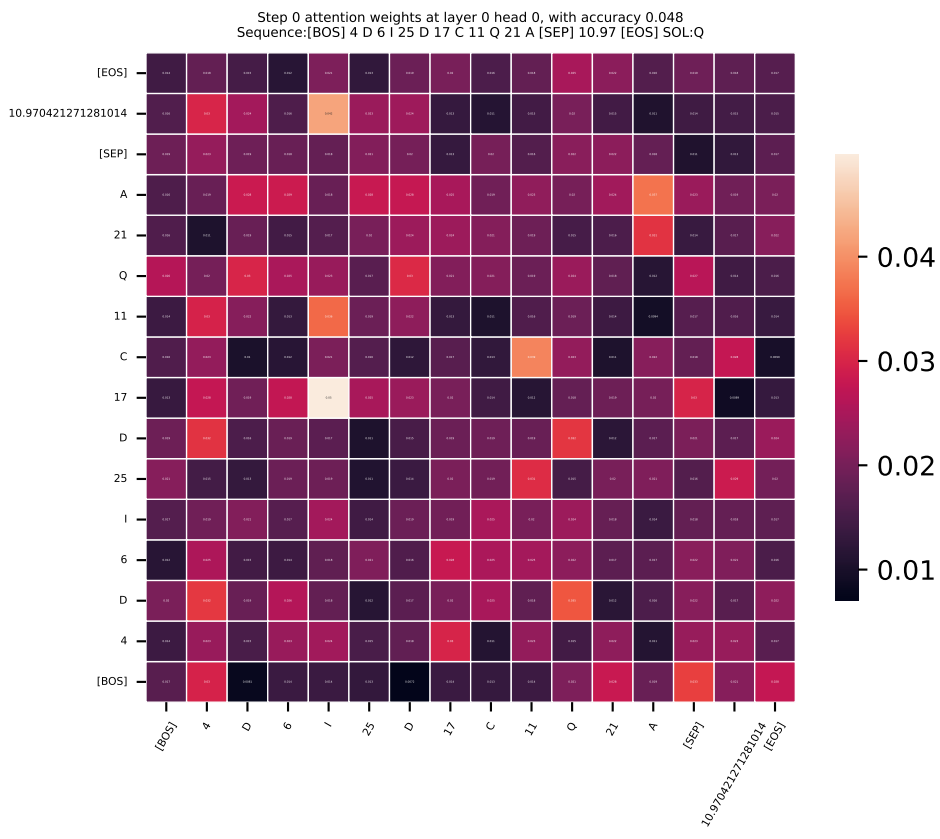
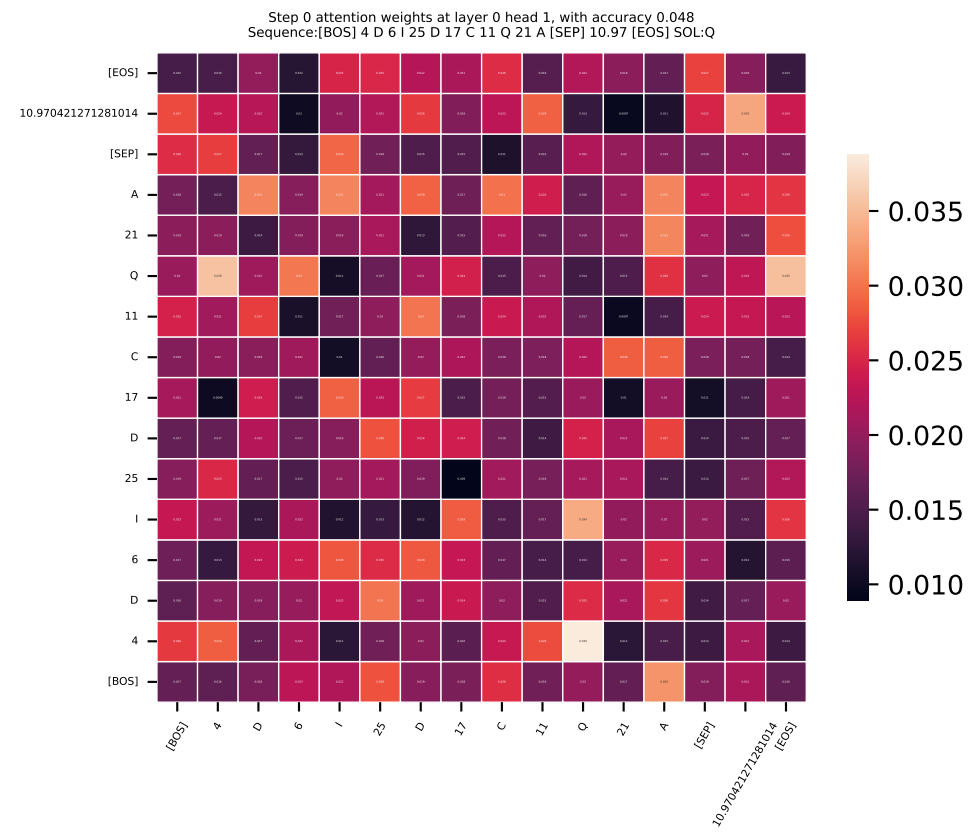
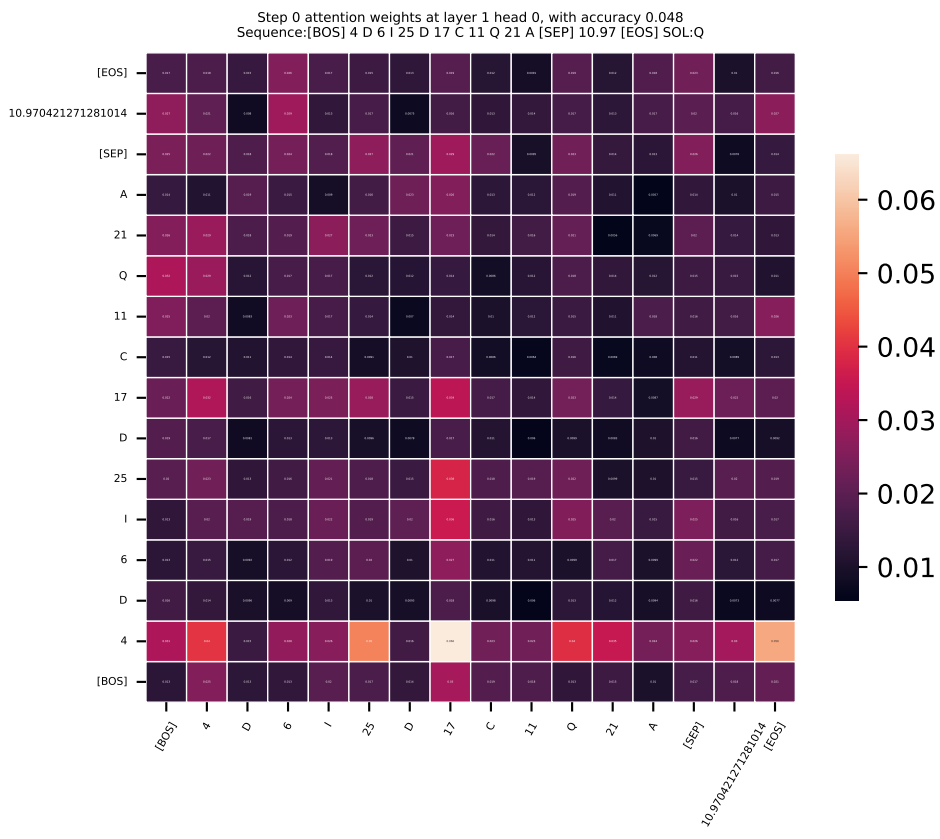
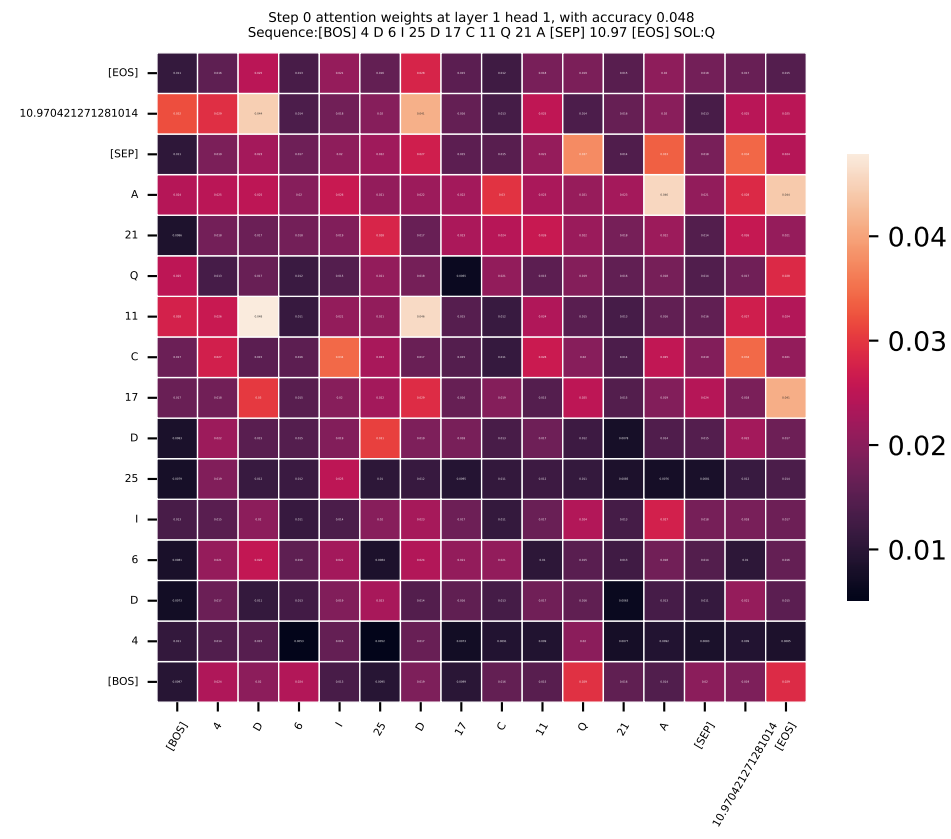
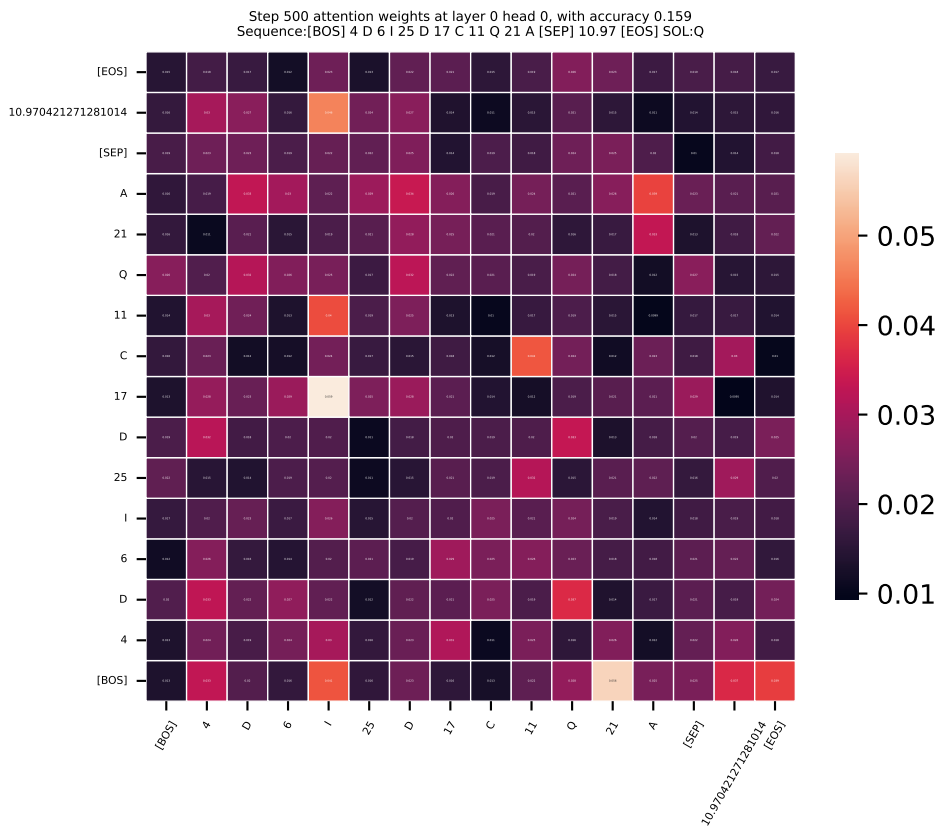
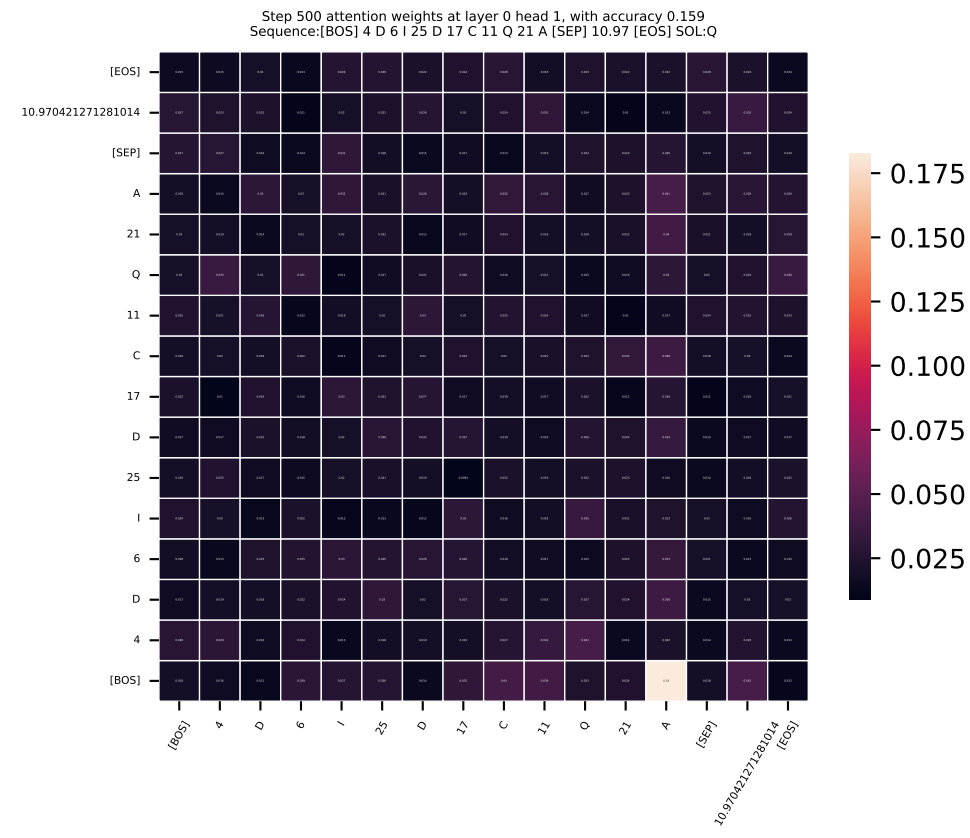

Step 0 attention weights at layer 0 head 0, with accuracy 0.048
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.04
0.03
0.02
0.01

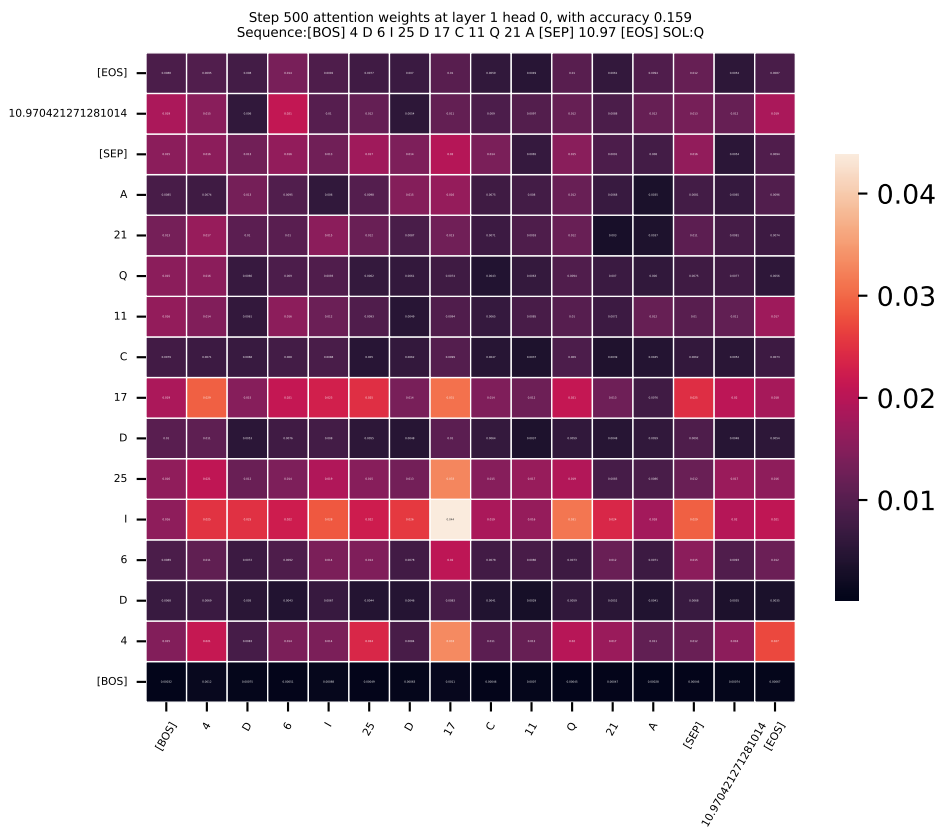

Step 0 attention weights at layer 0 head 1, with accuracy 0.048
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.035
0.030
0.025
0.020
0.015
0.010

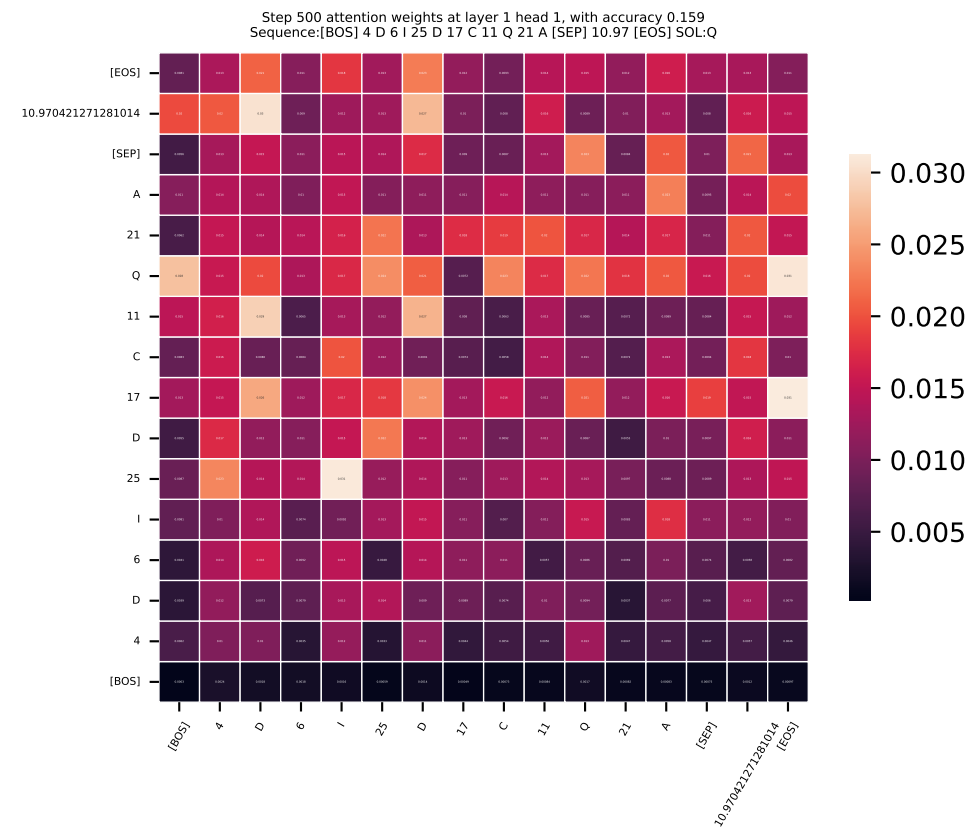

Step 0 attention weights at layer 1 head 0, with accuracy 0.048
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.06
0.05
0.04
0.03
0.02
0.01

