Model Evaluation

Model Statistics:

Model	Correct Answers	Partially Correct Answers	Wrong Answers
deepset/roberta-base-squad2	36 (90.0%)	1 (2.5%)	3 (7.5%)

Response Times Statistics:

Model	Average	Median	Min	Max
deepset/roberta-base-squad2	9.6	9.5	9.1	10.6

**All the response times are in seconds

Model answers confidence scores:

For deepset/roberta-base-squad2 model:

Range	Correct	Partially Correct	Wrong
0.0 - 0.1	5 (13.9%)	0 (0.0%)	1 (33.3%)
0.1 - 0.2	1 (2.8%)	0 (0.0%)	0 (0.0%)
0.2 - 0.3	5 (13.9%)	0 (0.0%)	0 (0.0%)
0.3 - 0.4	5 (13.9%)	0 (0.0%)	0 (0.0%)
0.4 - 0.5	6 (16.7%)	0 (0.0%)	2 (66.7%)
0.5 - 0.6	7 (19.4%)	0 (0.0%)	0 (0.0%)
0.6 - 0.7	1 (2.8%)	1 (100.0%)	0 (0.0%)
0.7 - 0.8	3 (8.3%)	0 (0.0%)	0 (0.0%)
0.8 - 0.9	1 (2.8%)	0 (0.0%)	0 (0.0%)
0.9 - 1.0	2 (5.6%)	0 (0.0%)	0 (0.0%)

Question Statistics:

36 questions were answered correctly by all models

3 questions were answered incorrectly by all models

36 questions were answered correctly by at least one model

36 questions were answered correctly by exactly one model

37 questions were answered correctly or partially correct by all models