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

Each box represents exactly 6 kappa values