

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

