XGB Model trained with samples created by text-davinci-003 annotated by gpt-3.5-turbo rating answers created by human experts used for training 0.5 0 0.4 0.3 0 QWK 0.2 0.1 50 800 1600 3200 100 200 400 Number of samples Each box represents exactly 6 kappa values