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 testing 0.45 -0.40 -0.35 -0.30 -0.30 Y 0.20 -0.15 0.10 -0.05 700 800 100 200 300 400 500 600 Amount of samples the model was trained with Each value represents the agreement of exactly 1576 ratings