XGB Model trained with samples created by text-davinci-003 annotated by gpt-3.5-turbo rating answers created by human experts used for testing 0.5 0.4 -Kappa E.0 0.2 0.1 0.0 -400 500 600 800 100 200 300 700 Amount of samples the model was trained with Each value represents the agreement of exactly 392 ratings