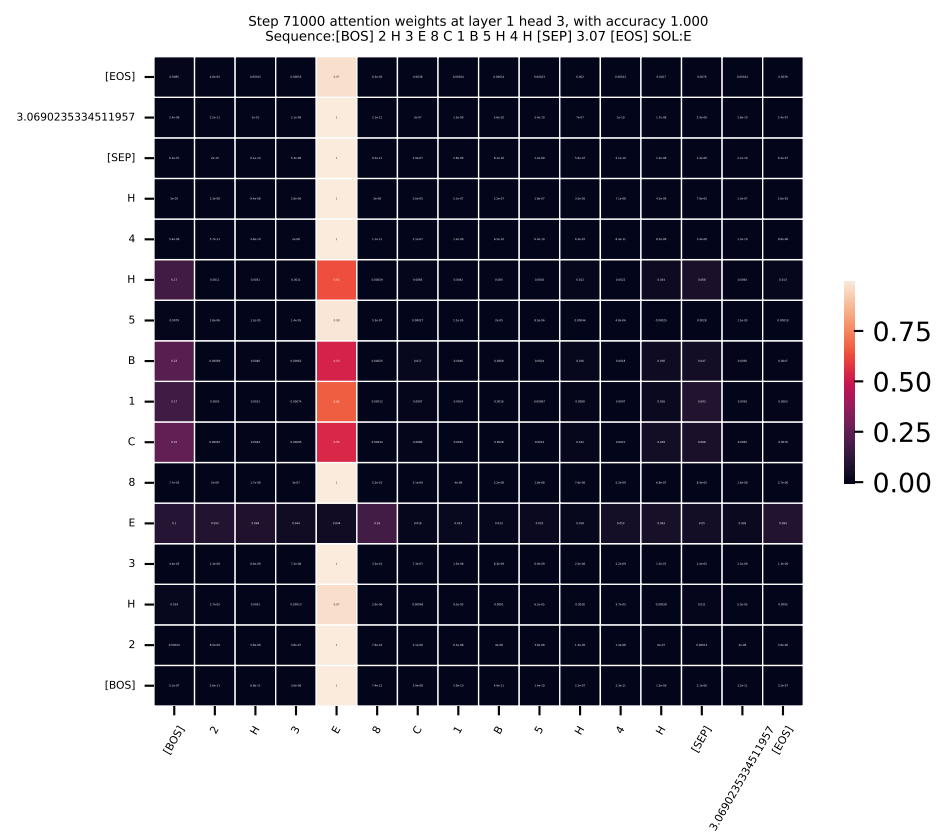

Step 71000 attention weights at layer 1 head 2, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0


Step 71000 attention weights at layer 1 head 3, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00


Step 71000 attention weights at layer 1 head 0, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25


Step 71000 attention weights at layer 1 head 1, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00


Step 71000 attention weights at layer 1 head 2, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
1.0
0.5
0.0


Step 71000 attention weights at layer 1 head 3, with accuracy 1.000
Sequence:[BOS] 2 H 3 E 8 C 1 B 5 H 4 H [SEP] 3.07 [EOS] SOL:E
[EOS]
3.0690235334511957
[SEP]
H
4
H
5
B
1
C
8
E
3
H
2
[BOS]
0.75
0.50
0.25
0.00