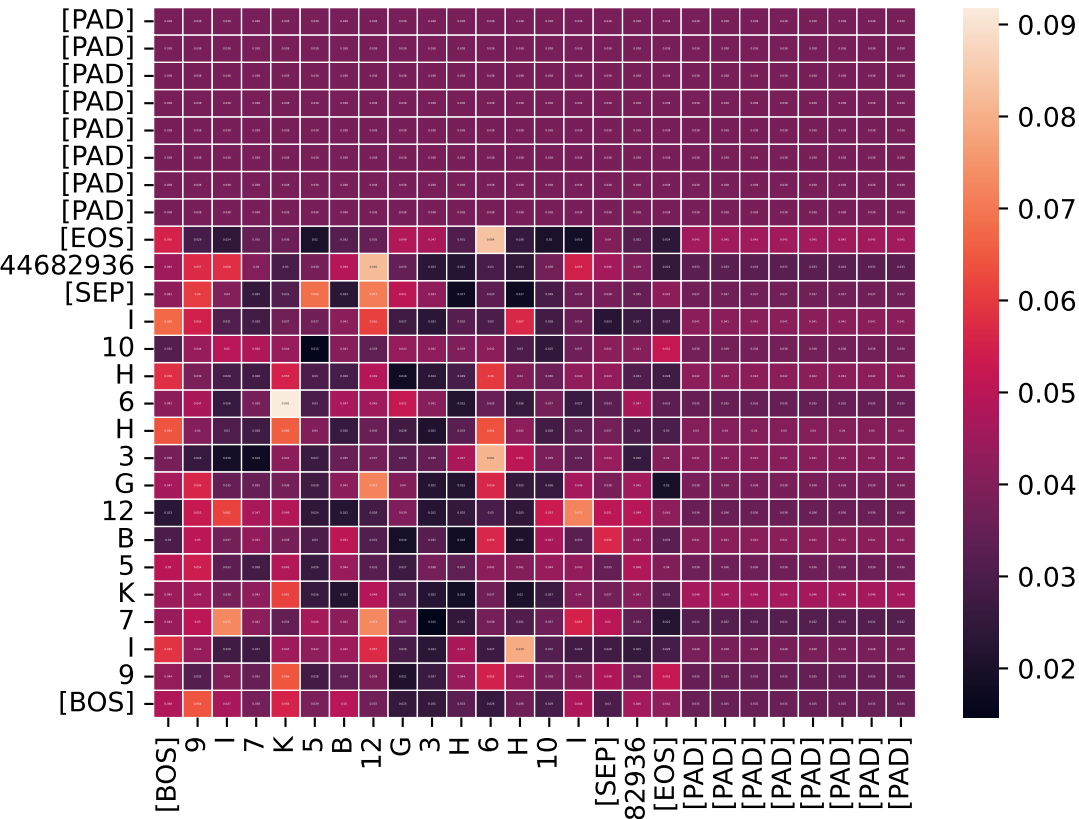
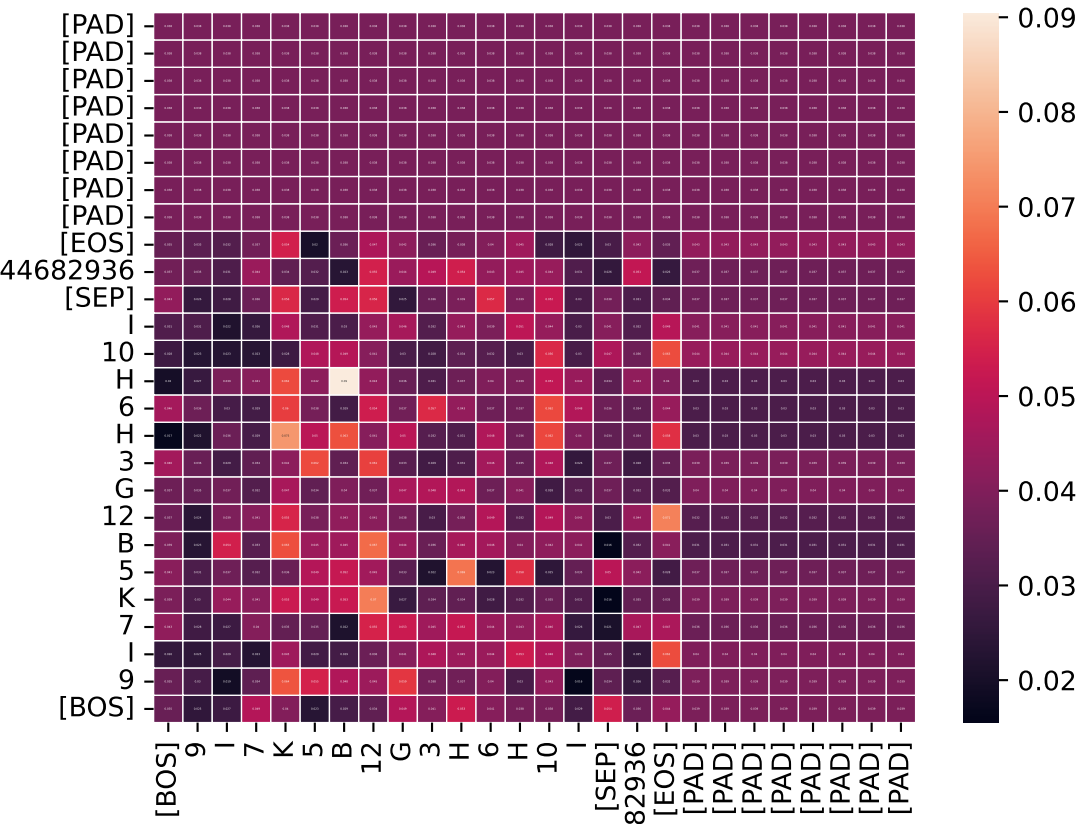


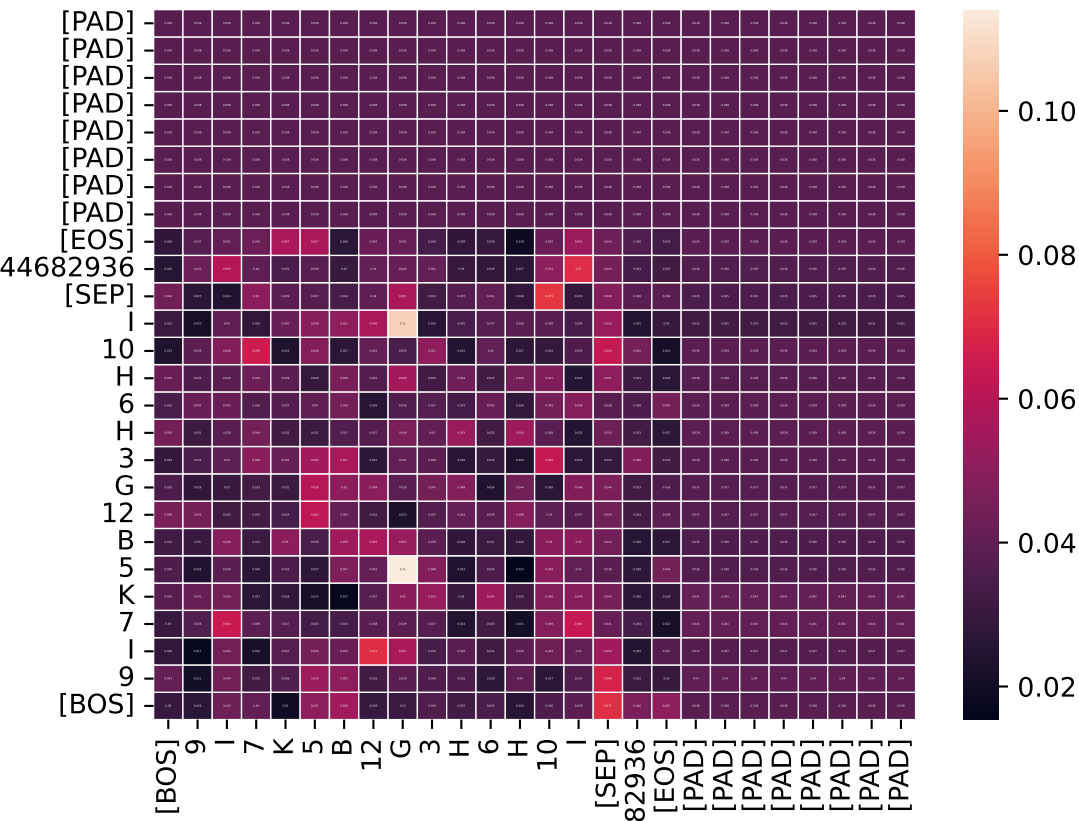
Step 0 attention weights at layer 0 head 1, with accuracy 0.064
 Sequence:[BOS] 9 I 7 K 5 B 12 G 3 H 6 H 10 I [SEP] 6.83 [EOS] SOL:K



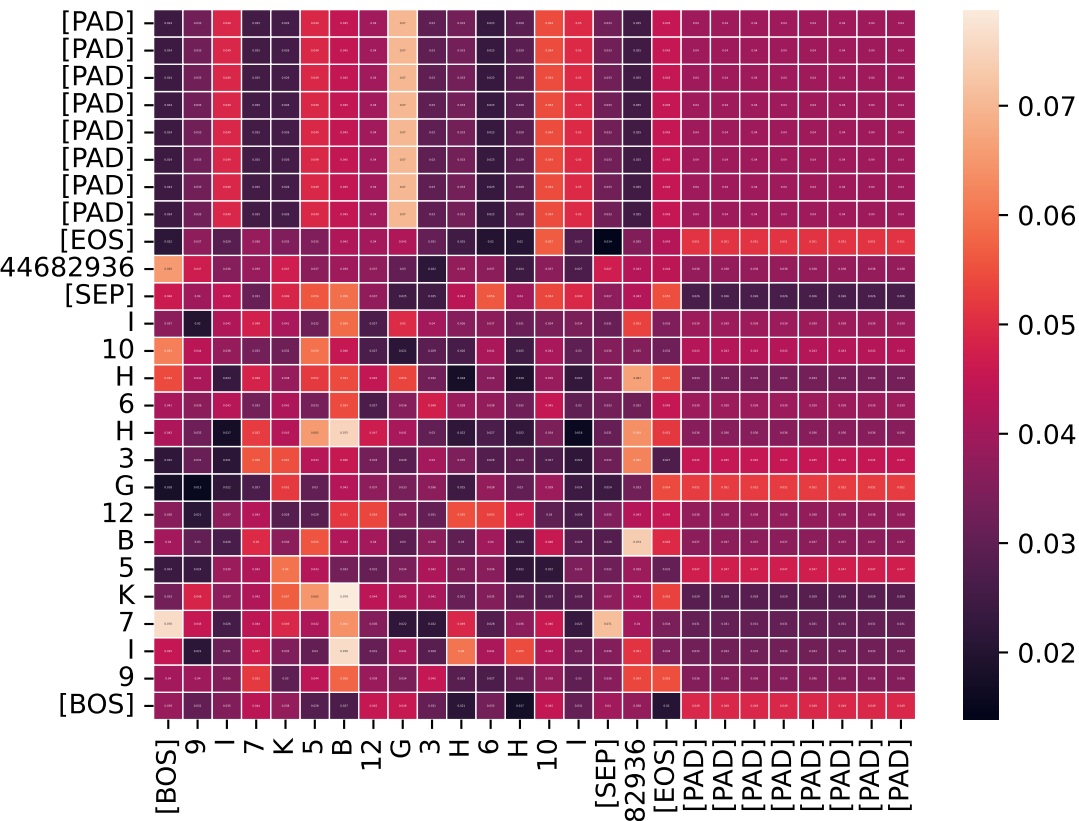
Step 0 attention weights at layer 0 head 2, with accuracy 0.064
 Sequence:[BOS] 9 I 7 K 5 B 12 G 3 H 6 H 10 I [SEP] 6.83 [EOS] SOL:K



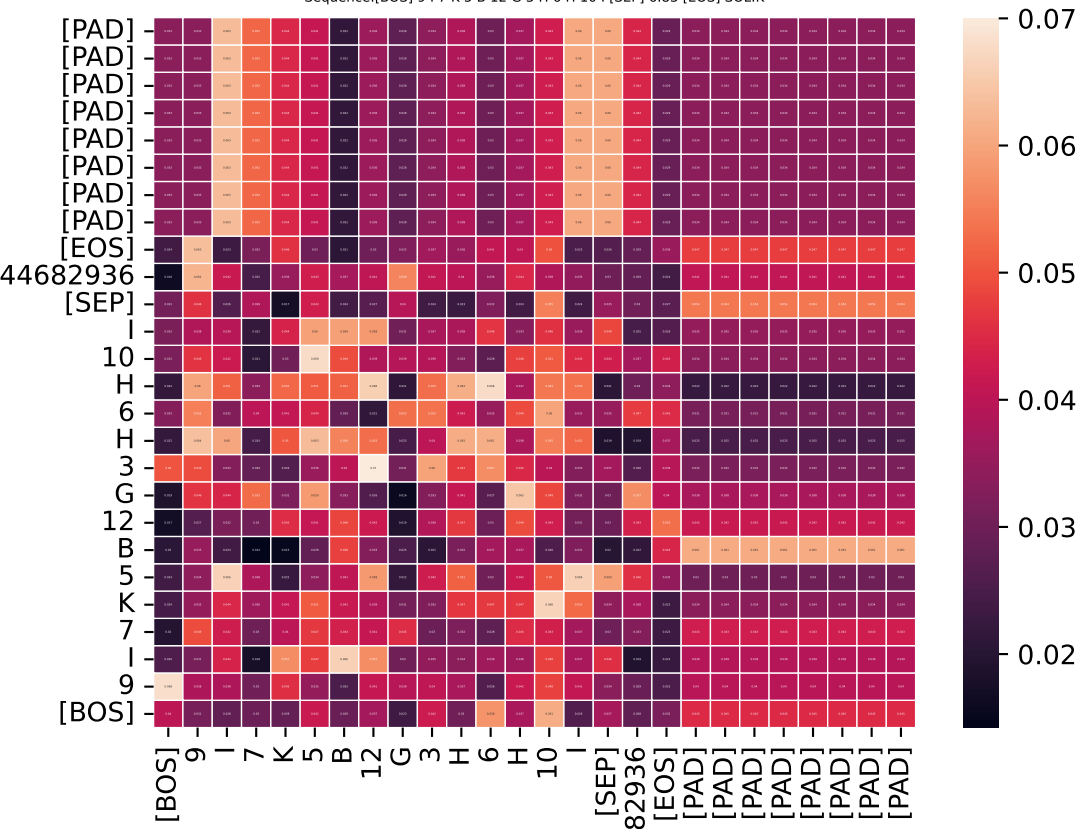
Step 0 attention weights at layer 0 head 3, with accuracy 0.064
 Sequence:[BOS] 9 1 7 K 5 B 12 G 3 H 6 H 10 1 [SEP] 6.83 [EOS] SOL:K



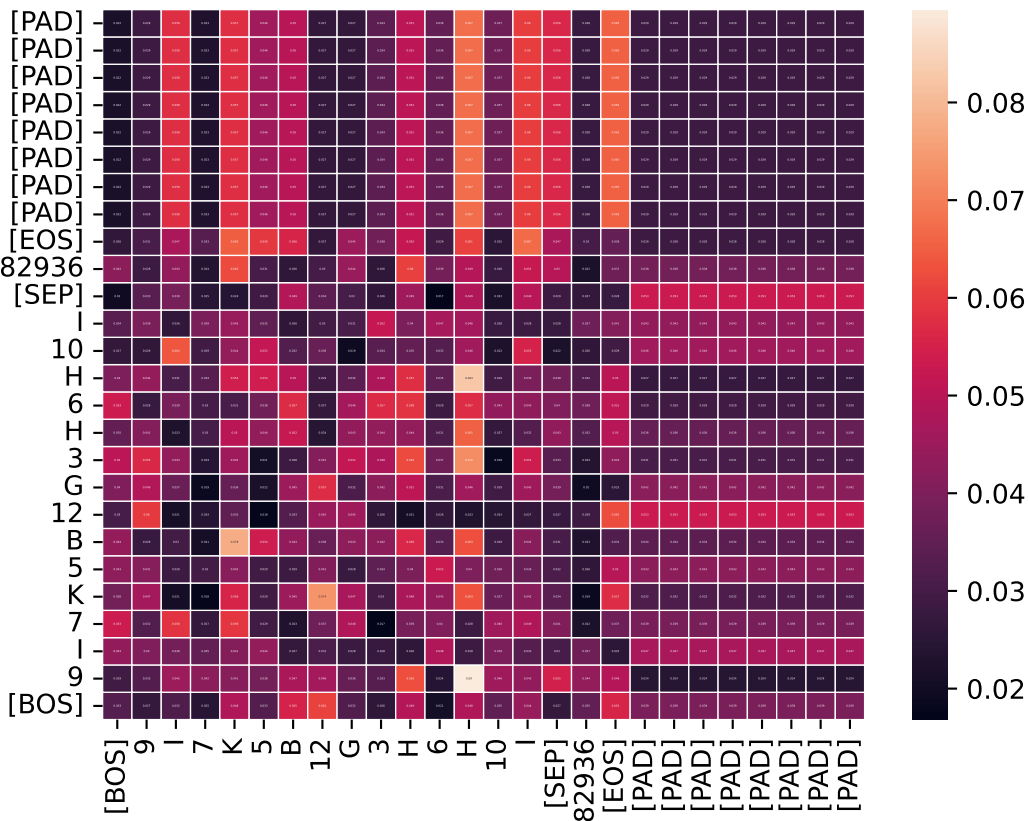
Step 0 attention weights at layer 1 head 0, with accuracy 0.064
 Sequence:[BOS] 9 I 7 K 5 B 12 G 3 H 6 H 10 I [SEP] 6.83 [EOS] SOL:K



Step 0 attention weights at layer 1 head 1, with accuracy 0.064
 Sequence:[BOS] 9 I 7 K 5 B 12 G 3 H 6 H 10 I [SEP] 6.83 [EOS] SOL:K

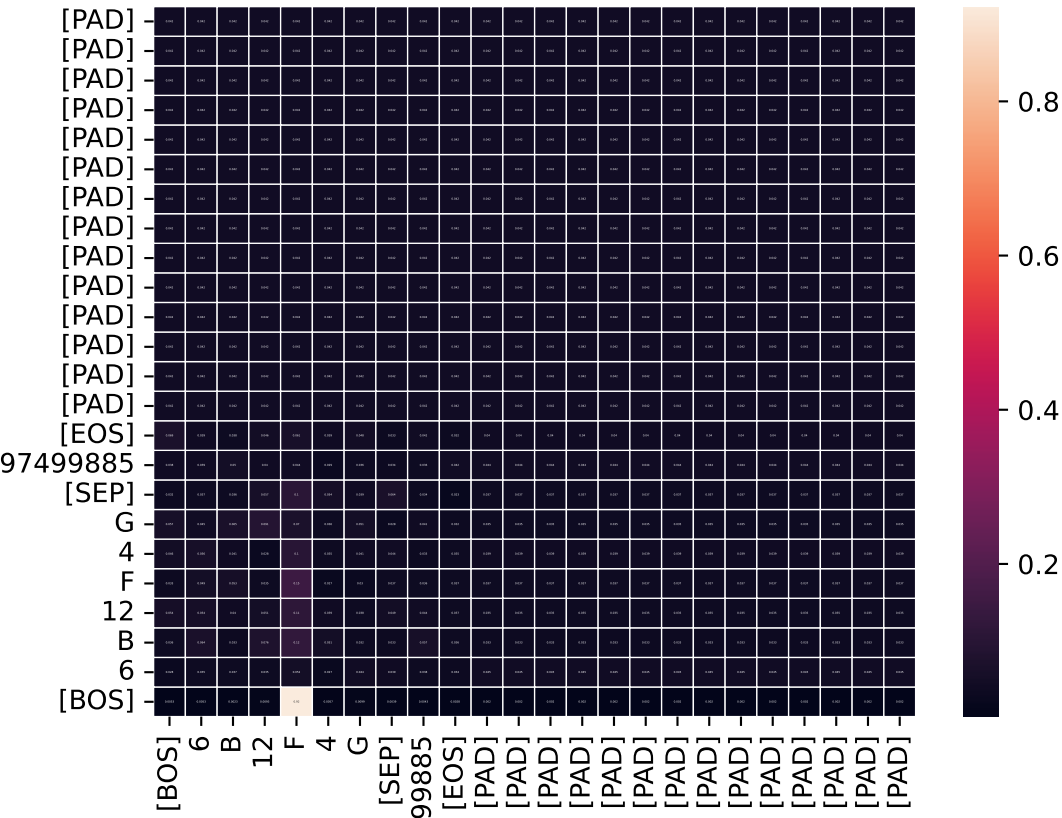


4468

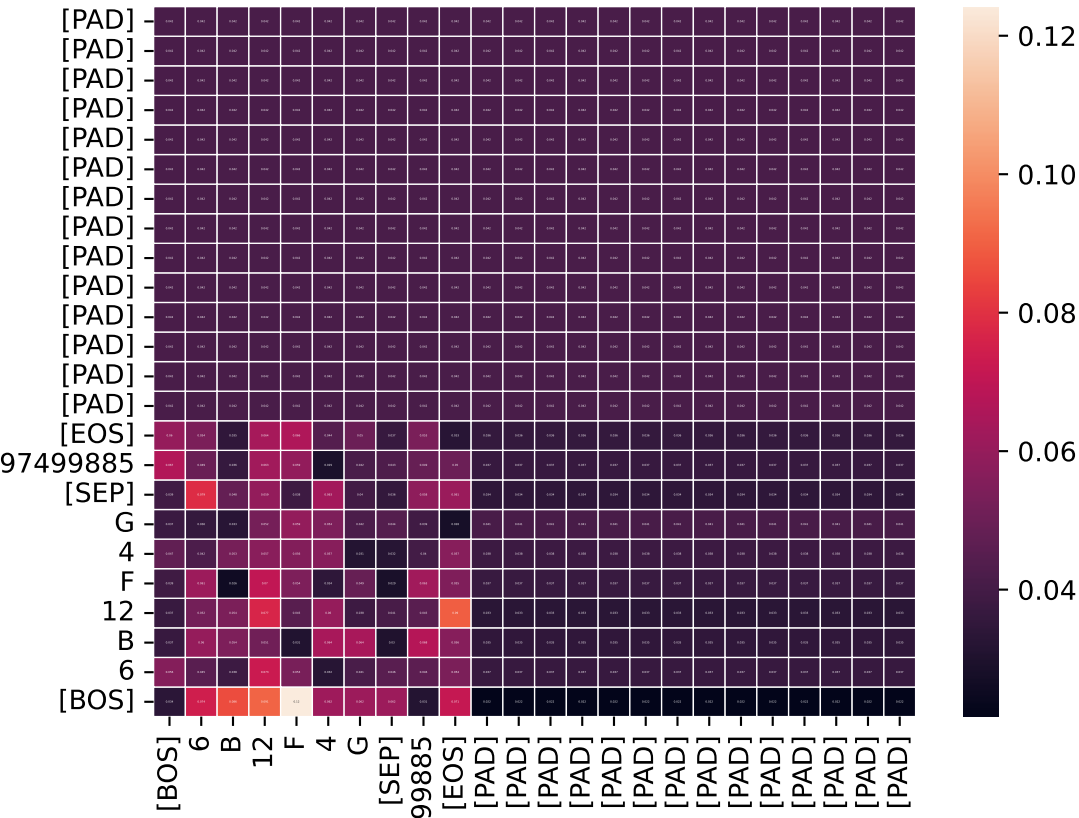


Step 500 attention weights at layer 0 head 0, with accuracy 0.344

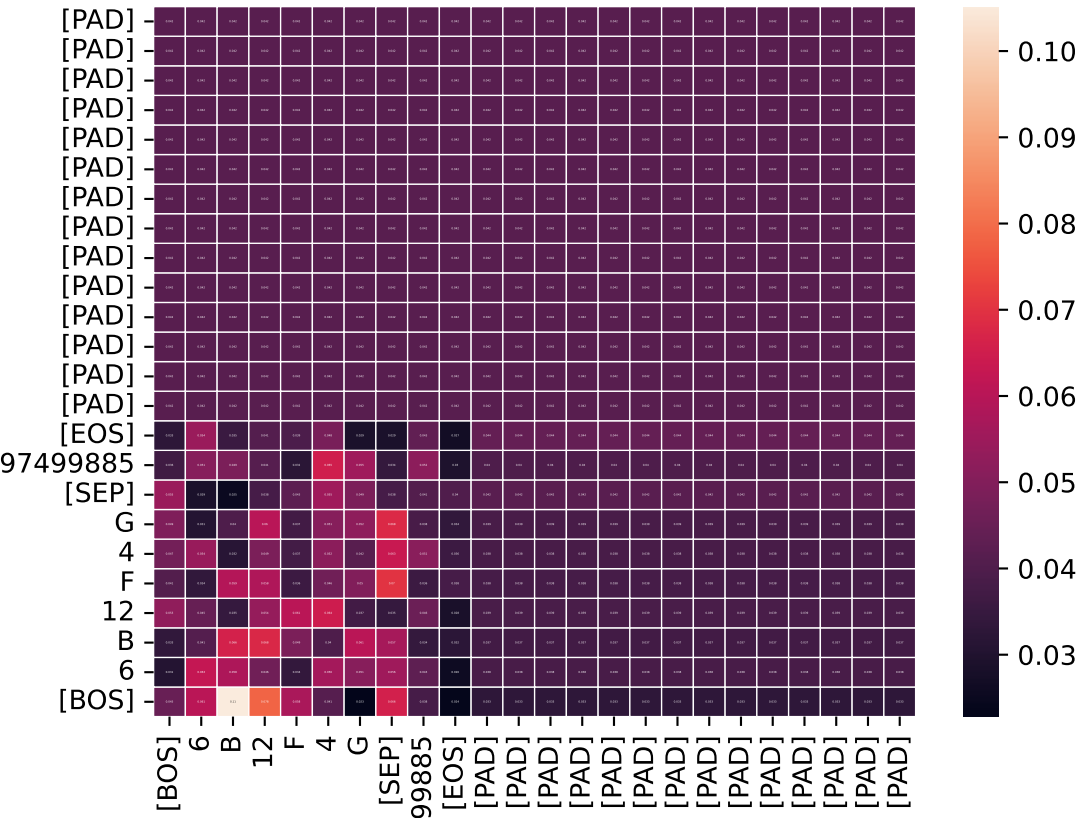
Sequence:[BOS] 6 B 12 F 4 G [SEP] 5.83 [EOS] SOL:B



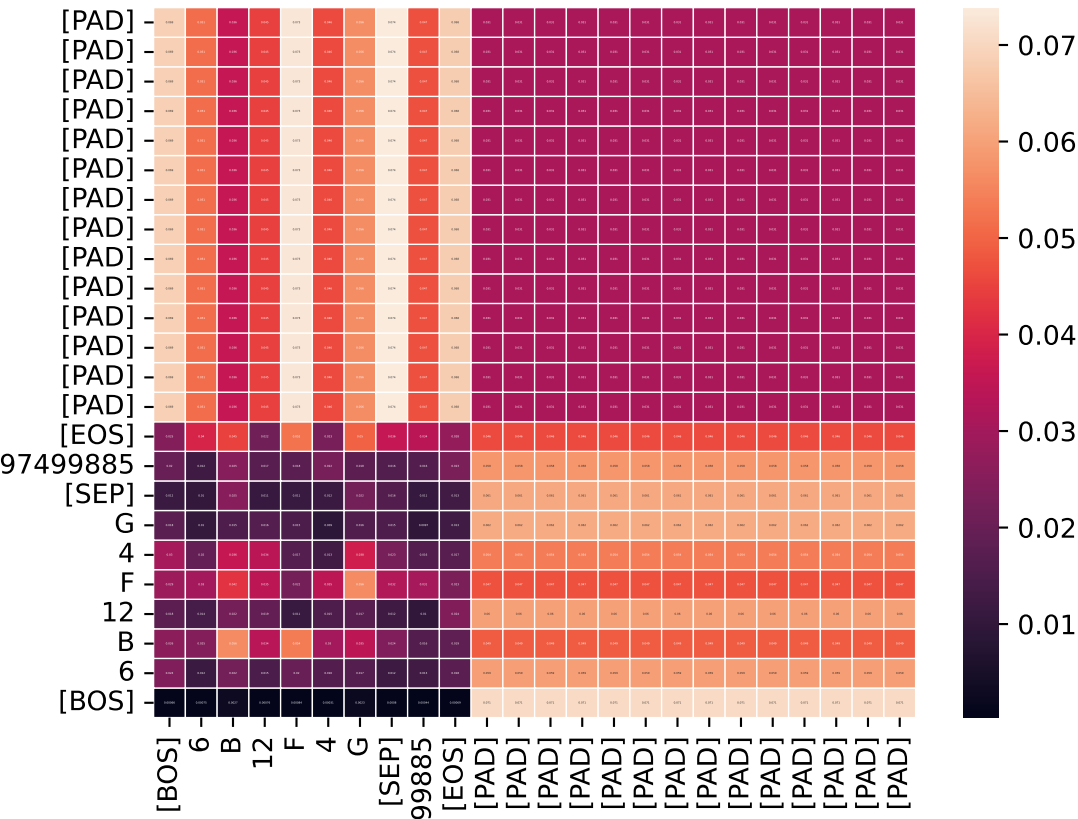
Step 500 attention weights at layer 0 head 2, with accuracy 0.344
 Sequence:[BOS] 6 B 12 F 4 G [SEP] 5.83 [EOS] SOL:B



