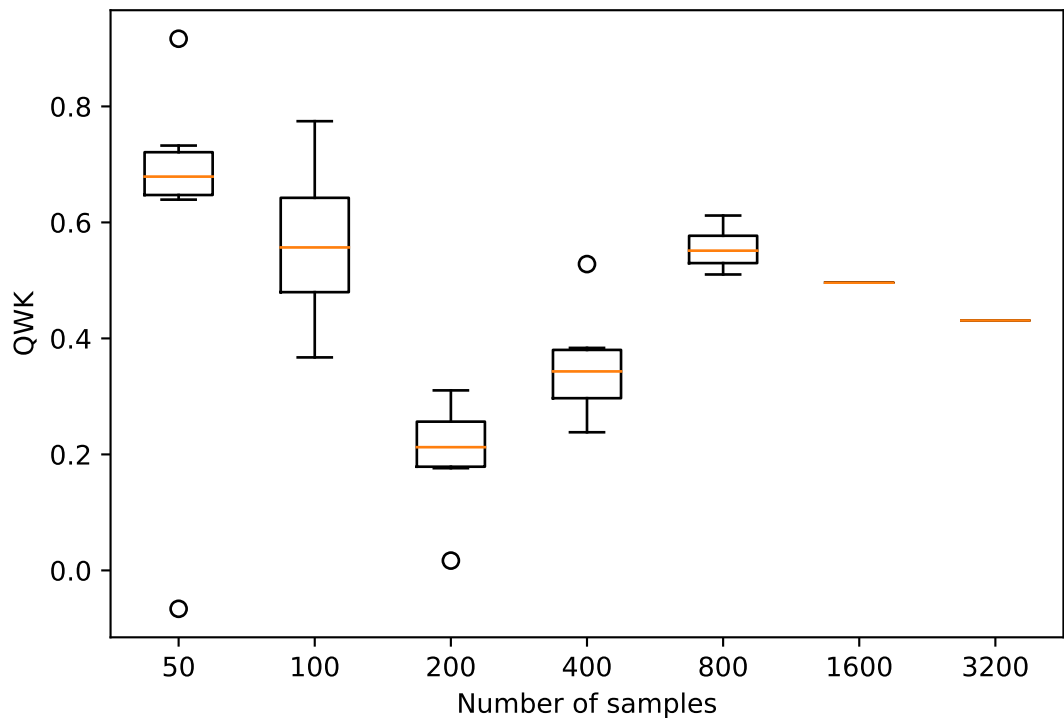


QWK of score_1 assigned to samples created by experts-testing by two BERT Models
One trained with samples created by davinci one trained with samples created by gpt4.



Each box represents exactly 6 kappa values