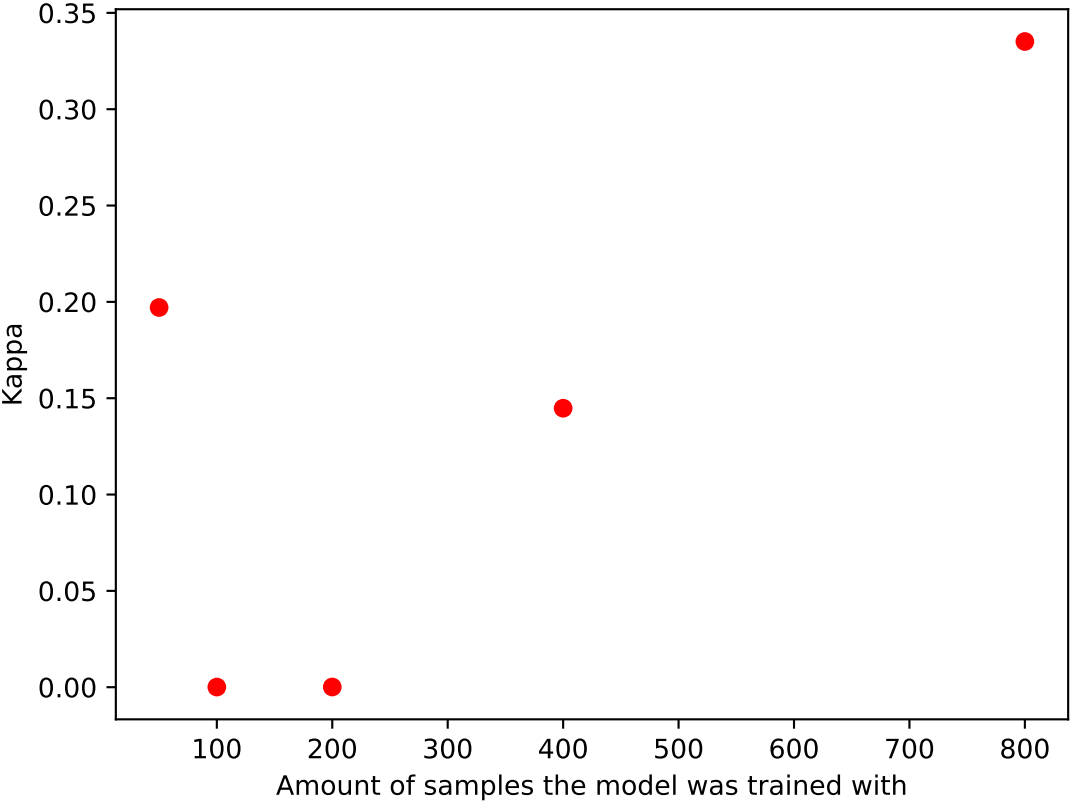QWK of score_1 assigned to samples created by gpt4-testing by two BERT Models
One trained with samples created by davinci one trained with samples created by experts.

Kappa

Amount of samples the model was trained with

Each value represents the agreement of exactly 1600 ratings