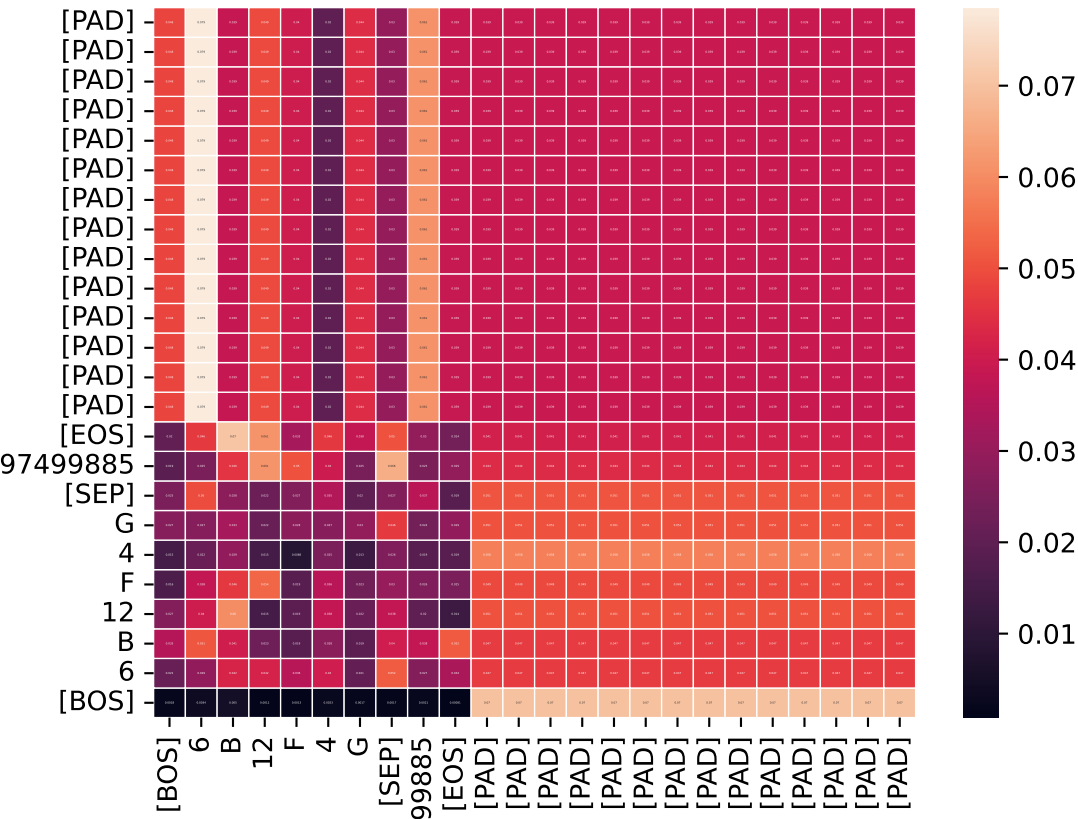
Step 500 attention weights at layer 0 head 3, with accuracy 0.344
Sequence:[BOS] 6 B 12 F 4 G [SEP] 5.83 [EOS] SOL:B



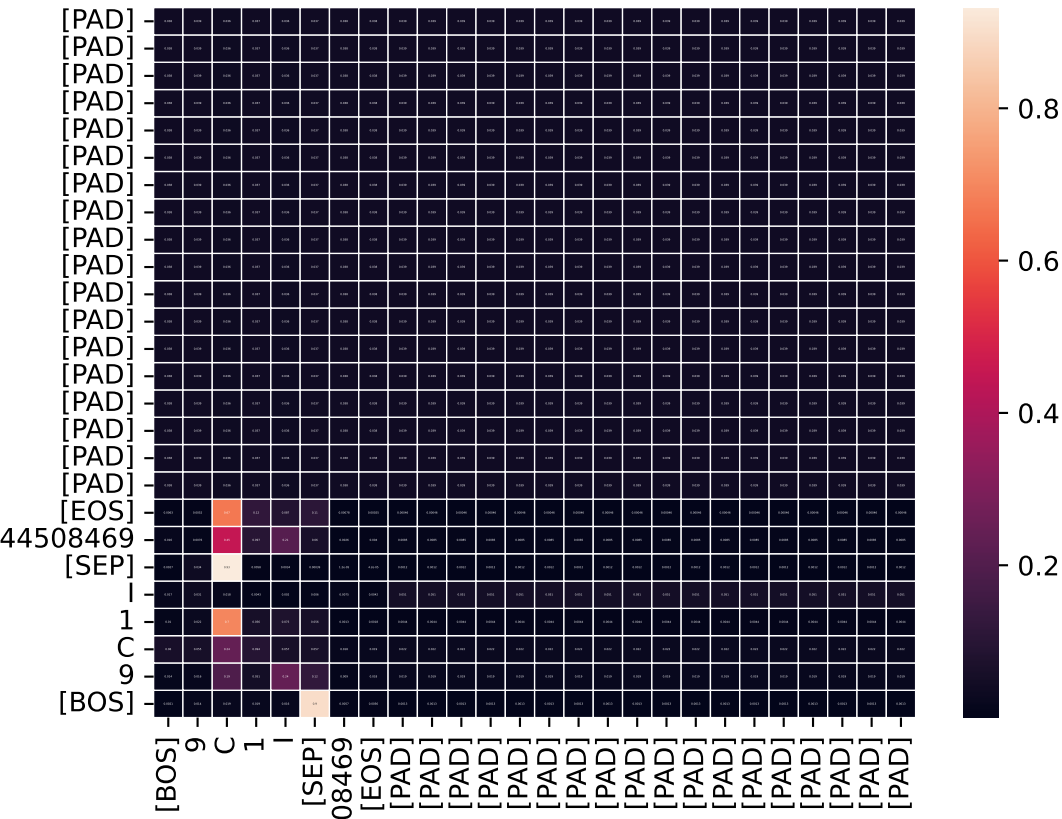
Step 500 attention weights at layer 1 head 2, with accuracy 0.344
Sequence:[BOS] 6 B 12 F 4 G [SEP] 5.83 [EOS] SOL:B



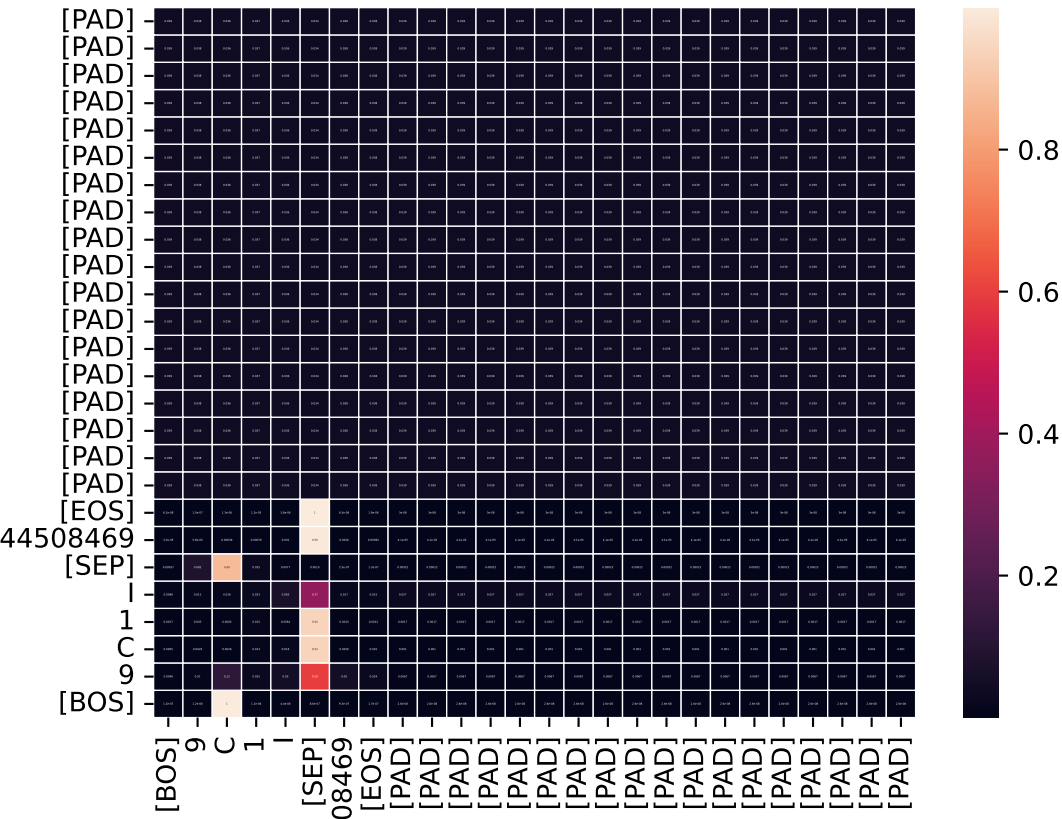
Step 500 attention weights at layer 1 head 3, with accuracy 0.344
Sequence:[BOS] 6 B 12 F 4 G [SEP] 5.83 [EOS] SOL:B



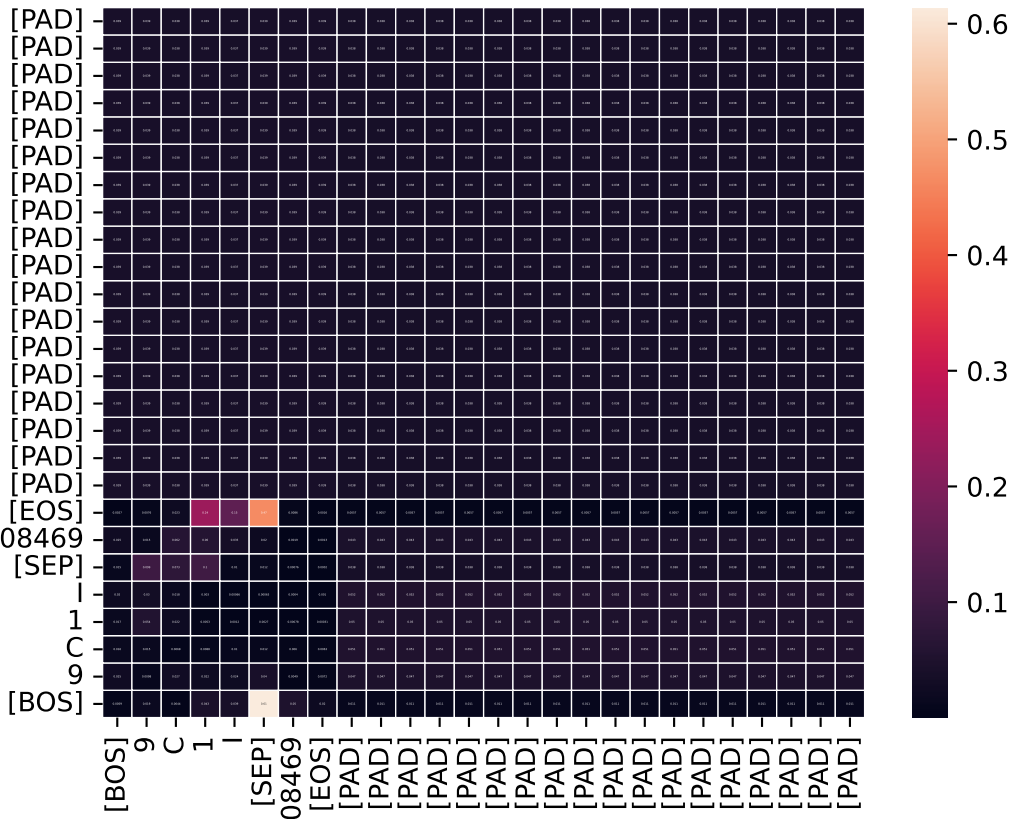
Step 9500 attention weights at layer 0 head 0, with accuracy 0.367
 Sequence:[BOS] 9 C 1 I [SEP] 8.82 [EOS] SOL:C



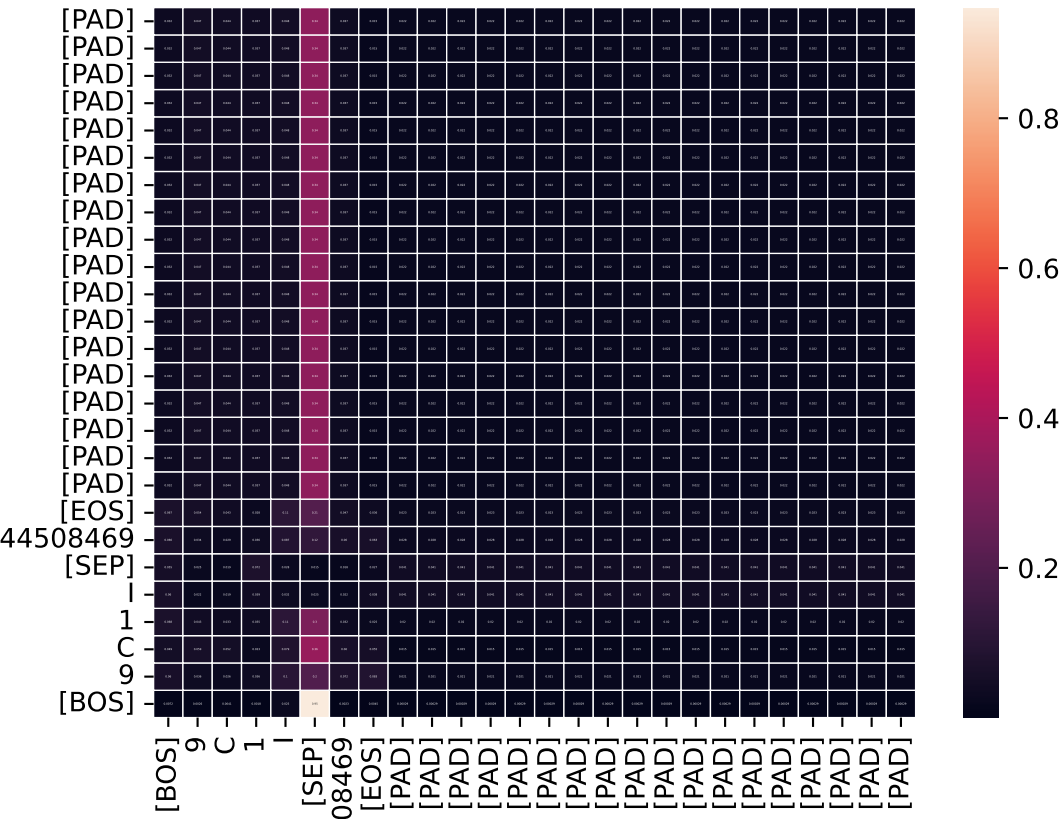
Step 9500 attention weights at layer 0 head 2, with accuracy 0.367
 Sequence:[BOS] 9 C 1 I [SEP] 8.82 [EOS] SOL:C



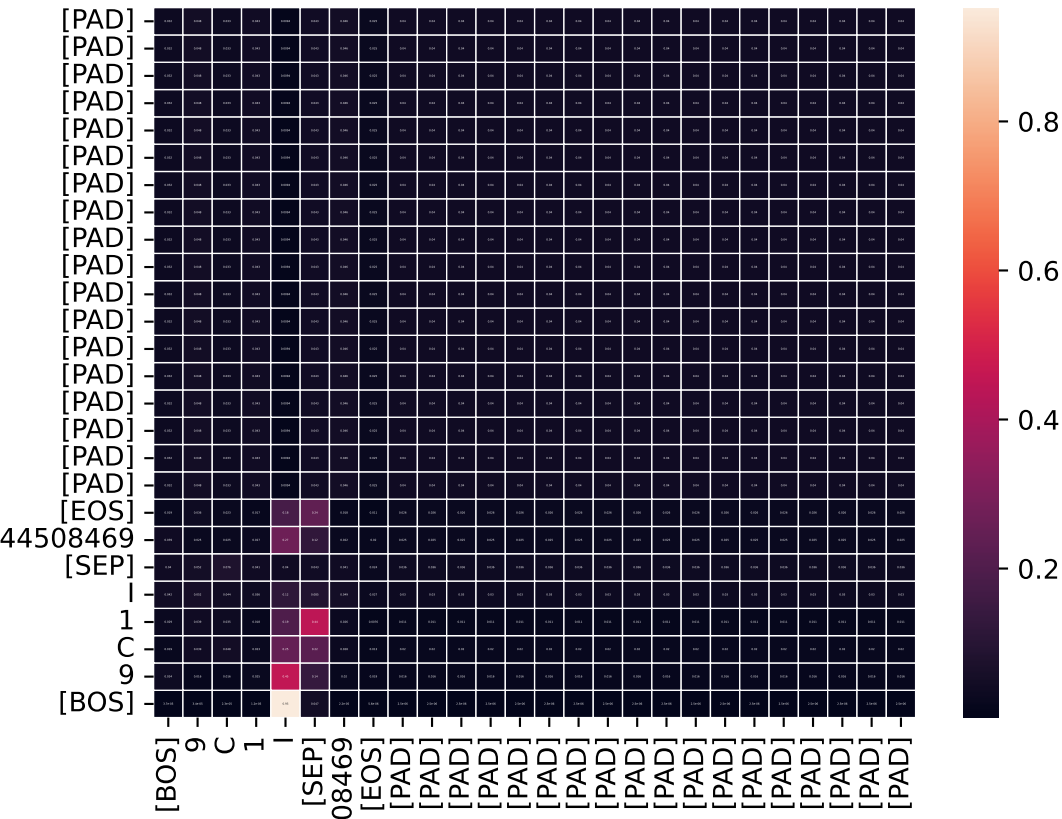
44508469



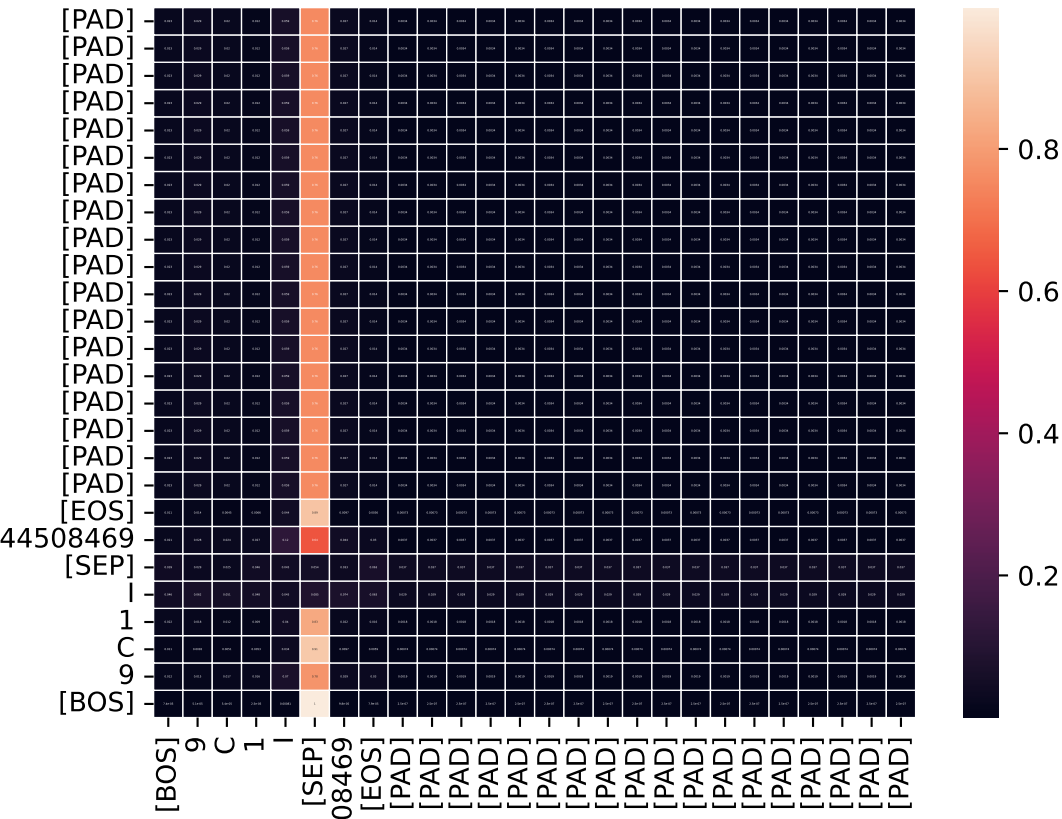
Step 9500 attention weights at layer 1 head 0, with accuracy 0.367
 Sequence:[BOS] 9 C 1 I [SEP] 8.82 [EOS] SOL:C



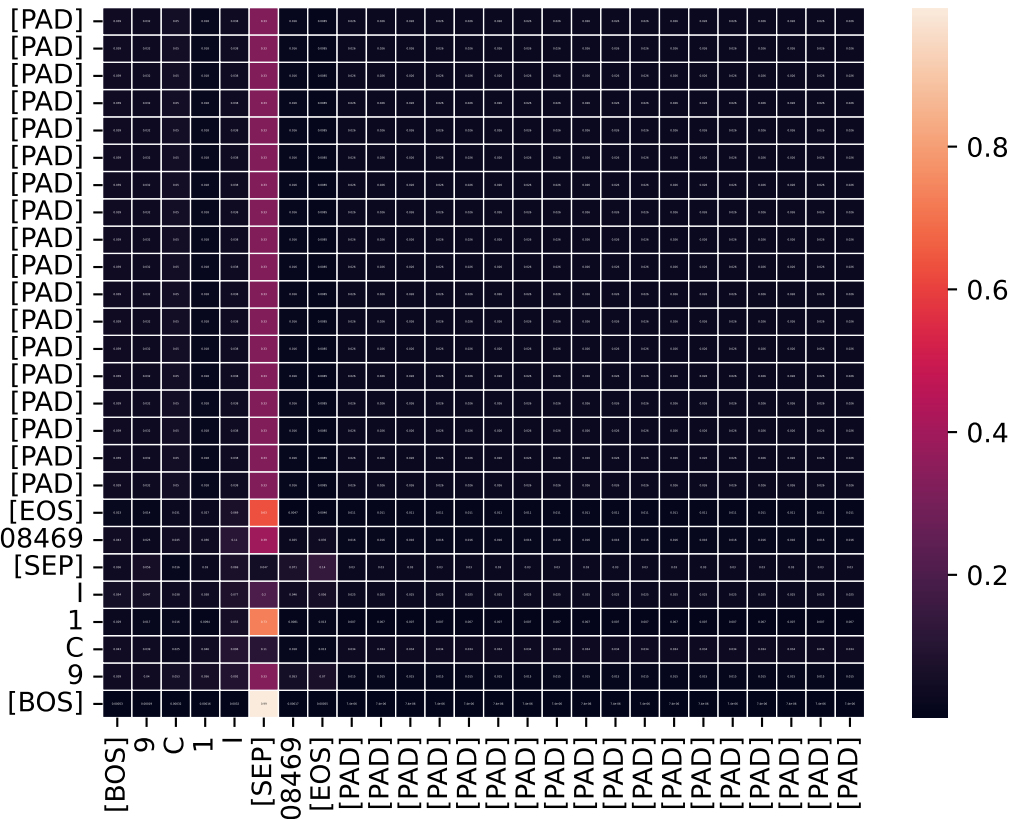
Step 9500 attention weights at layer 1 head 1, with accuracy 0.367
 Sequence:[BOS] 9 C 1 I [SEP] 8.82 [EOS] SOL:C



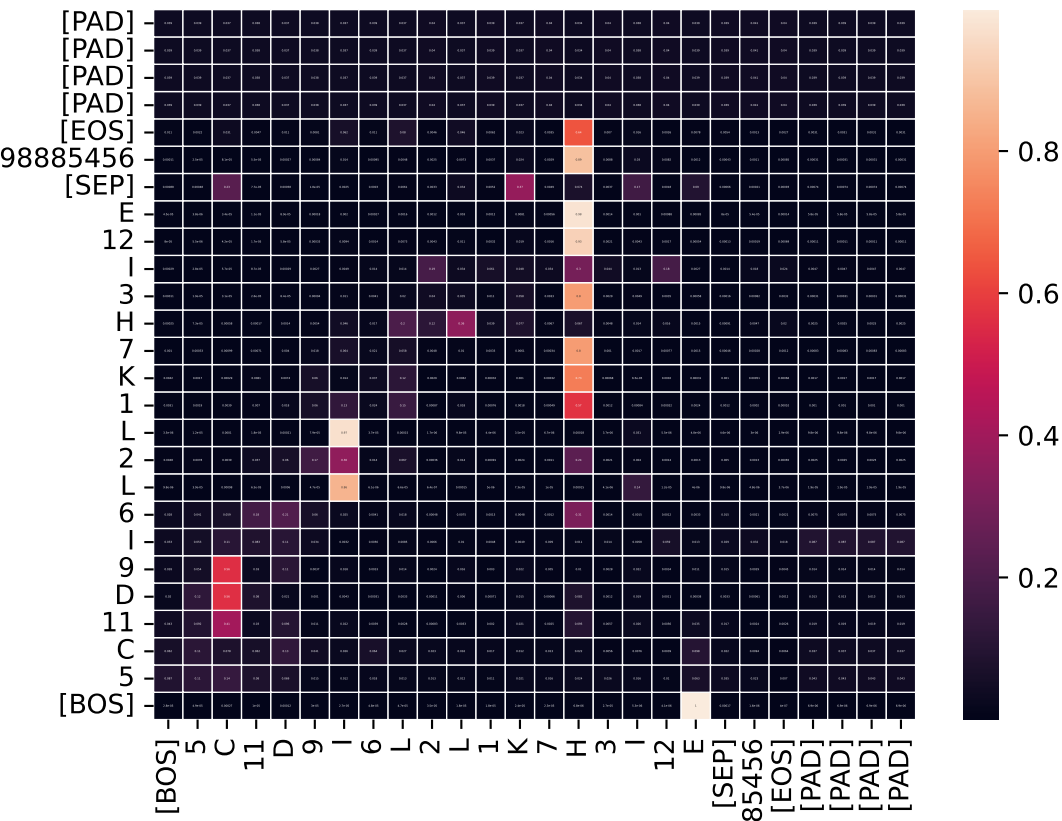
Step 9500 attention weights at layer 1 head 2, with accuracy 0.367
 Sequence:[BOS] 9 C 1 I [SEP] 8.82 [EOS] SOL:C



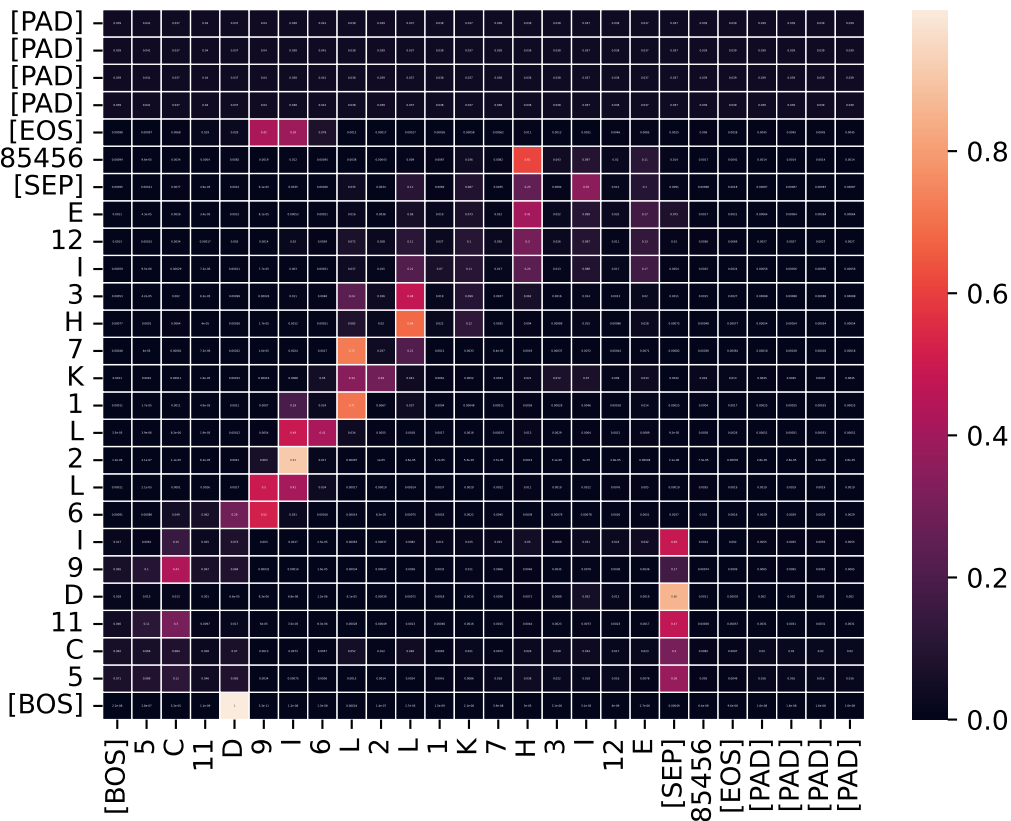
44508469



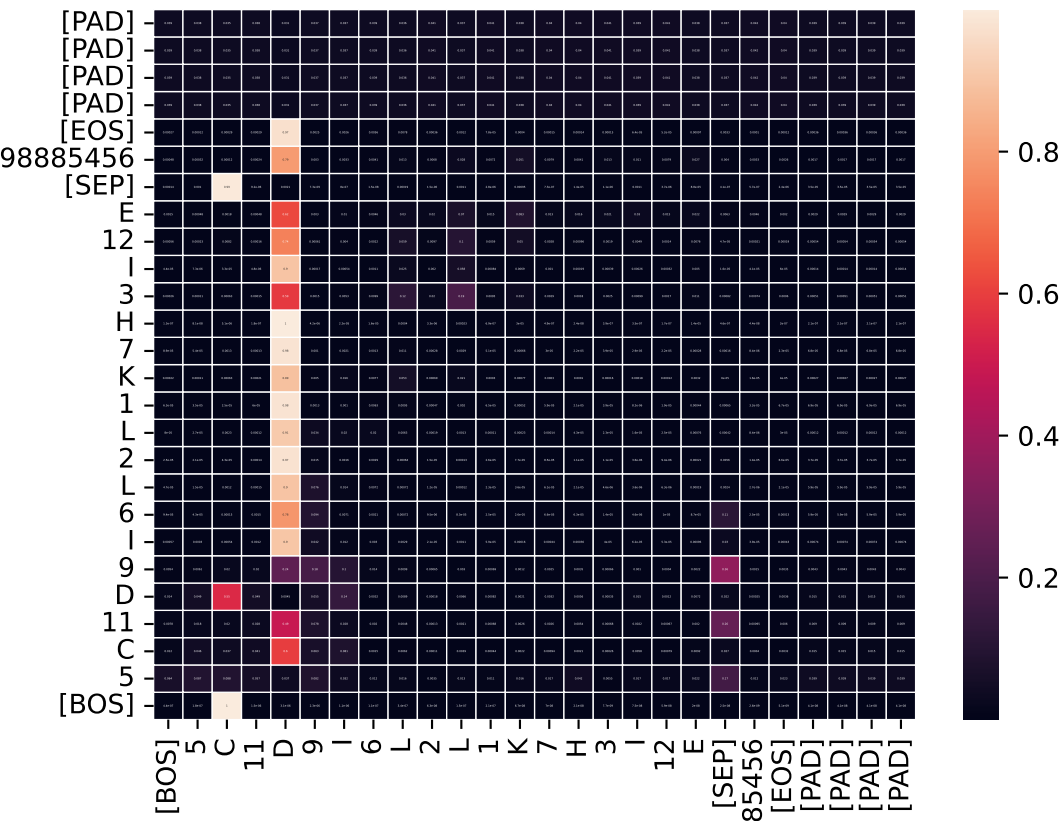
Step 10000 attention weights at layer 0 head 0, with accuracy 0.386
 Sequence:[BOS] 5 C 11 D 9 I 6 L 2 L 1 K 7 H 3 I 12 E [SEP] 1.83 [EOS] SOL:L



