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 0.35 -0.30 -0.25 -8 0.20 -0.15 -0.10 -0.05 -700 100 200 300 400 500 600 800 Amount of samples the model was trained with Each value represents the agreement of exactly 1576 ratings