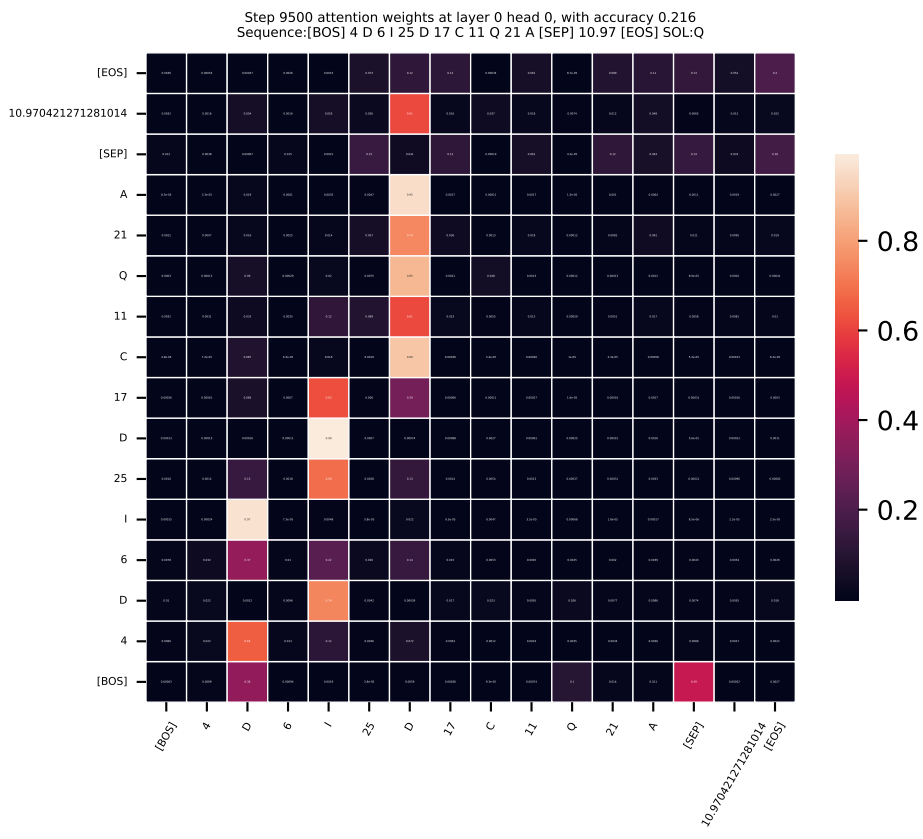

Step 0 attention weights at layer 1 head 1, with accuracy 0.048
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.04
0.03
0.02
0.01

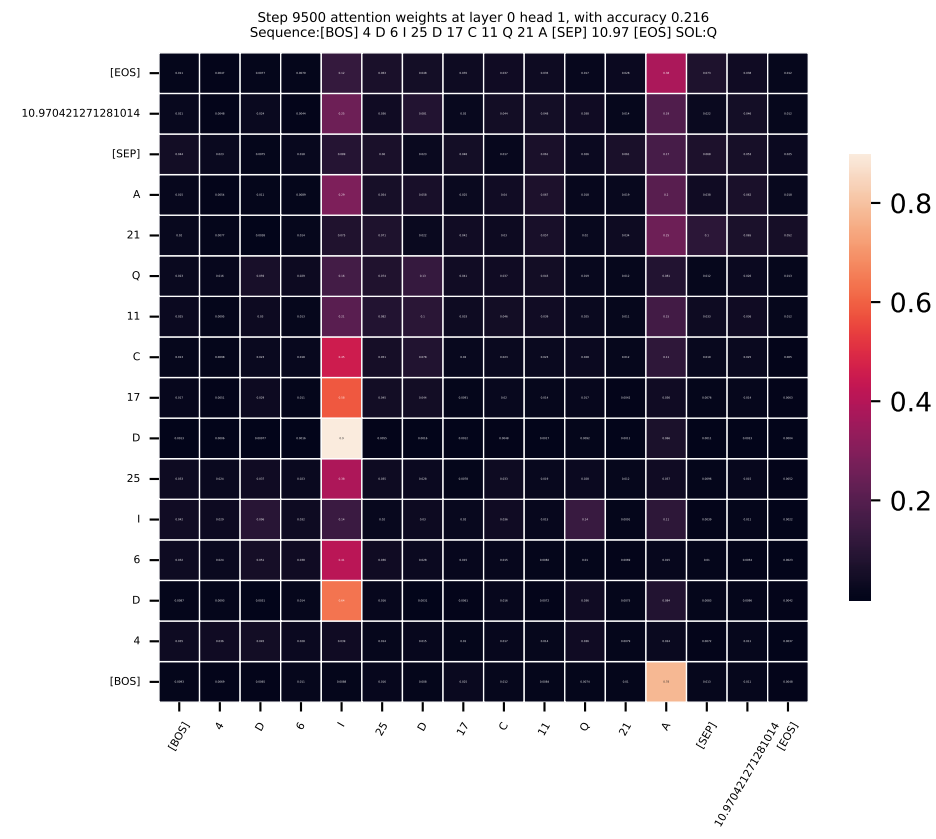

Step 500 attention weights at layer 0 head 0, with accuracy 0.159
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.05
0.04
0.03
0.02
0.01

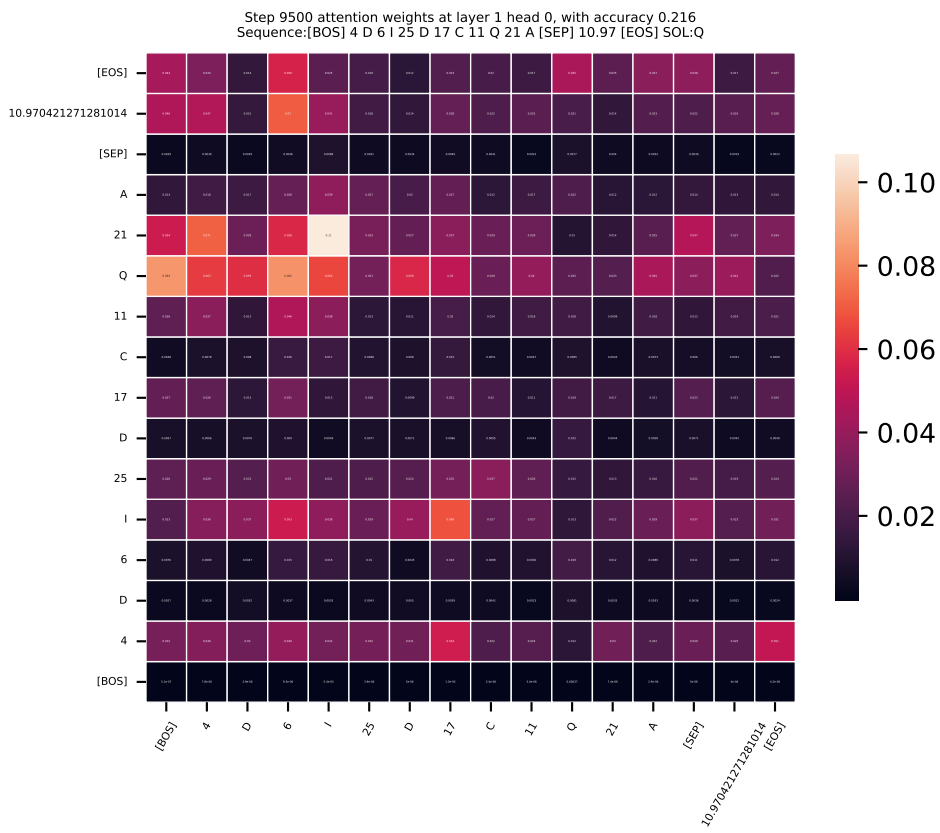

Step 500 attention weights at layer 0 head 1, with accuracy 0.159
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.175
0.150
0.125
0.100
0.075
0.050
0.025

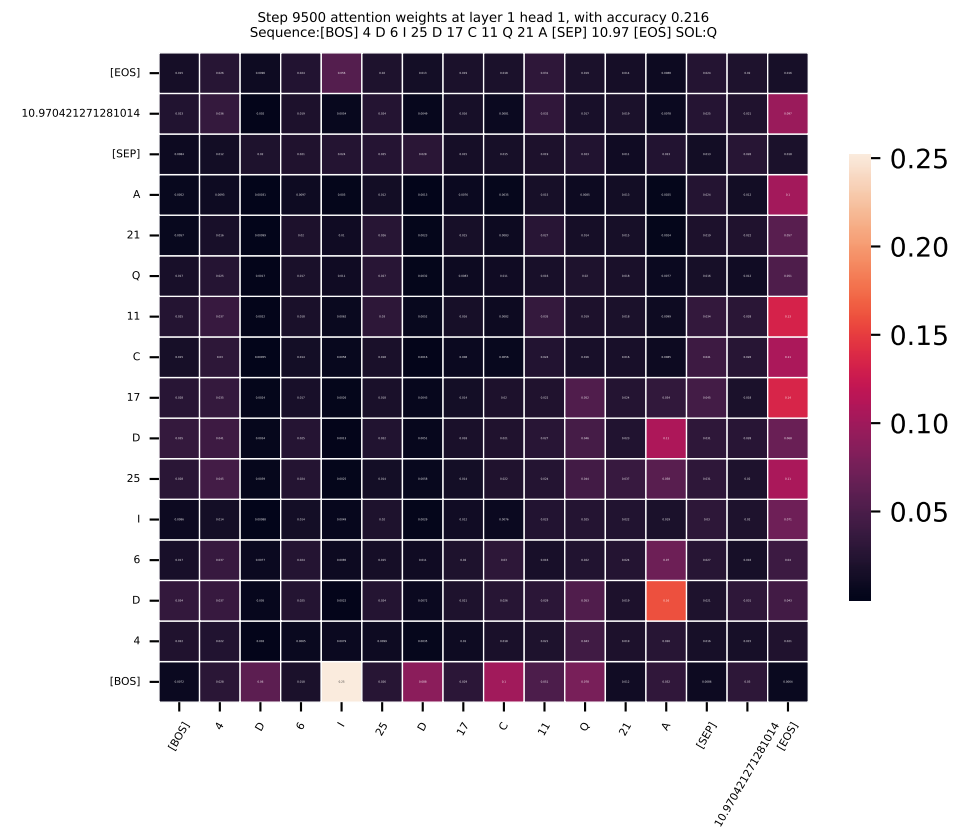

Step 500 attention weights at layer 1 head 0, with accuracy 0.159
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.04
0.03
0.02
0.01

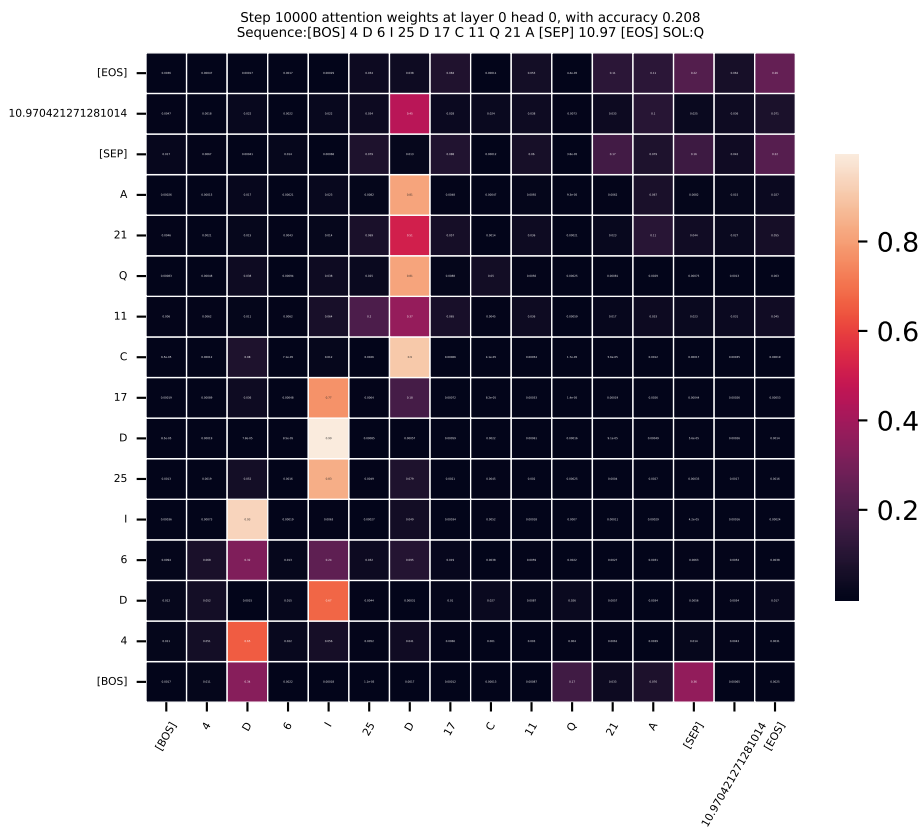

Step 500 attention weights at layer 1 head 1, with accuracy 0.159
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.030
0.025
0.020
0.015
0.010
0.005

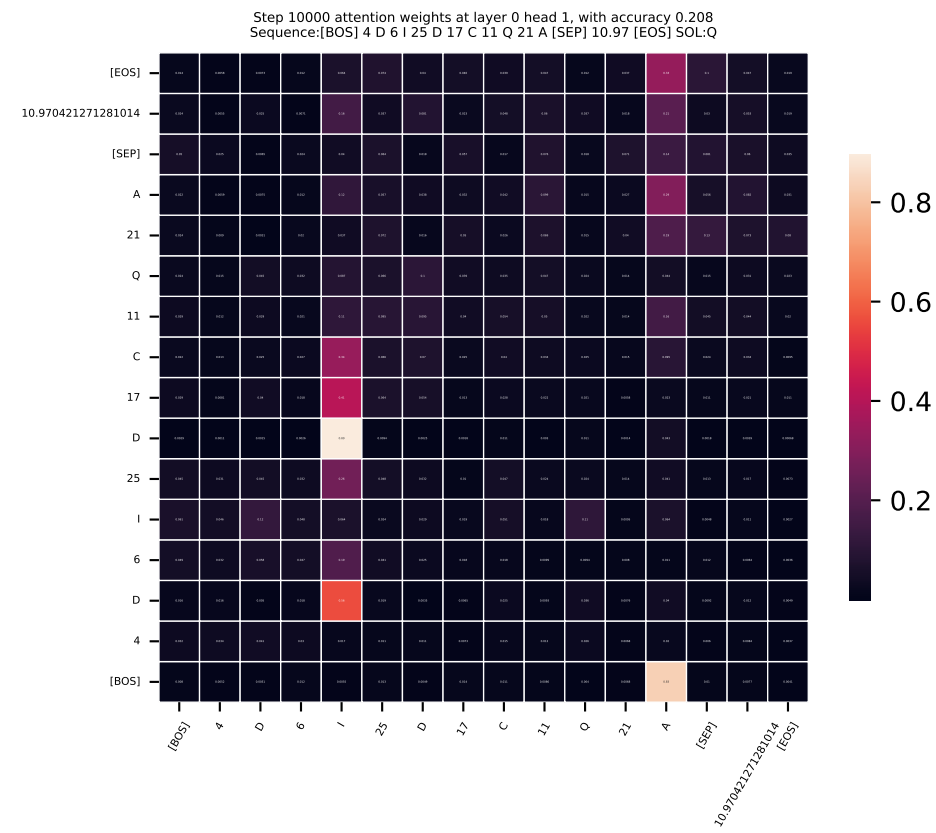

Step 9500 attention weights at layer 0 head 0, with accuracy 0.216
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.8
0.6
0.4
0.2

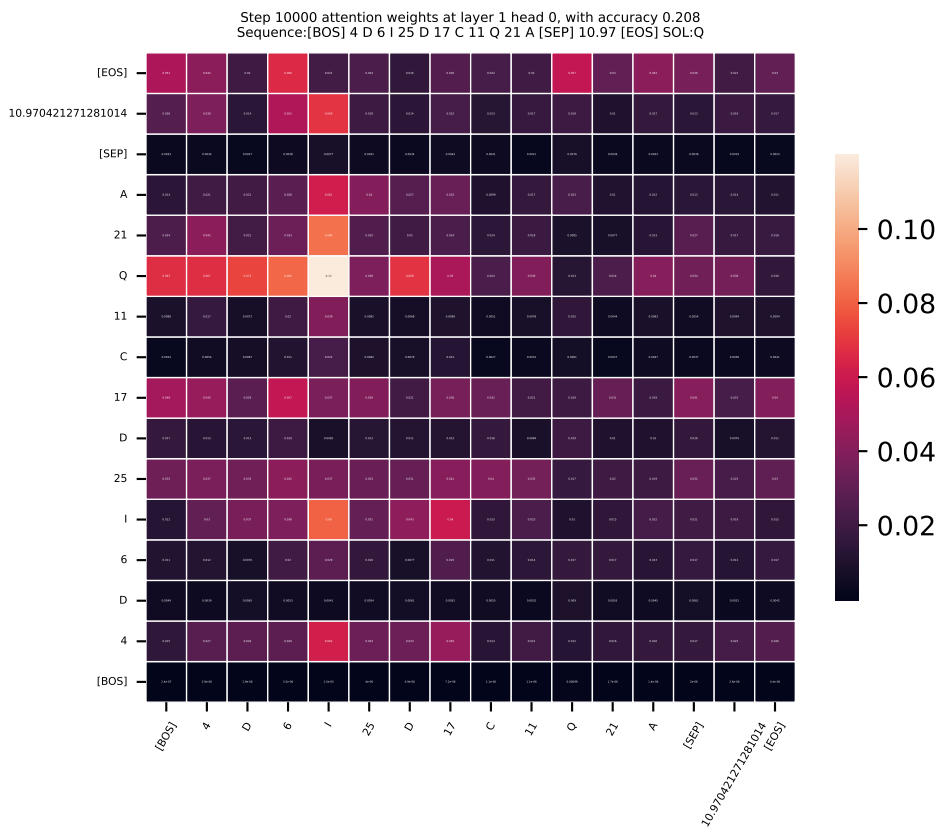

Step 9500 attention weights at layer 0 head 1, with accuracy 0.216
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.8
0.6
0.4
0.2