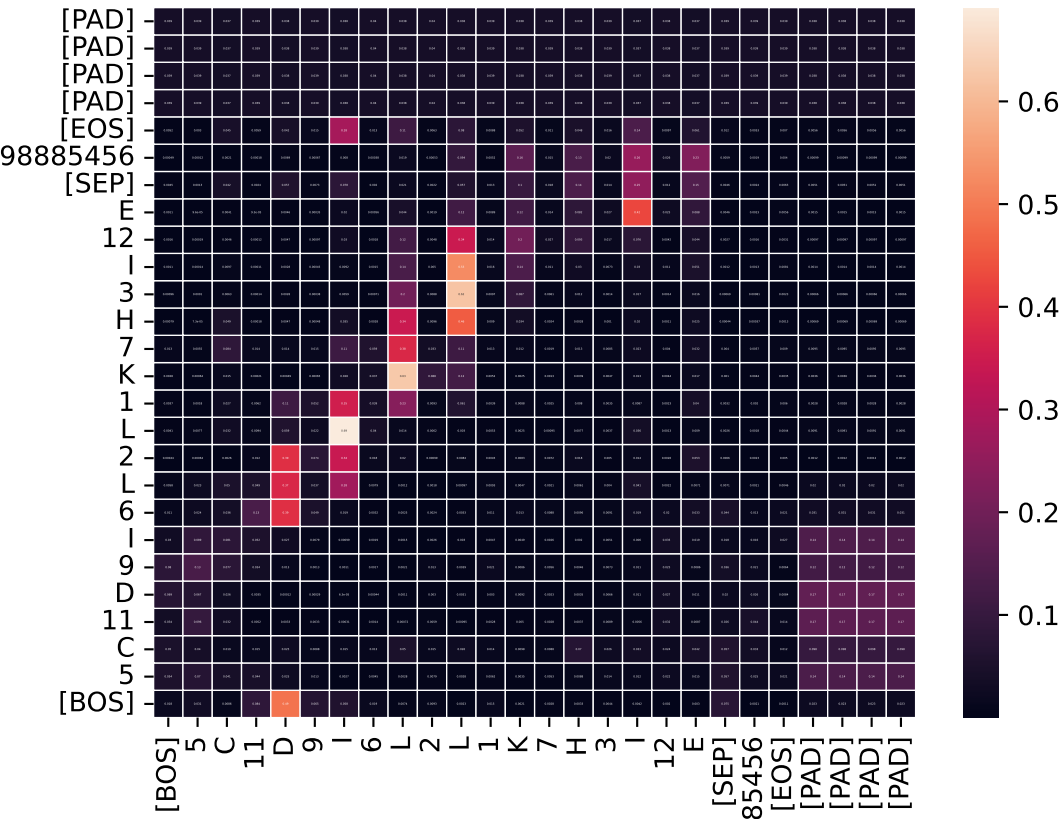
98885456



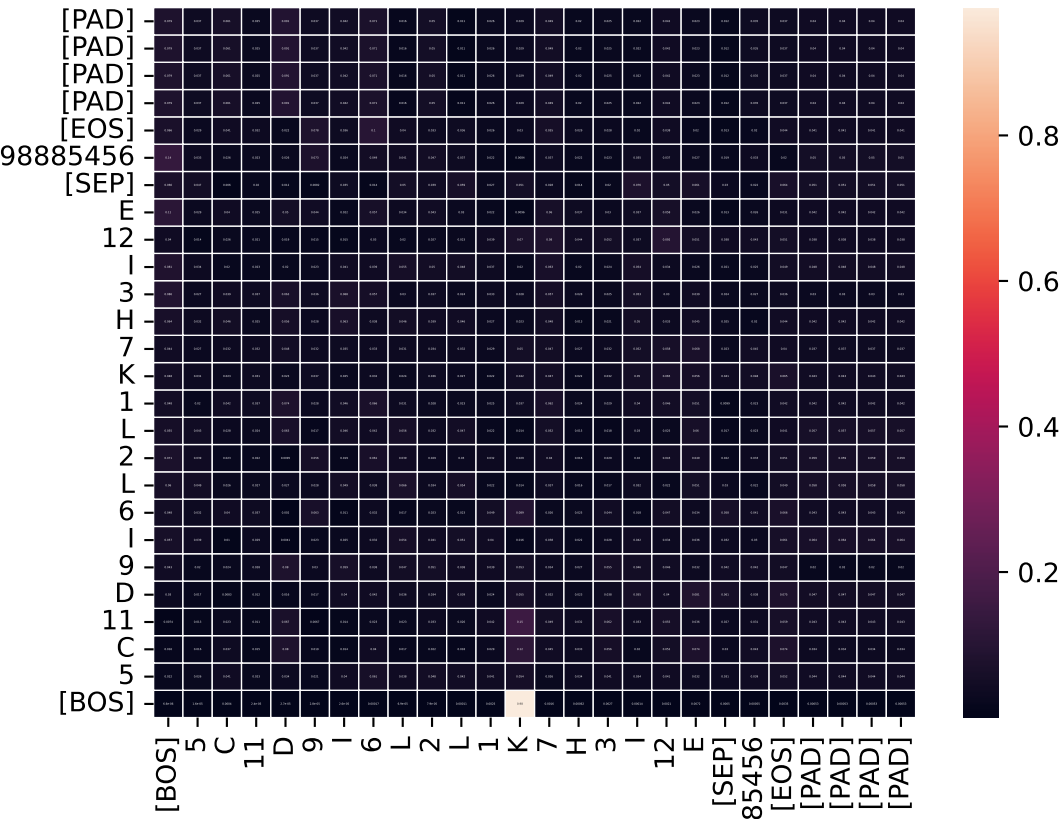
Step 10000 attention weights at layer 0 head 2, with accuracy 0.386
 Sequence:[BOS] 5 C 11 D 9 I 6 L 2 L 1 K 7 H 3 I 12 E [SEP] 1.83 [EOS] SOL:L



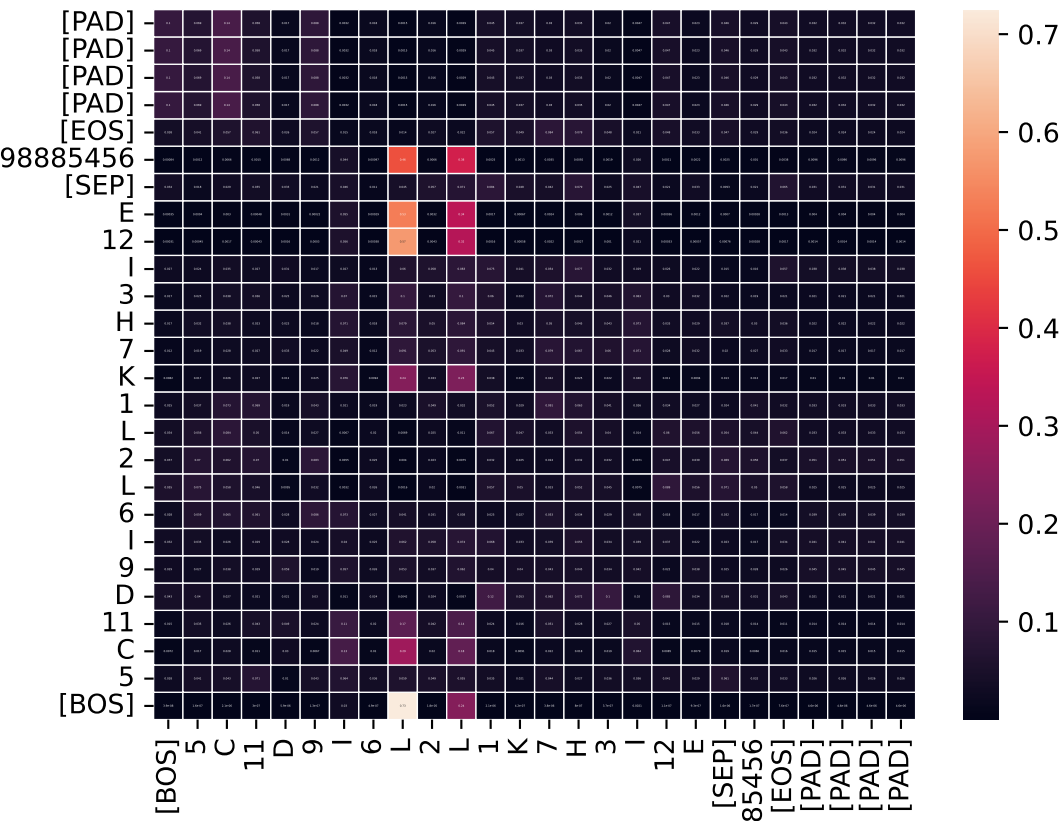
Step 10000 attention weights at layer 0 head 3, with accuracy 0.386
 Sequence:[BOS] 5 C 11 D 9 I 6 L 2 L 1 K 7 H 3 I 12 E [SEP] 1.83 [EOS] SOL:L



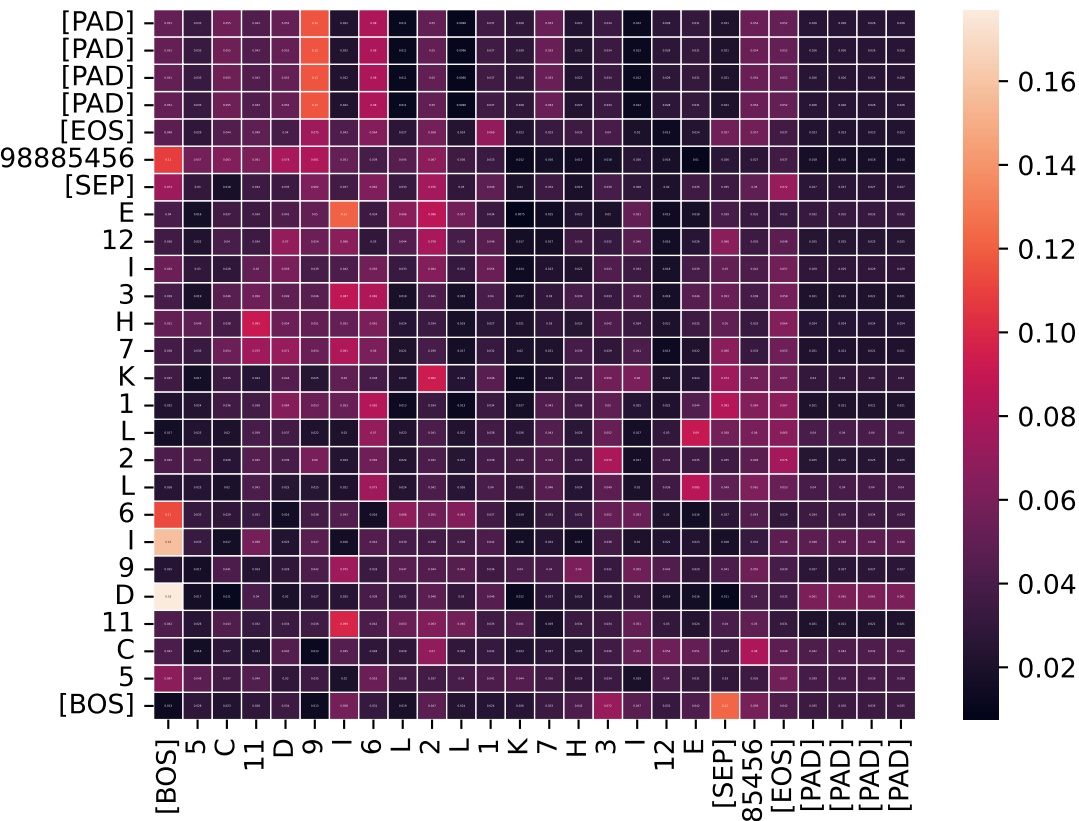
Step 10000 attention weights at layer 1 head 0, with accuracy 0.386
Sequence: [BOS] 5 C 11 D 9 16 L 2 L 1 K 7 H 3 112 E [SEP] 1.83 [EOS] SOL: L



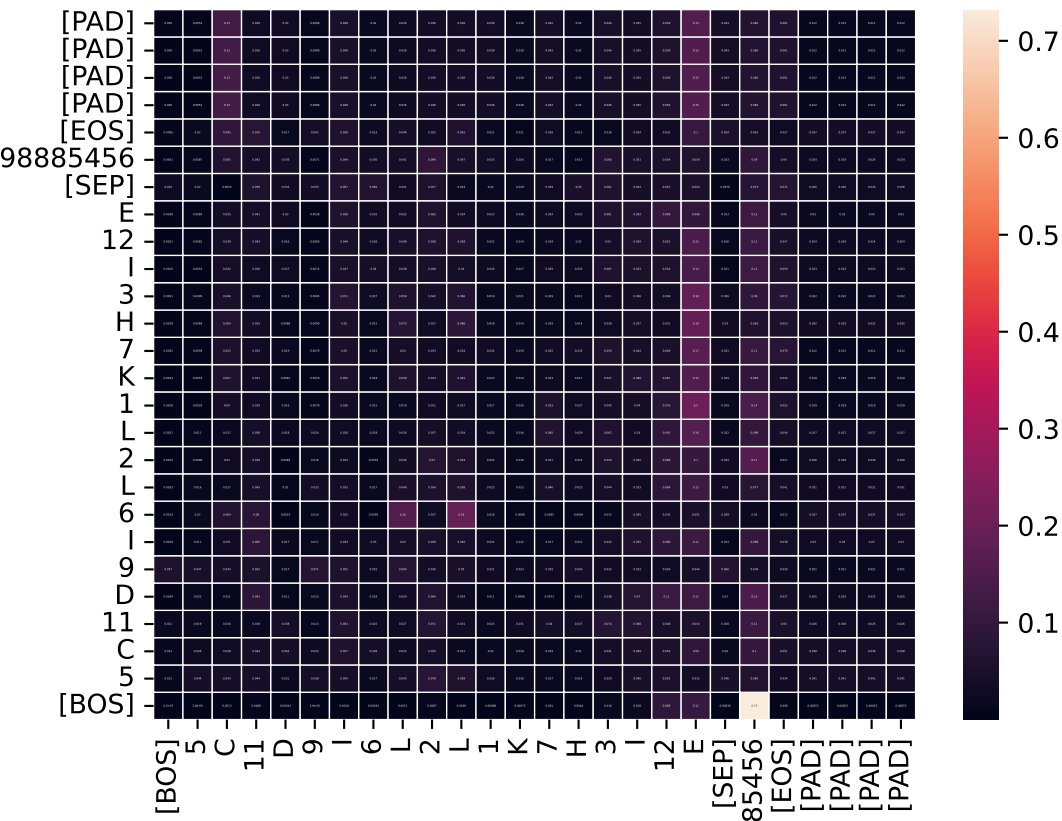
Step 10000 attention weights at layer 1 head 1, with accuracy 0.386
 Sequence:[BOS] 5 C 11 D 9 I 6 L 2 L 1 K 7 H 3 I 12 E [SEP] 1.83 [EOS] SOL:L



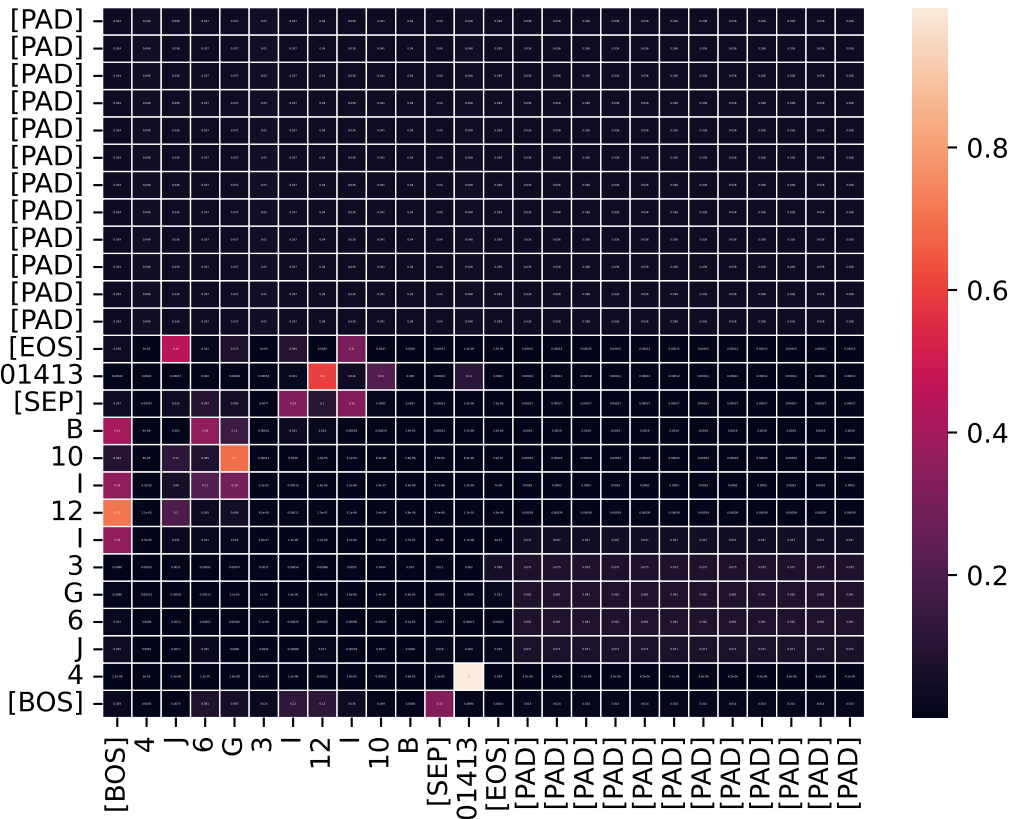
Step 10000 attention weights at layer 1 head 2, with accuracy 0.386
Sequence:[BOS] 5 C 11 D 9 1 6 L 2 L 1 K 7 H 3 1 12 E [SEP] 1.83 [EOS] SOL:L



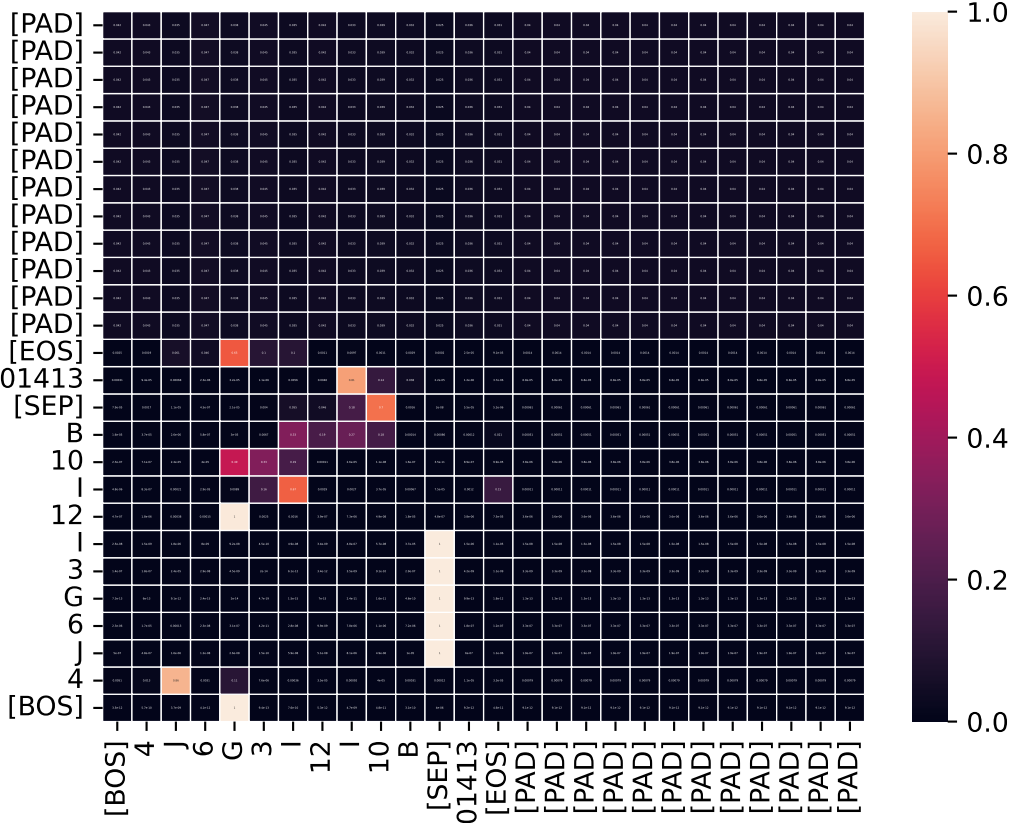
Step 10000 attention weights at layer 1 head 3, with accuracy 0.386
 Sequence:[BOS] 5 C 11 D 9 1 6 L 2 L 1 K 7 H 3 1 12 E [SEP] 1.83 [EOS] SOL:L



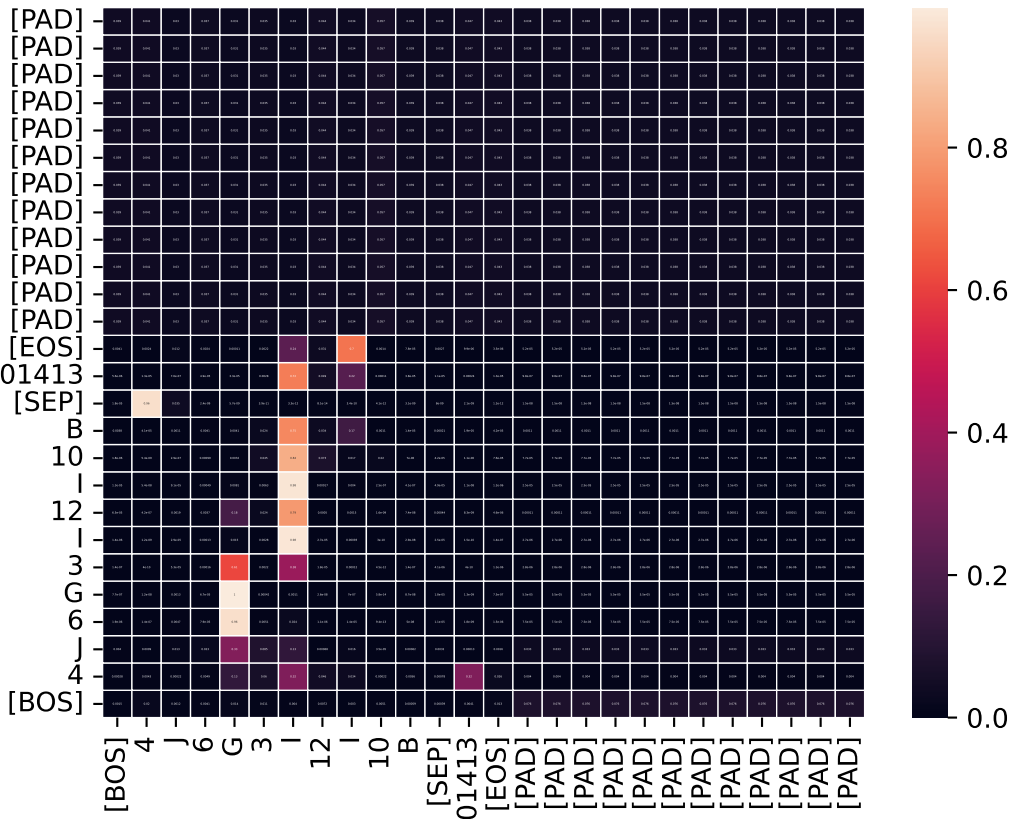
760



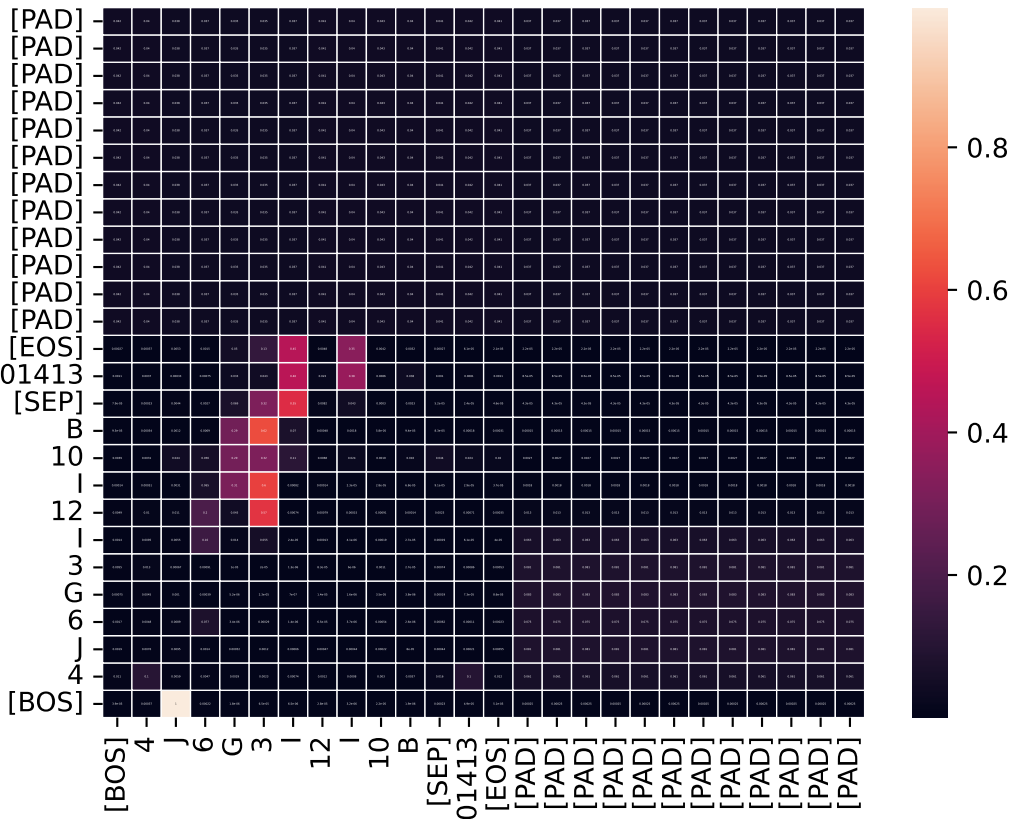
760



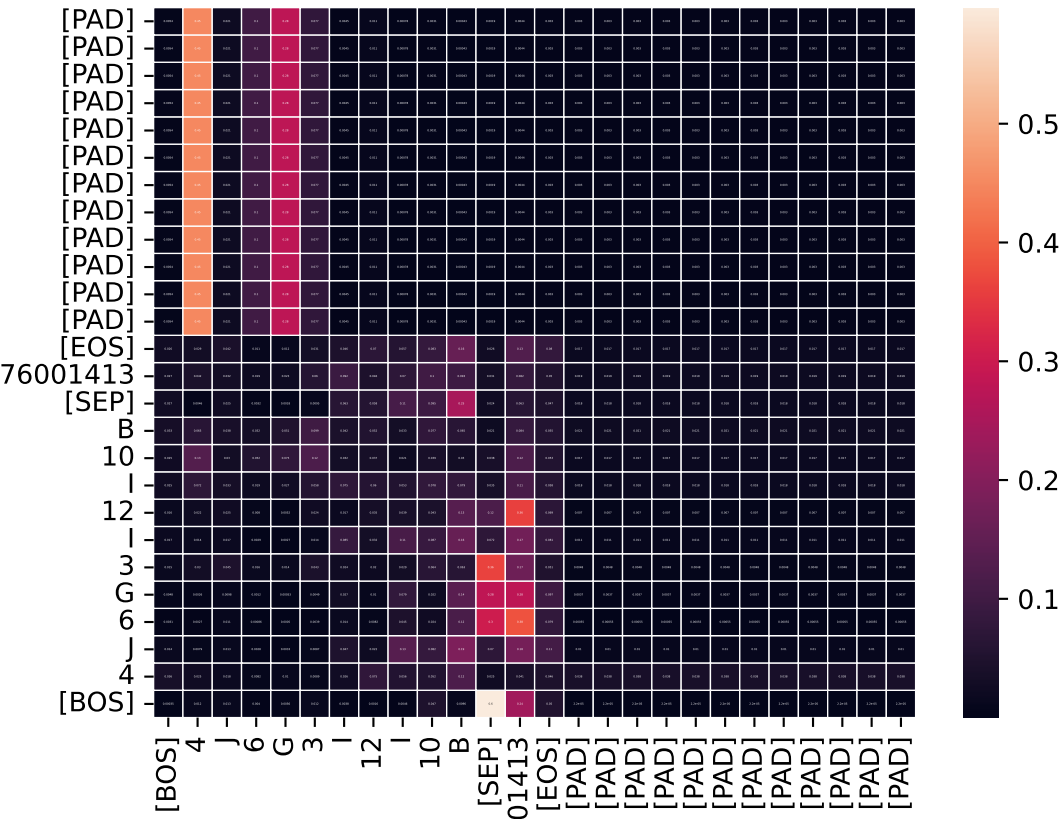
76001413



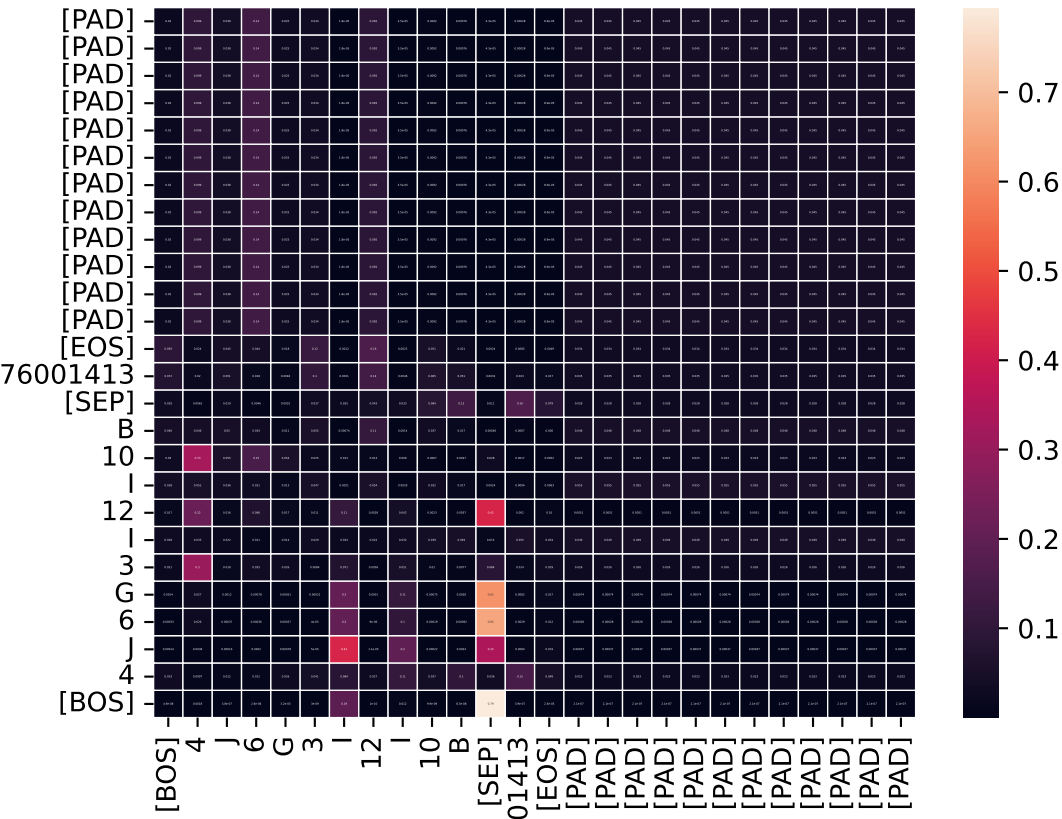
76001413



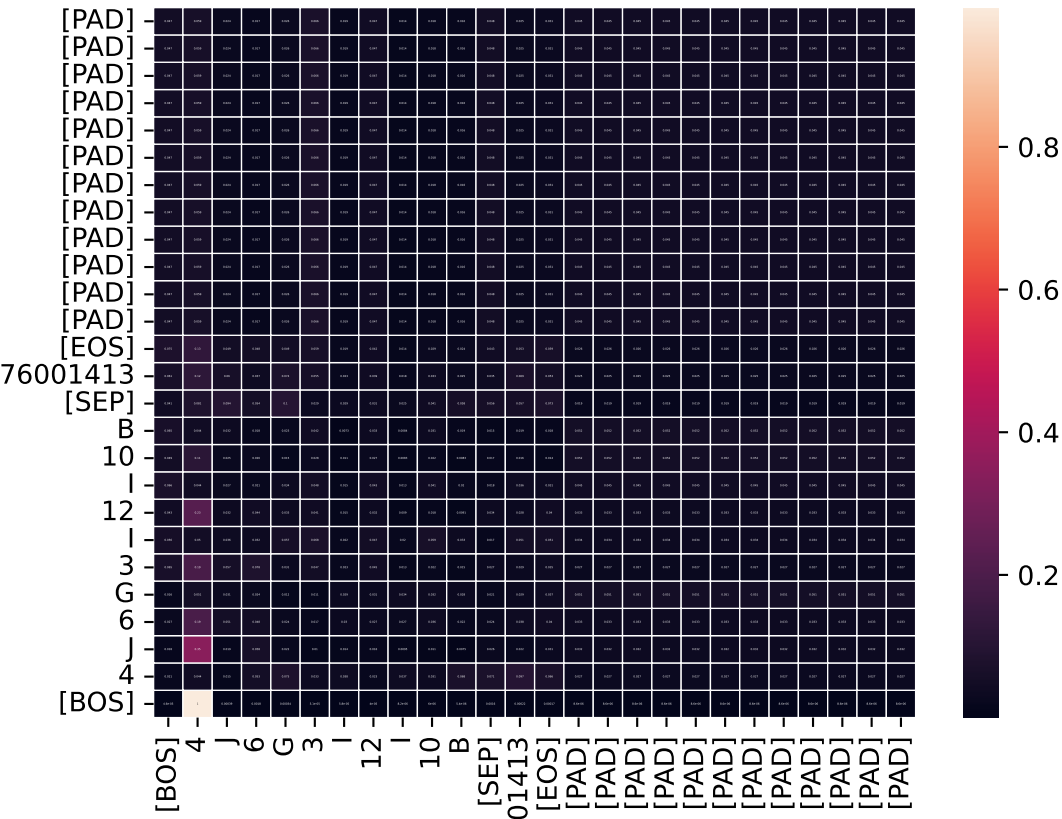
Step 18000 attention weights at layer 1 head 0, with accuracy 0.588
 Sequence:[BOS] 4 J 6 G 3 I 12 I 10 B [SEP] 3.98 [EOS] SOL:]



