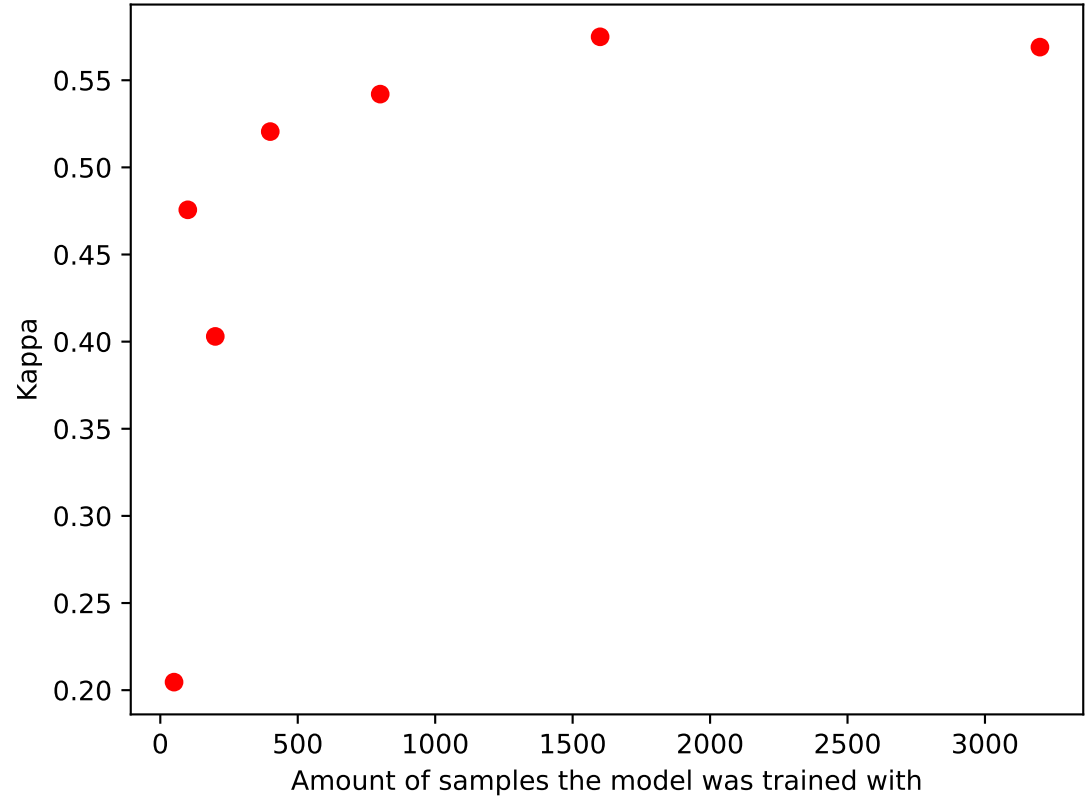XGB Model trained with samples created by gpt4 rating answers created by human experts used for training

Kappa

Amount of samples the model was trained with

Each value represents the agreement of exactly 3200 ratings