XGB Model trained with samples created by gpt4 rating answers created by text-davinci-003 used for testing 0.20 0.15 0 0.10 0.05 0.00 0 -0.05-0.10100 200 1600 3200 50 400 800 Number of samples Each box represents exactly 6 kappa values