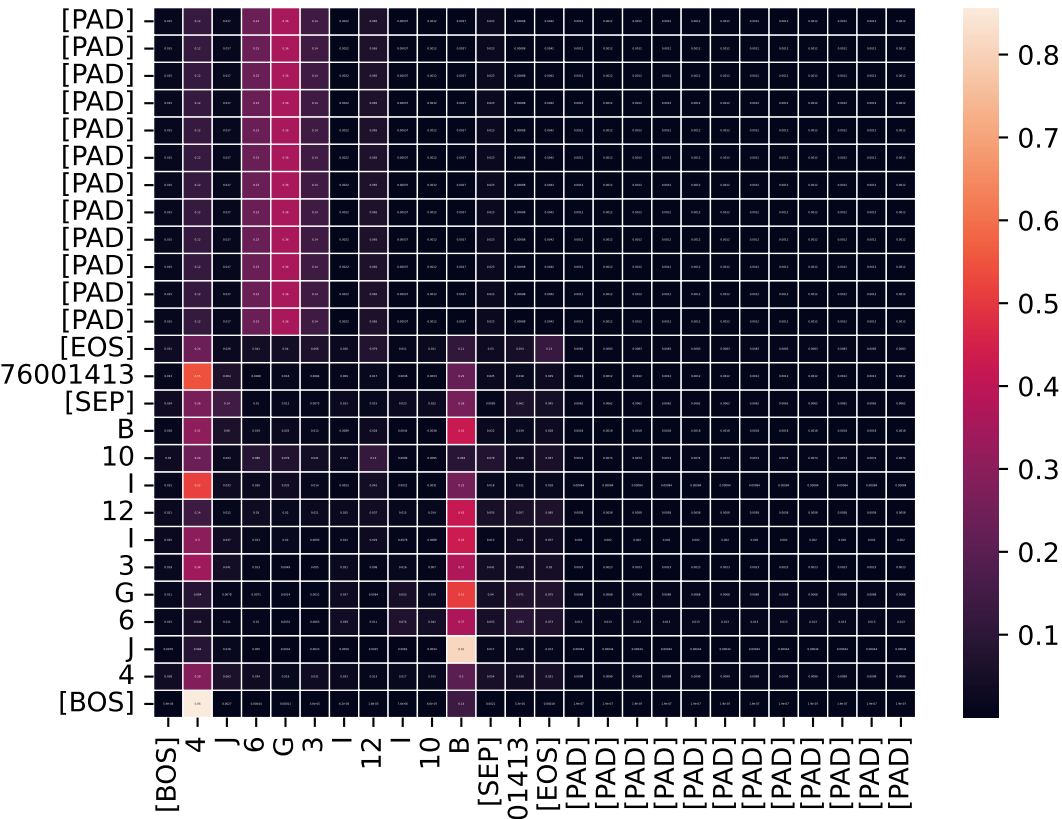
Step 18000 attention weights at layer 1 head 1, with accuracy 0.588
Sequence:[BOS] 4 J 6 G 3 I 12 I 10 B [SEP] 3.98 [EOS] SOL:]



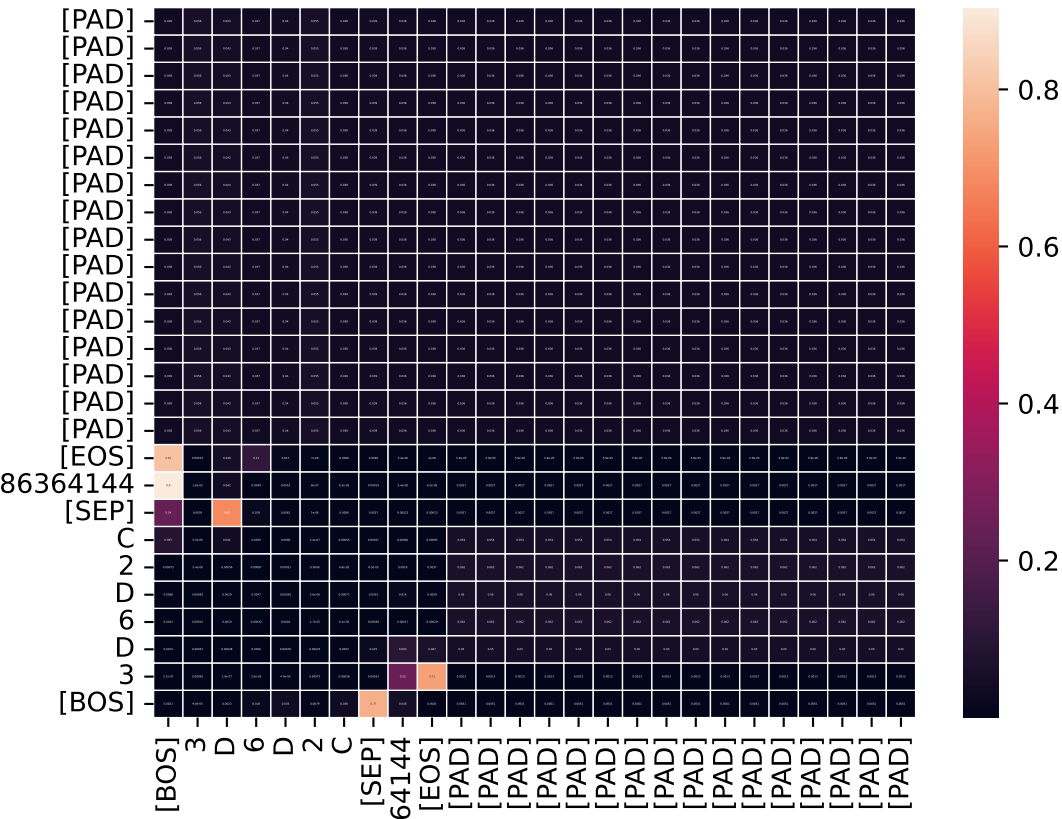
Step 18000 attention weights at layer 1 head 2, with accuracy 0.588
 Sequence:[BOS] 4 J 6 G 3 I 12 I 10 B [SEP] 3.98 [EOS] SOL:]



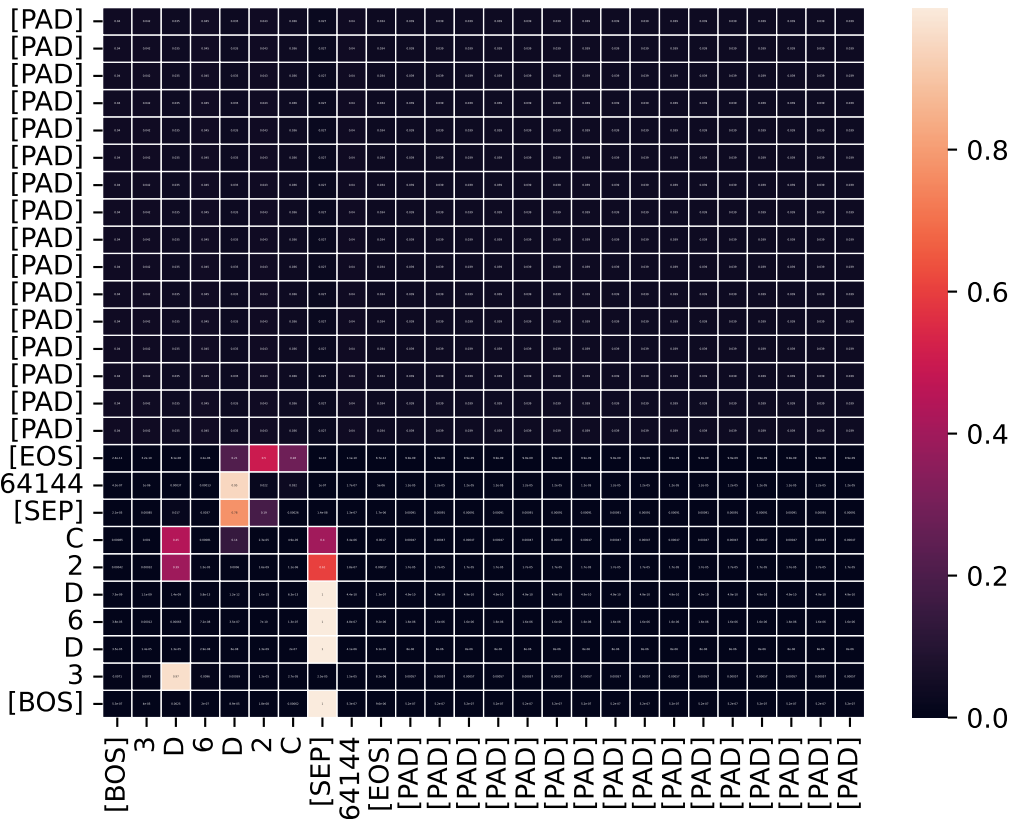
Step 18000 attention weights at layer 1 head 3, with accuracy 0.588
Sequence:[BOS] 4 J 6 G 3 I 12 I 10 B [SEP] 3.98 [EOS] SOL:]



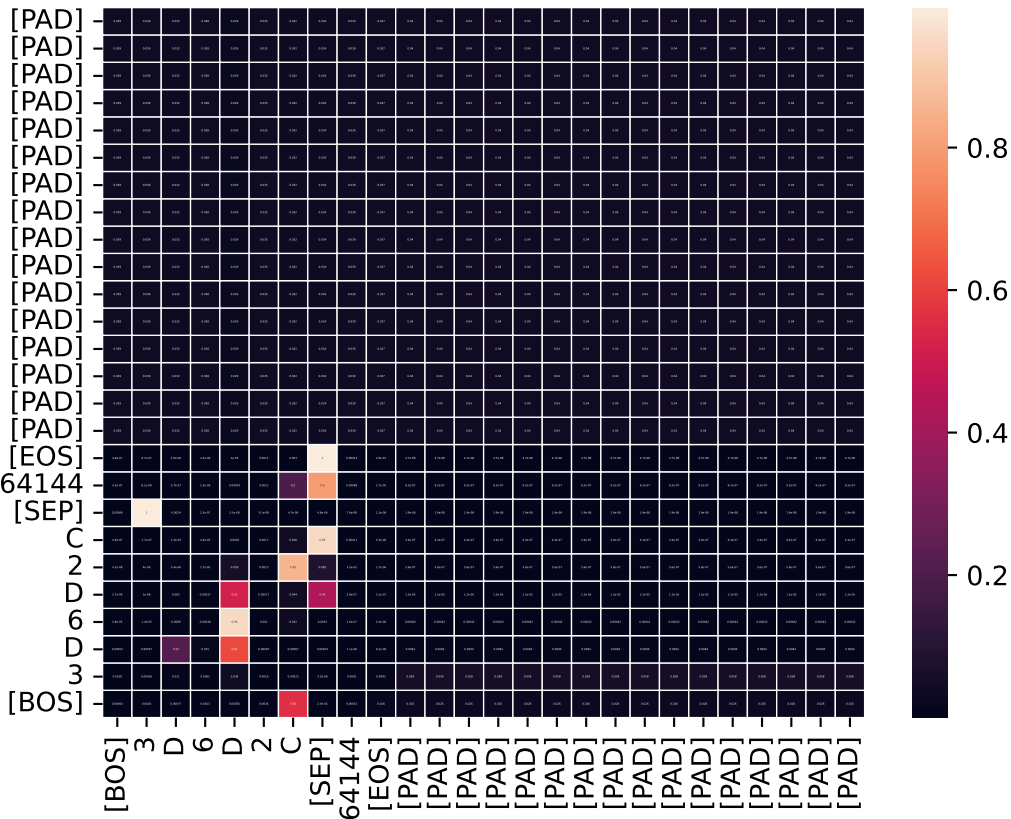
Step 18500 attention weights at layer 0 head 0, with accuracy 0.785
 Sequence:[BOS] 3 D 6 D 2 C [SEP] 6.14 [EOS] SOL:D

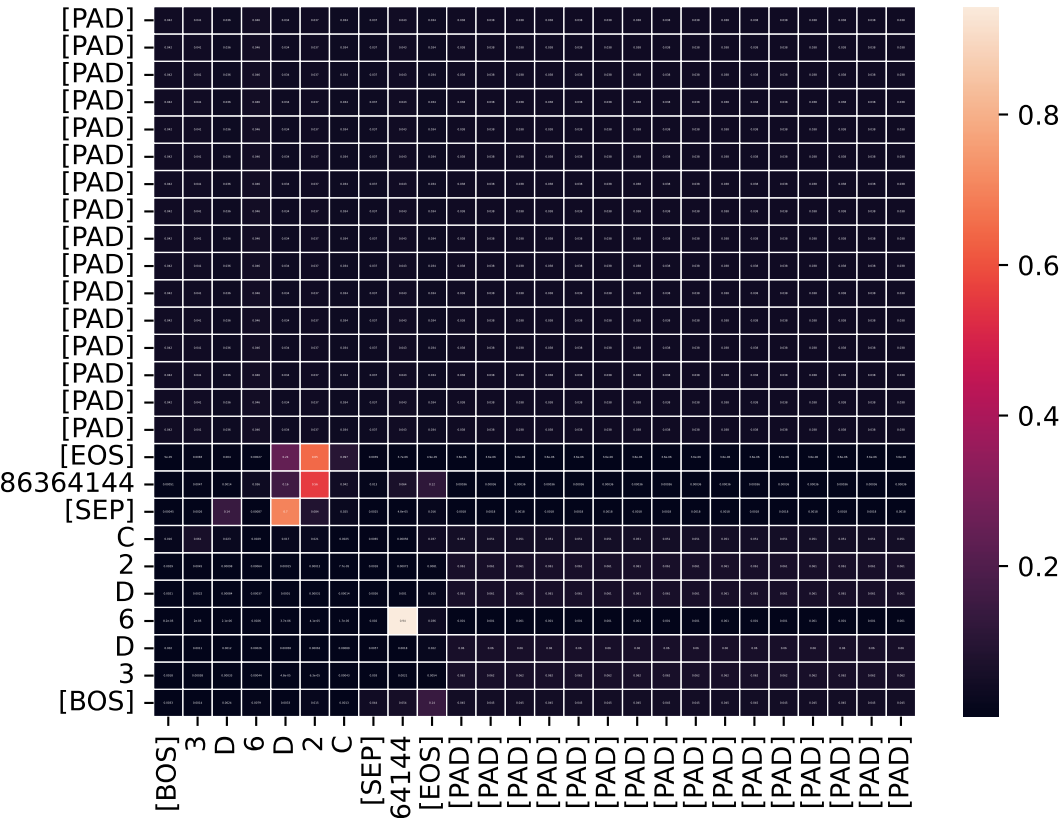


86364144

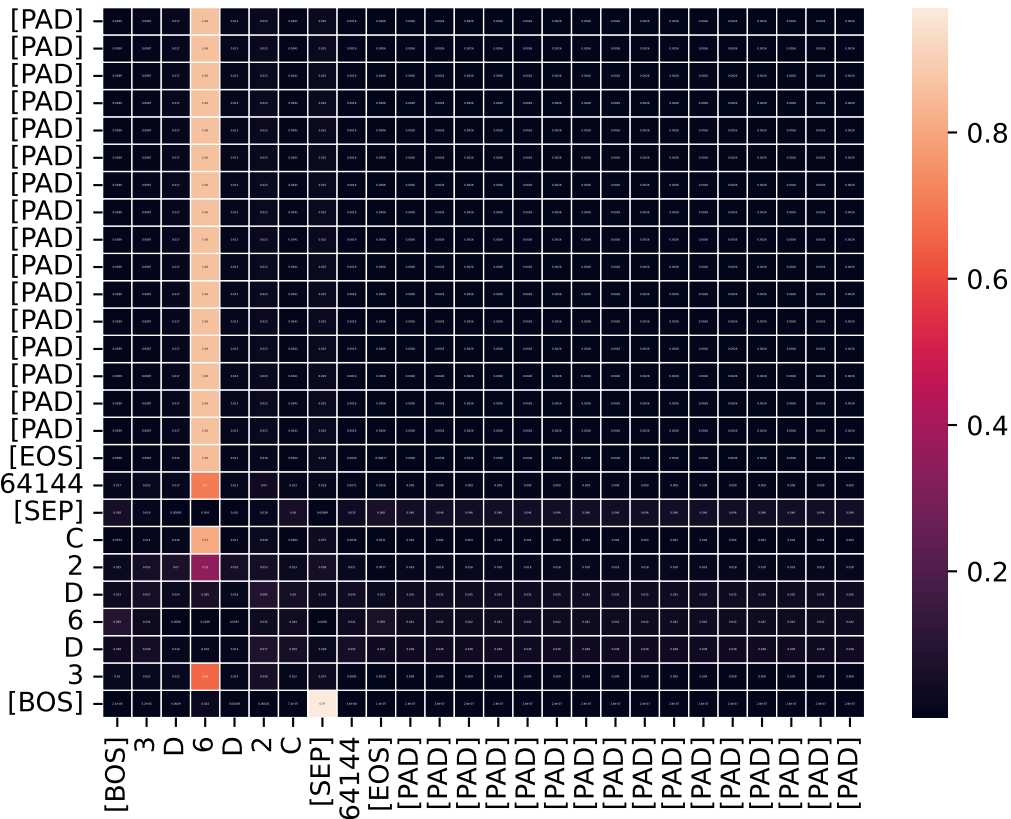


86364144

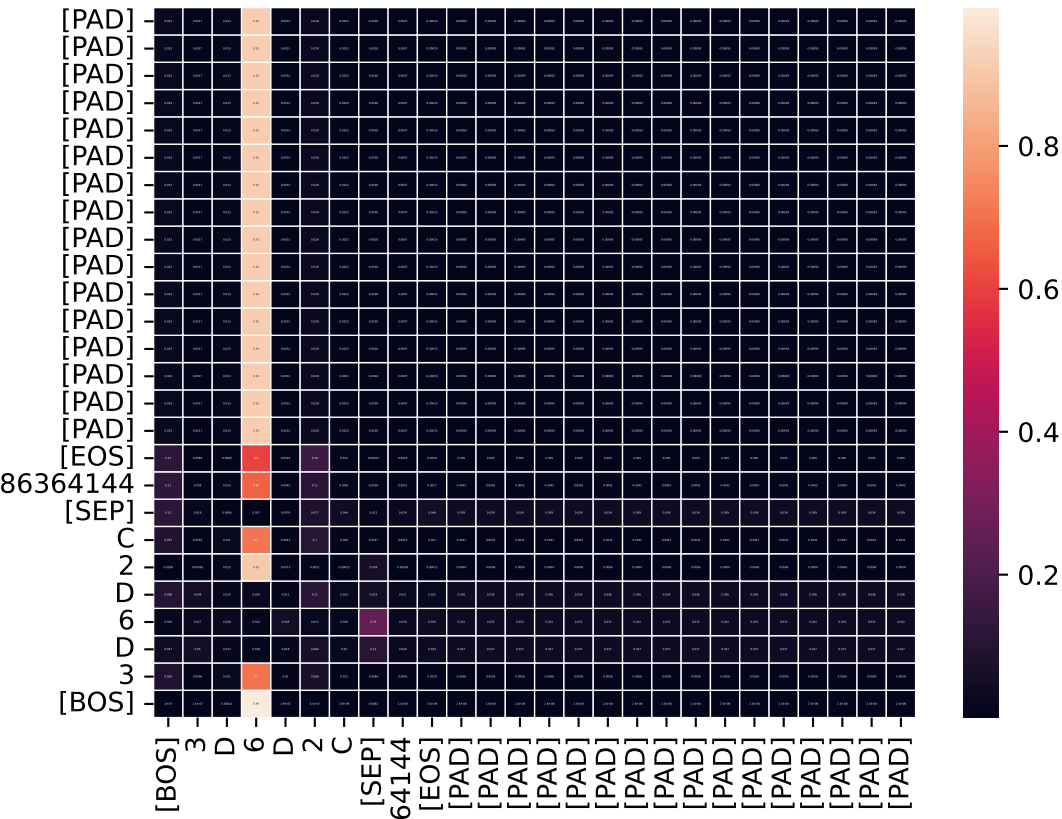


[illegible]

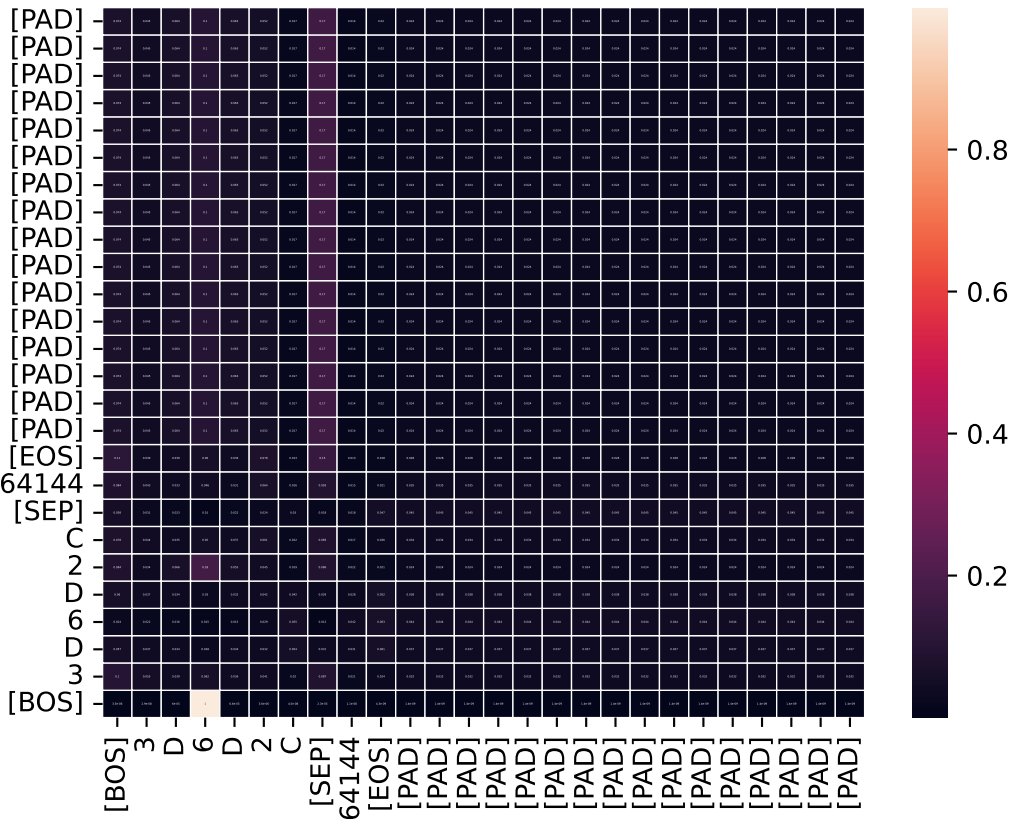
86364144



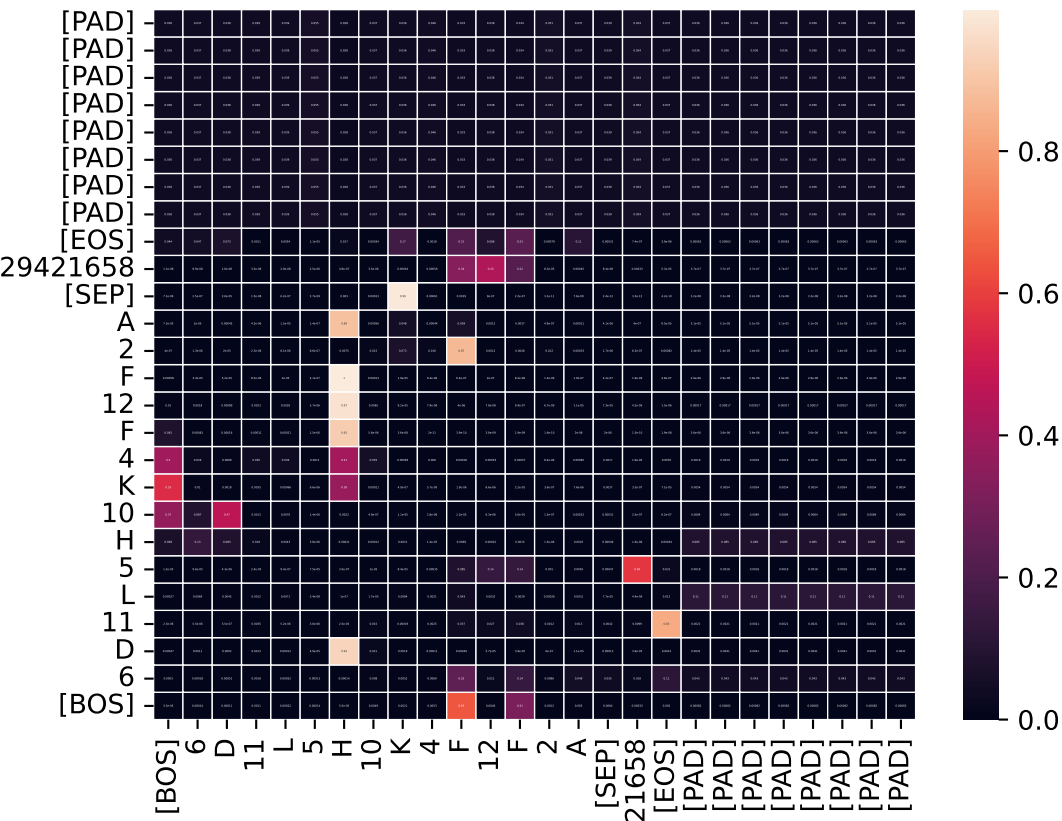
Step 18500 attention weights at layer 1 head 1, with accuracy 0.785
Sequence:[BOS] 3 D 6 D 2 C [SEP] 6.14 [EOS] SOL:D