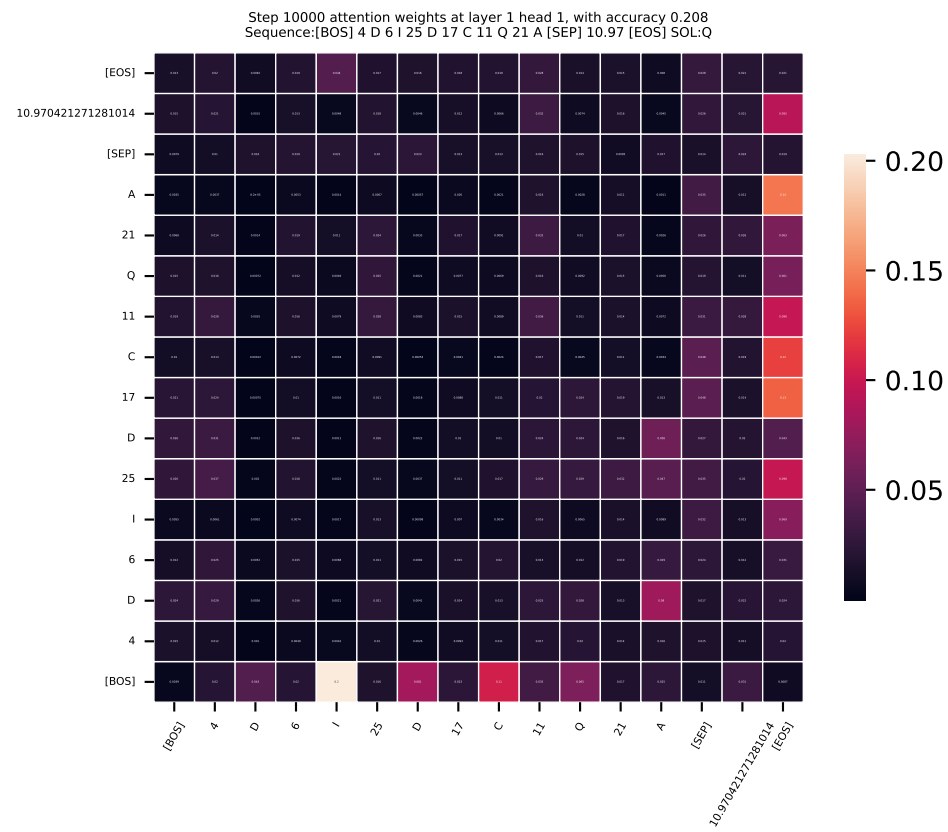

Step 9500 attention weights at layer 1 head 0, with accuracy 0.216
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS]
4
D
6
I
25
D
17
C
11
Q
21
A
[SEP]
10.970421271281014
[EOS]
0.10
0.08
0.06
0.04
0.02

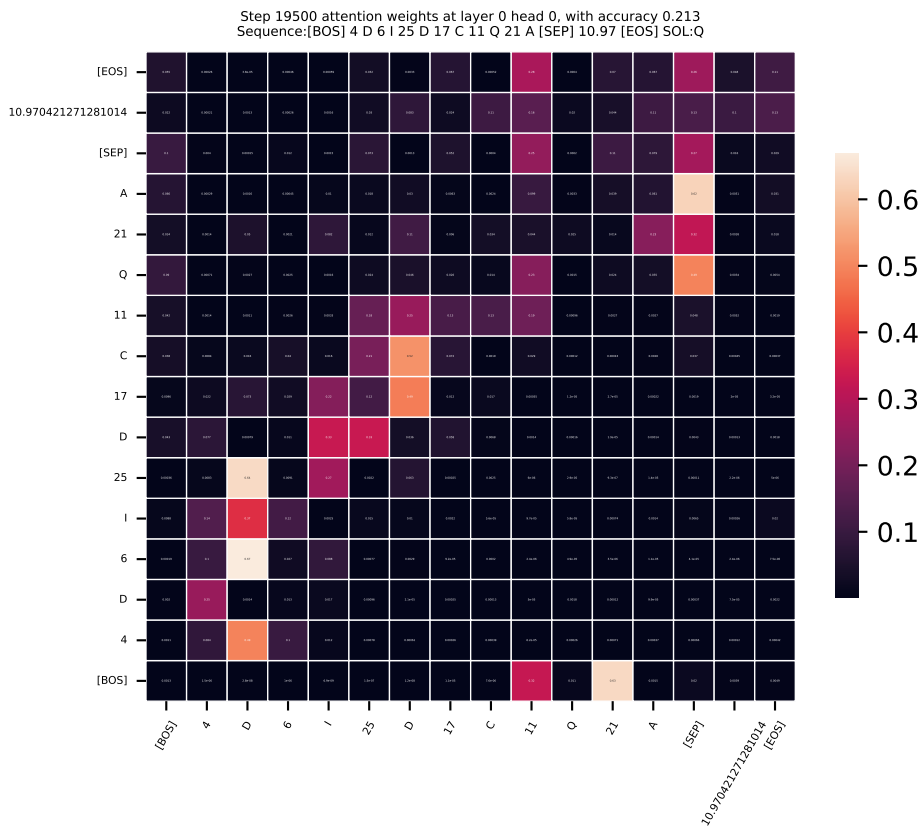

Step 9500 attention weights at layer 1 head 1, with accuracy 0.216
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS]
4
D
6
I
25
D
17
C
11
Q
21
A
[SEP]
10.970421271281014
[EOS]
0.25
0.20
0.15
0.10
0.05

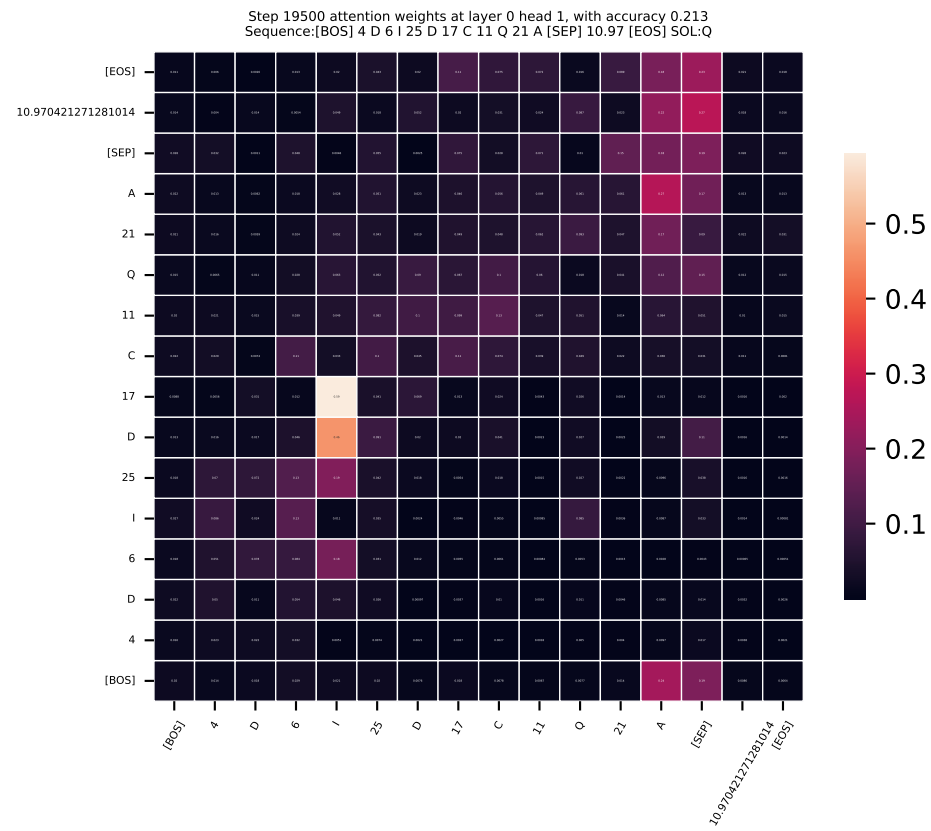

Step 10000 attention weights at layer 0 head 0, with accuracy 0.208
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.8
0.6
0.4
0.2

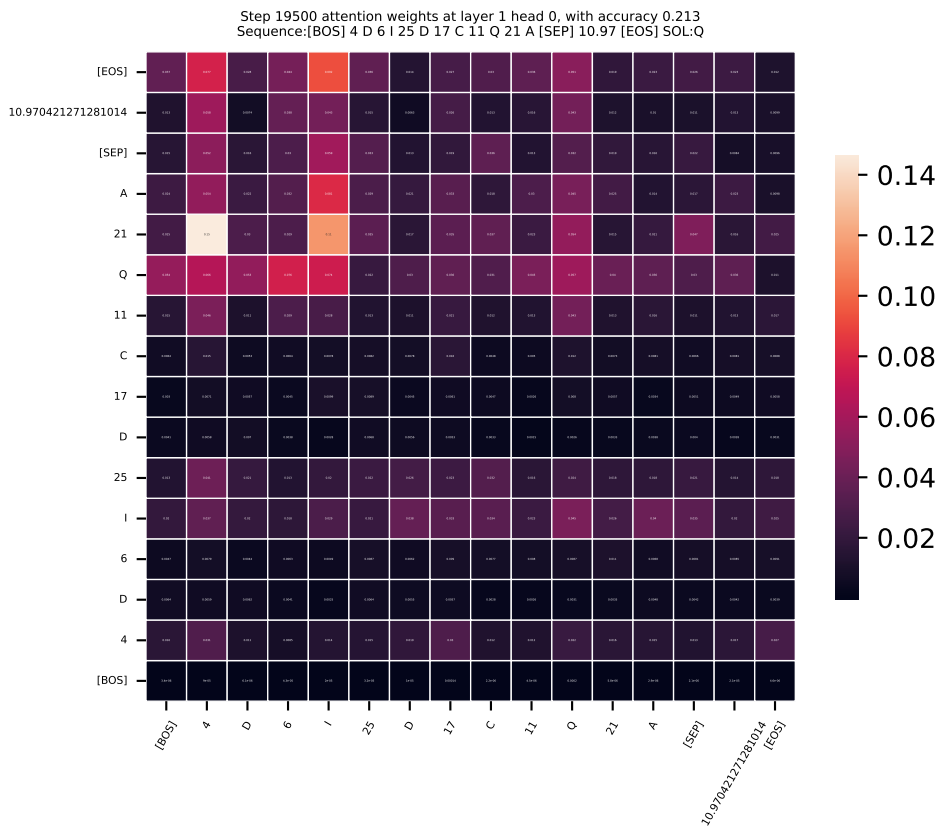

Step 10000 attention weights at layer 0 head 1, with accuracy 0.208
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.8
0.6
0.4
0.2

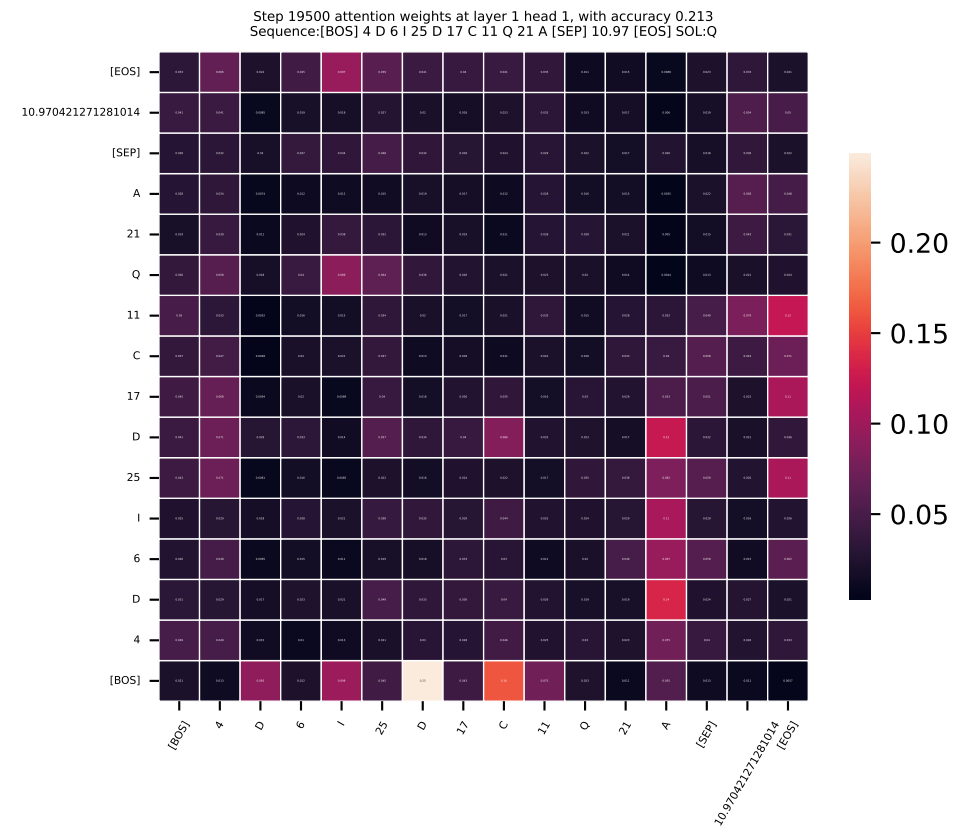

Step 10000 attention weights at layer 1 head 0, with accuracy 0.208
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS]
4
D
6
I
25
D
17
C
11
Q
21
A
[SEP]
10.970421271281014
[EOS]
0.10
0.08
0.06
0.04
0.02

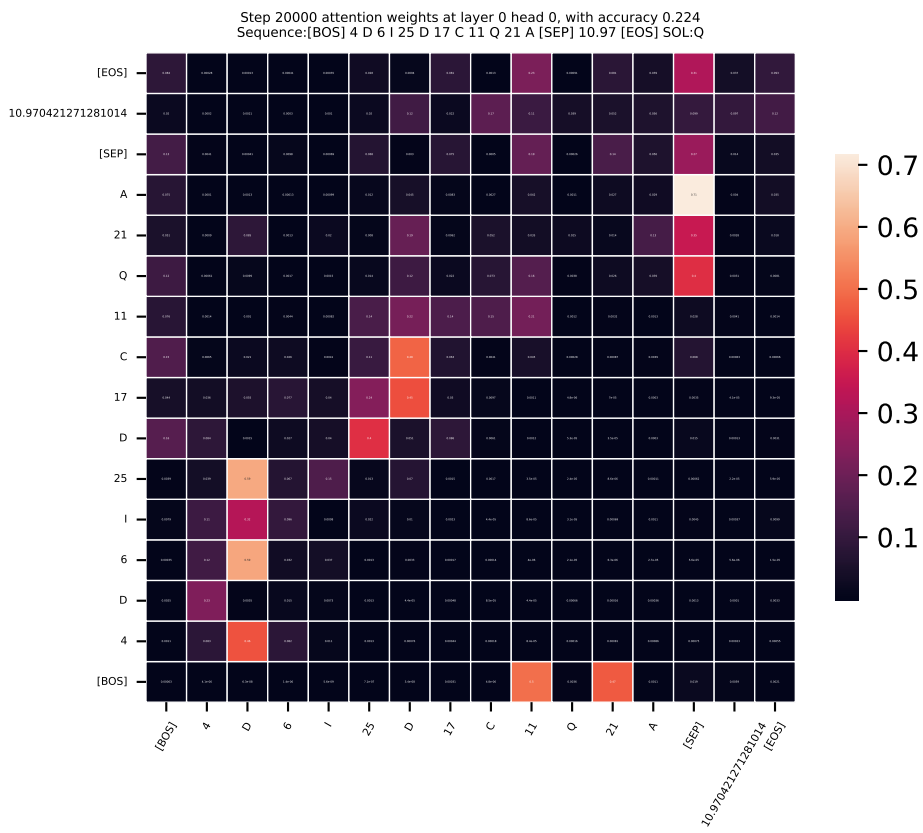

Step 10000 attention weights at layer 1 head 1, with accuracy 0.208
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS]
4
D
6
I
25
D
17
C
11
Q
21
A
[SEP]
10.970421271281014
[EOS]
0.20
0.15
0.10
0.05

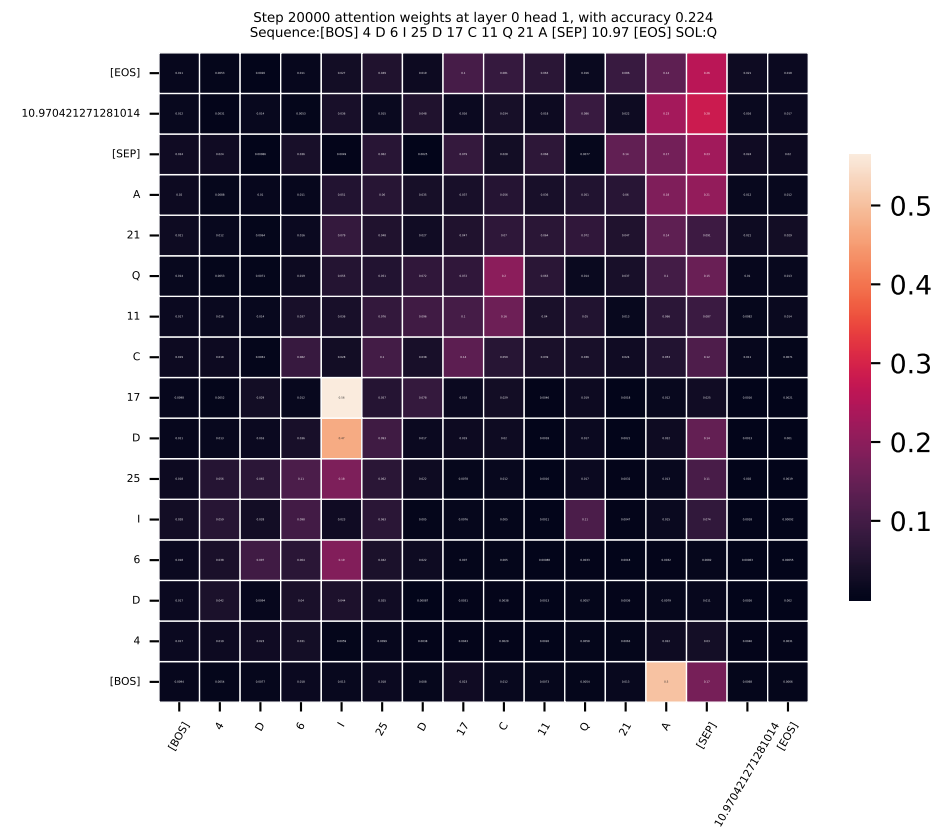

Step 19500 attention weights at layer 0 head 0, with accuracy 0.213
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.6
0.5
0.4
0.3
0.2
0.1

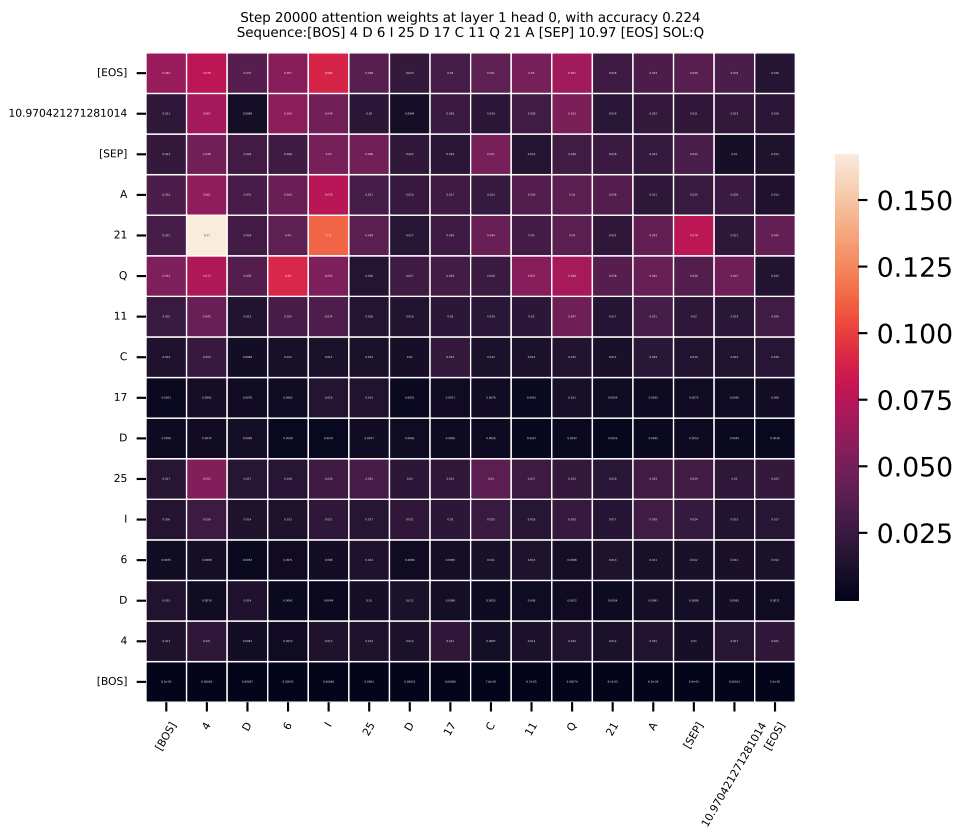

Step 19500 attention weights at layer 0 head 1, with accuracy 0.213
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.5
0.4
0.3
0.2
0.1

