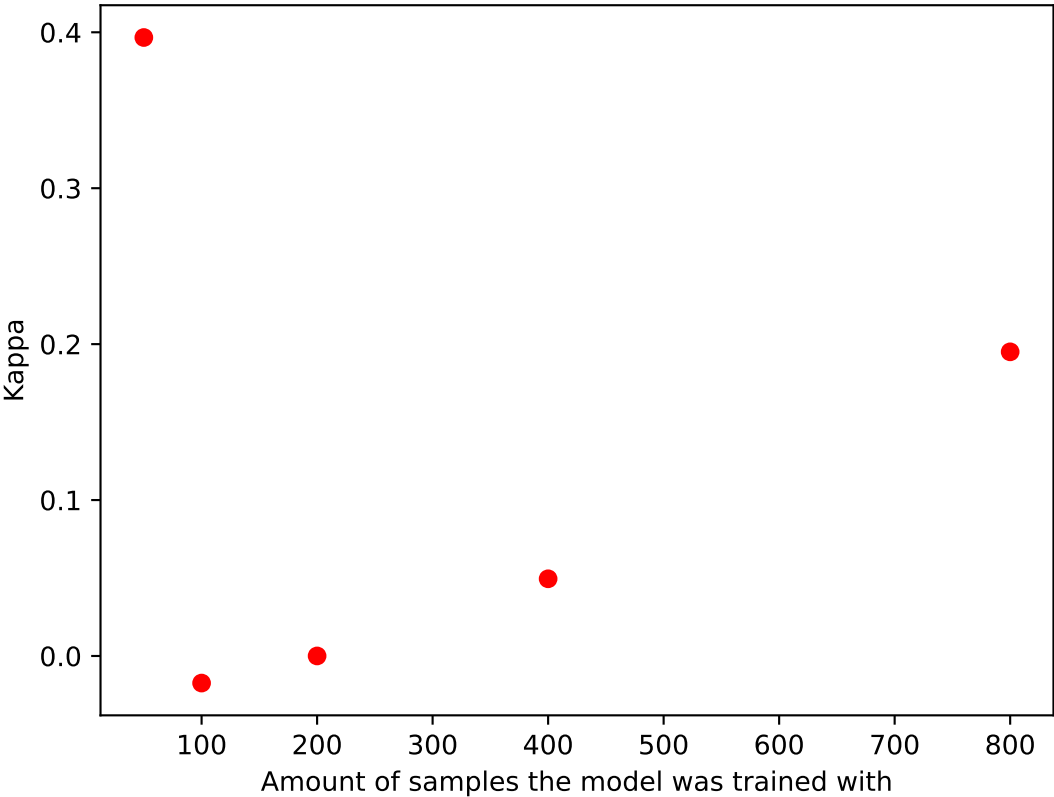


QWK of score\_1 assigned to samples created by davinci-testing by two BERT Models  
One trained with samples created by experts one trained with samples created by gpt4.



Each value represents the agreement of exactly 1577 ratings