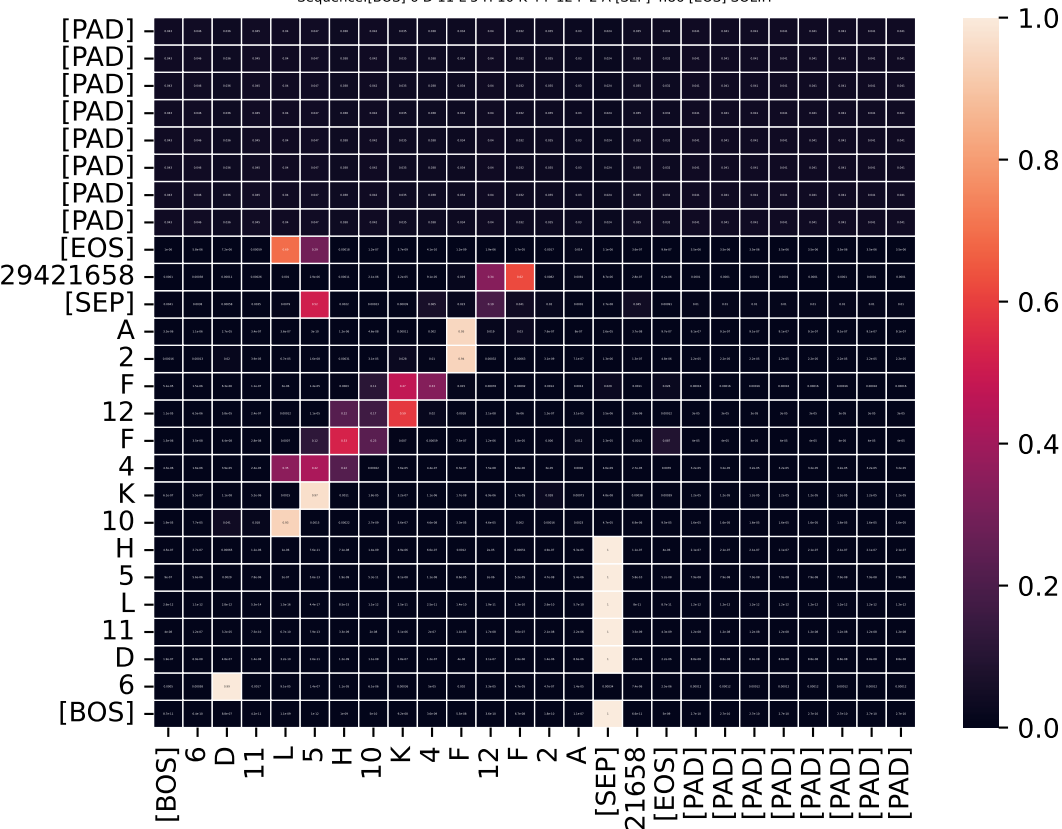
86364144



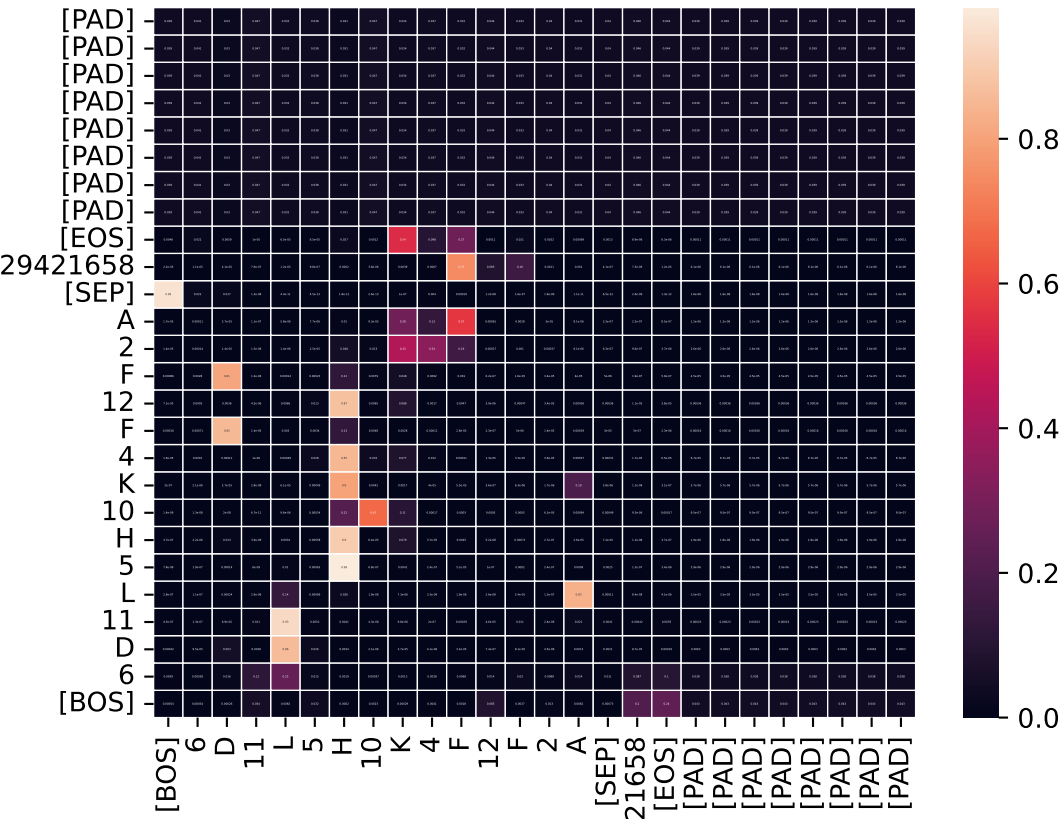
Step 19500 attention weights at layer 0 head 0, with accuracy 0.805
Sequence:[BOS] 6 D 11 L 5 H 10 K 4 F 12 F 2 A [SEP] 4.86 [EOS] SOL:1



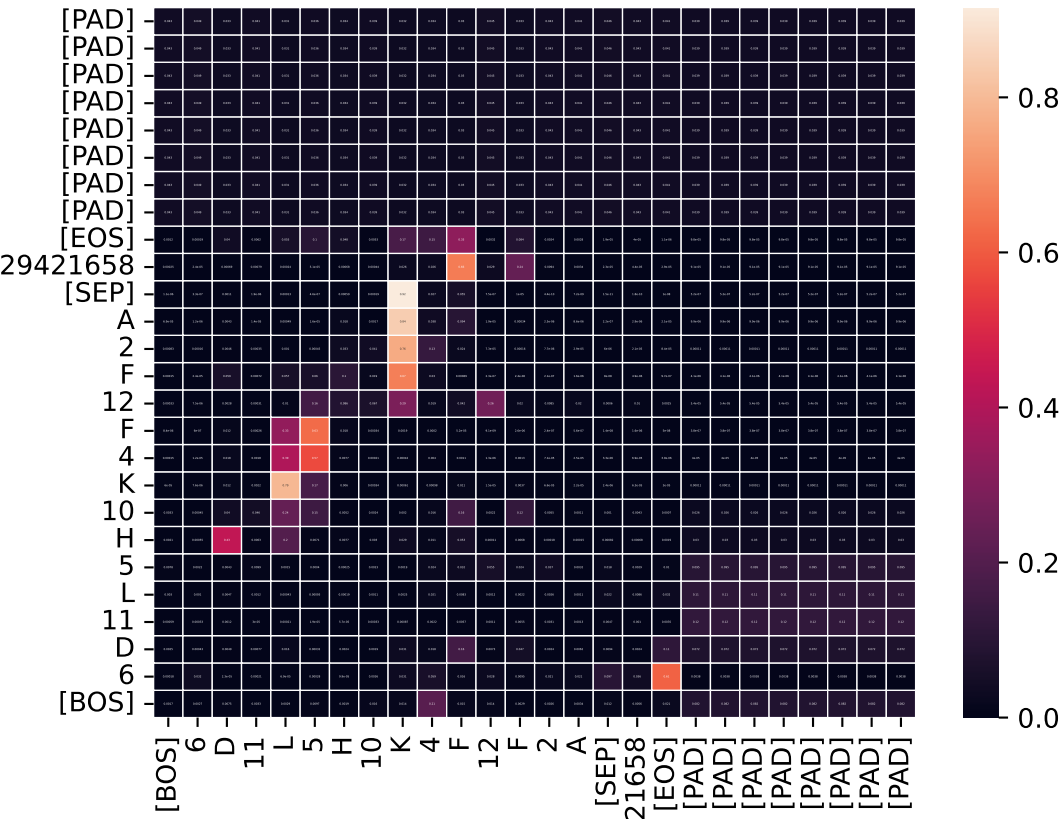
Step 19500 attention weights at layer 0 head 1, with accuracy 0.805
Sequence:[BOS] 6 D 11 L 5 H 10 K 4 F 12 F 2 A [SEP] 4.86 [EOS] SOL:H



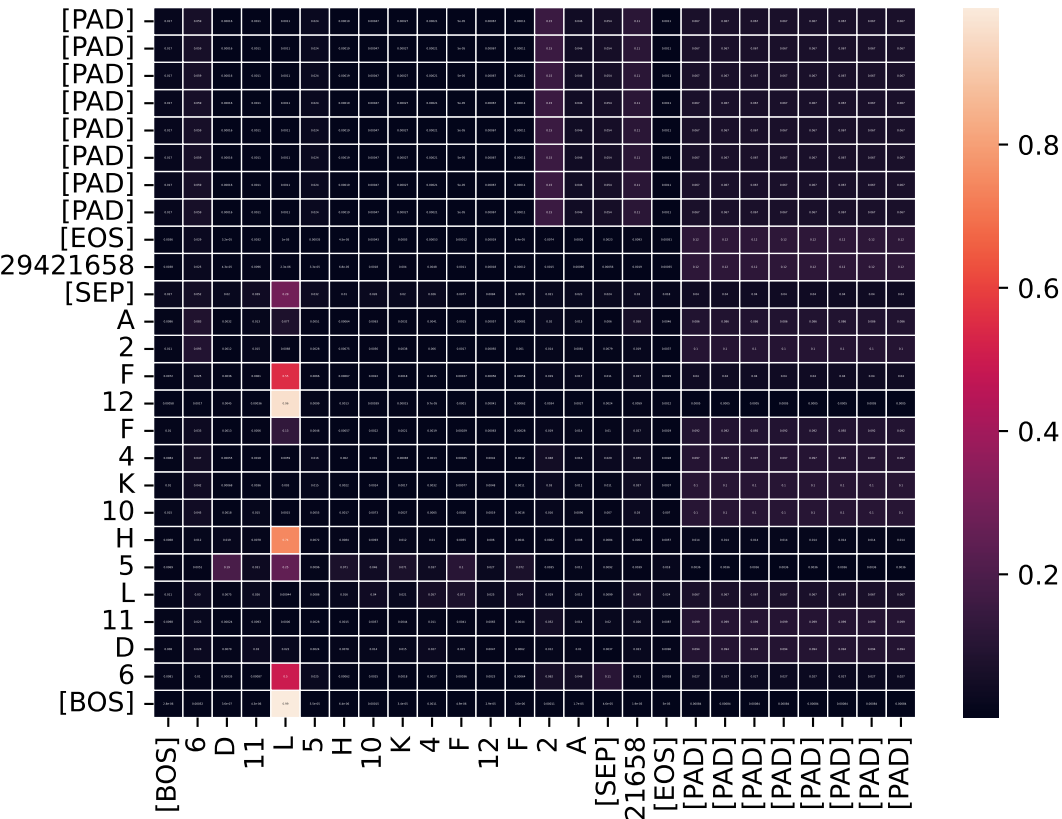
Step 19500 attention weights at layer 0 head 2, with accuracy 0.805
Sequence:[BOS] 6 D 11 L 5 H 10 K 4 F 12 F 2 A [SEP] 4.86 [EOS] SOL:H



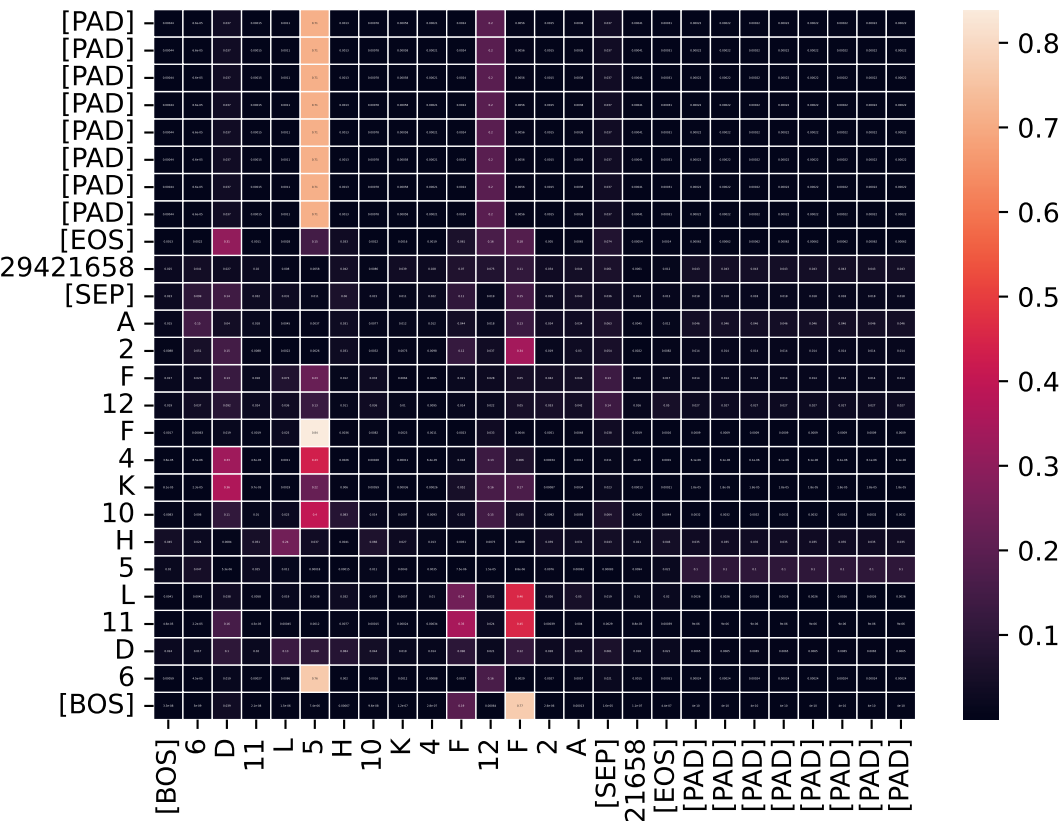
Step 19500 attention weights at layer 0 head 3, with accuracy 0.805
Sequence:[BOS] 6 D 11 L 5 H 10 K 4 F 12 F 2 A [SEP] 4.86 [EOS] SOL:H



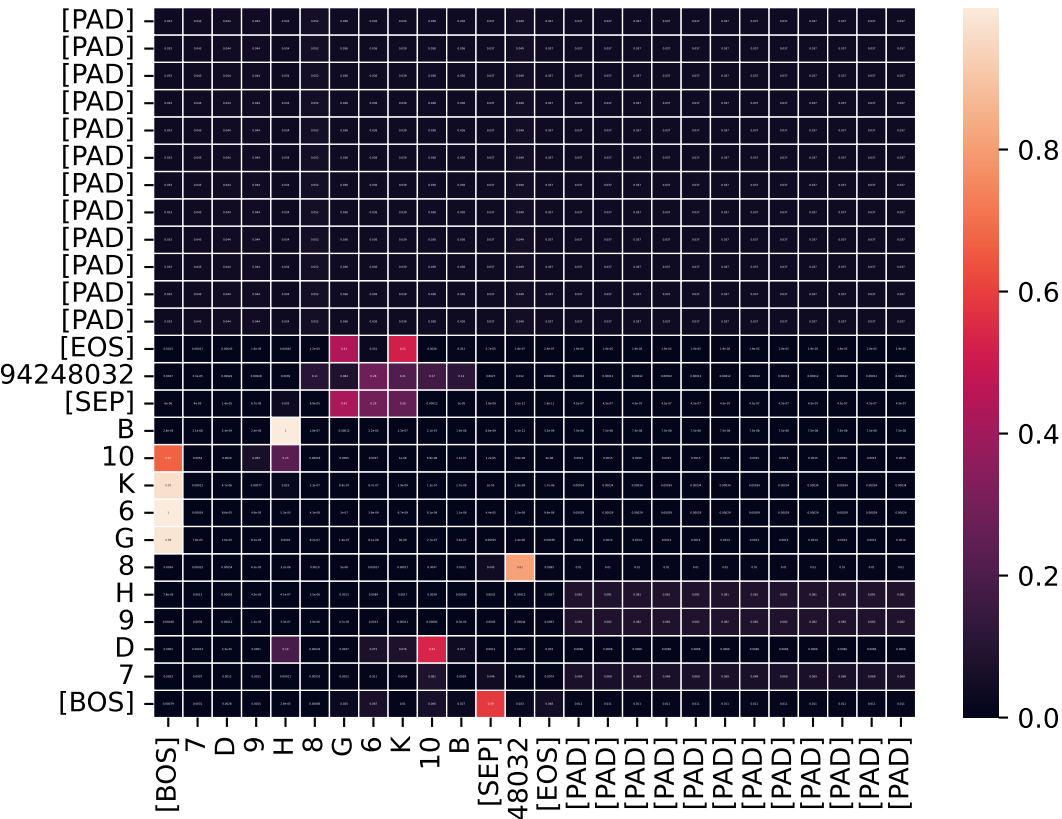
Step 19500 attention weights at layer 1 head 1, with accuracy 0.805
Sequence:[BOS] 6 D 11 L 5 H 10 K 4 F 12 F 2 A [SEP] 4.86 [EOS] SOL:H



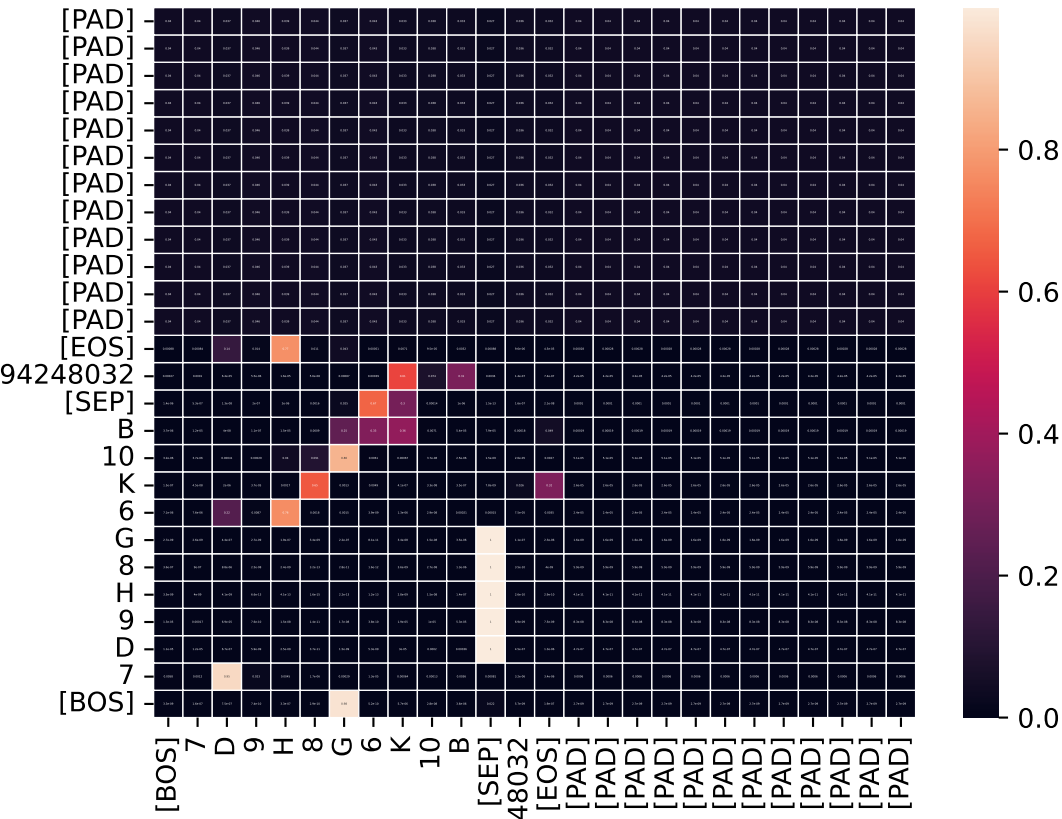
Step 19500 attention weights at layer 1 head 3, with accuracy 0.805
 Sequence:[BOS] 6 D 11 L 5 H 10 K 4 F 12 F 2 A [SEP] 4.86 [EOS] SOL:H

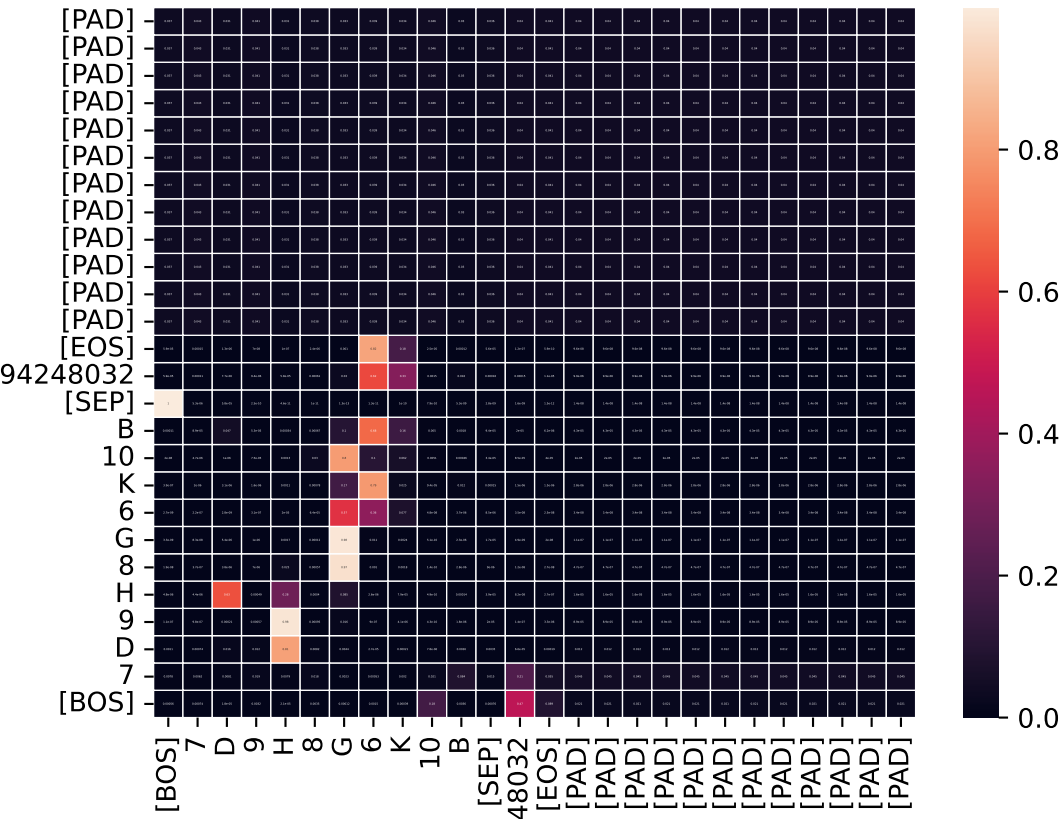


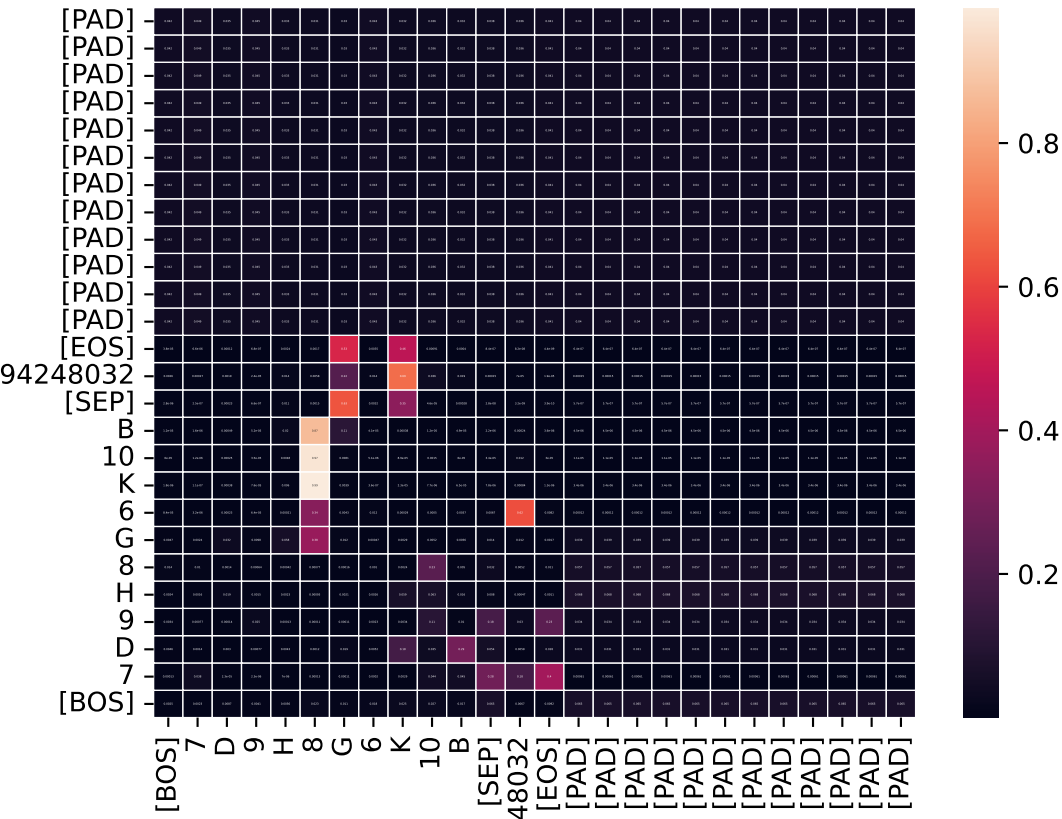
Step 20000 attention weights at layer 0 head 0, with accuracy 0.835
Sequence:[BOS] 7 D 9 H 8 G 6 K 10 B [SEP] 8.0 [EOS] SOL:G



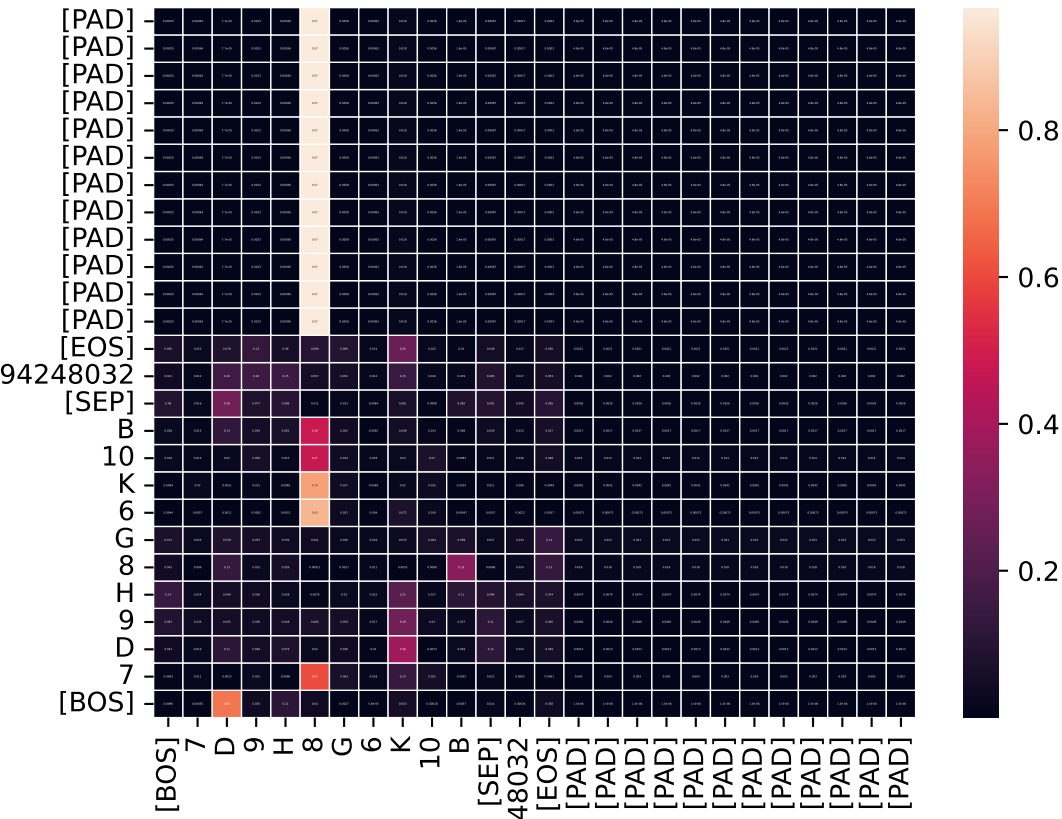
Step 20000 attention weights at layer 0 head 1, with accuracy 0.835
Sequence:[BOS] 7 D 9 H 8 G 6 K 10 B [SEP] 8.0 [EOS] SOL:G



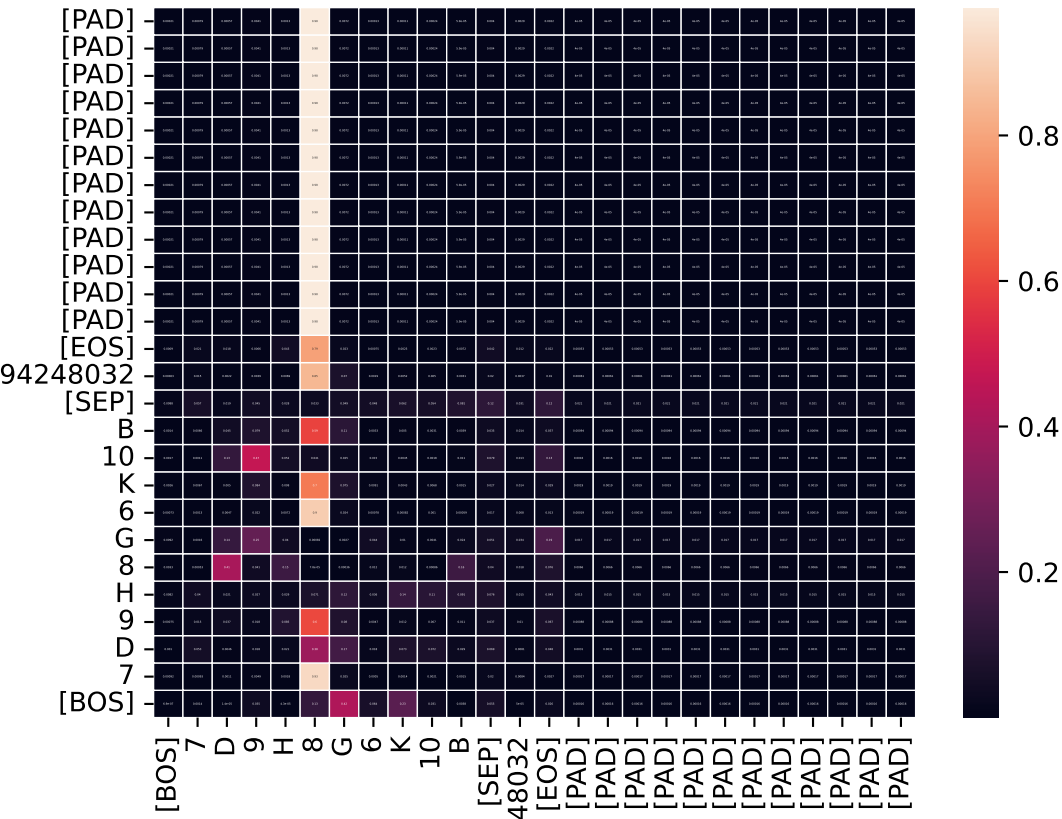


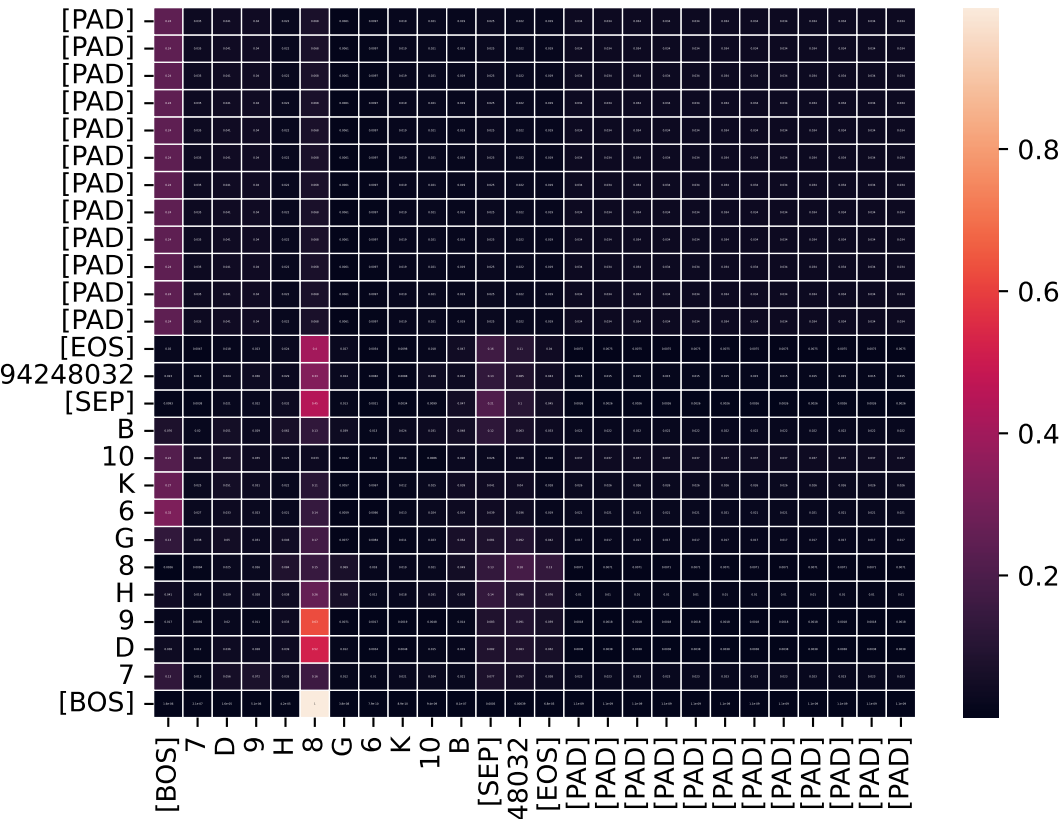


Step 20000 attention weights at layer 1 head 0, with accuracy 0.835
Sequence:[BOS] 7 D 9 H 8 G 6 K 10 B [SEP] 8.0 [EOS] SOL:G

