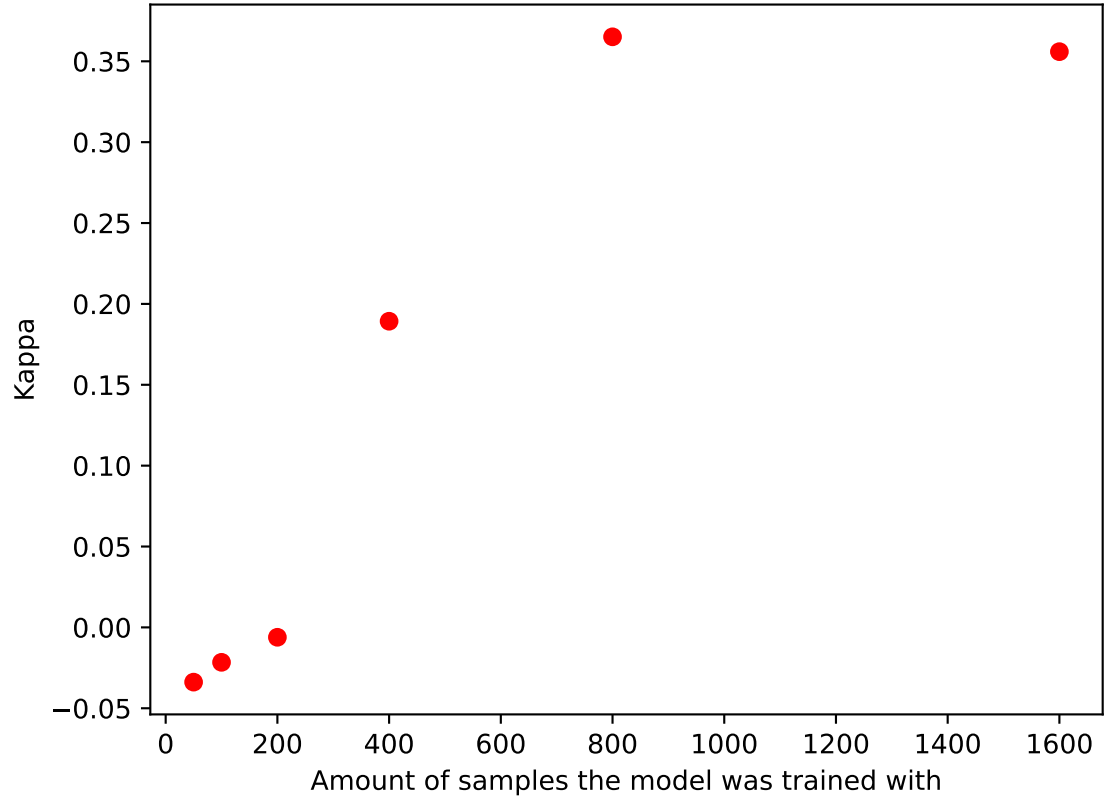XGB Model trained with samples created by human experts rating answers created by text-davinci-003 used for testing

Kappa

Amount of samples the model was trained with

Each value represents the agreement of exactly 1577 ratings