XGB Model trained with samples created by text-davinci-003 annotated by gpt-3.5-turbo rating answers created by gpt4 used for testing 0.5 -0.4 0.3 Kappa 0.2 0.1 0.0 -200 300 400 500 600 700 800 100 Amount of samples the model was trained with Each value represents the agreement of exactly 1600 ratings