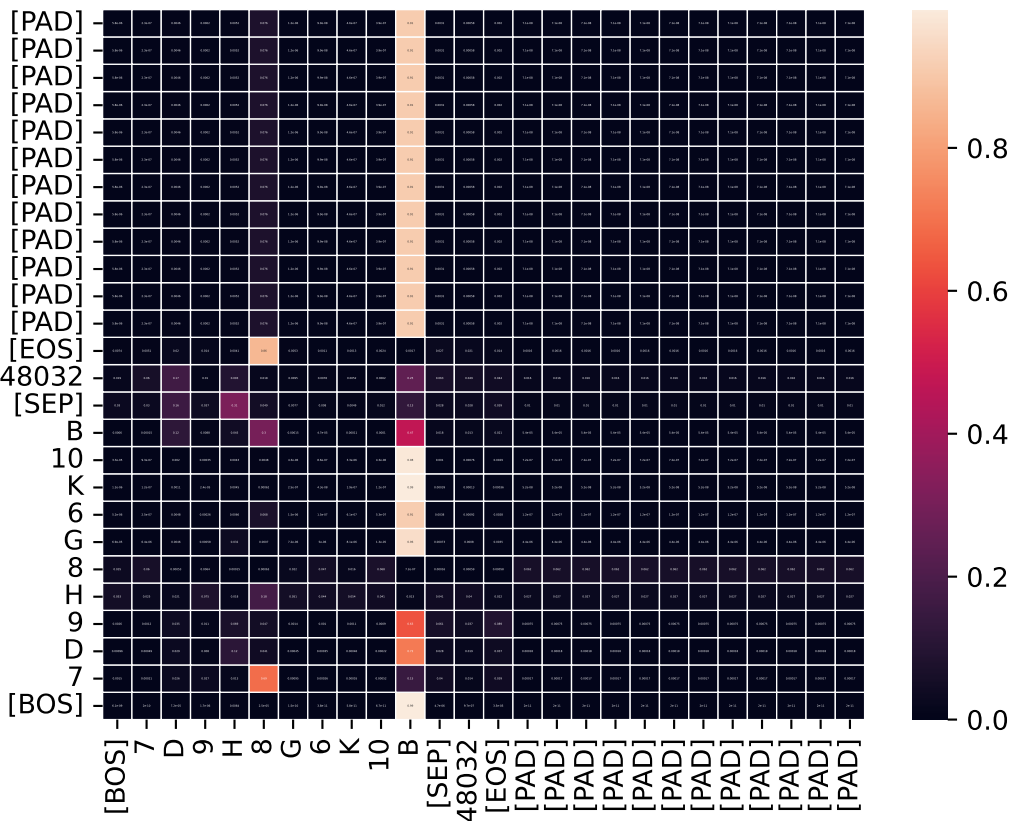


Step 20000 attention weights at layer 1 head 1, with accuracy 0.835
Sequence:[BOS] 7 D 9 H 8 G 6 K 10 B [SEP] 8.0 [EOS] SOL:G

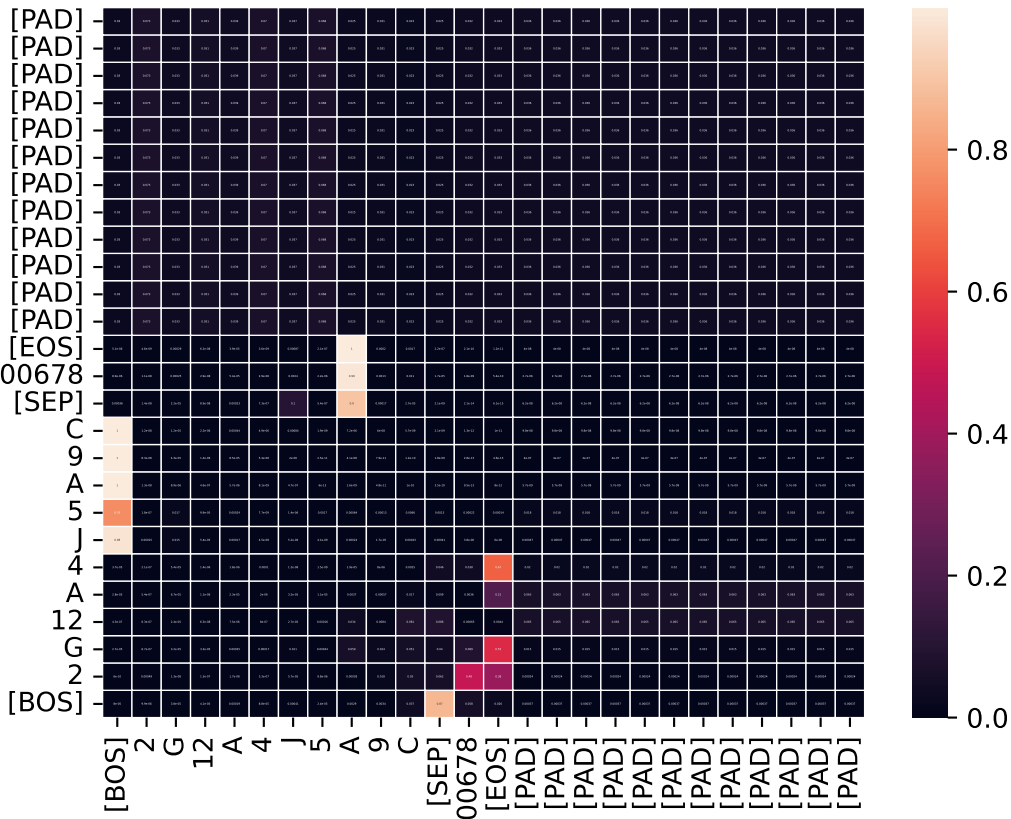




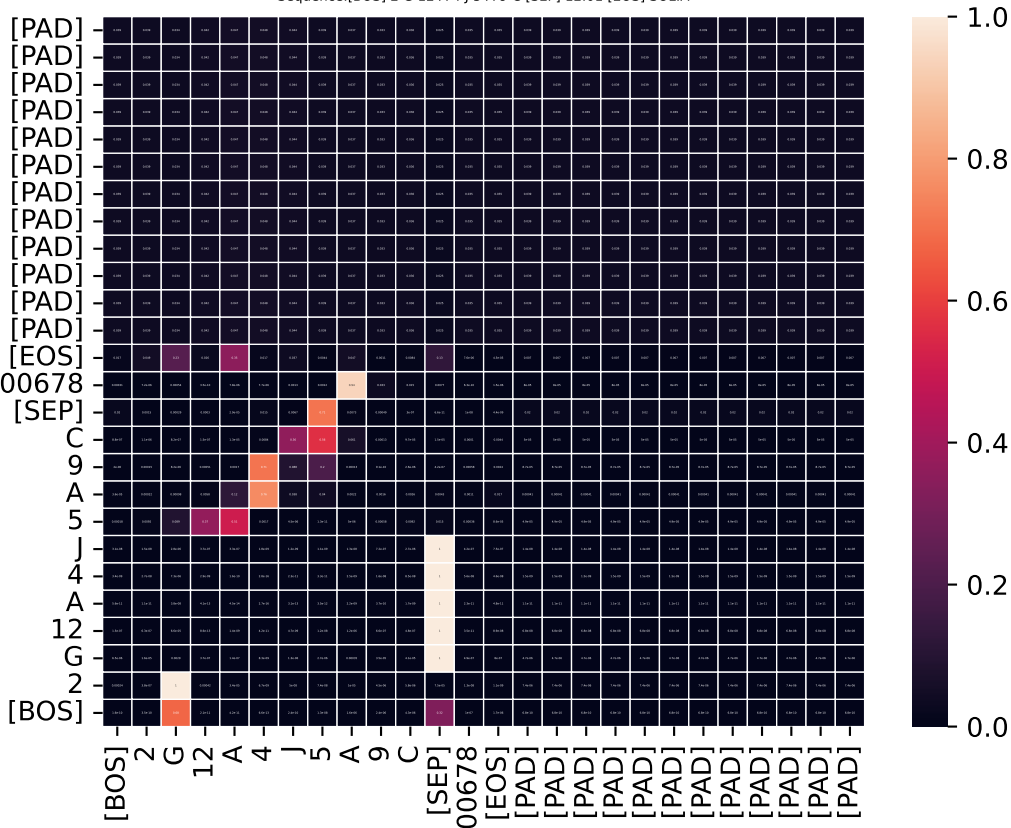
942



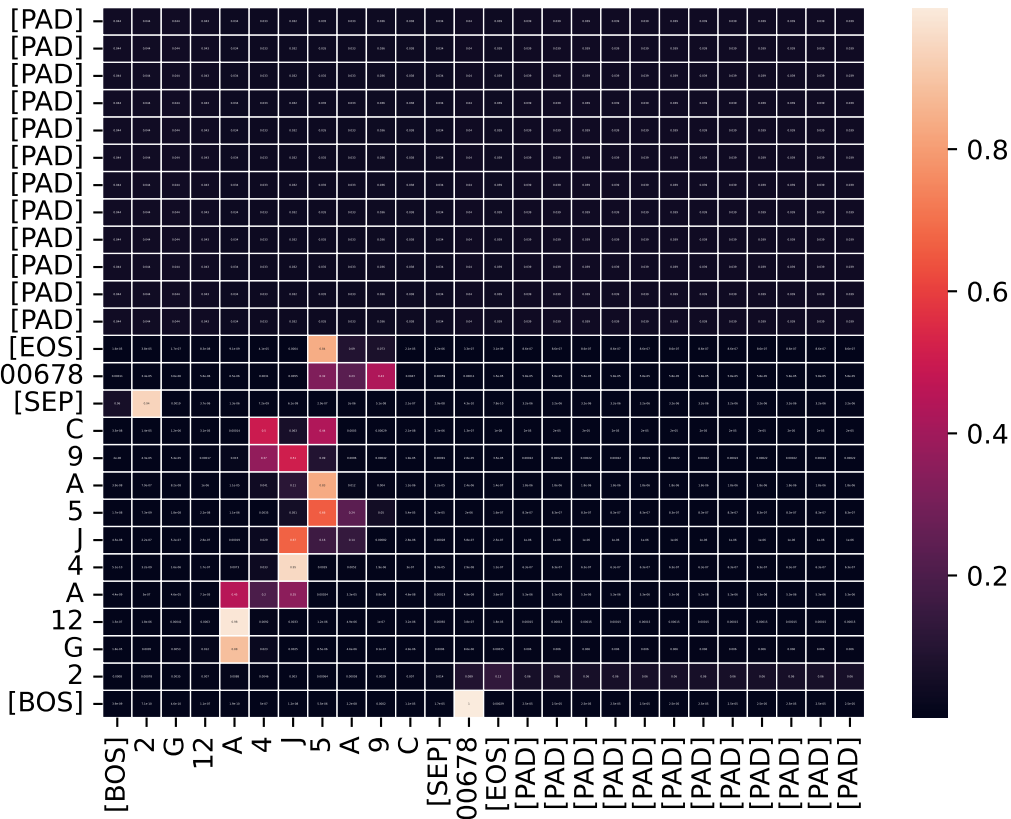
20200678
[SEP]



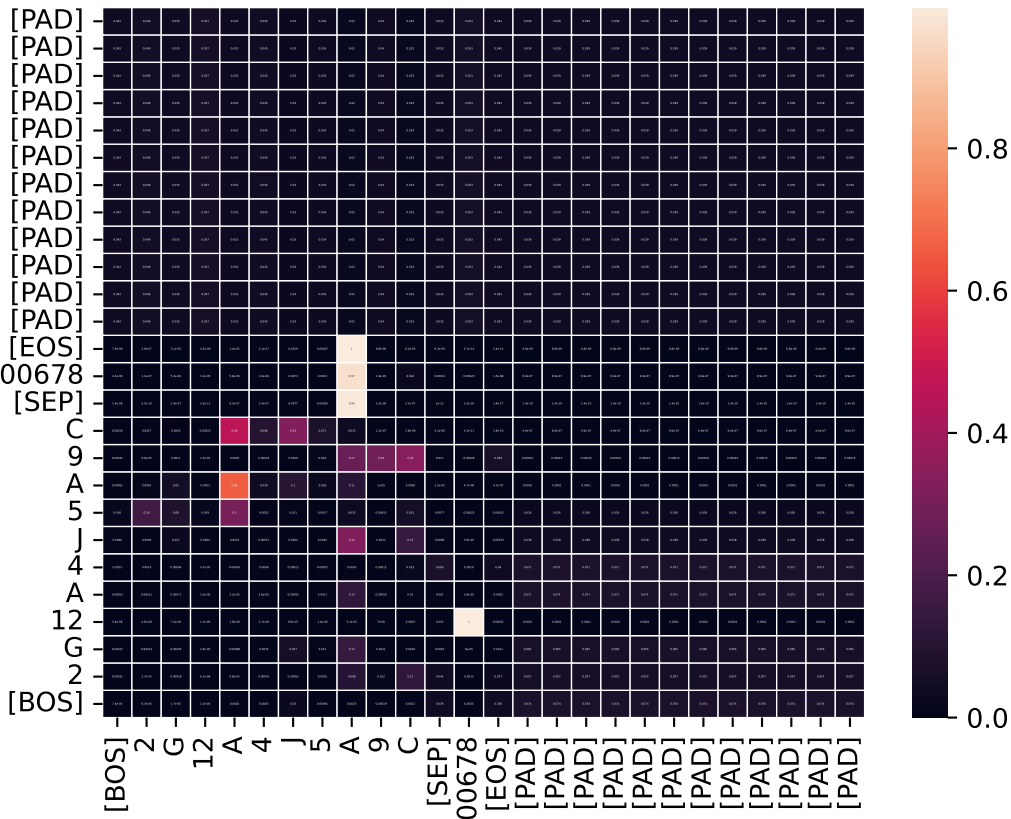
20200678



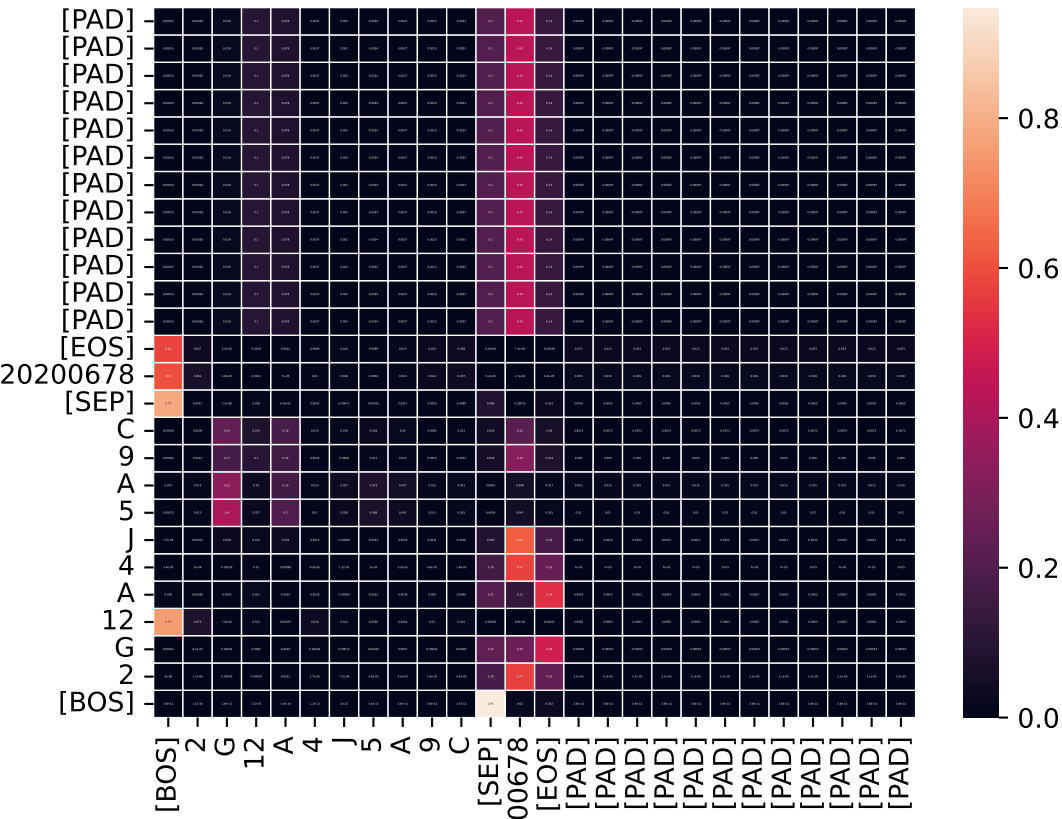
20200678



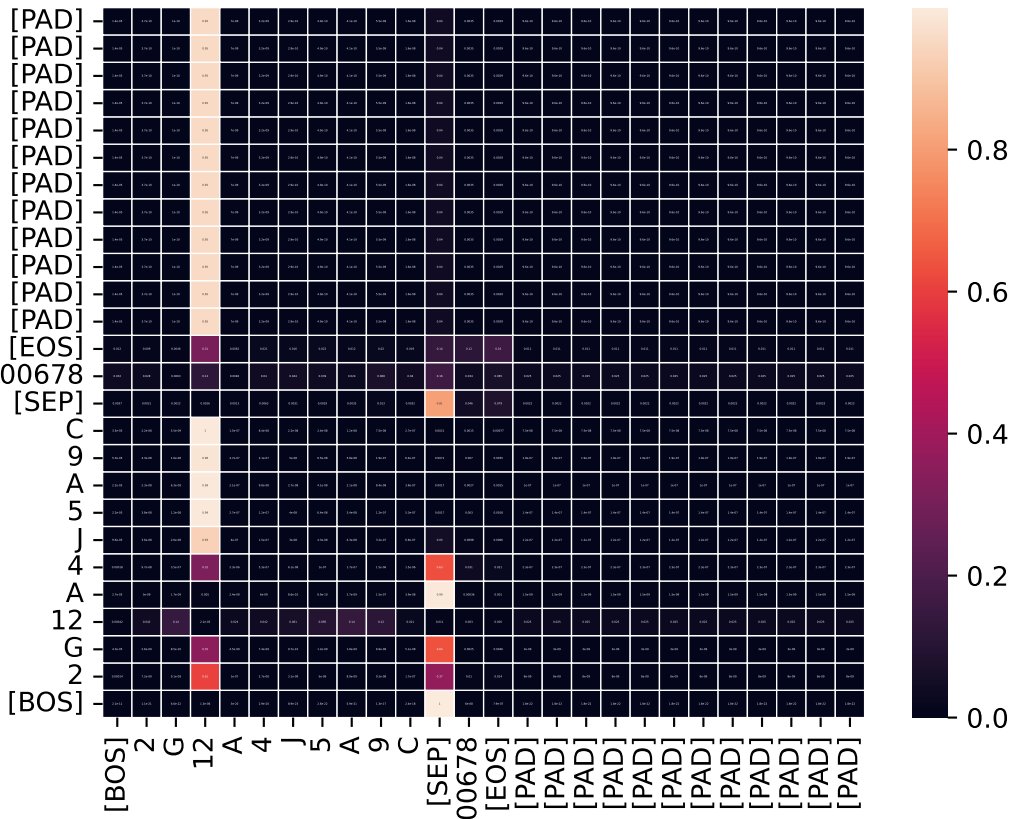
20200678



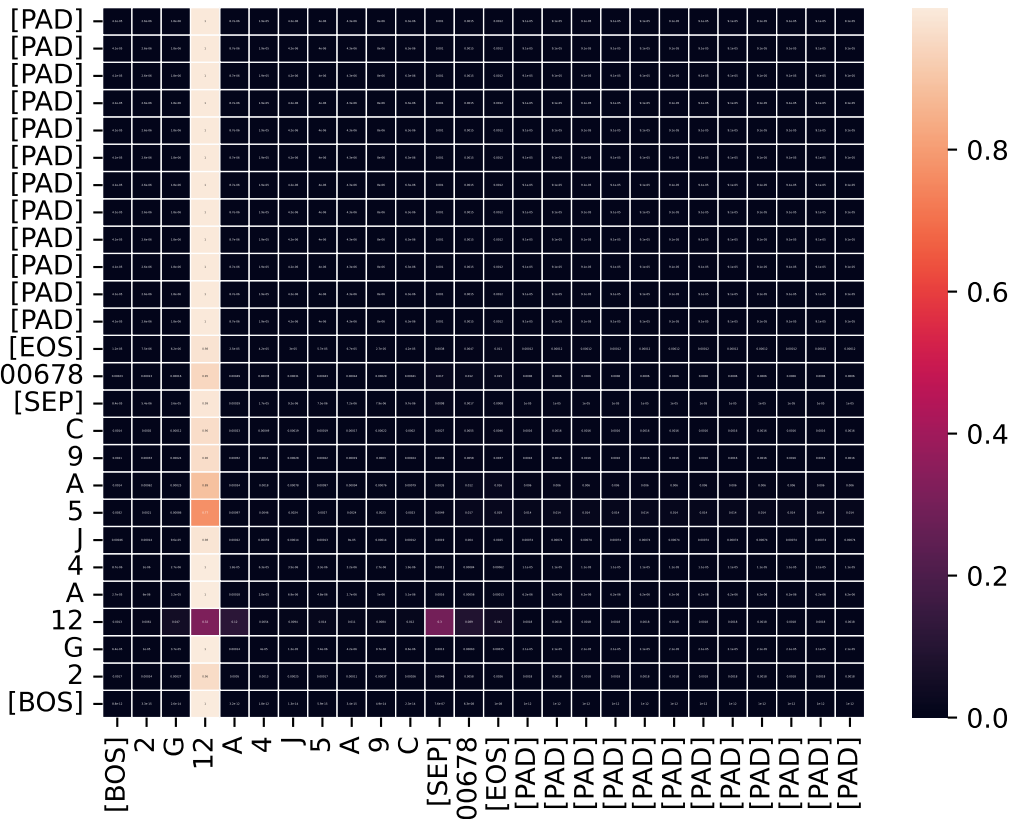
Step 29500 attention weights at layer 1 head 0, with accuracy 0.890
 Sequence:[BOS] 2 G 12 A 4 J 5 A 9 C [SEP] 12.01 [EOS] SOL:A