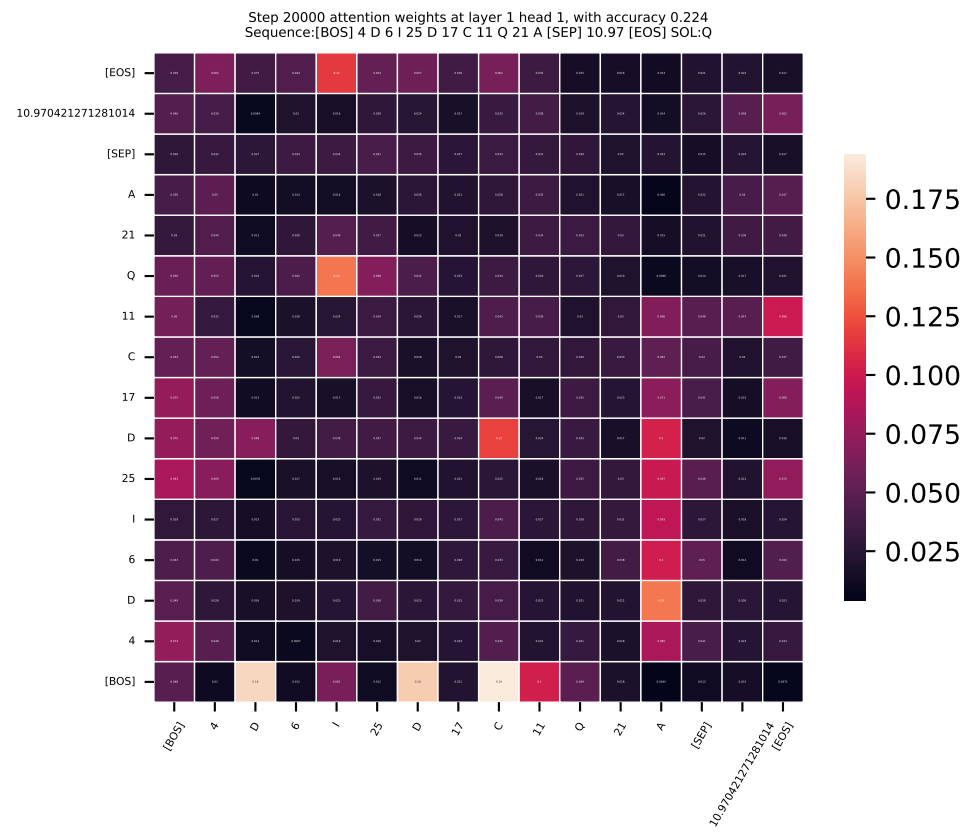

Step 19500 attention weights at layer 1 head 0, with accuracy 0.213
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS]
4
D
6
I
25
D
17
C
11
Q
21
A
[SEP]
10.970421271281014
[EOS]
0.14
0.12
0.10
0.08
0.06
0.04
0.02

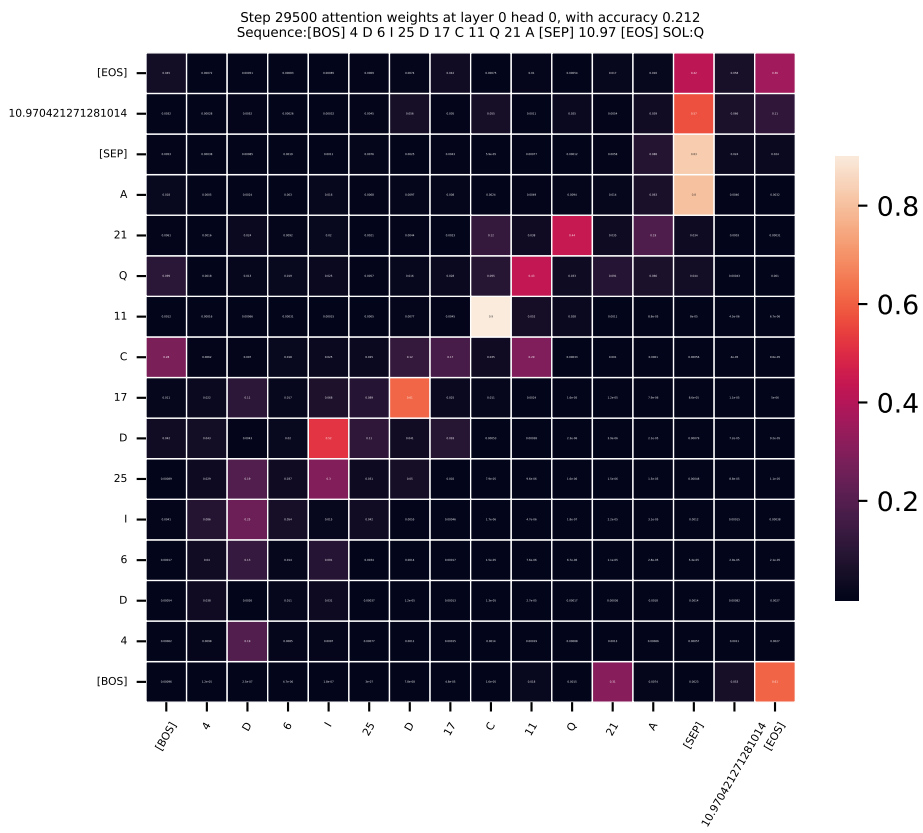

Step 19500 attention weights at layer 1 head 1, with accuracy 0.213
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS]
4
D
6
I
25
D
17
C
11
Q
21
A
[SEP]
10.970421271281014
[EOS]
0.20
0.15
0.10
0.05

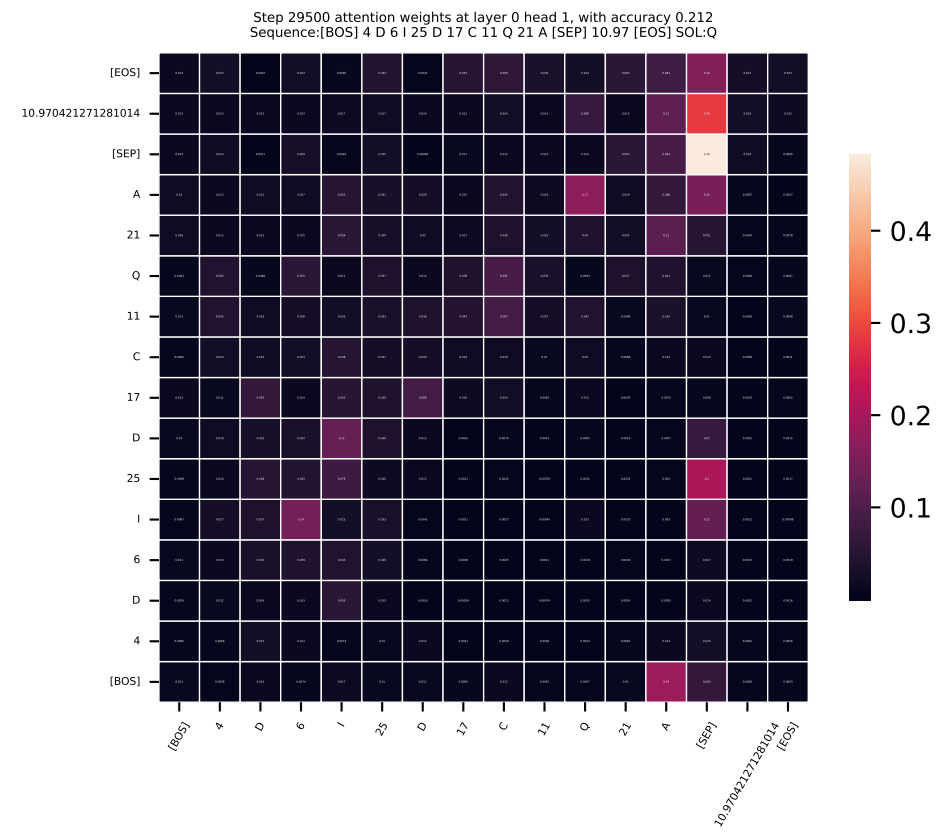

Step 20000 attention weights at layer 0 head 0, with accuracy 0.224
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.7
0.6
0.5
0.4
0.3
0.2
0.1

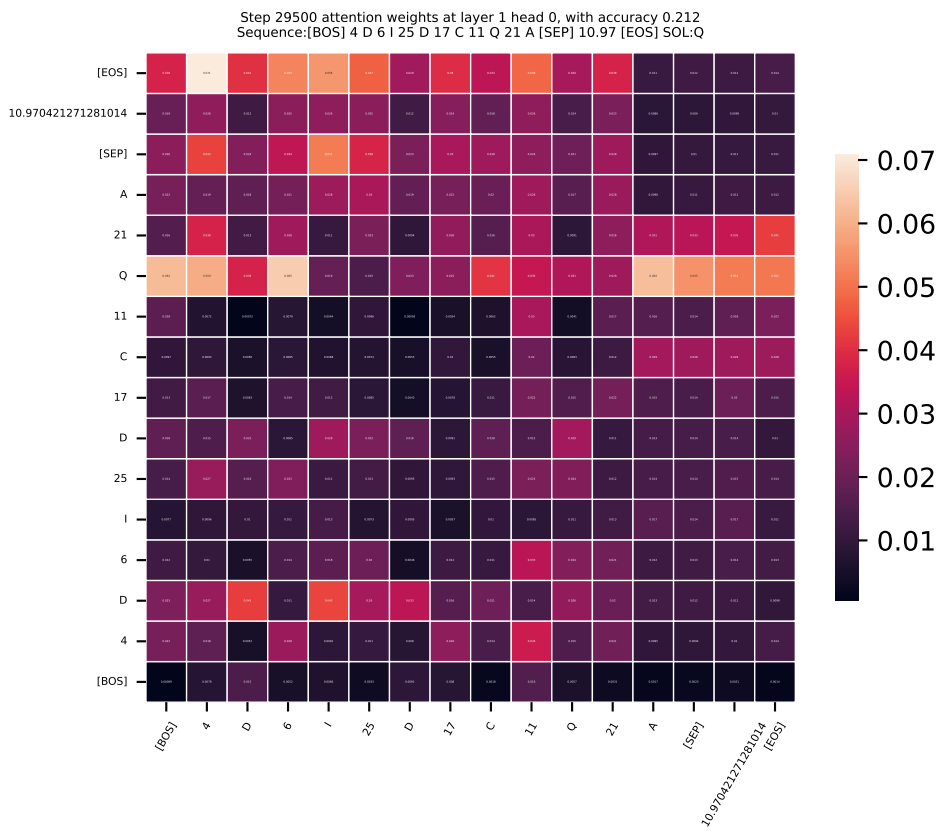

Step 20000 attention weights at layer 0 head 1, with accuracy 0.224
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.5
0.4
0.3
0.2
0.1

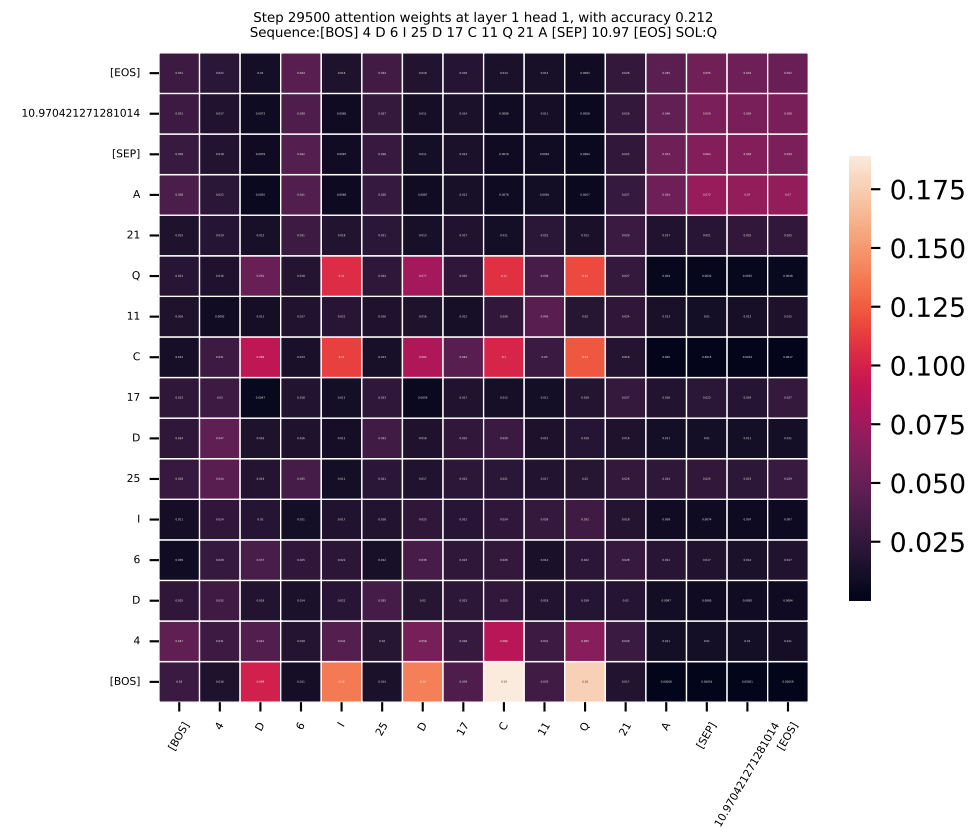

Step 20000 attention weights at layer 1 head 0, with accuracy 0.224
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.150
0.125
0.100
0.075
0.050
0.025

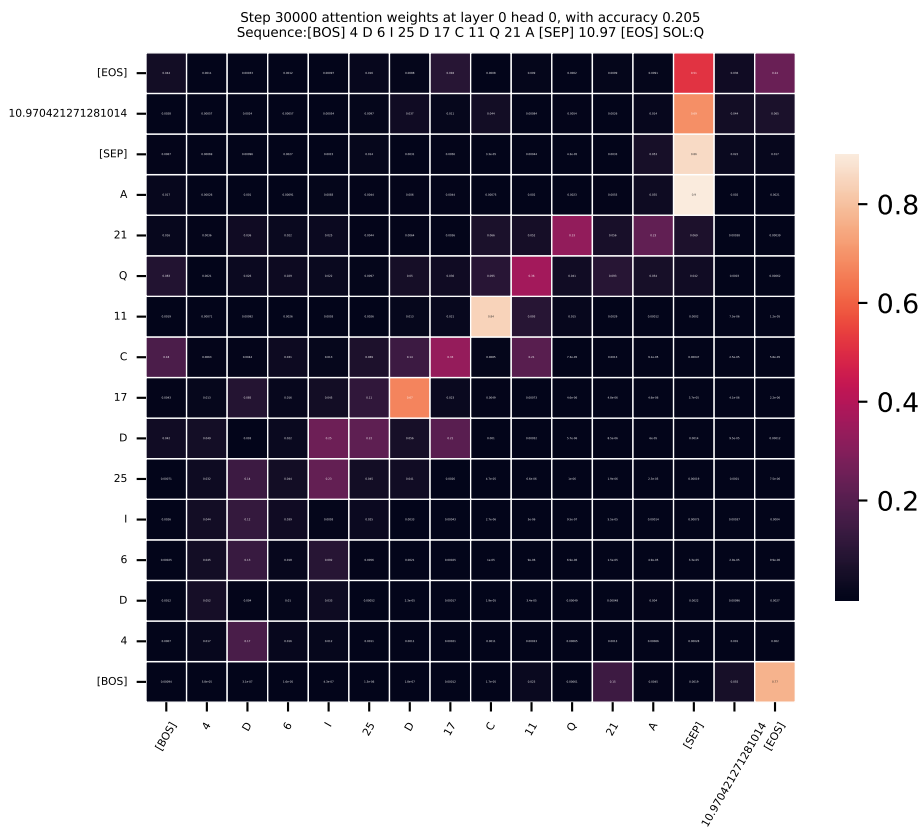

Step 20000 attention weights at layer 1 head 1, with accuracy 0.224
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.175
0.150
0.125
0.100
0.075
0.050
0.025

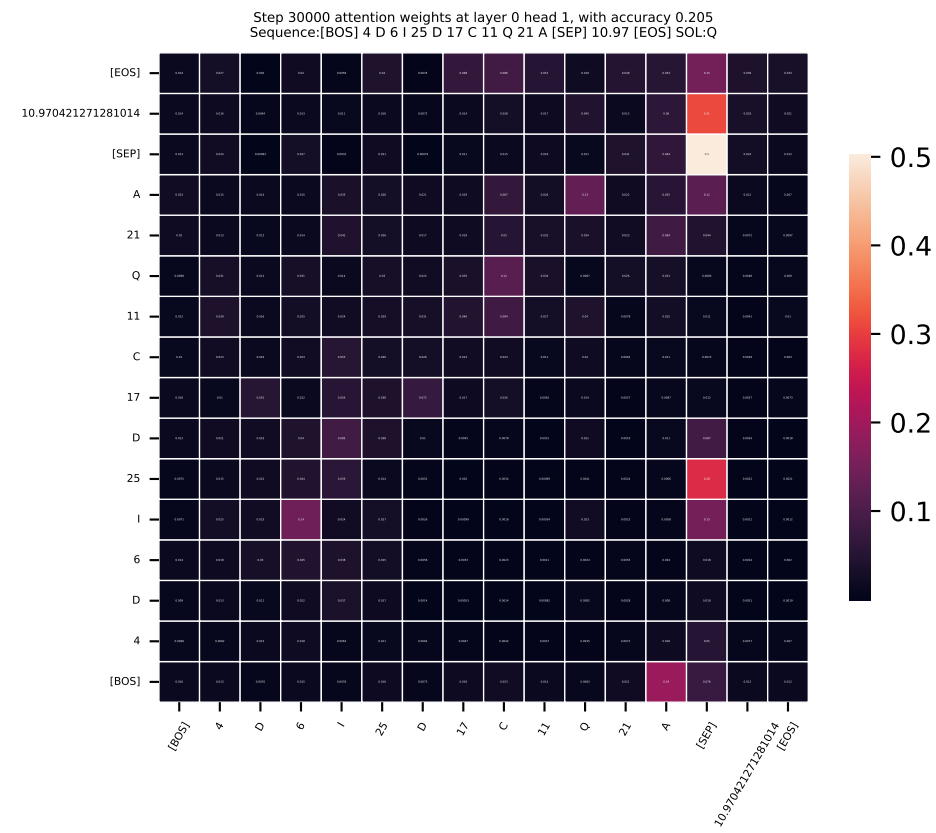

Step 29500 attention weights at layer 0 head 0, with accuracy 0.212
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS]
4
D
6
I
25
D
17
C
11
Q
21
A
[SEP]
10.970421271281014
[EOS]
0.8
0.6
0.4
0.2

