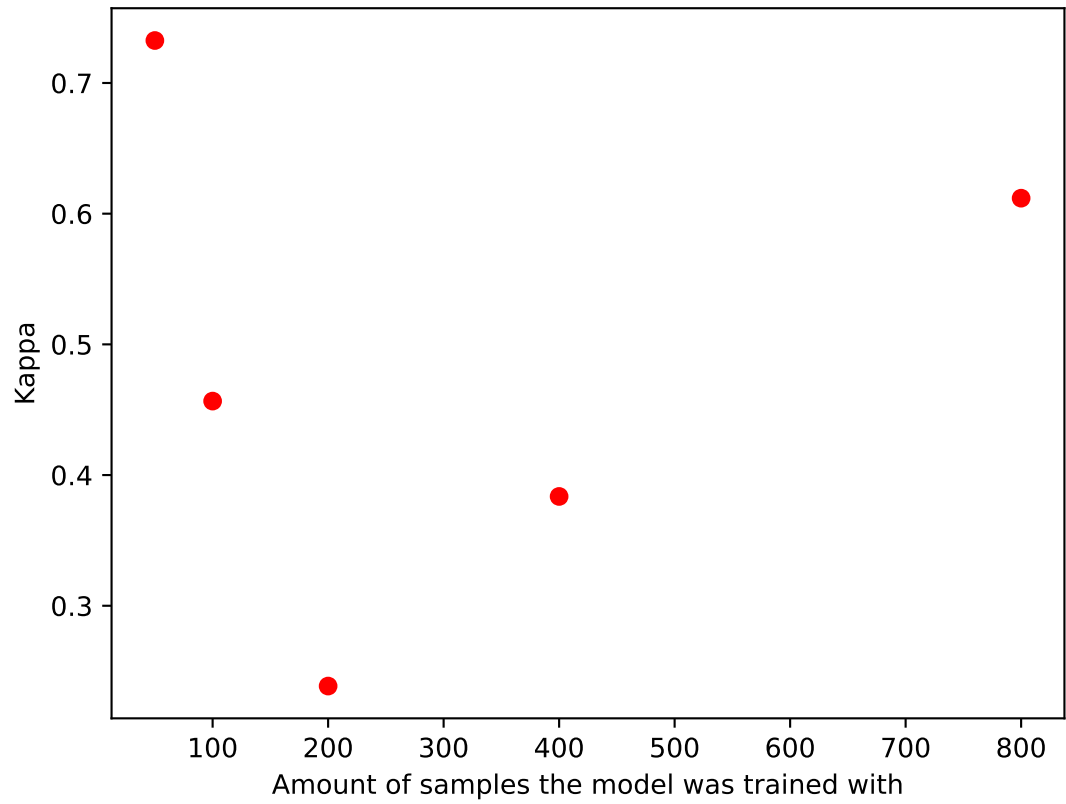


QWK of score\_1 assigned to samples created by experts-testing by two BERT Models  
One trained with samples created by davinci one trained with samples created by gpt4.



Each value represents the agreement of exactly 392 ratings