

The QWK for score_1 when comparing two XGB models versus the QWK for score_1 between a different pair of XGB models.
The first pair was trained with samples created by gpt4 and human experts,
the second pair was trained with samples created by human experts and text-davinci-003