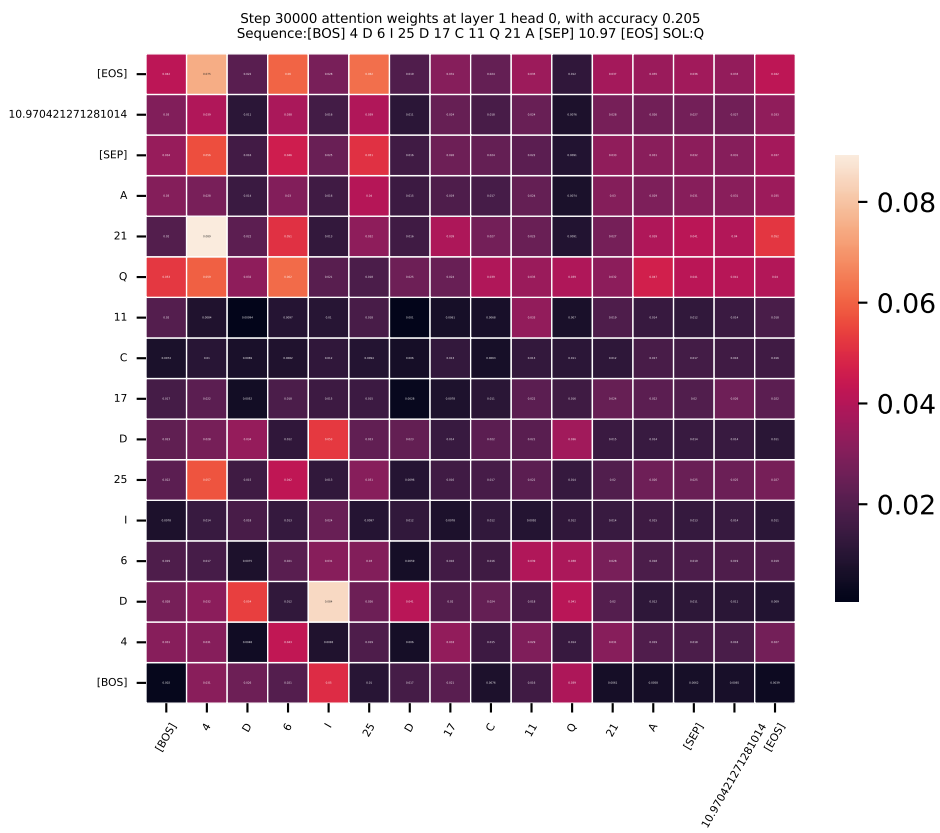

Step 29500 attention weights at layer 0 head 1, with accuracy 0.212
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS]
4
D
6
I
25
D
17
C
11
Q
21
A
[SEP]
10.970421271281014
[EOS]
0.4
0.3
0.2
0.1

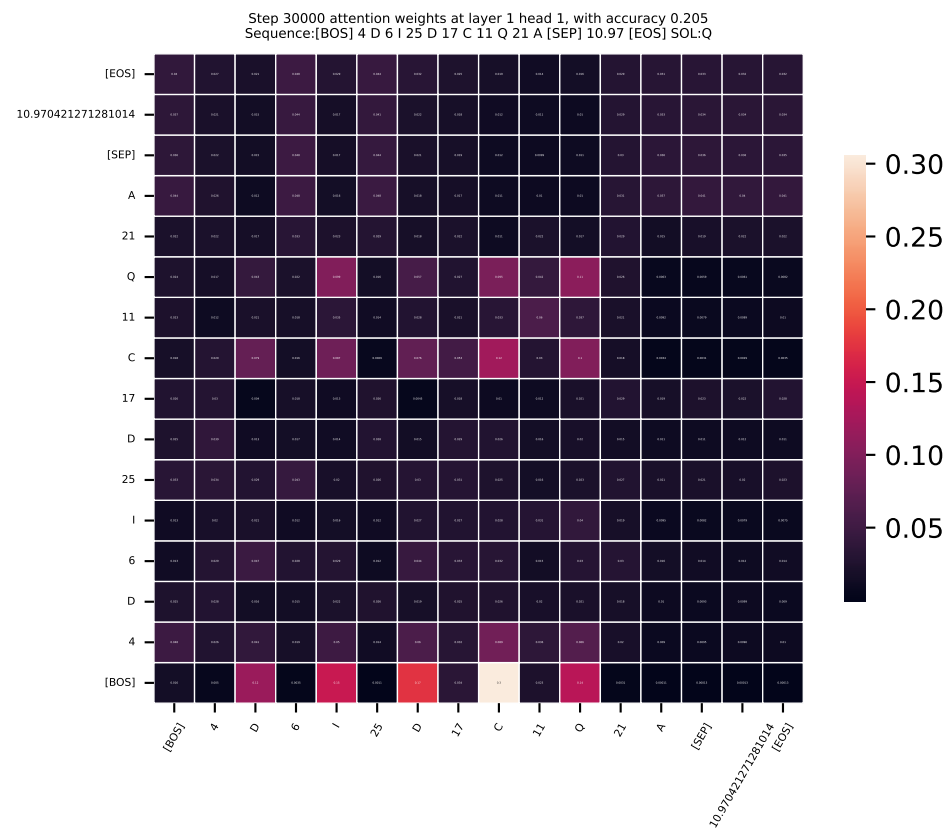

Step 29500 attention weights at layer 1 head 0, with accuracy 0.212
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.07
0.06
0.05
0.04
0.03
0.02
0.01

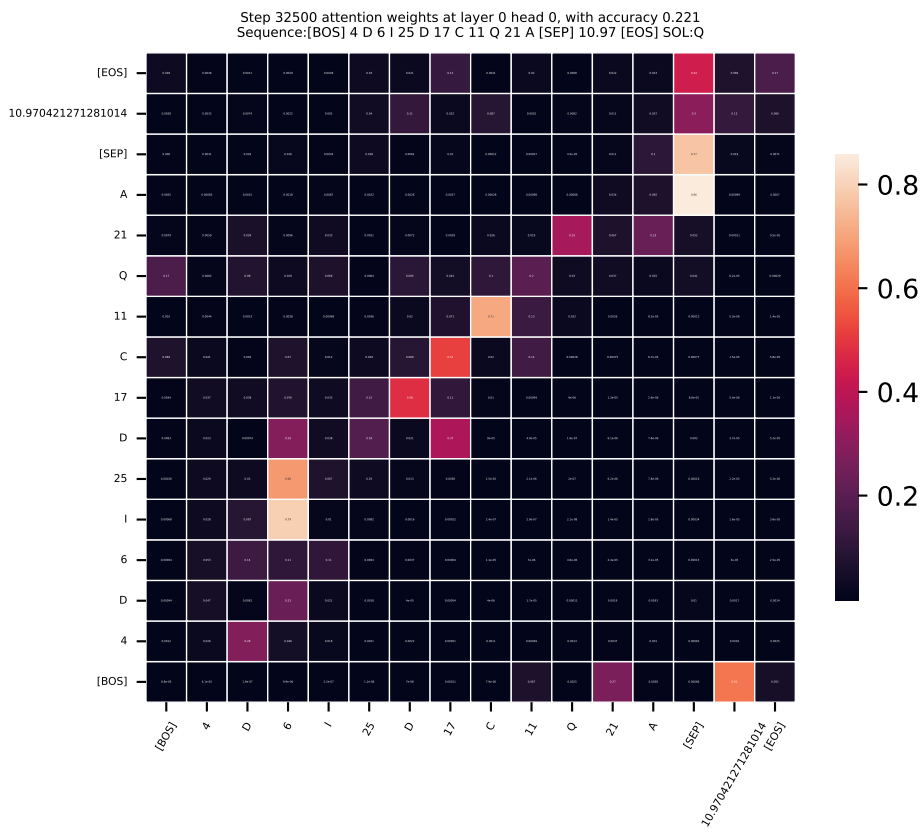

Step 29500 attention weights at layer 1 head 1, with accuracy 0.212
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.175
0.150
0.125
0.100
0.075
0.050
0.025

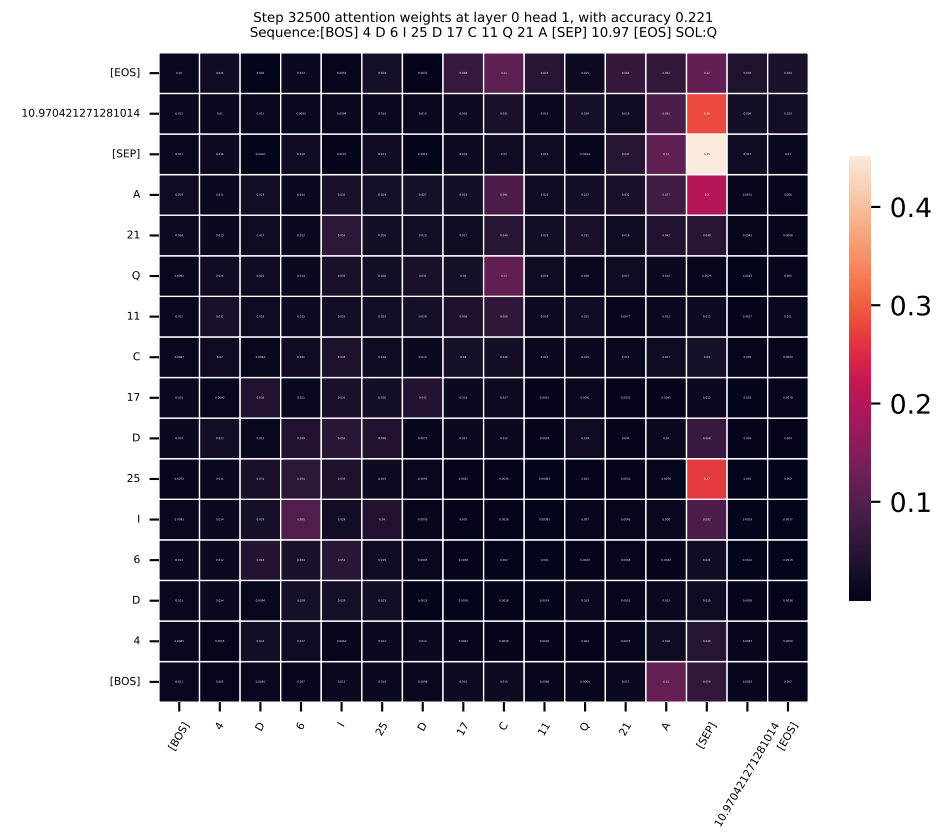

Step 30000 attention weights at layer 0 head 0, with accuracy 0.205
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.8
0.6
0.4
0.2

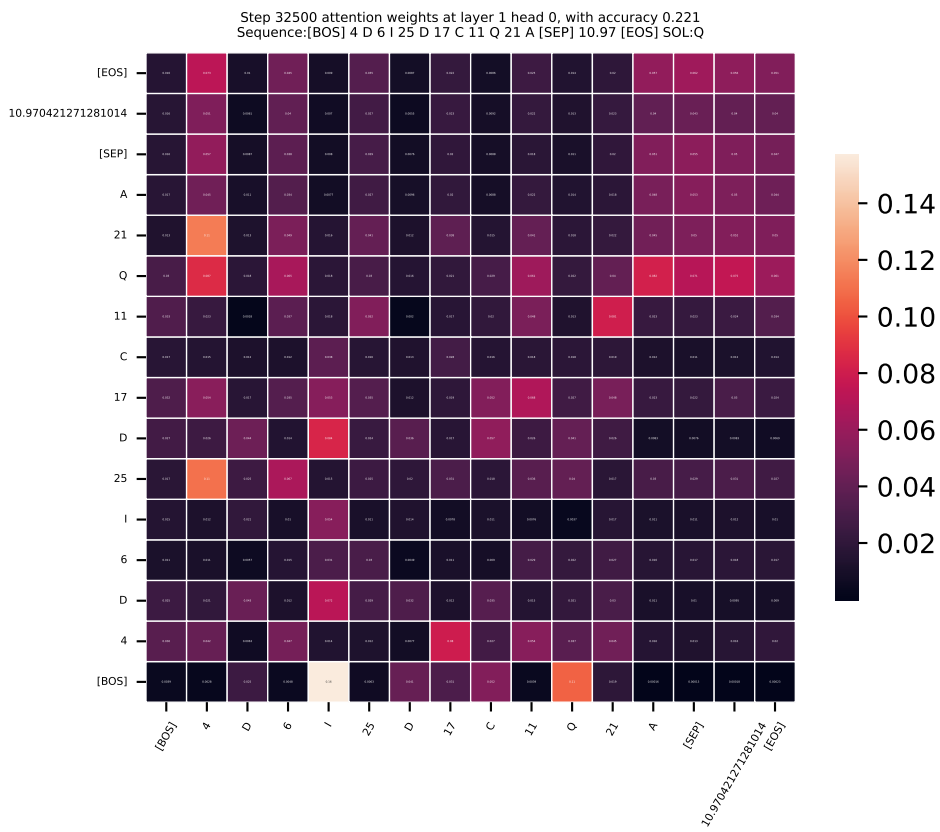

Step 30000 attention weights at layer 0 head 1, with accuracy 0.205
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.5
0.4
0.3
0.2
0.1

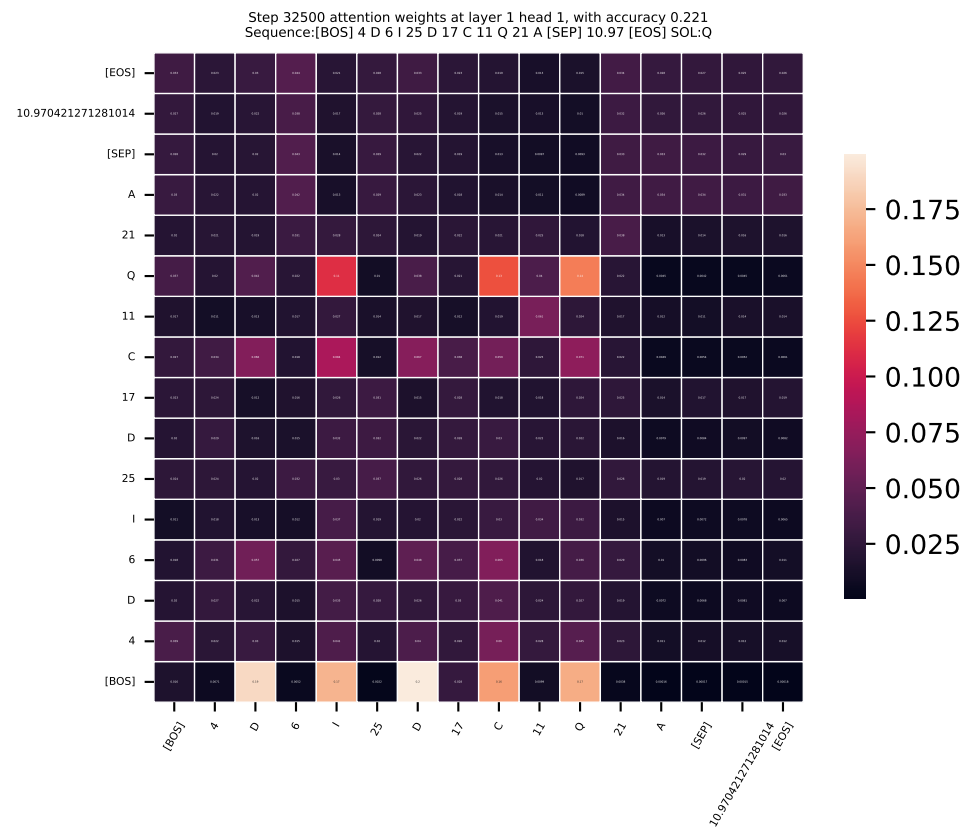

Step 30000 attention weights at layer 1 head 0, with accuracy 0.205
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS]
4
D
6
I
25
D
17
C
11
Q
21
A
[SEP]
10.970421271281014
[EOS]
0.08
0.06
0.04
0.02

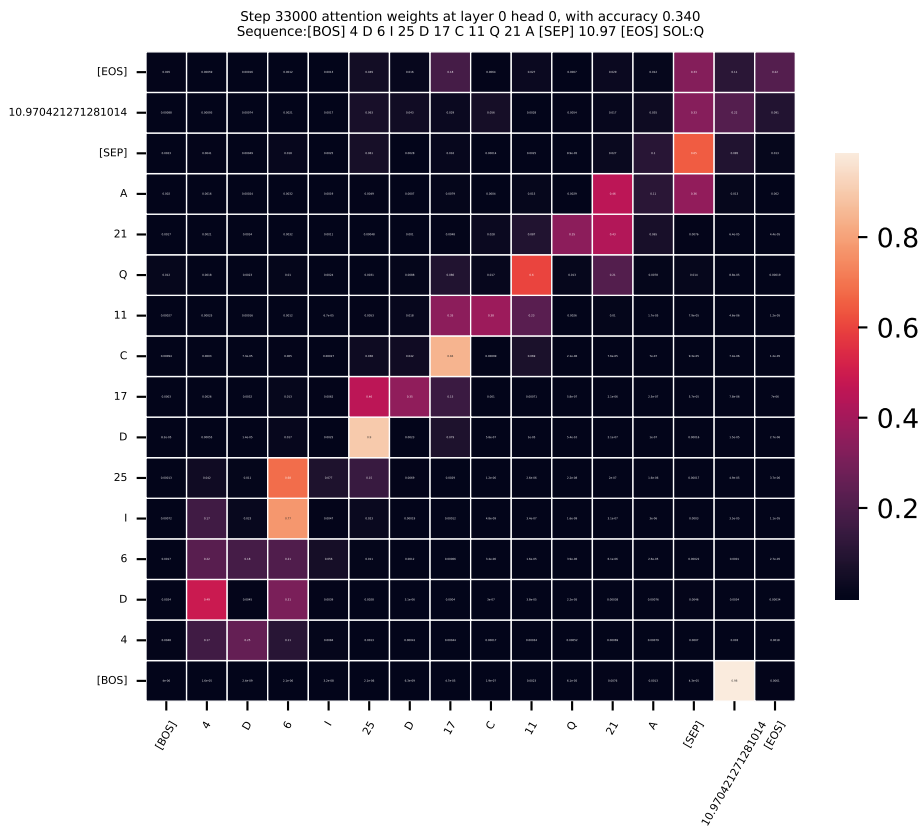

Step 30000 attention weights at layer 1 head 1, with accuracy 0.205
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS]
4
D
6
I
25
D
17
C
11
Q
21
A
[SEP]
10.970421271281014
[EOS]
0.30
0.25
0.20
0.15
0.10
0.05

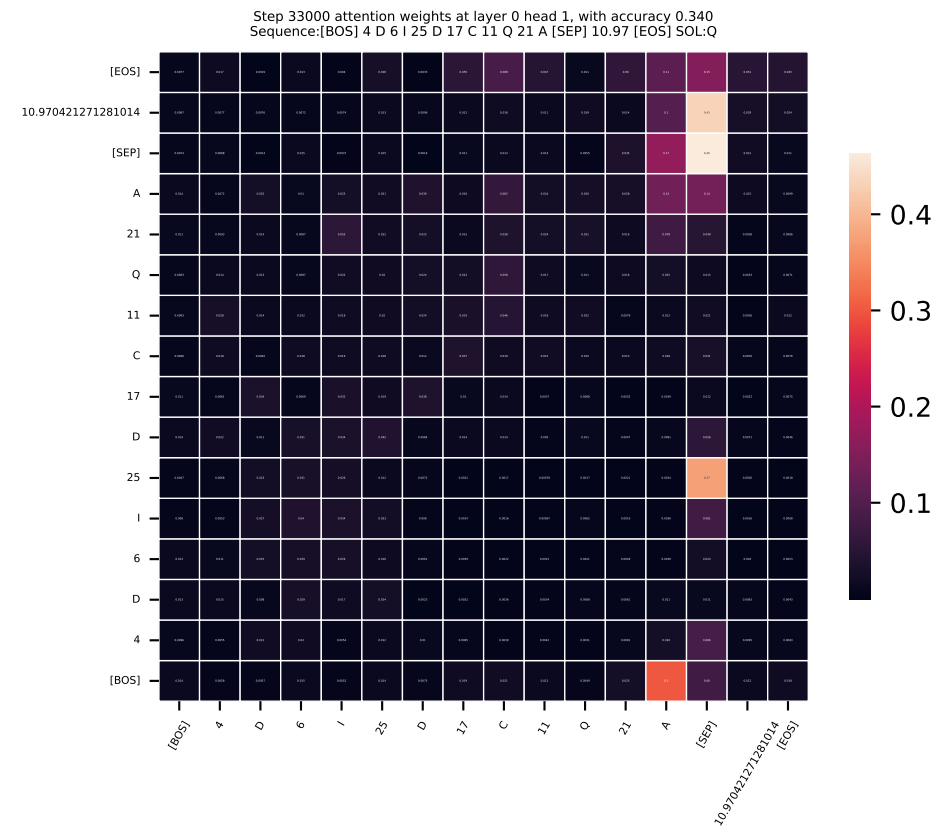

Step 32500 attention weights at layer 0 head 0, with accuracy 0.221
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS]
4
D
6
I
25
D
17
C
11
Q
21
A
[SEP]
10.970421271281014
[EOS]
0.8
0.6
0.4
0.2

