XGB Model trained with samples created by text-davinci-003 rating answers created by human experts used for testing 0.5 0.4 Kappa E.0 0.2 0.1 500 1000 1500 2000 2500 3000 Amount of samples the model was trained with Each value represents the agreement of exactly 392 ratings