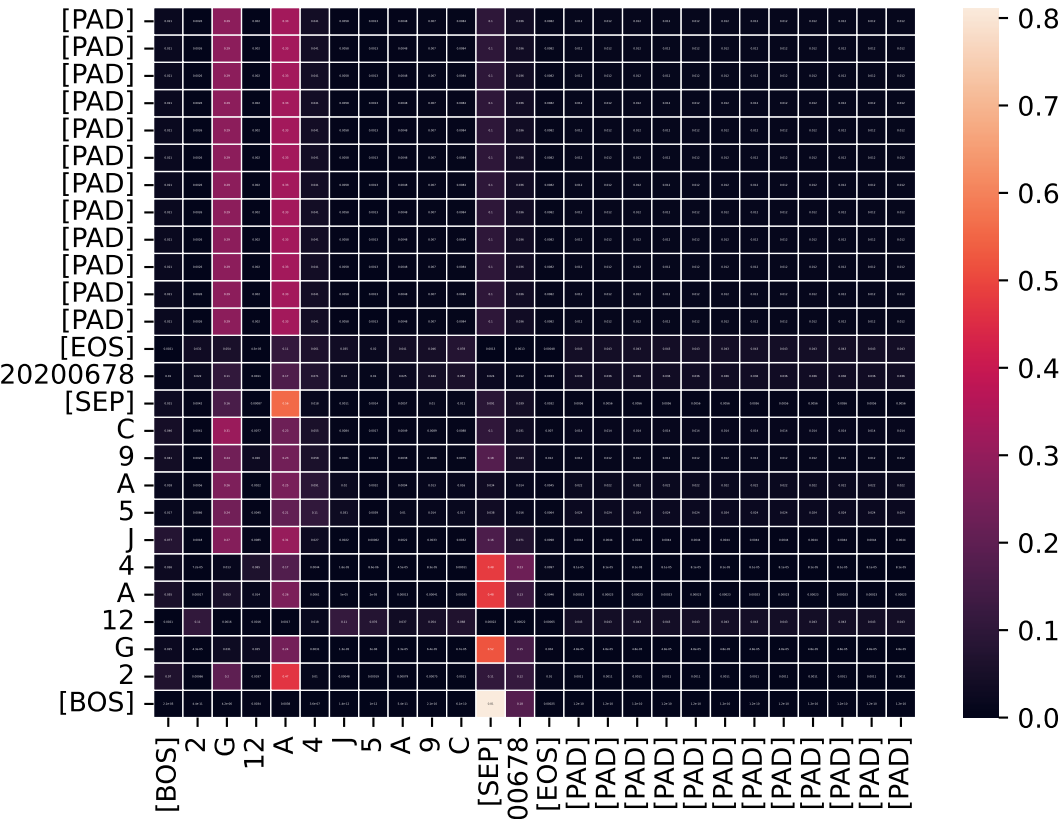


20200678

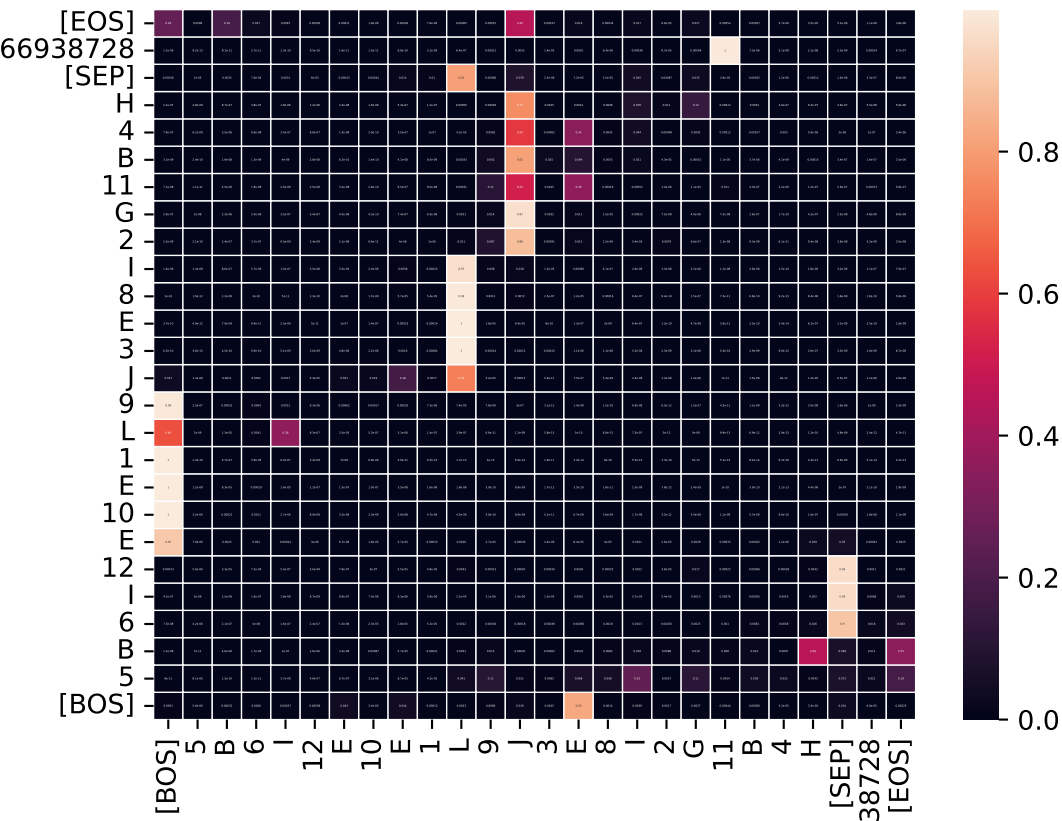


20200678

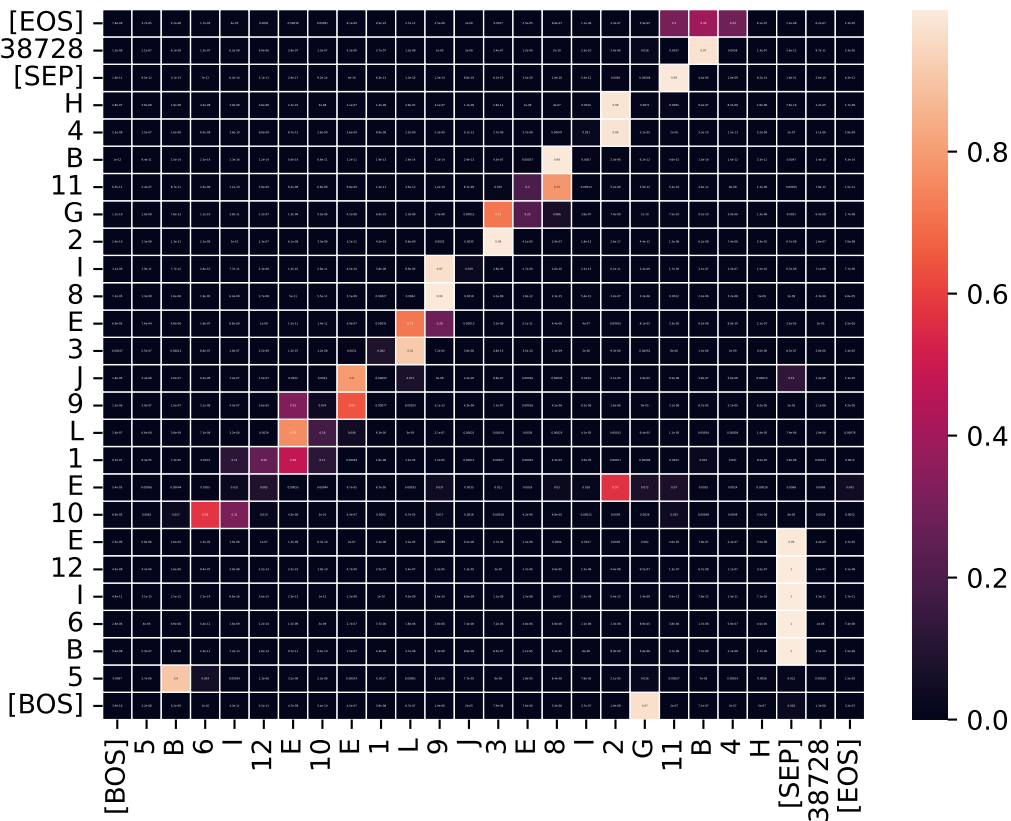




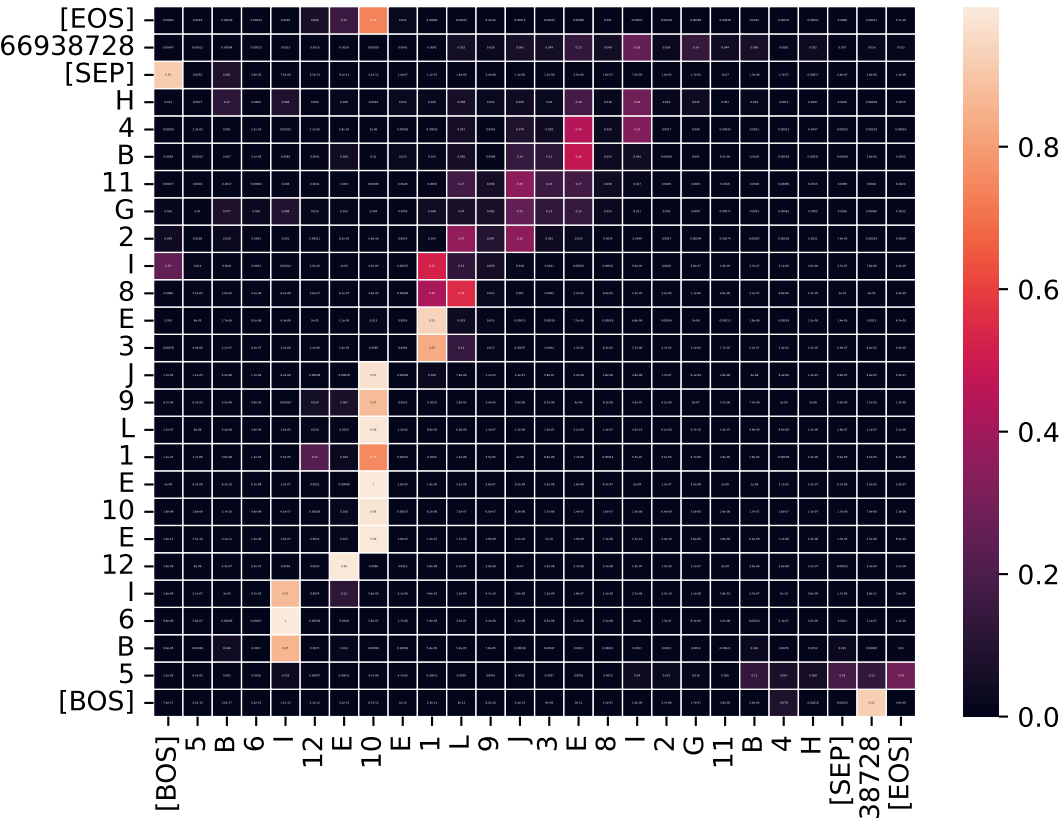
Step 30000 attention weights at layer 0 head 0, with accuracy 0.941
Sequence: [BOS] 5 B 6 I 12 E 10 E 1 L 9 J 3 E 8 I 2 G 11 B 4 H [SEP] 10.88 [EOS] SOL:B



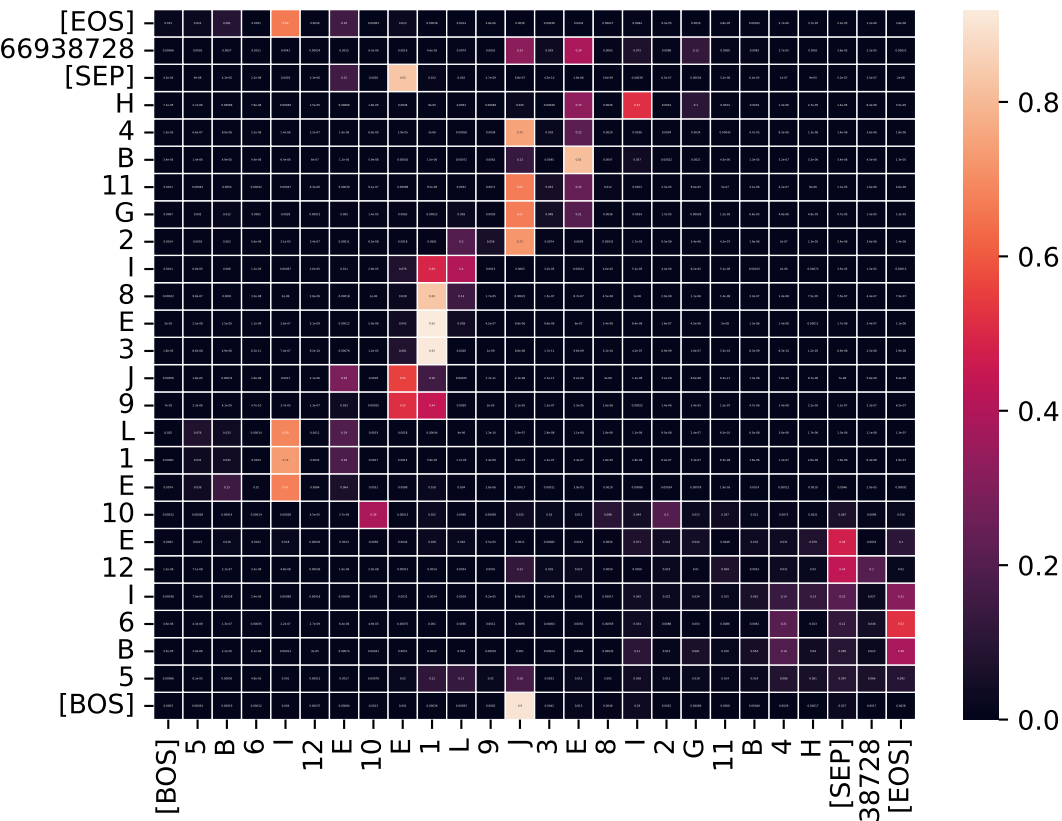
[EOS]
66938728



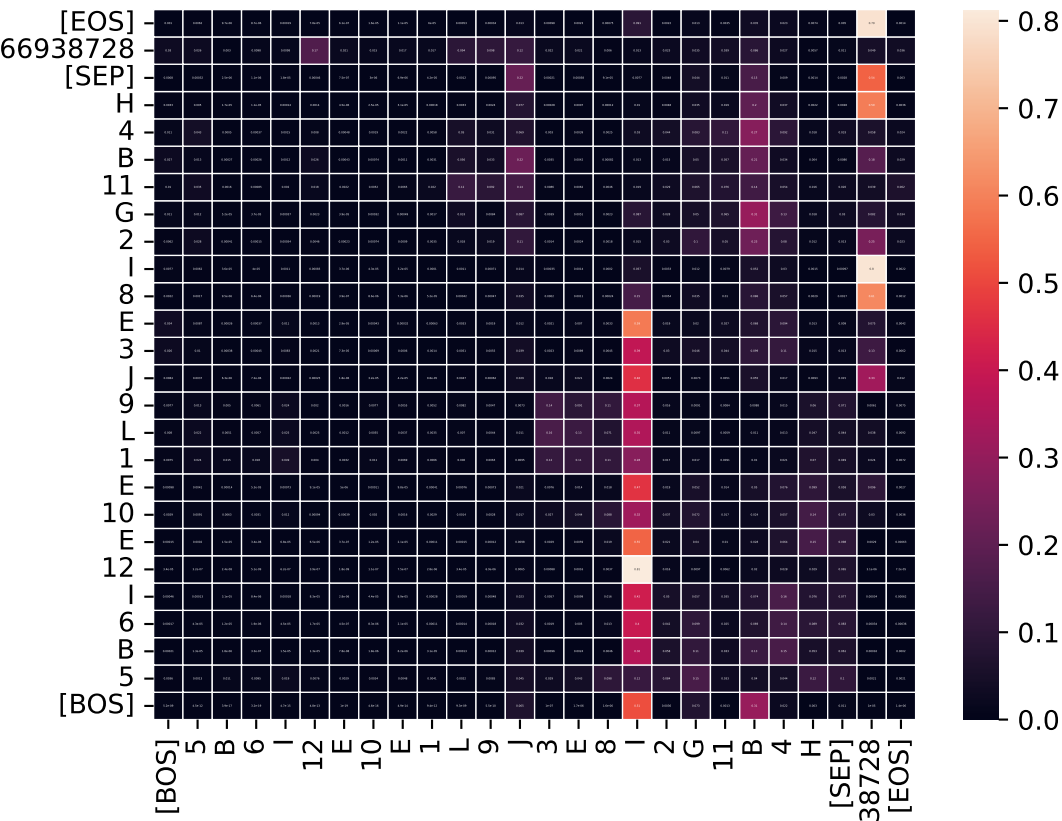
Step 30000 attention weights at layer 0 head 2, with accuracy 0.941
 Sequence: [BOS] 5 B 6 I 12 E 10 E 1 L 9 J 3 E 8 I 2 G 11 B 4 H [SEP] 10.88 [EOS] SOL:B



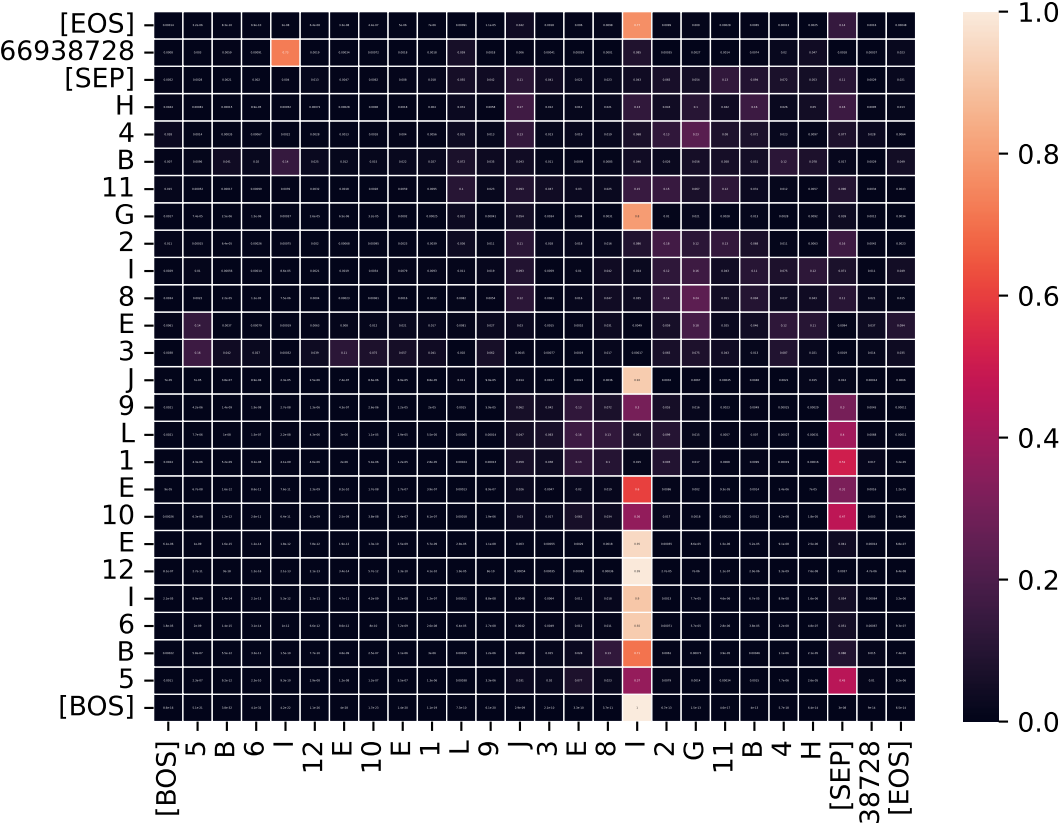
Step 30000 attention weights at layer 0 head 3, with accuracy 0.941
 Sequence: [BOS] 5 B 6 I 12 E 10 E 1 L 9 J 3 E 8 I 2 G 11 B 4 H [SEP] 10.88 [EOS] SOL:B



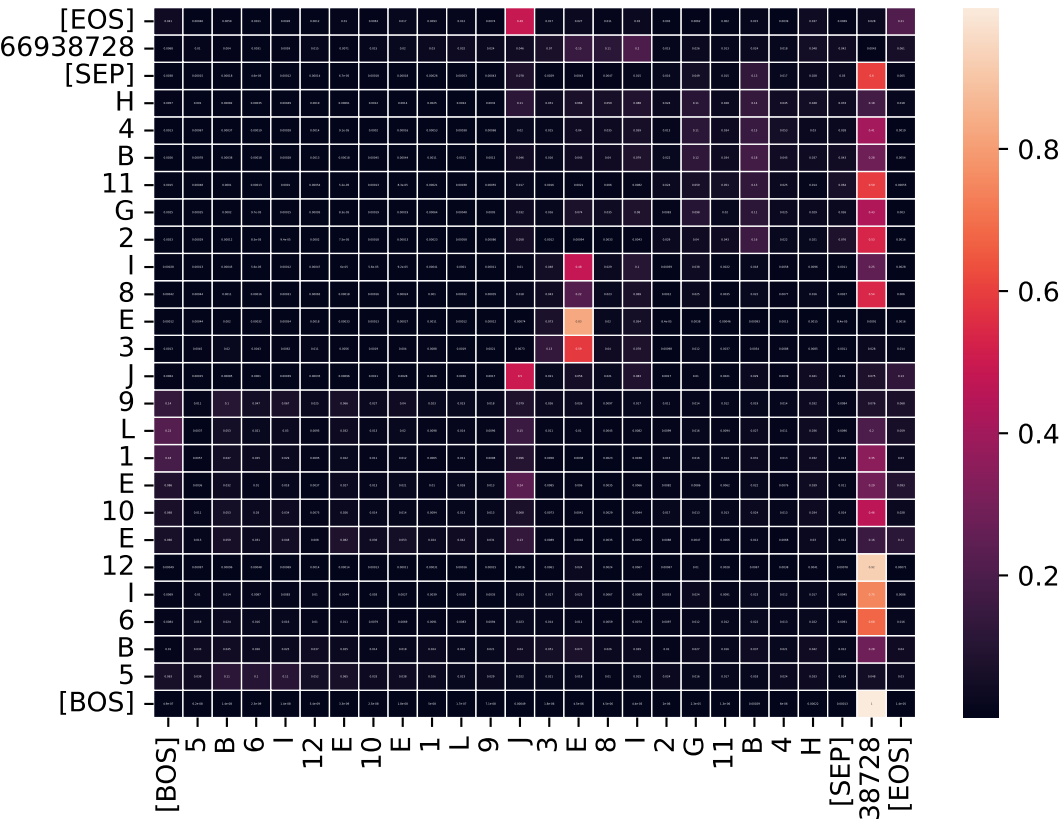
Step 30000 attention weights at layer 1 head 0, with accuracy 0.941
 Sequence:[BOS] 5 B 6 I 12 E 10 E 1 L 9 J 3 E 8 I 2 G 11 B 4 H [SEP] 10.88 [EOS] SOL:B



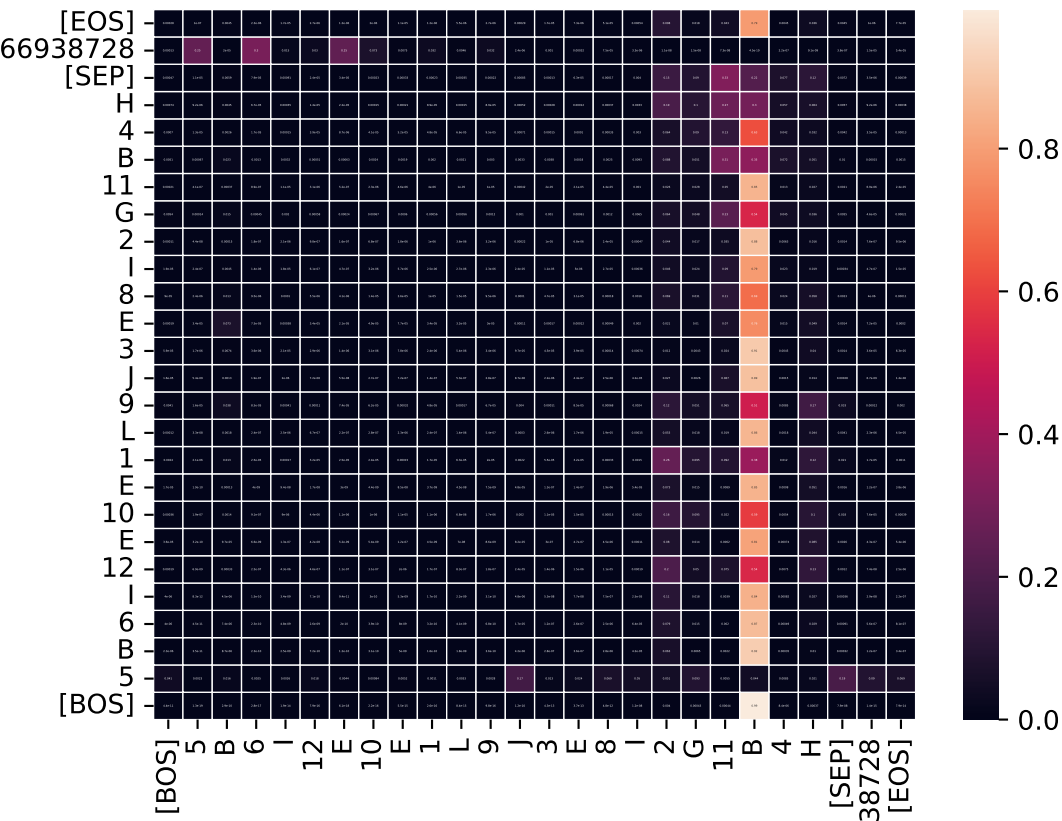
Step 30000 attention weights at layer 1 head 1, with accuracy 0.941
 Sequence:[BOS] 5 B 6 I 12 E 10 E 1 L 9 J 3 E 8 I 2 G 11 B 4 H [SEP] 10.88 [EOS] SOL:B



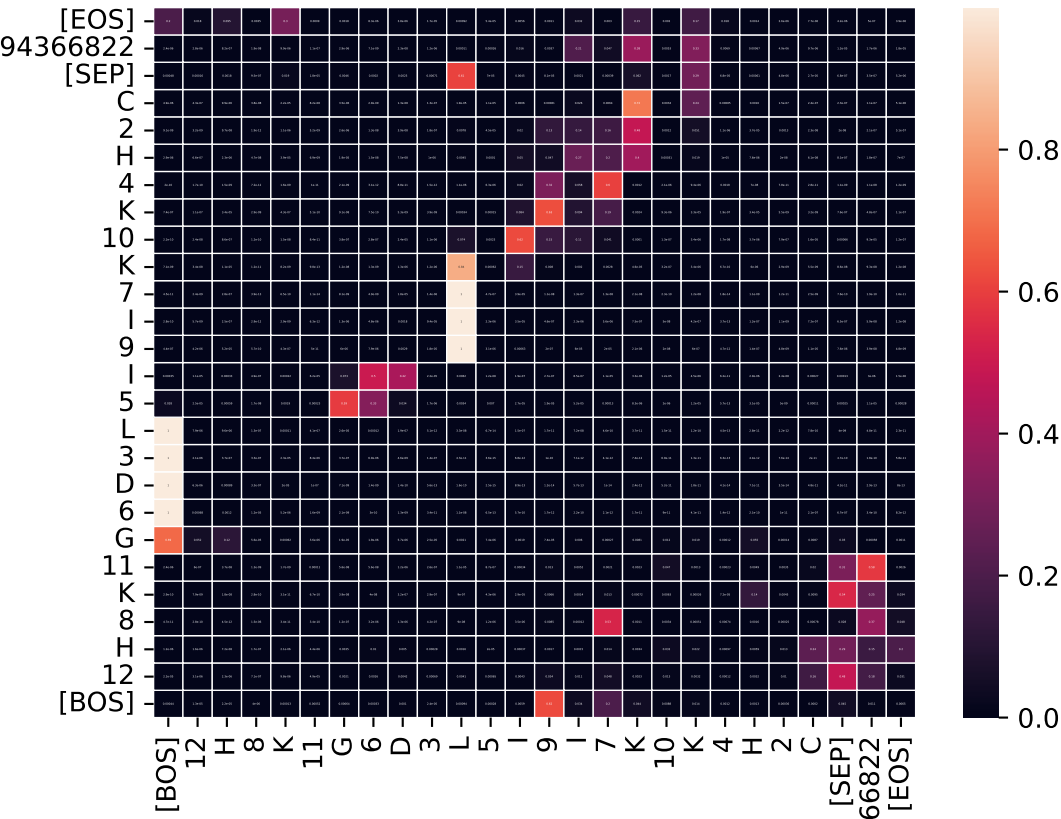
Step 30000 attention weights at layer 1 head 2, with accuracy 0.941
 Sequence: [BOS] 5 B 6 I 12 E 10 E 1 L 9 J 3 E 8 I 2 G 11 B 4 H [SEP] 10.88 [EOS] SOL:B



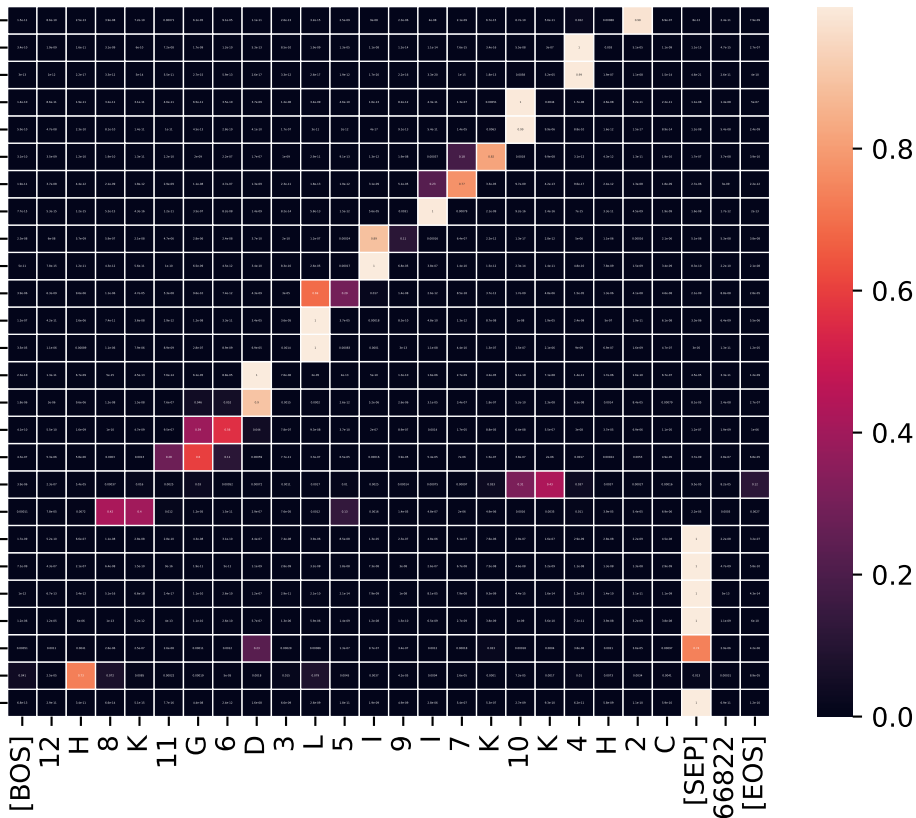
Step 30000 attention weights at layer 1 head 3, with accuracy 0.941
Sequence: [BOS] 5 B 6 I 12 E 10 E 1 L 9 J 3 E 8 I 2 G 11 B 4 H [SEP] 10.88 [EOS] SOL:B



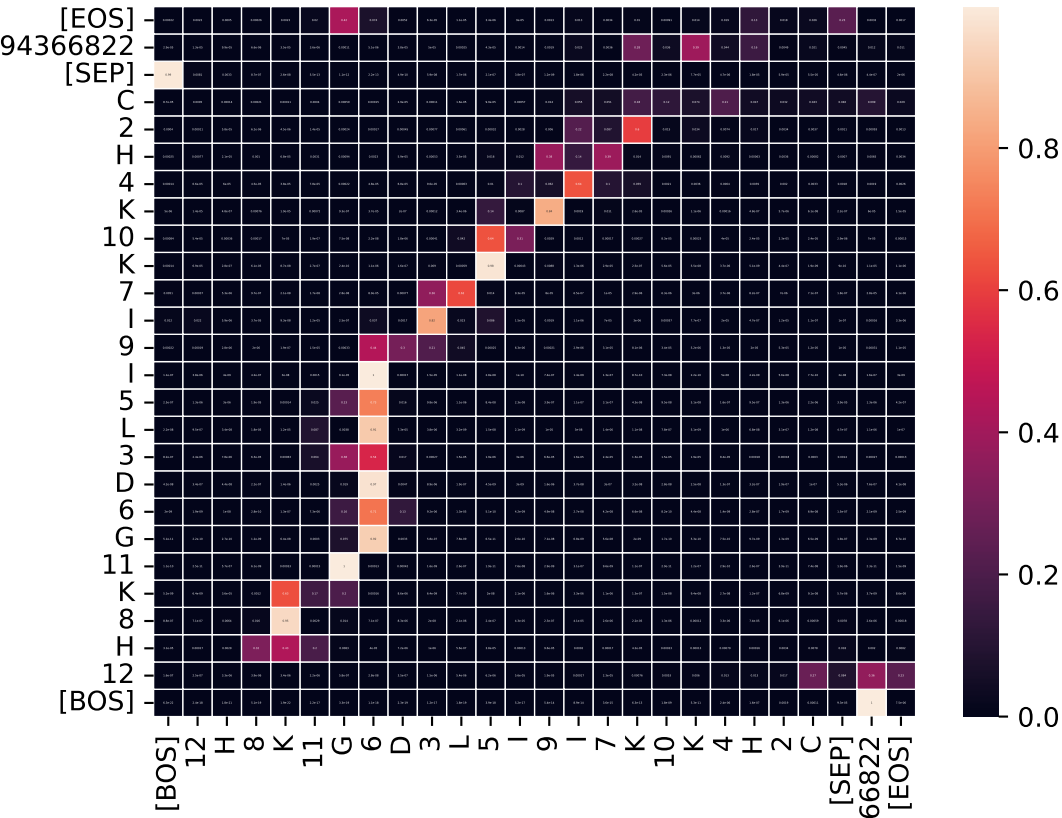
Step 32500 attention weights at layer 0 head 0, with accuracy 0.983
 Sequence:[BOS] 12 H 8 K 11 G 6 D 3 L 5 I 9 I 7 K 10 K 4 H 2 C [SEP] 8.95 [EOS] SOL:I



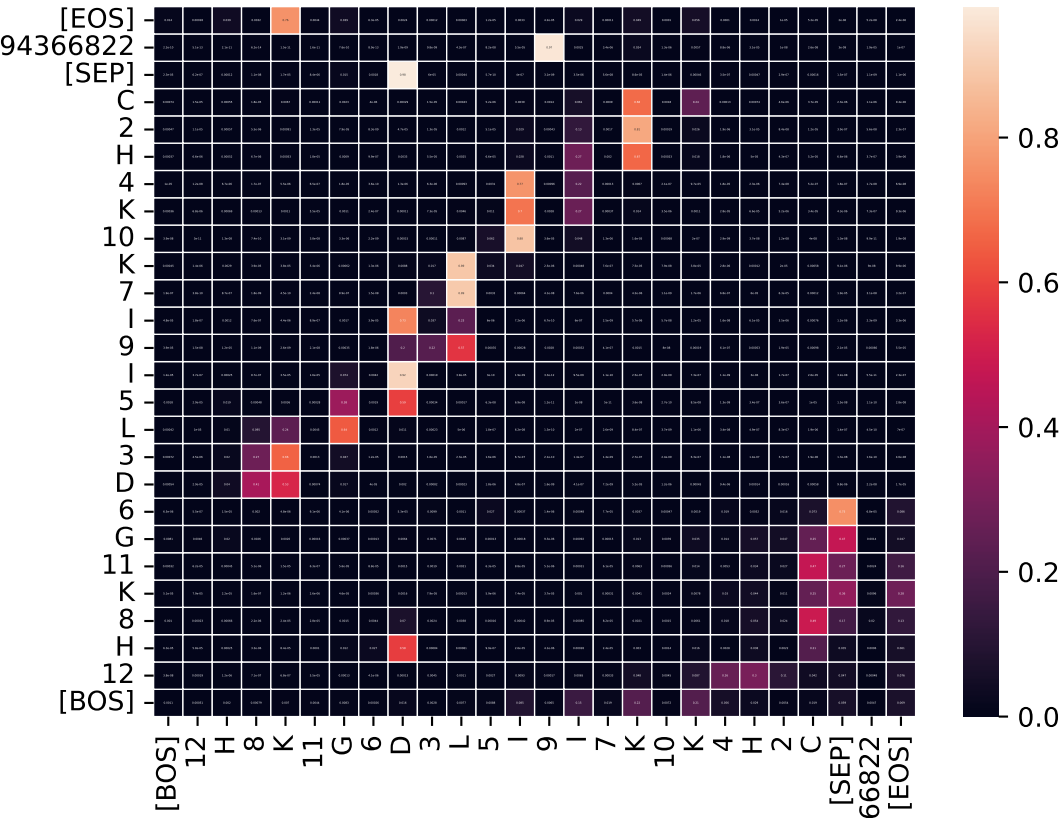
[EOS]	94366822
[SEP]	C
	2
	H
	4
	K
	10
	K
	7
	I
	9
	I
	5
	L
	3
	D
	6
	G
	11
	K
	8
	H
	12
[BOS]	