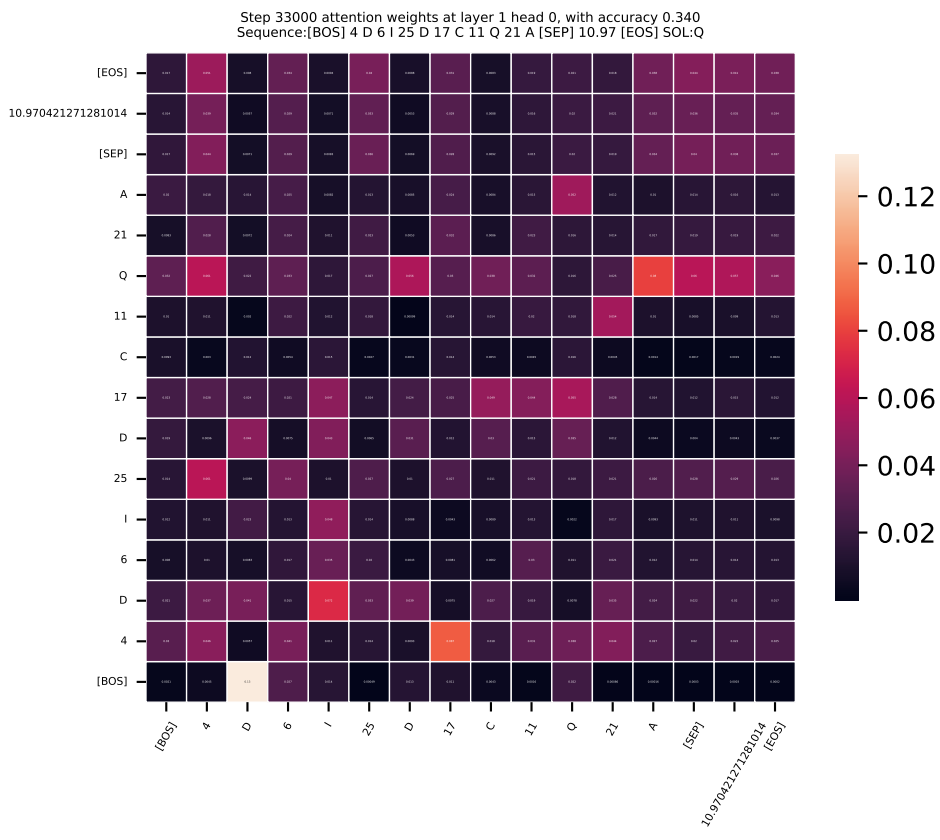

Step 32500 attention weights at layer 0 head 1, with accuracy 0.221
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS]
4
D
6
I
25
D
17
C
11
Q
21
A
[SEP]
10.970421271281014
[EOS]
0.4
0.3
0.2
0.1

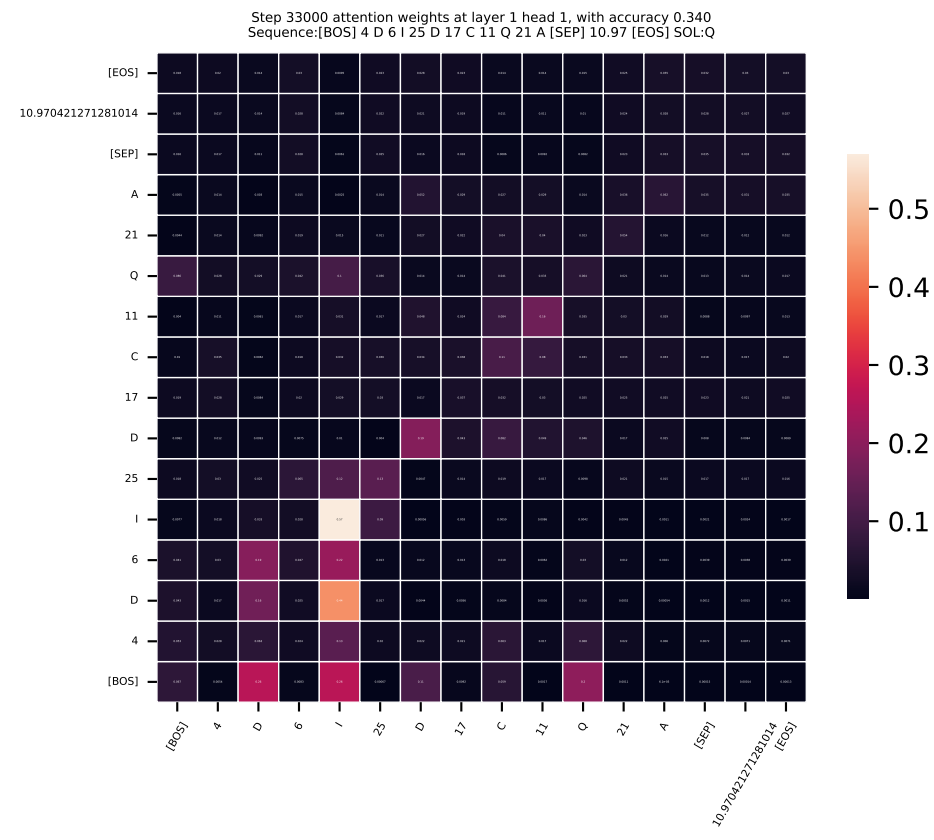

Step 32500 attention weights at layer 1 head 0, with accuracy 0.221
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.14
0.12
0.10
0.08
0.06
0.04
0.02

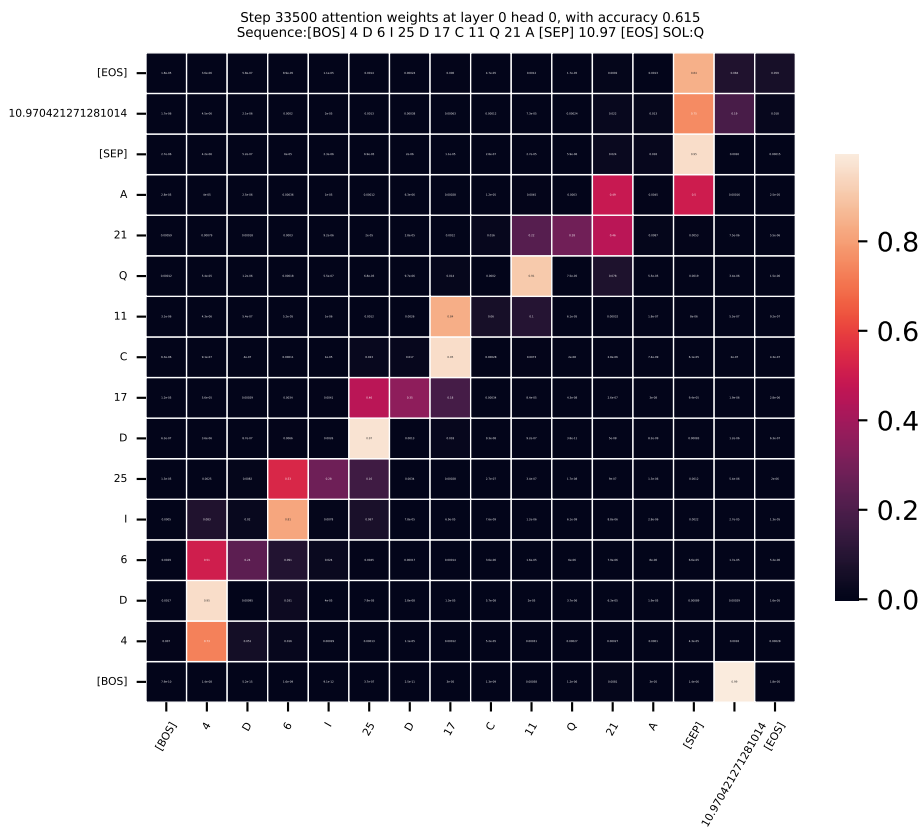

Step 32500 attention weights at layer 1 head 1, with accuracy 0.221
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.175
0.150
0.125
0.100
0.075
0.050
0.025

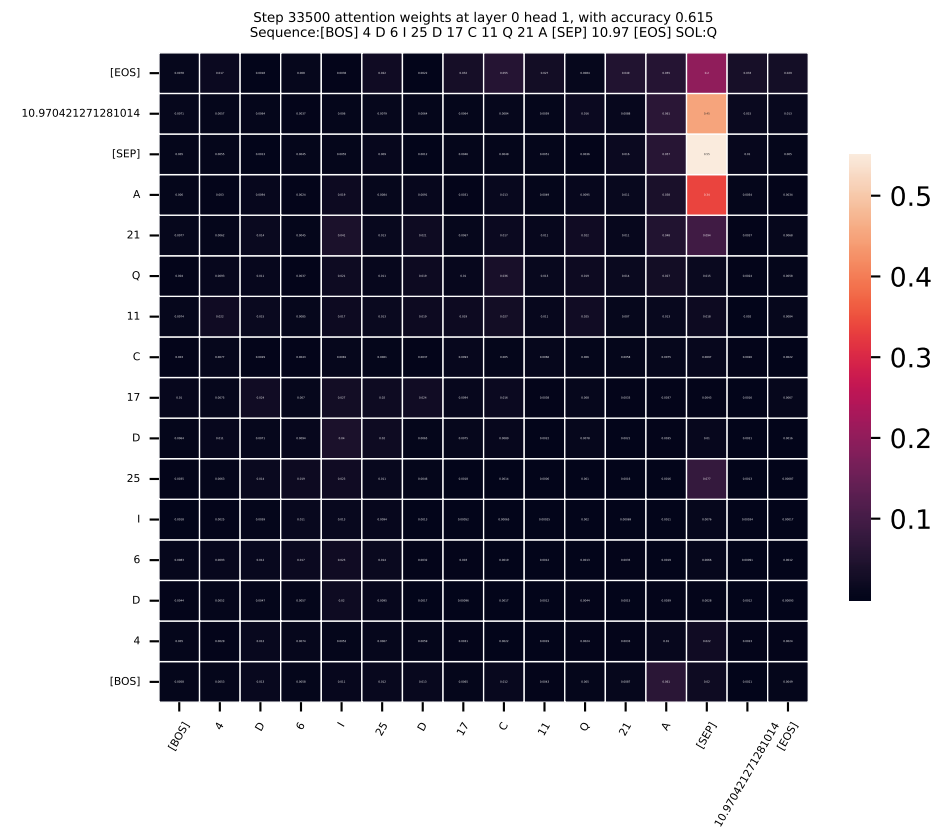

Step 33000 attention weights at layer 0 head 0, with accuracy 0.340
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS]
4
D
6
I
25
D
17
C
11
Q
21
A
[SEP]
10.970421271281014
[EOS]
0.8
0.6
0.4
0.2

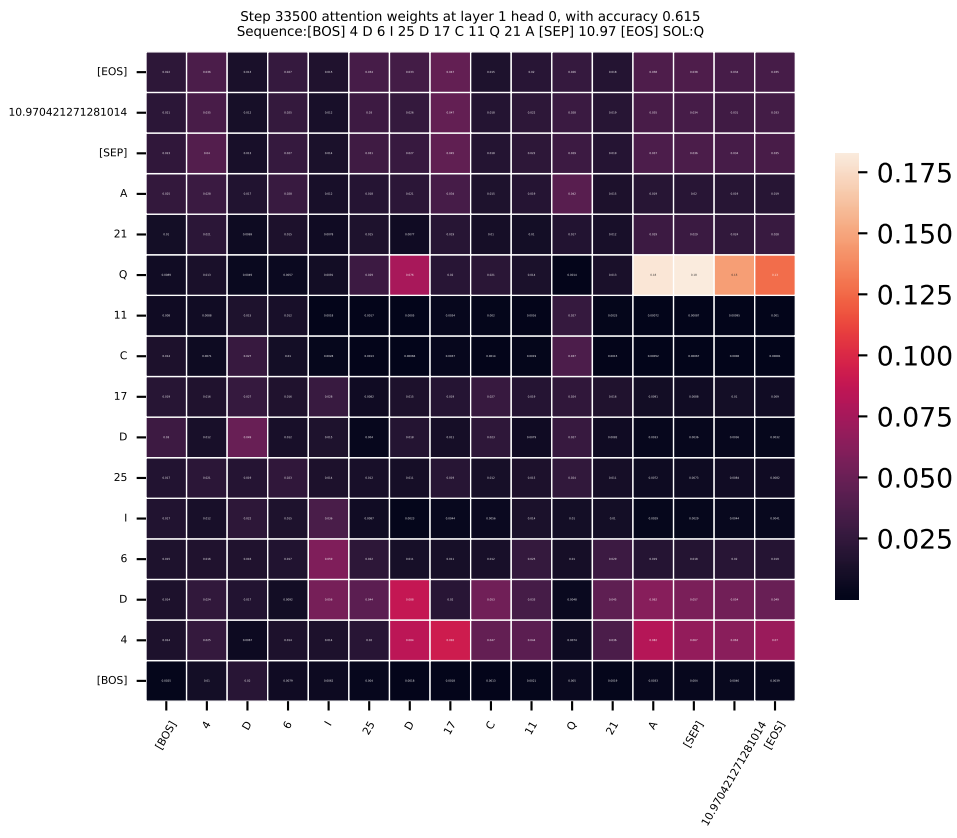

Step 33000 attention weights at layer 0 head 1, with accuracy 0.340
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS]
4
D
6
I
25
D
17
C
11
Q
21
A
[SEP]
10.970421271281014
[EOS]
0.4
0.3
0.2
0.1

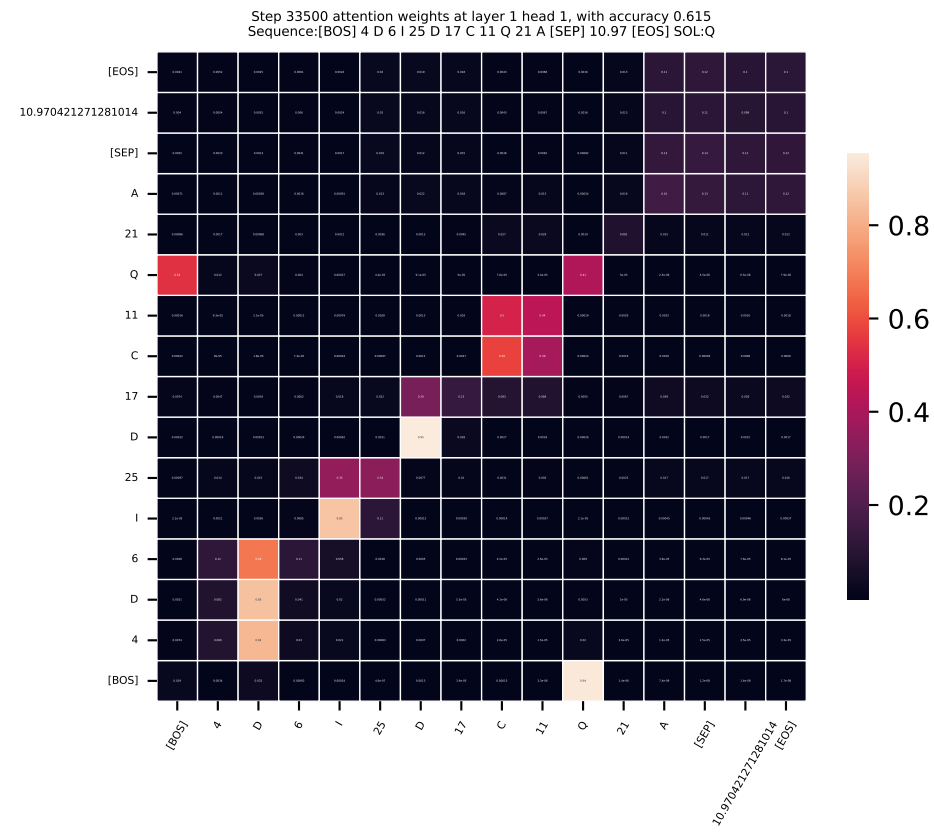

Step 33000 attention weights at layer 1 head 0, with accuracy 0.340
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.12
0.10
0.08
0.06
0.04
0.02

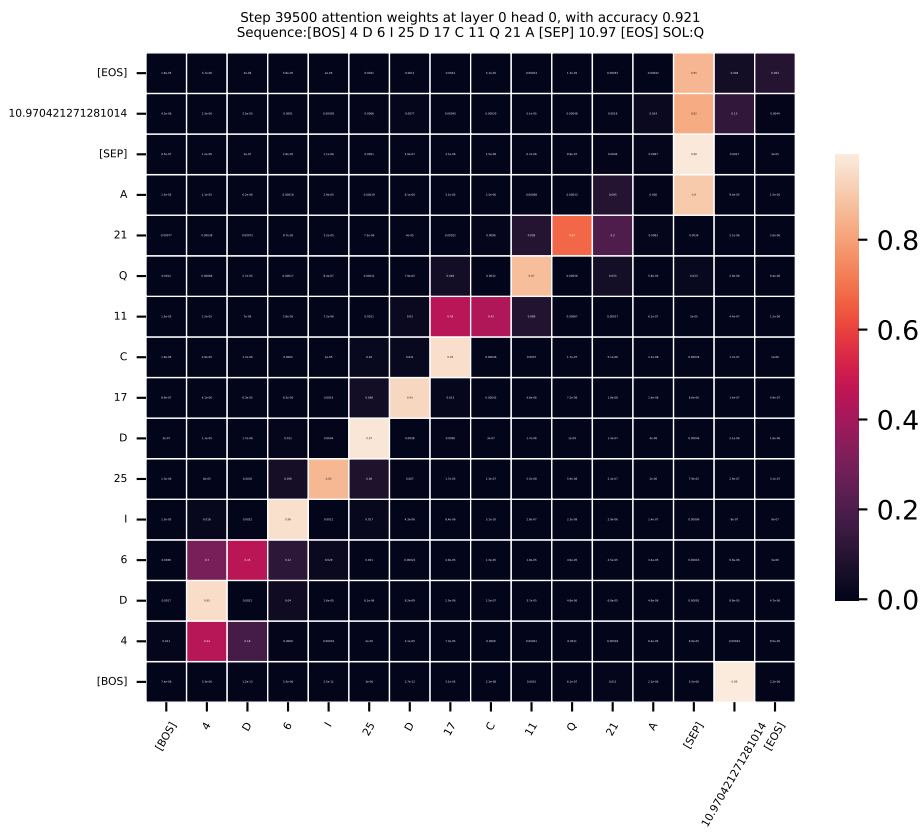

Step 33000 attention weights at layer 1 head 1, with accuracy 0.340
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.5
0.4
0.3
0.2
0.1

