XGB Model trained with samples created by text-davinci-003 rating answers created by text-davinci-003 used for training 0.9 0.8 0.7 -₩ 0.6 0.5 -0.4 0.3 100 200 400 800 1600 3200 50 Number of samples Each box represents exactly 6 kappa values