BERT Model trained with samples created by text-davinci-003 annotated by gpt-3.5-turbo rating answers created by human experts used for training 0.40 -0.35 0.30 0.25 ¥ 0.20 0.15 0.10 0.05 -50 800 1600 100 200 400 3200 Number of samples

Each box represents exactly 6 kappa values