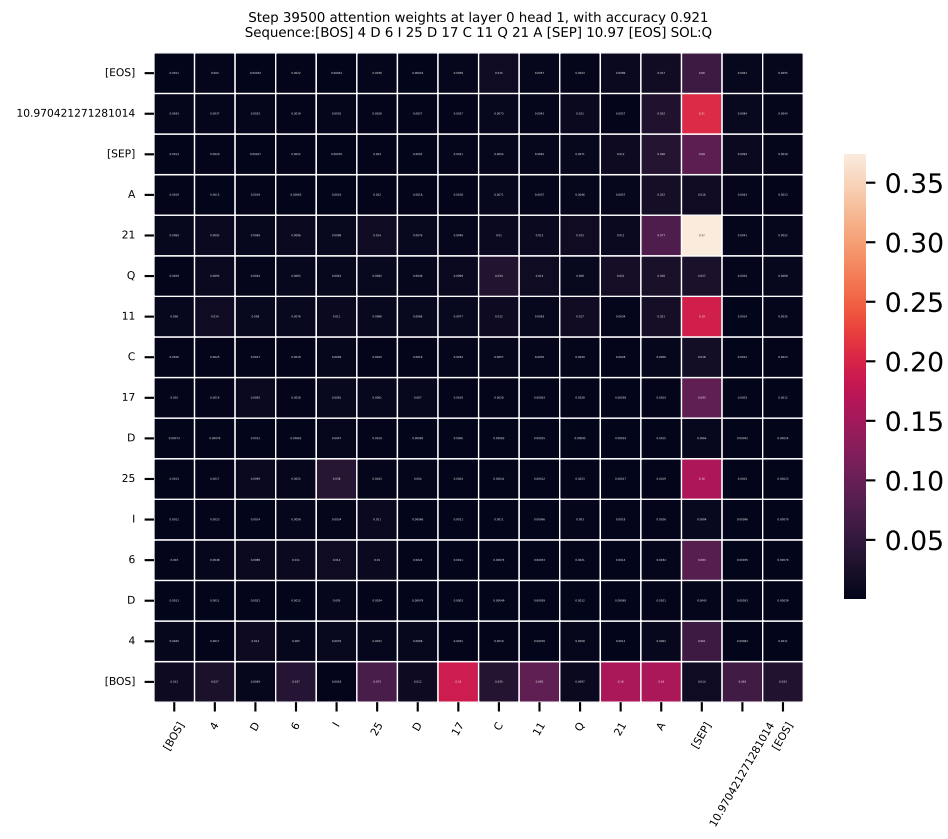

Step 33500 attention weights at layer 0 head 0, with accuracy 0.615
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS]
4
D
6
I
25
D
17
C
11
Q
21
A
[SEP]
10.970421271281014
[EOS]
0.8
0.6
0.4
0.2
0.0

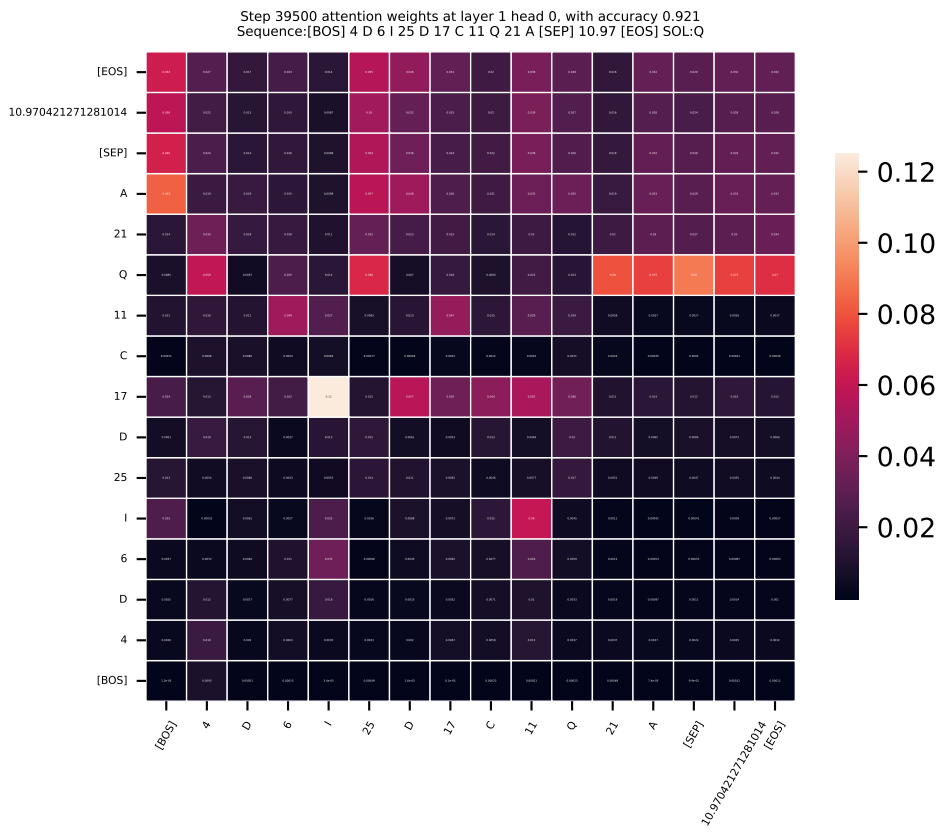

Step 33500 attention weights at layer 0 head 1, with accuracy 0.615
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS]
4
D
6
I
25
D
17
C
11
Q
21
A
[SEP]
10.970421271281014
[EOS]
0.5
0.4
0.3
0.2
0.1

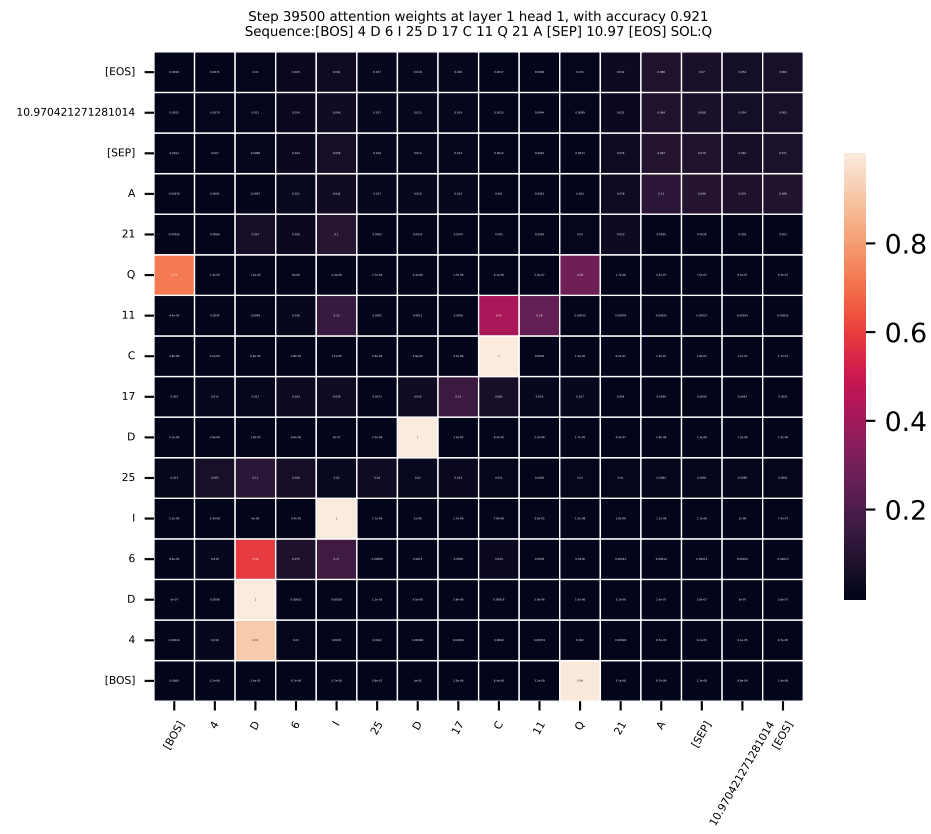

Step 33500 attention weights at layer 1 head 0, with accuracy 0.615
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS]
4
D
6
I
25
D
17
C
11
Q
21
A
[SEP]
10.970421271281014
[EOS]
0.175
0.150
0.125
0.100
0.075
0.050
0.025

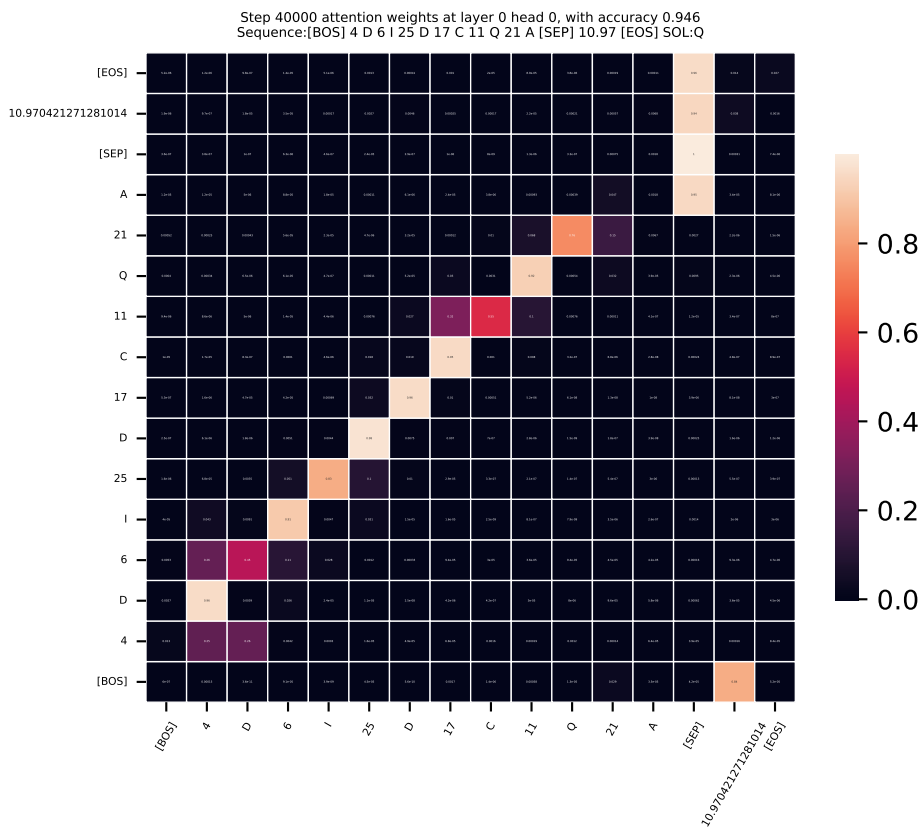

Step 33500 attention weights at layer 1 head 1, with accuracy 0.615
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS]
4
D
6
I
25
D
17
C
11
Q
21
A
[SEP]
10.970421271281014
[EOS]
0.8
0.6
0.4
0.2

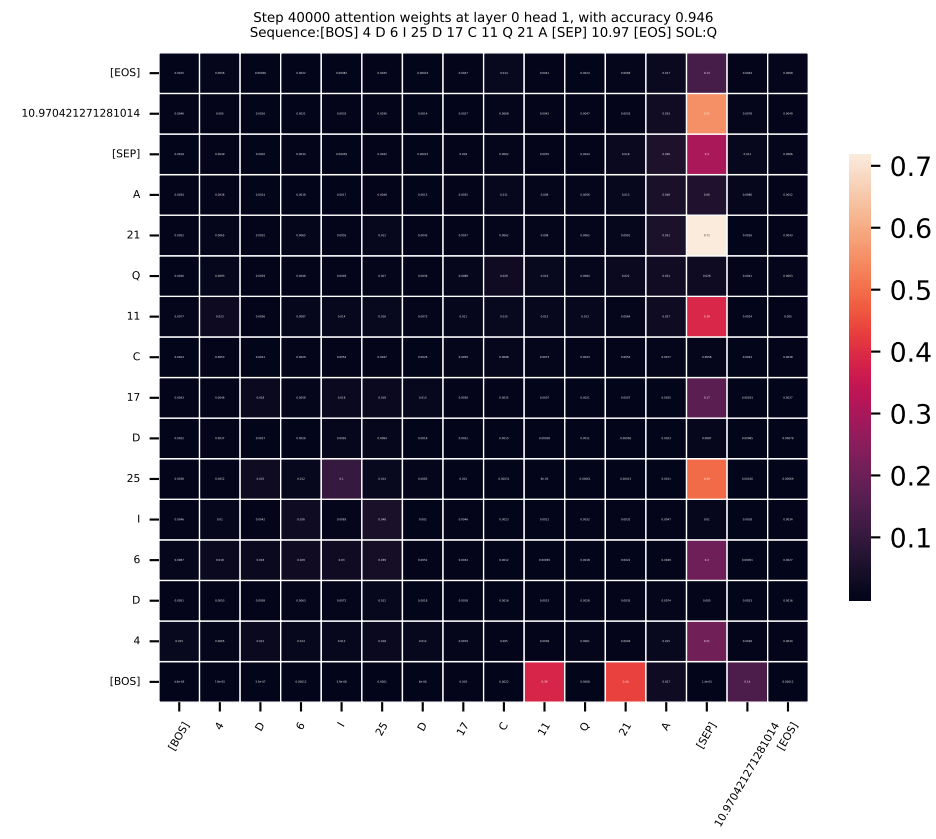

Step 39500 attention weights at layer 0 head 0, with accuracy 0.921
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.8
0.6
0.4
0.2
0.0

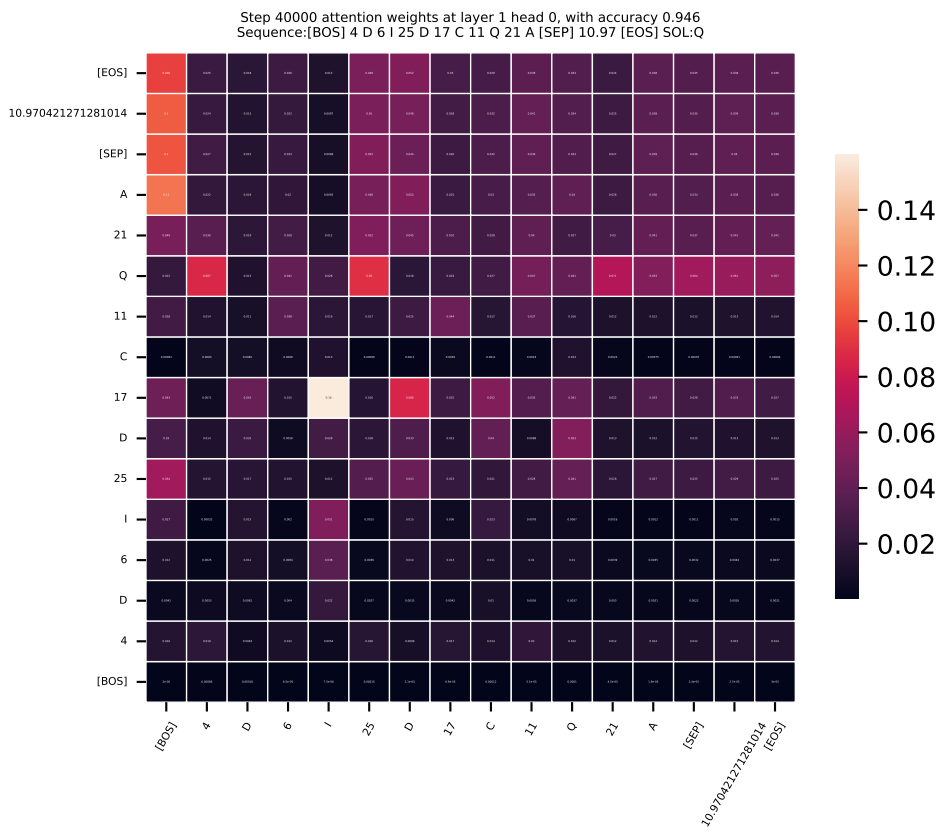

Step 39500 attention weights at layer 0 head 1, with accuracy 0.921
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.35
0.30
0.25
0.20
0.15
0.10
0.05

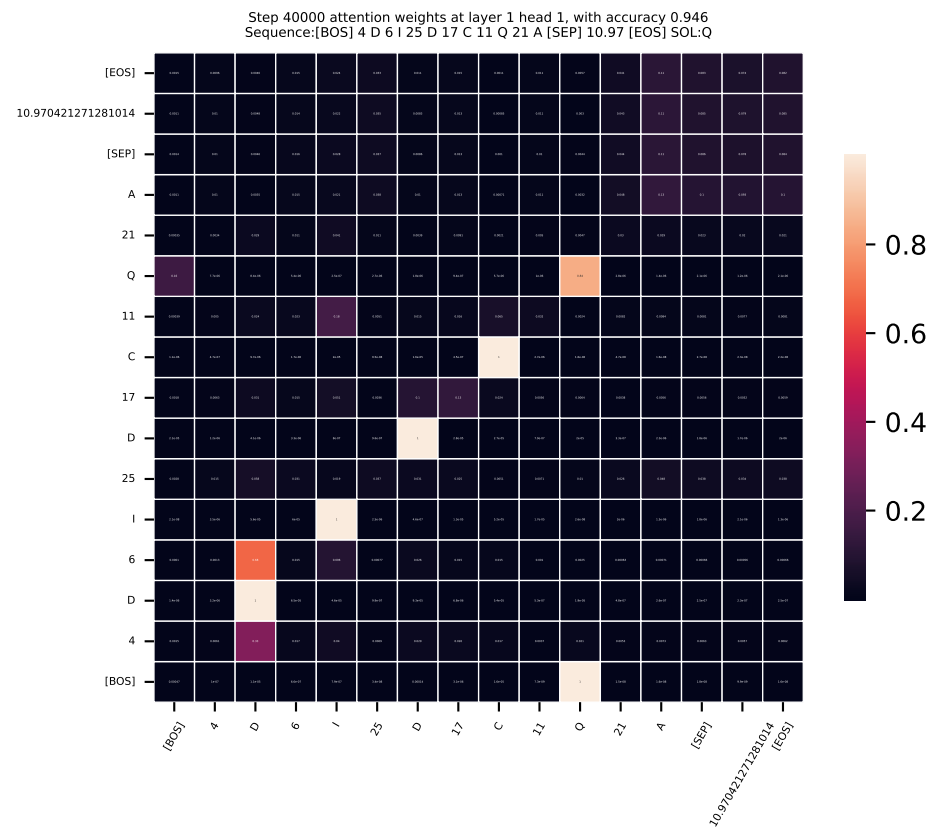

Step 39500 attention weights at layer 1 head 0, with accuracy 0.921
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS]
4
D
6
I
25
D
17
C
11
Q
21
A
[SEP]
10.970421271281014
[EOS]
0.12
0.10
0.08
0.06
0.04
0.02

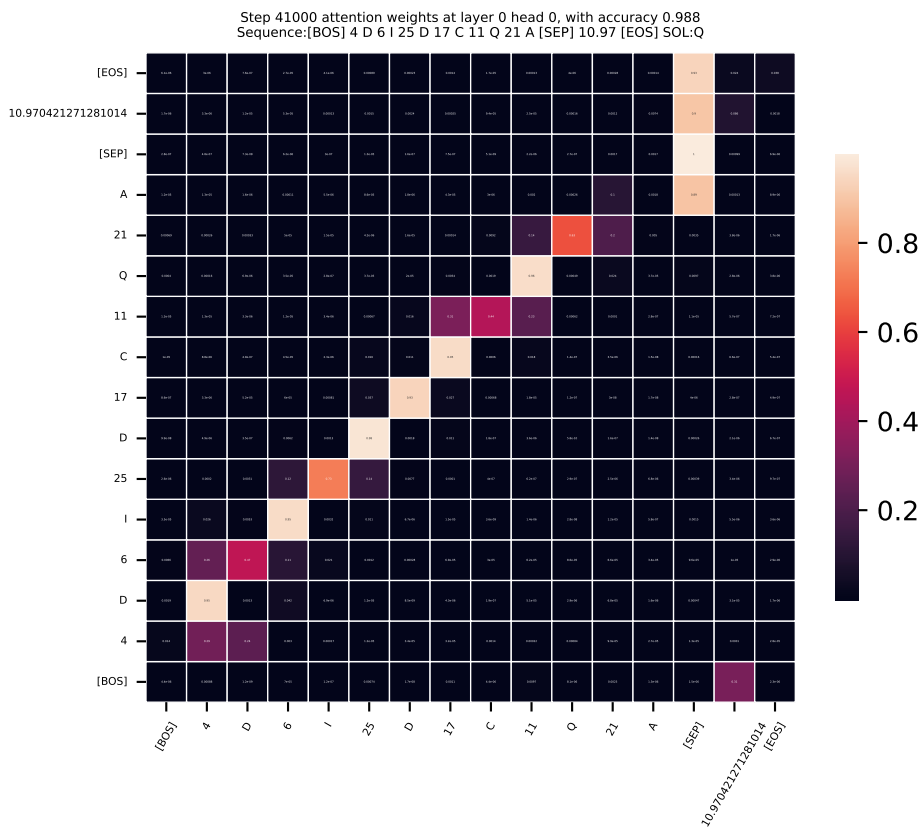

Step 39500 attention weights at layer 1 head 1, with accuracy 0.921
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS]
4
D
6
I
25
D
17
C
11
Q
21
A
[SEP]
10.970421271281014
[EOS]
0.8
0.6
0.4
0.2

