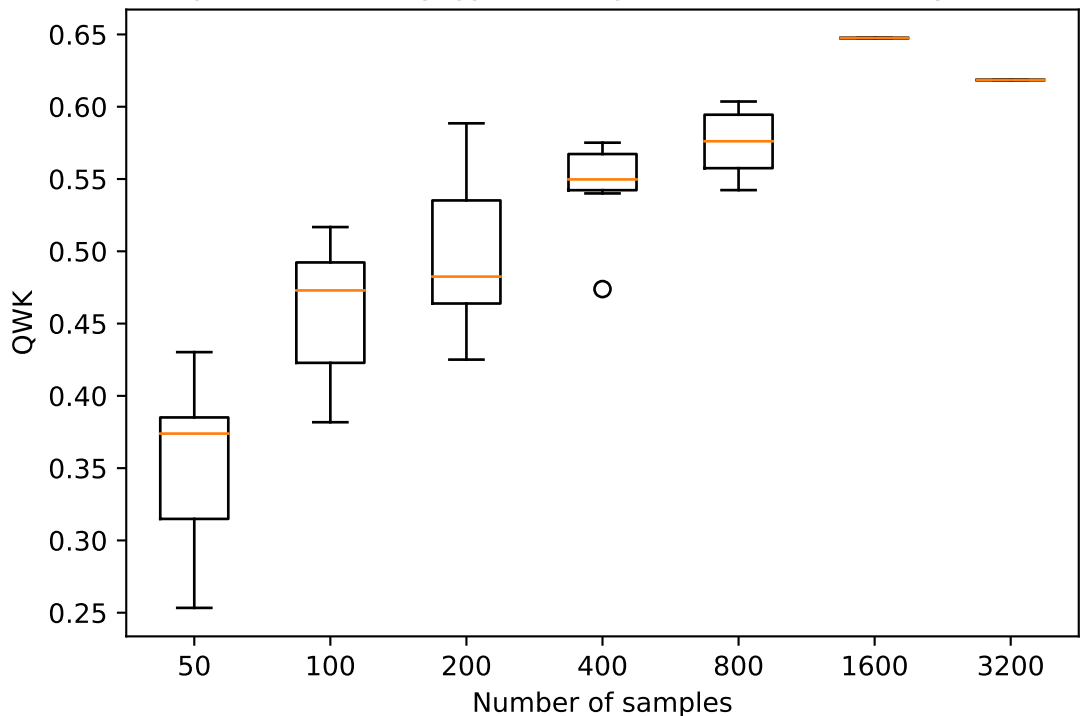XGB Model trained with samples created by gpt4 rating answers created by human experts used for testing

Each box represents exactly 6 kappa values