

# Model Evaluation

## Model Statistics:

<i>Model</i>	<i>Correct Answers</i>	<i>Partially Correct Answers</i>	<i>Wrong Answers</i>
<i>deepset/roberta-base-squad2</i>	33 (82.5%)	0 (0.0%)	7 (17.5%)

## Response Times Statistics:

<i>Model</i>	<i>Average</i>	<i>Median</i>	<i>Min</i>	<i>Max</i>
<i>deepset/roberta-base-squad2</i>	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:

<i>Range</i>	<i>Correct</i>	<i>Partially Correct</i>	<i>Wrong</i>
<i>0.0 - 0.1</i>	3 (9.1%)	0 (0%)	3 (42.9%)
<i>0.1 - 0.2</i>	1 (3.0%)	0 (0%)	0 (0.0%)
<i>0.2 - 0.3</i>	5 (15.2%)	0 (0%)	0 (0.0%)
<i>0.3 - 0.4</i>	5 (15.2%)	0 (0%)	0 (0.0%)
<i>0.4 - 0.5</i>	6 (18.2%)	0 (0%)	2 (28.6%)
<i>0.5 - 0.6</i>	6 (18.2%)	0 (0%)	1 (14.3%)
<i>0.6 - 0.7</i>	1 (3.0%)	0 (0%)	1 (14.3%)
<i>0.7 - 0.8</i>	3 (9.1%)	0 (0%)	0 (0.0%)
<i>0.8 - 0.9</i>	1 (3.0%)	0 (0%)	0 (0.0%)
<i>0.9 - 1.0</i>	2 (6.1%)	0 (0%)	0 (0.0%)

**Question Statistics:**

- 33 questions were answered correctly by all models
- 7 questions were answered incorrectly by all models
- 33 questions were answered correctly by at least one model
- 33 questions were answered correctly by exactly one model
- 33 questions were answered correctly or partially correct by all models