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 0.50 -0.45 -0.40 Kappa - 55.0 0.30 0.25 0.20 0.15 -400 500 800 100 200 300 600 700 Amount of samples the model was trained with Each value represents the agreement of exactly 392 ratings