XGB Model trained with samples created by gpt4 rating answers created by gpt4 used for training 0.95 0.90 ¥ 0.85 0.80 0.75 0.70 100 400 800 1600 3200 50 200 Number of samples Each box represents exactly 6 kappa values