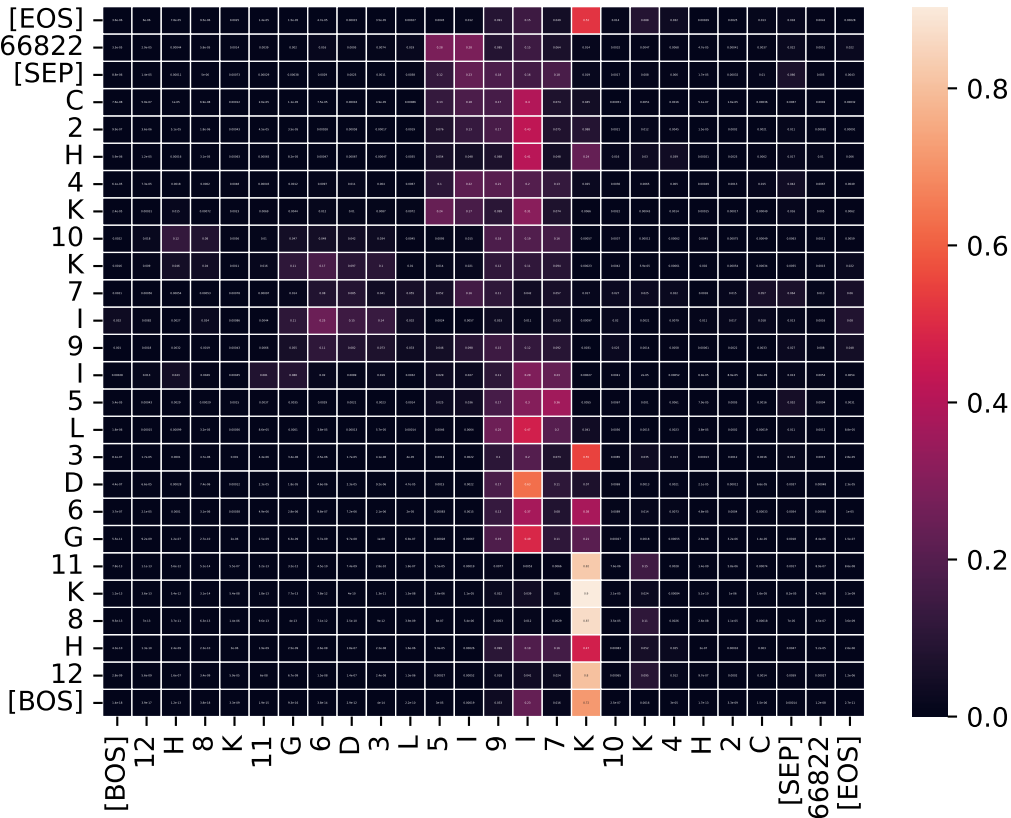
Step 32500 attention weights at layer 0 head 2, with accuracy 0.983
 Sequence:[BOS] 12 H 8 K 11 G 6 D 3 L 5 I 9 I 7 K 10 K 4 H 2 C [SEP] 8.95 [EOS] SOL:I



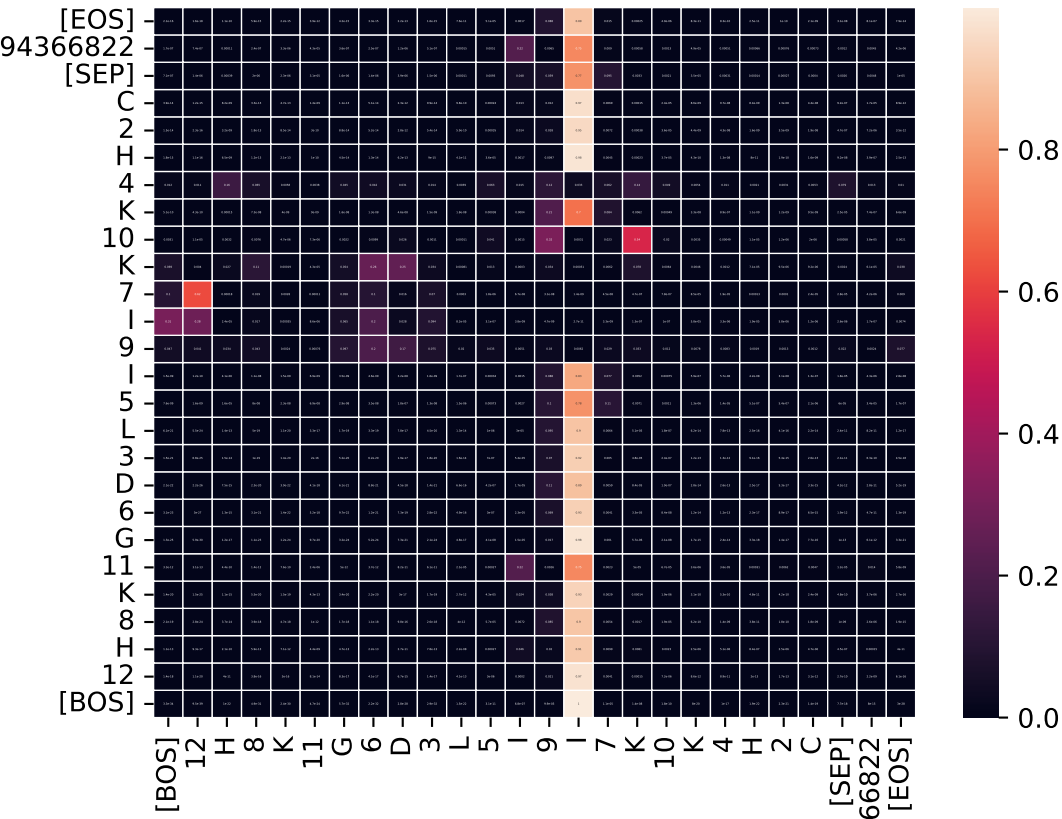
Step 32500 attention weights at layer 0 head 3, with accuracy 0.983
 Sequence:[BOS] 12 H 8 K 11 G 6 D 3 L 5 I 9 I 7 K 10 K 4 H 2 C [SEP] 8.95 [EOS] SOL:I



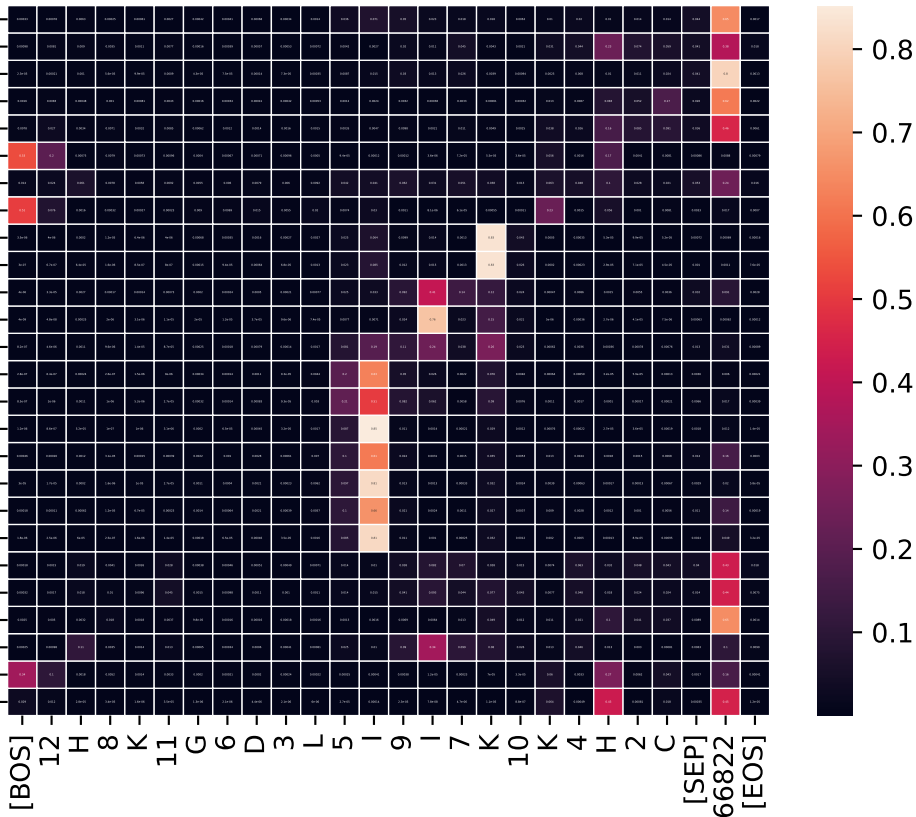
[EOS]	94366822
[SEP]	C
	2
	H
	4
	K
	10
	K
	7
	I
	9
	I
	5
	L
	3
	D
	6
	G
	11
	K
	8
	H
	12
[BOS]	



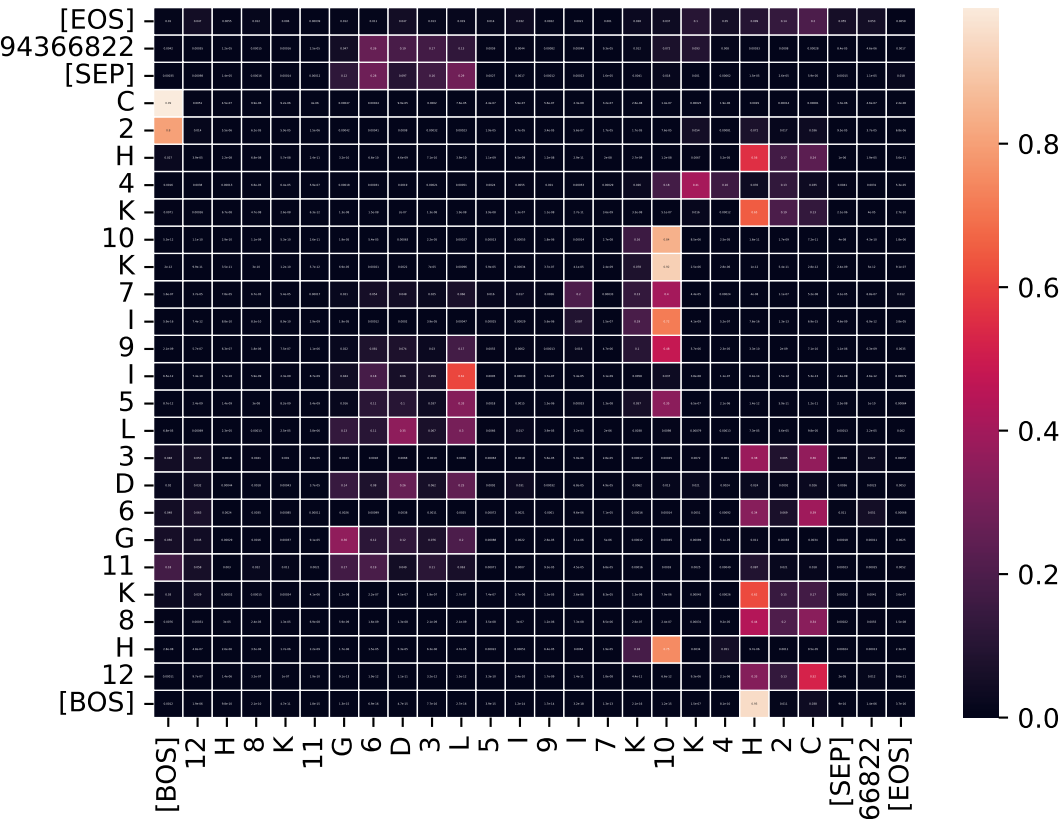
Step 32500 attention weights at layer 1 head 1, with accuracy 0.983
Sequence:[BOS] 12 H 8 K 11 G 6 D 3 L 5 I 9 I 7 K 10 K 4 H 2 C [SEP] 8.95 [EOS] SOL:I



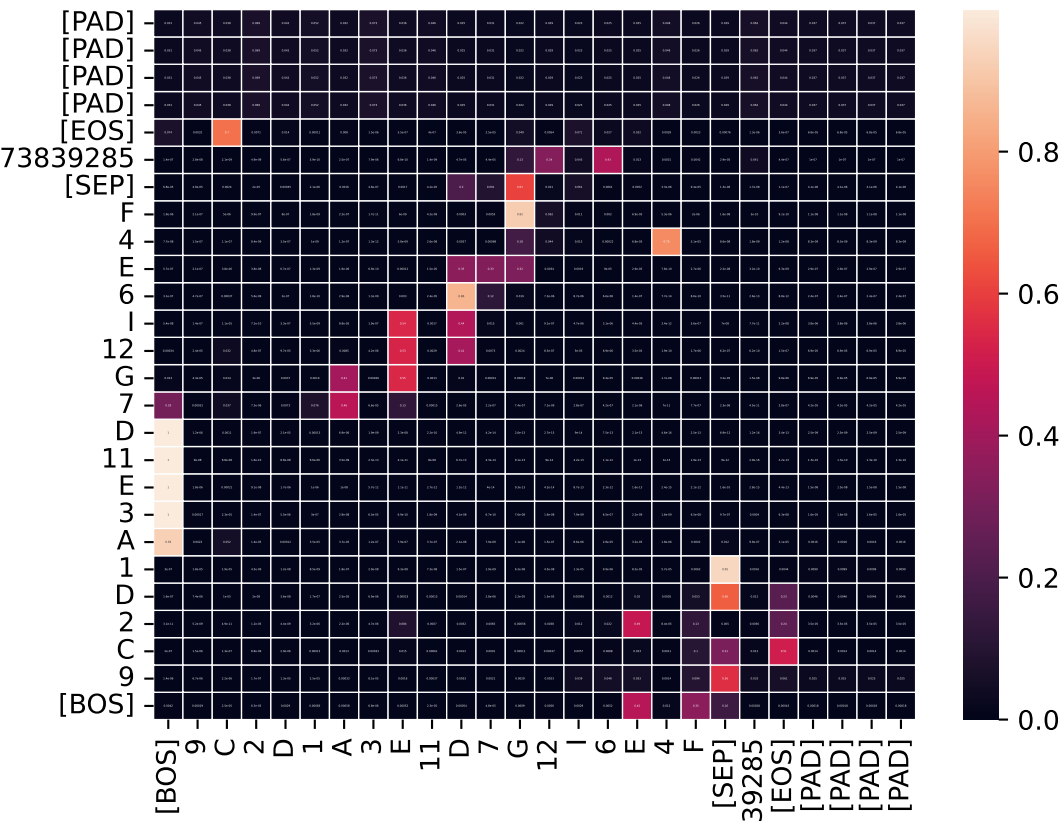
[EOS]	94366822
[SEP]	C
	2
	H
	4
	K
	10
	K
	7
	I
	9
	I
	5
	L
	3
	D
	6
	G
	11
	K
	8
	H
	12
[BOS]	



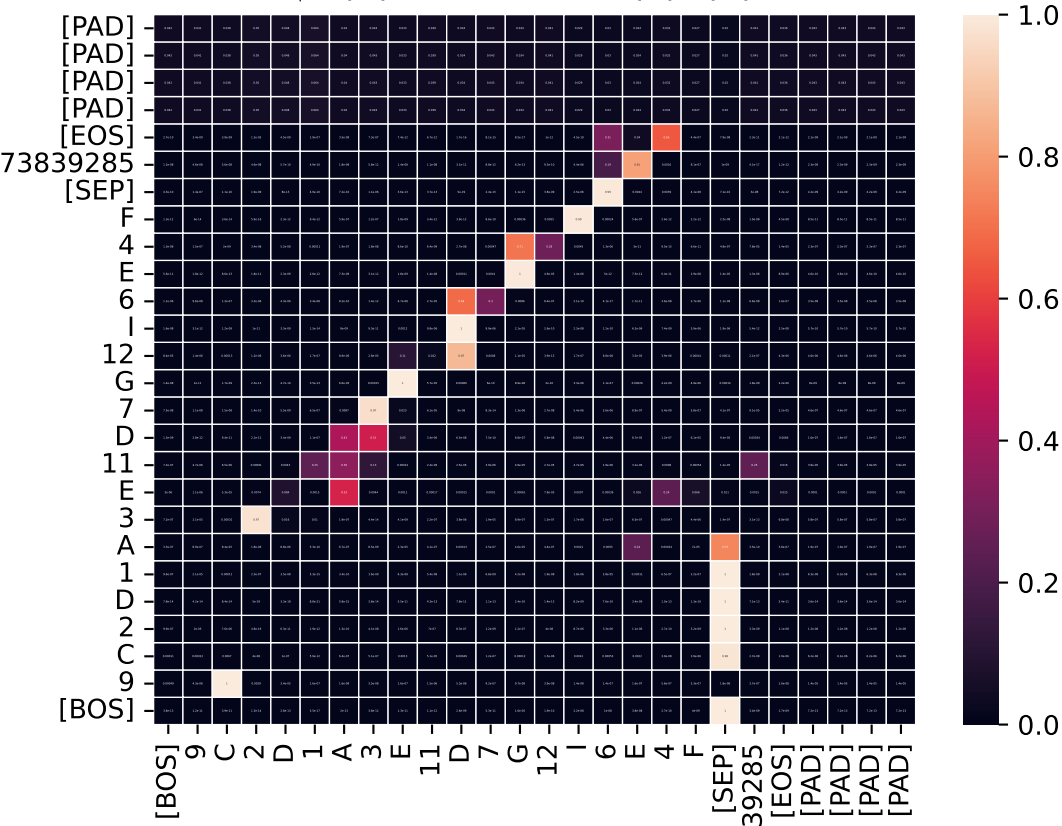
Step 32500 attention weights at layer 1 head 3, with accuracy 0.983
 Sequence:[BOS] 12 H 8 K 11 G 6 D 3 L 5 I 9 I 7 K 10 K 4 H 2 C [SEP] 8.95 [EOS] SOL:I



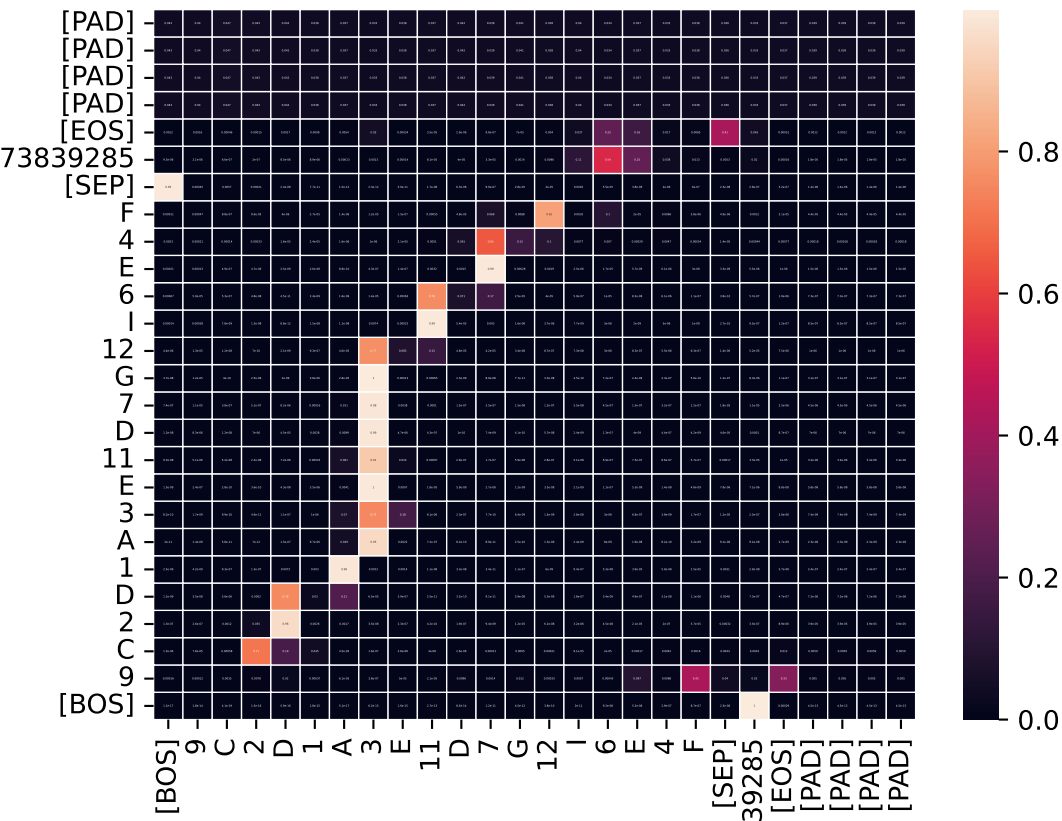
Step 33000 attention weights at layer 0 head 0, with accuracy 0.997
Sequence: [BOS] 9 C 2 D 1 A 3 E 11 D 7 G 12 I 6 E 4 F [SEP] 2.86 [EOS] SOL:E



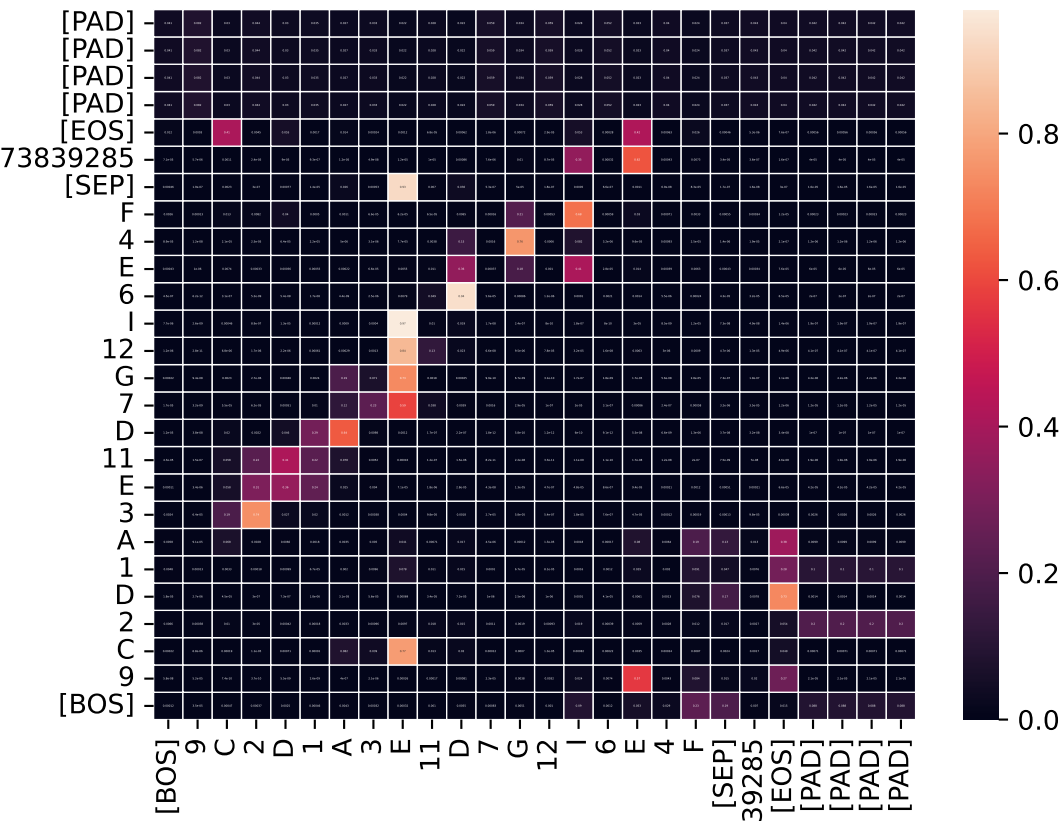
Step 33000 attention weights at layer 0 head 1, with accuracy 0.997
Sequence: [BOS] 9 C 2 D 1 A 3 E 11 D 7 G 12 I 6 E 4 F [SEP] 2.86 [EOS] SOL:E



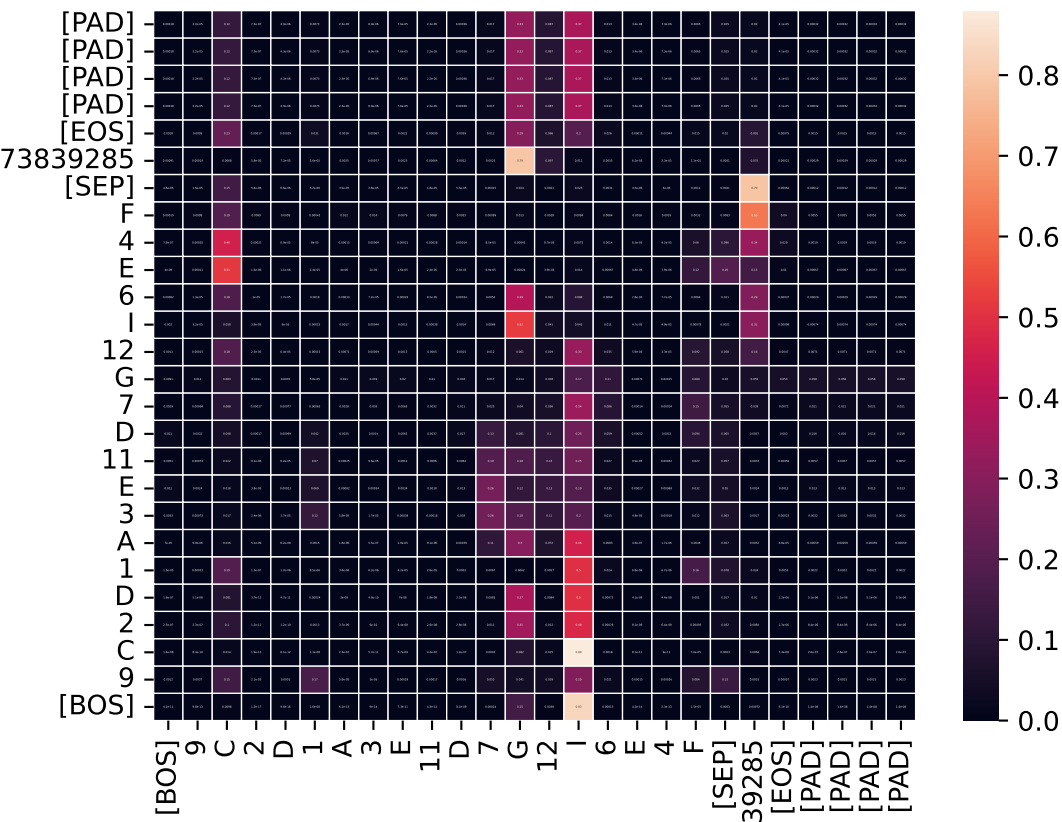
Step 33000 attention weights at layer 0 head 2, with accuracy 0.997
Sequence: [BOS] 9 C 2 D 1 A 3 E 11 D 7 G 12 I 6 E 4 F [SEP] 2.86 [EOS] SOL:E



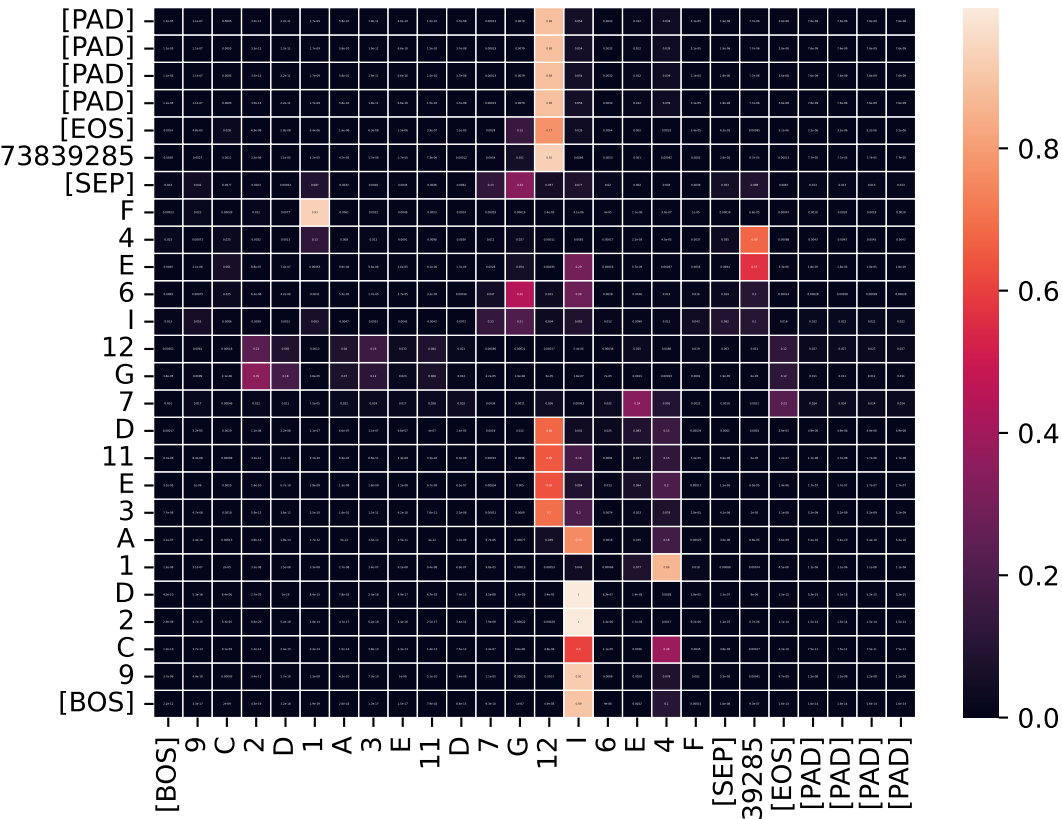
Step 33000 attention weights at layer 0 head 3, with accuracy 0.997
Sequence: [BOS] 9 C 2 D 1 A 3 E 11 D 7 G 12 I 6 E 4 F [SEP] 2.86 [EOS] SOL:E



Step 33000 attention weights at layer 1 head 0, with accuracy 0.997
Sequence: [BOS] 9 C 2 D 1 A 3 E 11 D 7 G 12 I 6 E 4 F [SEP] 2.86 [EOS] SOL:E



Step 33000 attention weights at layer 1 head 1, with accuracy 0.997
Sequence:[BOS] 9 C 2 D 1 A 3 E 11 D 7 G 12 I 6 E 4 F [SEP] 2.86 [EOS] SOL:E



Step 33000 attention weights at layer 1 head 2, with accuracy 0.997
Sequence:[BOS] 9 C 2 D 1 A 3 E 11 D 7 G 12 I 6 E 4 F [SEP] 2.86 [EOS] SOL:E