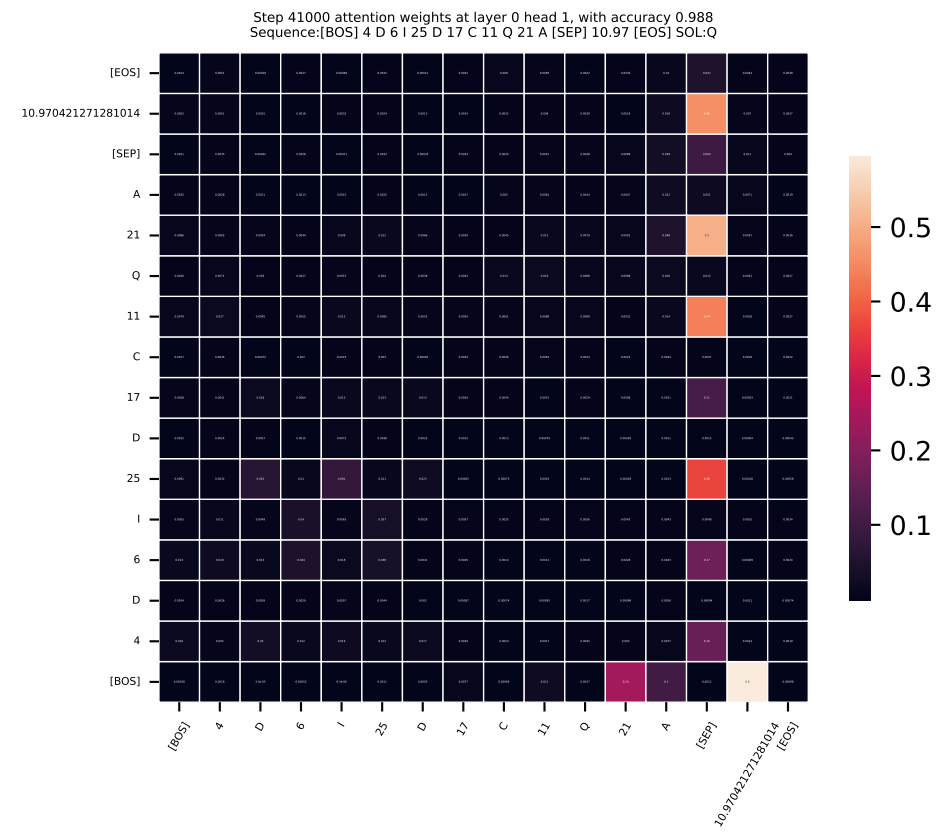

Step 40000 attention weights at layer 0 head 0, with accuracy 0.946
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS]
4
D
6
I
25
D
17
C
11
Q
21
A
[SEP]
10.970421271281014
[EOS]
0.8
0.6
0.4
0.2
0.0

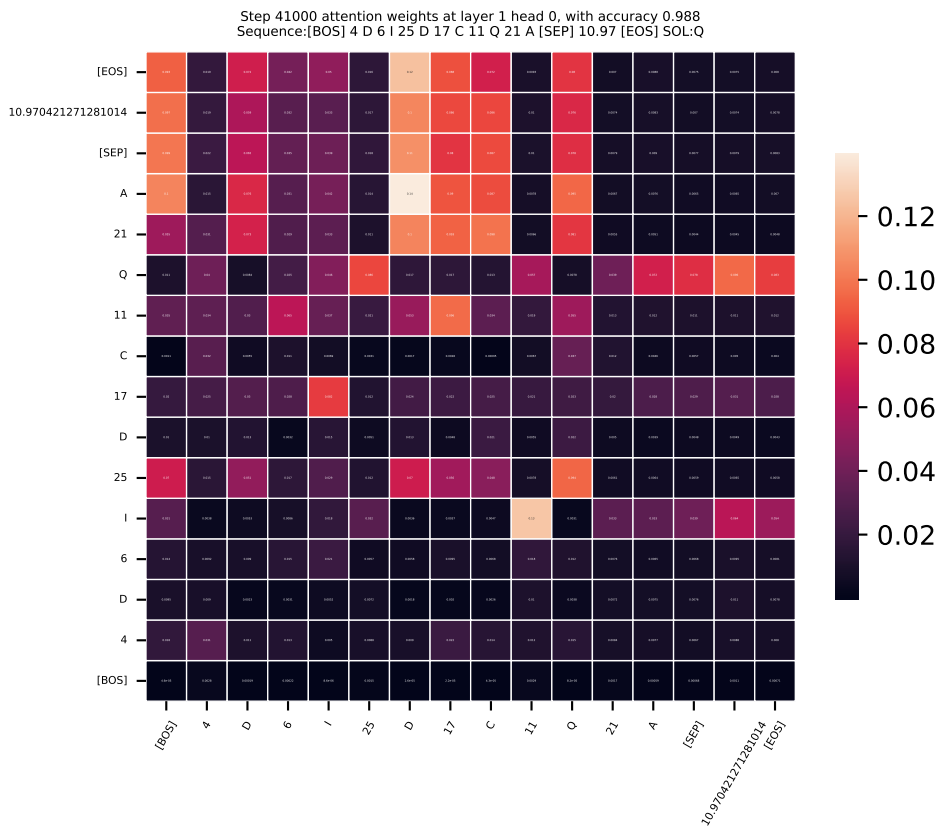

Step 40000 attention weights at layer 0 head 1, with accuracy 0.946
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS]
4
D
6
I
25
D
17
C
11
Q
21
A
[SEP]
10.970421271281014
[EOS]
0.7
0.6
0.5
0.4
0.3
0.2
0.1

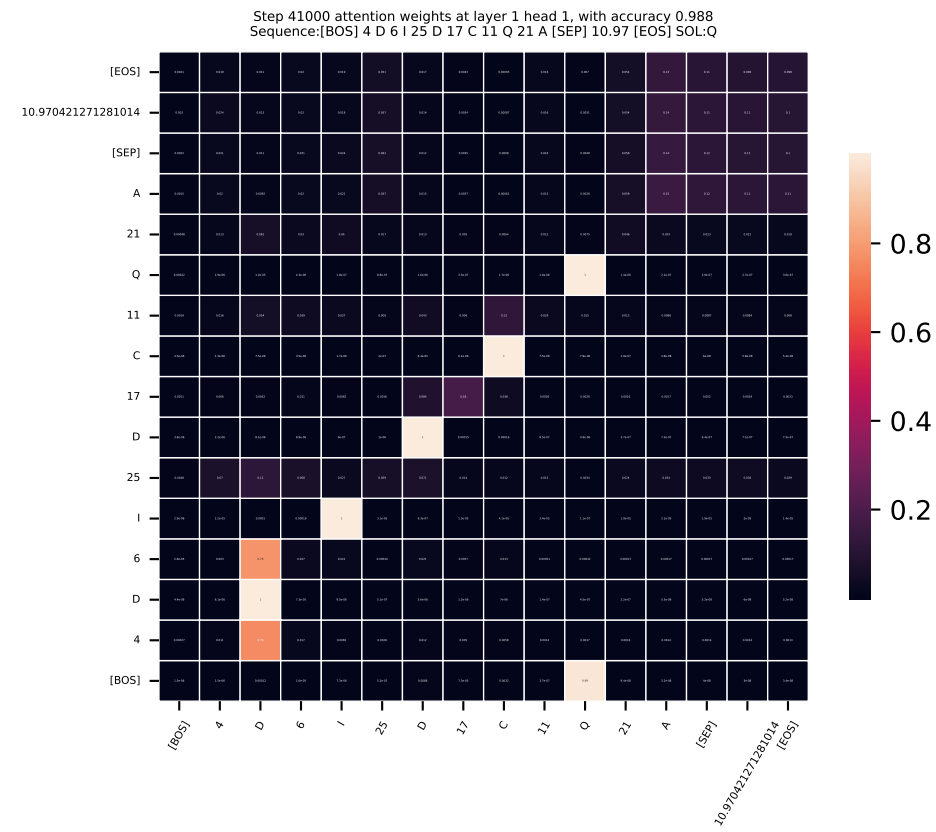

Step 40000 attention weights at layer 1 head 0, with accuracy 0.946
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
0.14
0.12
0.10
0.08
0.06
0.04
0.02

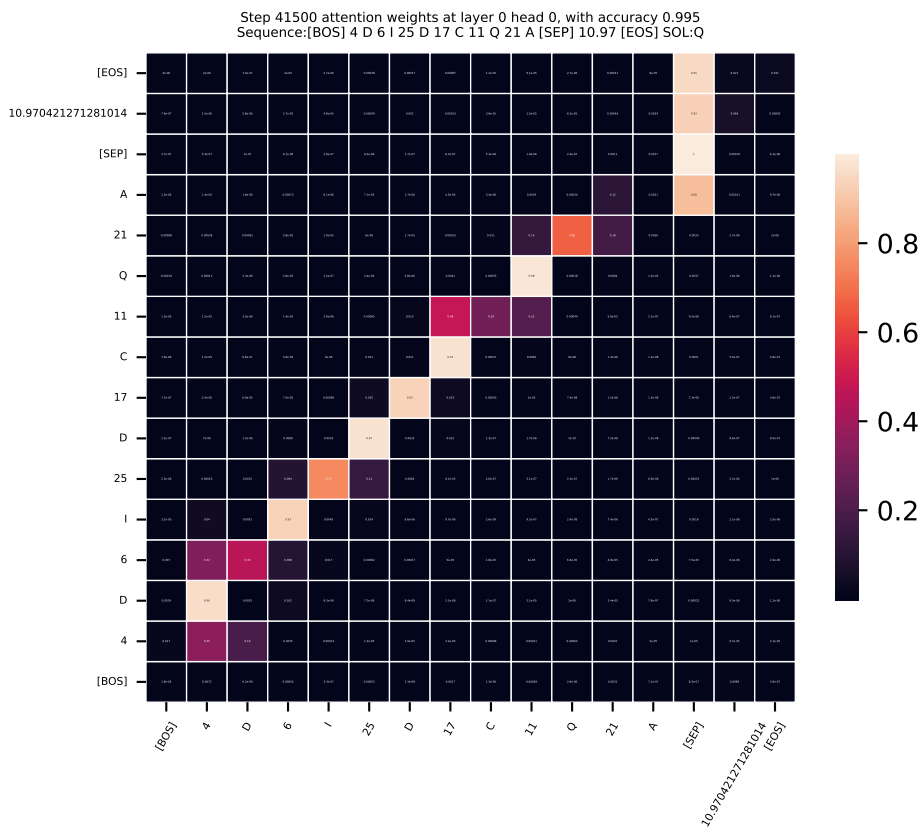

Step 40000 attention weights at layer 1 head 1, with accuracy 0.946
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
0.8
0.6
0.4
0.2

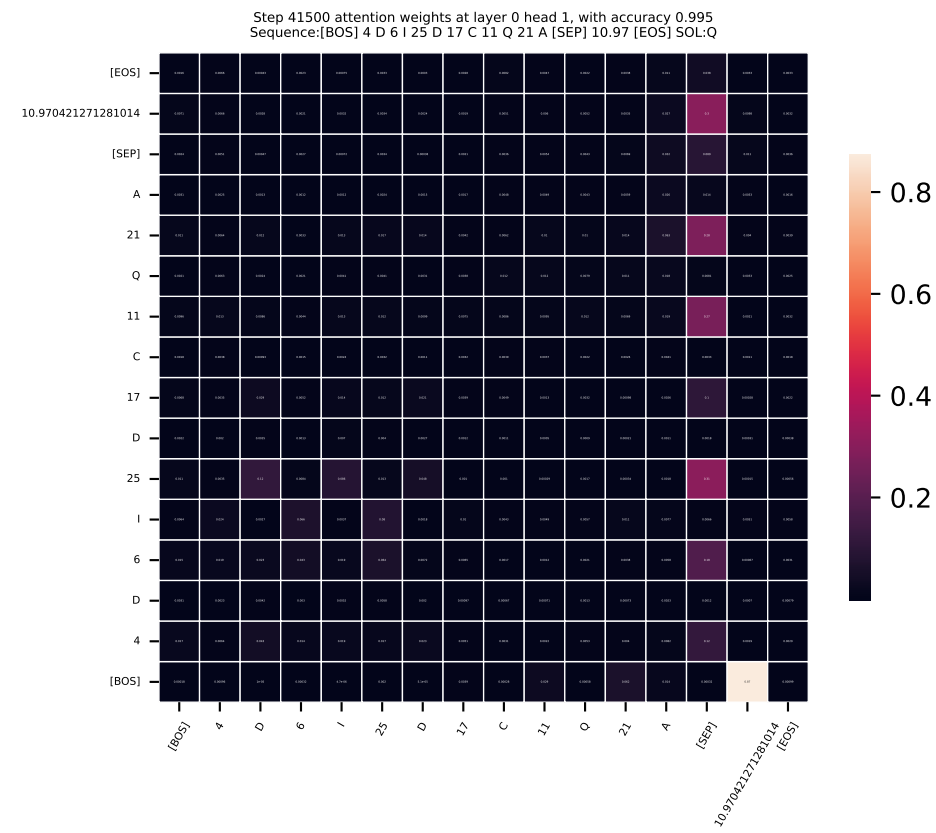

Step 41000 attention weights at layer 0 head 0, with accuracy 0.988
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
0.8
0.6
0.4
0.2

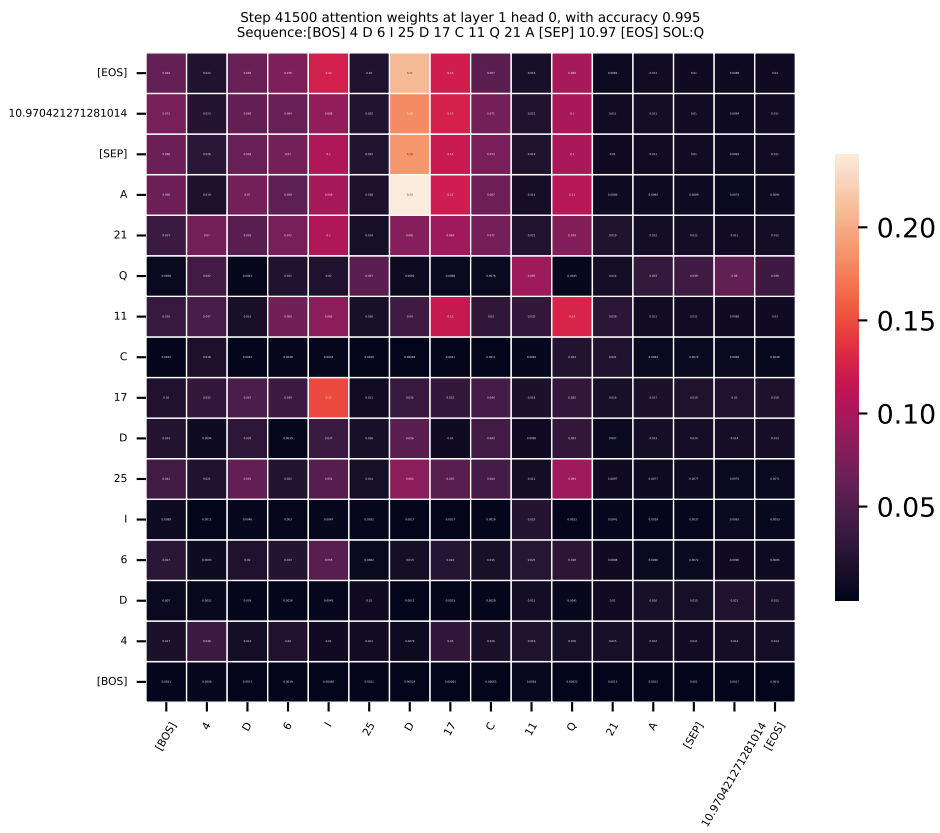

Step 41000 attention weights at layer 0 head 1, with accuracy 0.988
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
0.5
0.4
0.3
0.2
0.1

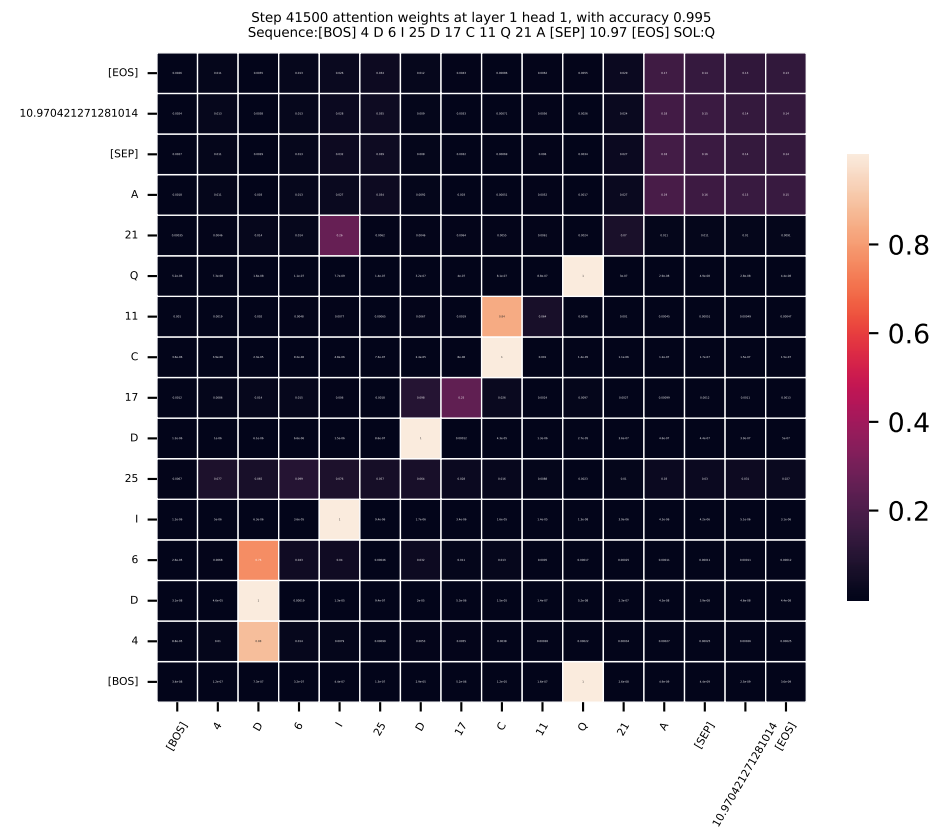

Step 41000 attention weights at layer 1 head 0, with accuracy 0.988
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS]
4
D
6
I
25
D
17
C
11
Q
21
A
[SEP]
10.970421271281014
[EOS]
0.12
0.10
0.08
0.06
0.04
0.02


Step 41000 attention weights at layer 1 head 1, with accuracy 0.988
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS]
4
D
6
I
25
D
17
C
11
Q
21
A
[SEP]
10.970421271281014
[EOS]
0.8
0.6
0.4
0.2


Step 41500 attention weights at layer 0 head 0, with accuracy 0.995
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS]
4
D
6
I
25
D
17
C
11
Q
21
A
[SEP]
10.970421271281014
[EOS]
0.8
0.6
0.4
0.2


Step 41500 attention weights at layer 0 head 1, with accuracy 0.995
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS]
4
D
6
I
25
D
17
C
11
Q
21
A
[SEP]
10.970421271281014
[EOS]
0.8
0.6
0.4
0.2


Step 41500 attention weights at layer 1 head 0, with accuracy 0.995
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.20
0.15
0.10
0.05


Step 41500 attention weights at layer 1 head 1, with accuracy 0.995
Sequence:[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.97 [EOS] SOL:Q
[EOS]
10.970421271281014
[SEP]
A
21
Q
11
C
17
D
25
I
6
D
4
[BOS]
[BOS] 4 D 6 I 25 D 17 C 11 Q 21 A [SEP] 10.970421271281014 [EOS]
0.8
0.6
0.4
0.2