BERT Model trained with samples created by text-davinci-003 annotated by gpt-3.5-turbo rating answers created by human experts used for testing 0.35 0.30 0.25 е ф 0.20 0.15 0.10 0.05 -500 1000 1500 2000 2500 3000 Amount of samples the model was trained with

Each value represents the agreement of exactly 392 ratings