XGB 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 training 0.9 -0.8 -Kappa - 2.0 0.6 0.5 -1500 2500 3000 500 1000 2000 Amount of samples the model was trained with Each value represents the agreement of exactly 6401 ratings