XGB Model trained with samples created by gpt4 rating answers created by text-davinci-003 used for testing 0.6 -0.5 0.4 -0 ¥ 0.3 0.2 0.1 0.0 -100 200 1600 3200 50 400 800 Number of samples

Each box represents exactly 6 kappa values