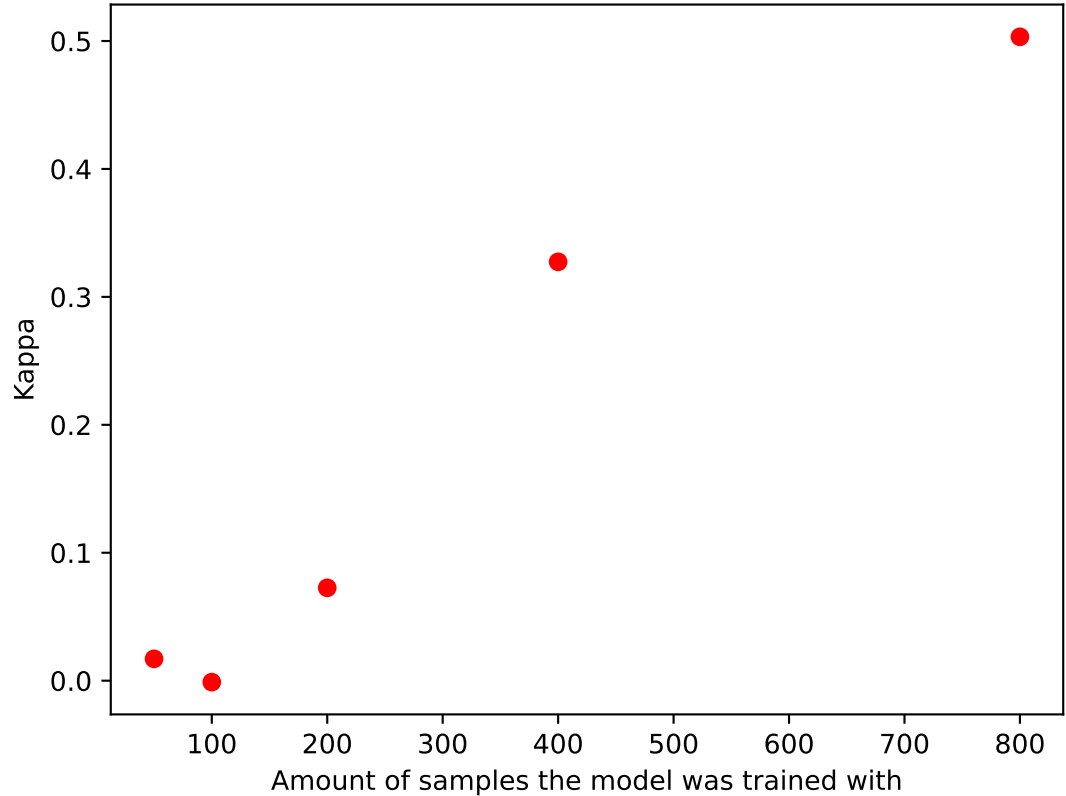XGB Model trained with samples created by text-davinci-003 annotated by gpt-3.5-turbo rating answers created by gpt4 used for testing

Kappa

Amount of samples the model was trained with

Each value represents the agreement of exactly 1600 ratings