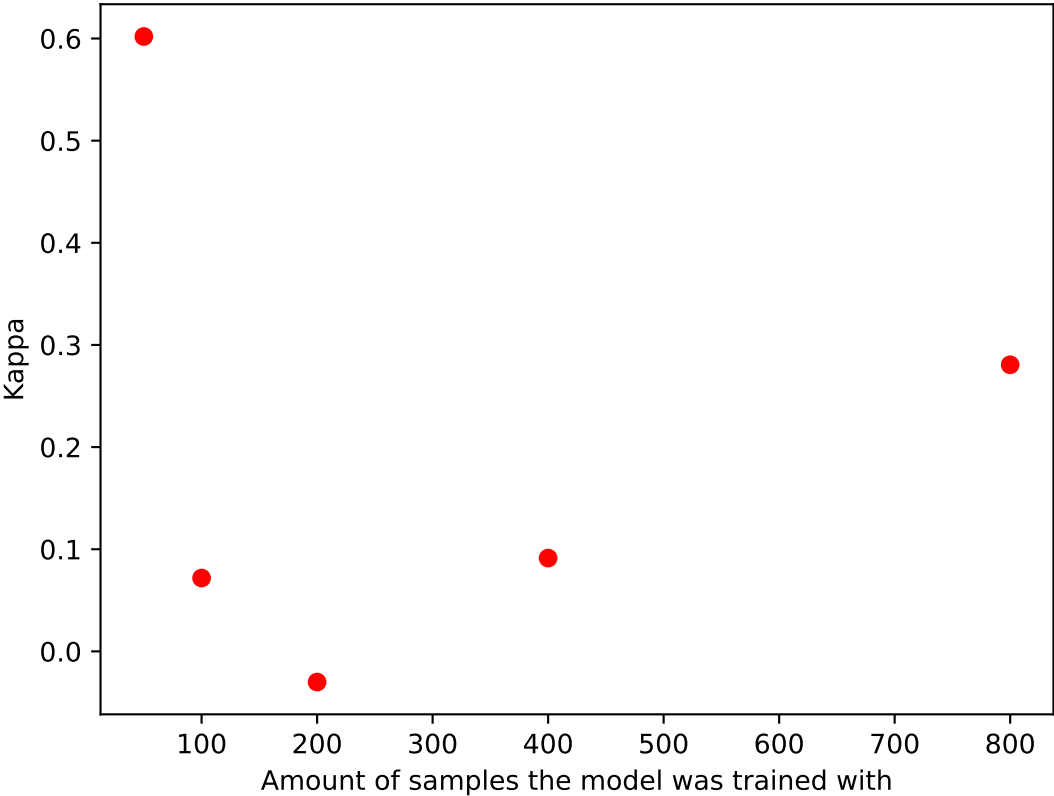


QWK of score_1 assigned to samples created by experts-testing by two BERT Models
One trained with samples created by experts one trained with samples created by gpt4.



Each value represents the agreement of exactly 392 ratings