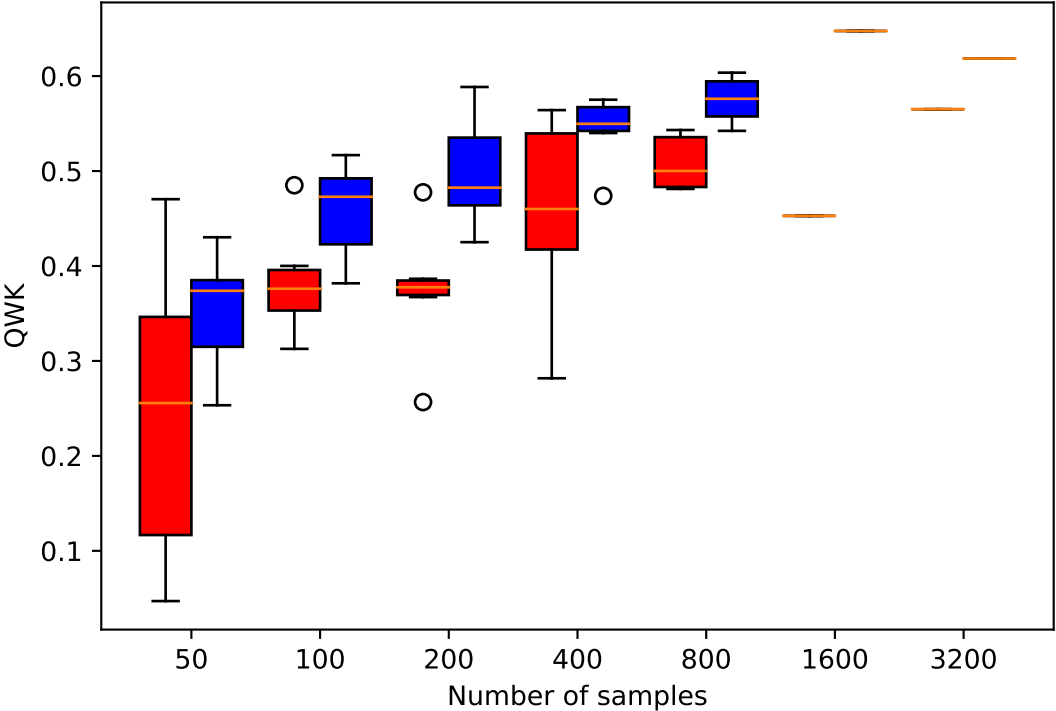


QWK of two XGB Models with the source of the data. One trained with samples created by text-davinci-003 One trained with samples created by gpt4.  
Both rated answers created by human experts used for testing



Each box represents exactly 6 kappa values  
text-davinci-003 is represented by red  
gpt4 is represented by blue