BERT Model trained with samples created by text-davinci-003 annotated by gpt-3.5-turbo rating answers created by text-davinci-003 annotated by gpt-3.5-turbo used for testing 0.150 -0.125 0.100 -0.075 0.050 -0.025 -0.000 --0.025 --0.050 -100 200 300 400 500 600 700 800 Amount of samples the model was trained with

Each value represents the agreement of exactly 1576 ratings