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 training 0.9 -0.8 -0.7 -X 0.6 -0.5 0.4 0.3 -0.2 100 200 400 800 1600 3200 50 Number of samples

Each box represents exactly 6 kappa values